

---

# GLOSSÁRIO DE ESTATÍSTICA

---

JANILSON PINHEIRO DE ASSIS  
ROBERTO PEQUENO DE SOUSA  
CARLOS TADEU DOS SANTOS DIAS

---

$\bar{Y}$	$\prod_{i=1}^n x_i$	$\cap$	$P(W = w_i) = \frac{m}{n}$
$\mathbf{A}$	$\beta$	$\mathbf{P}$	$\sum_{i=1}^n y_i^2$
$\alpha$	$\mathbf{S}$	$\sigma$	$\mathbf{W}$
$\Gamma$	$\sum_{i=1}^n (x_i - \bar{x})^2$	$\chi^2$	$\theta$
$\int_a^b f(x) dx$	$\mathbf{X}$	$\pm$	$\mu$
$\approx$	$\geq$	$\partial$	$\infty$
$\Lambda$	$\pi$	$\Omega$	$\Delta$

JANILSON PINHEIRO DE ASSIS  
ROBERTO PEQUENO DE SOUSA  
CARLOS TADEU DOS SANTOS DIAS

## **GLOSSÁRIO DE ESTATÍSTICA**

©2019. Direitos Morais reservados aos autores: Janilson Pinheiro de Assis, Roberto Pequeno de Sousa, Carlos Tadeu dos Santos Dias. Direitos Patrimoniais cedidos à Editora da Universidade Federal Rural do Semi-Árido (EdUFERSA). Não é permitida a reprodução desta obra podendo incorrer em crime contra a propriedade intelectual previsto no Art. 184 do Código Penal Brasileiro. Fica facultada a utilização da obra para fins educacionais, podendo a mesma ser lida, citada e referenciada. Editora signatária da Lei n. 10.994, de 14 de dezembro de 2004 que disciplina o Depósito Legal.

**Reitor**

José de Arimateia de Matos

**Vice-Reitor**

José Domingos Fontenele Neto

**Coordenador Editorial**

Pacelli Costa

**Conselho Editorial**

Pacelli Costa, Walter Martins Rodrigues, Francisco Franciné Maia Júnior, Rafael Castelo Guedes Martins, Keina Cristina S. Sousa, Antonio Ronaldo Gomes Garcia, Auristela Crisanto da Cunha, Janilson Pinheiro de Assis, Luís Cesar de Aquino Lemos Filho, Rodrigo Silva da Costa e Valquíria Melo Souza Correia.

**Equipe Técnica**

Francisca Nataligeuza Maia de Fontes (Secretária), José Arimateia da Silva (Designer Gráfico).

**Dados Internacionais da Catalogação na Publicação (CIP)  
Editora Universitária (EdUFERSA)**

A848G	Assis, Janilson Pinheiro de. Glossário de estatística / Janilson Pinheiro de Assis, Roberto Pequeno de Sousa, Carlos Tadeu dos Santos Dias – Mossoró: EdUFERSA, 2019. 901p.  ISBN: 978-85-5757-104-4  1. Estatística. 2. Glossário. 3. Dicionário. 4. Termos estatísticos. I. Assis, Janilson Pinheiro de. II. Sousa, Roberto Pequeno de. III. Dias, Carlos Tadeu dos Santos. IV. Título.
EdUFERSA	CDD – 519.03

Bibliotecário-Documentalista  
Pacelli Costa (CRB15-658)

Editora filiada:



**Para Nossas Esposas**

Wilbea Maria Teixeira de Oliveira

Rita de Cássia Bezerra de Farias

Elvina Maria de Lima Dias





# SUMÁRIO

---

APRESENTAÇÃO .....	9
PREFÁCIO .....	11
Letra A .....	13
Letra B .....	89
Letra C .....	97
Letra D .....	163
Letra E .....	243
Letra F .....	339
Letra G .....	373
Letra H .....	397
Letra I .....	413
Letra J .....	439
Letra K .....	441
Letra L .....	443
Letra M .....	453
Letra N .....	515
Letra O .....	525
Letra P .....	535
Letra Q .....	587
Letra R .....	593
Letra S .....	627
Letra T .....	647
Letra U .....	723
Letra V .....	727
Letra W .....	777
Letra X .....	779
Letra Y .....	781
Letra Z .....	783
APÊNDICE .....	785
REFERÊNCIAS .....	883



# APRESENTAÇÃO

---

A crescente necessidade de se extrair dos dados de observação informações compreensíveis para qualquer pessoa nas diferentes áreas da atividade humana, tem levado cada vez mais a utilização da estatística nos mais diversos domínios da ciência, exigindo um bom conhecimento dos seus métodos e técnicas. Em consequência, esta disciplina, ou melhor, essa ciência encontra-se atualmente integrada com vários níveis de aprofundamento nos planos curriculares dos diferentes ciclos de formação do ensino superior em áreas muito diversas, e como ferramenta indispensável aos pesquisadores nas diferentes áreas das ciências agrárias, das ciências físicas, ciências biológicas, ciências sociais, na engenharia, na indústria em geral como por exemplo indústria naval, indústria automotiva, indústria aeronáutica, no comércio, dentre outras.

Dessa forma o conhecimento e a aplicação da estatística na prática exigem do pesquisador, além do conhecimento teórico e da resolução de problemas e casos de estudo, o significado verbal dos diferentes termos, verbetes, símbolos, equações, leis, teoremas, definições, conceitos, dentre outros, utilizado quando se emprega essa importante ferramenta de uso tão frequente pelos pesquisadores em geral.

Sendo assim, a elaboração deste glossário vem colaborar com os usuários da estatística para tornar os métodos estatísticos paramétricos e não-paramétricos, univariados e multivariados, mais fáceis de serem utilizados, na própria análise estatística, bem como na interpretação das análises, na compreensão dos resultados, melhorando consequentemente as discussões e conclusões racionais obtidas após as análises dos dados experimentais.

Enfim este glossário foi elaborado para que didaticamente fossem expostos e publicados um conteúdo para consulta nas diferentes áreas do conhecimento quais sejam ciências agrárias, ciências biológicas, ciências exatas e tecnológicas, engenharias, zootecnia, medicina veterinária, administração, ciências da computação, ciências contábeis, em pesquisas de marketing, pesquisas de opinião pública, pesquisas de Survey, hidrologia, meteorologia, ecologia, engenharia florestal, genética, melhoramento genético animal e vegetal, probabilidade, matemática, econometria, na estatística Bayesiana, na análise de sobrevivência, além de setores relacionados a medicina, saúde pública ou coletiva como bioestatística e epidemiologia, modelagem, simulação, metodologia científica ou de pesquisa, geoestatística, métodos de análises para avaliação sensorial de alimentos, controle estatístico de qualidade, dentre outras. Sendo também uma

leitura básica e indispensável para aqueles que pretendem aprofundar seus conhecimentos em estatística e métodos quantitativos em geral, seja em áreas aplicadas ou de desenvolvimento teóricos mais avançados.

Sendo assim esperamos que esta obra sirva de forma direta e indireta aos pesquisadores das mais diferentes áreas do conhecimento humano, bem como aos estudantes do ensino básico e fundamental, aos estudantes de graduação e de pós-graduação, aos técnicos que utilizam estatística rotineiramente ou não, e finalmente ao público em geral, que deseja obter informações sobre o uso da ciência estatística.

Mossoró-RN, Brasil, Julho de 2019.

**Janilson Pinheiro de Assis**

Professor Titular

Engenheiro Agrônomo

**Roberto Pequeno de Sousa**

Professor Associado

Engenheiro Agrícola

**Carlos Tadeu dos Santos Dias**

Professor Titular

Engenheiro Agrônomo

## PREFÁCIO

As nossas primeiras palavras para explicar por que este livro foi escrito, e por que o escrevemos da maneira como ele se apresenta, se deve a necessidade de preencher uma lacuna até hoje sem iniciativas concretas para por no mercado editorial um livro dicionário impresso em português do Brasil, para servir de guia de consultas a todos aqueles que utilizam a estatística como ferramenta de trabalho nas diversas áreas da pesquisa científica, no comércio, no mercado financeiro, na indústria, nas universidades e nas escolas de uma maneira geral, pois em algum momento de suas análises e interpretações necessitamos tirar dúvidas sobre termos, verbetes, símbolos, nomenclaturas estatística, dentre outros. Ele não surgiu evidentemente de nenhuma inspiração momentânea. Pelo contrário, é fruto de anos de trabalho e experiência adquiridos lecionando as disciplinas de estatística e estatística experimental na Escola Superior de Agricultura de Mossoró (ESAM) e atualmente Universidade federal Rural do Semi-árido (UFERSA), em Mossoró, RN, bem como na Escola Superior de Agricultura “Luiz de Queiróz” (ESALQ), em Piracicaba, SP, durante mais de vinte e nove anos. Sendo assim o que nos ampara e dá respaldo é a experiência de mais de vinte e nove anos dedicados à orientação e à elaboração da análise estatística de muitos trabalhos de pesquisa, tanto teses, dissertações, monografias, como artigos científicos para publicação, e alguns de nossa própria autoria como de outros pesquisadores, muitos destes bastante experientes em pesquisa, mas comumente pouco versados em Estatística.

A verdade é que como não dispomos no Brasil de um glossário voltado para a área de estatística publicado em português, vimos a necessidade de elaborar esta obra onde descrevemos o maior número possível de verbetes e termos mais comuns usados na estatística que achamos razoável para uma obra dessa natureza.

Vale salientar que na década de 70 do século passado a Fundação Instituto Brasileiro de Geografia e Estatística, publicou uma excelente obra intitulado “Dicionário Brasileiro de Estatística”, mas que não teve muita divulgação e também não foi reeditado ou reimpresso, além de ficar restrito a poucos, é tanto que hoje é raro encontrar-se algum exemplar disponível para consulta ou aquisição.

Neste glossário o leitor encontrará palavras iniciando desde a vogal A, como por exemplo, a palavra amostra, até a consoante Z como a palavra zona de aceitação, muitas vezes repetidas com outra descrição, mas com o mesmo significado. Os verbetes foram inseridos usando como critério a letra inicial do termo descrito, e não a ordem alfabética, isto porque muitos termos são fórmulas, equações, dentre outros, o

que torna sem sentido este critério de inserção ou descrição dos termos do glossário, ou seja, as palavras foram alojadas em capítulos definidos e iniciados pelas letras A, B, C, D, E, ... , dentre outras.

Acreditamos que esse glossário sirva como um guia de cabeceira para todos aqueles pesquisadores, leitores ou mesmo curiosos que são mais familiarizados com a estatística ou mesmo aqueles mais avessos a essa disciplina, para que ele ajude a desvendar dúvidas e corrigir erros, enfim colaborar de forma consistente nos trabalhos conduzidos nas mais diferentes áreas do conhecimento humano.

Todos os erros que porventura existam são de inteira responsabilidade dos autores, e os mesmos agradecem as críticas e sugestões enviadas, pois as mesmas serão bem vindas e só ajudaram no aperfeiçoamento e aprimoramento desta obra tão valiosa para a ciência.

Mossoró-RN, Brasil, Julho de 2019

**Professor Dr. Janilson Pinheiro de Assis**

(janilson@ufersa.edu.br)

**Professor Dr. Roberto Pequeno de Sousa**

(rpequeno@ufersa.edu.br)

**Professor Dr. Carlos Tadeu dos Santos Dias**

(ctsdias@usp.br)



**#A:** Cardinal do conjunto A. Corresponde ao número de elementos que compõe um determinado conjunto finito e discreto A.

$A^c$ : Complementar do conjunto (evento) A. Também simbolizado por  $\bar{A} = A' = CA = \sim \tilde{A}$

$A^c = \bar{A} = \tilde{A} = A' = \sim \tilde{A} = CA$ : Conjunto complementar de um conjunto A.

$\bar{A}$ : Complementar do conjunto (evento) A.

$a_3$ : Coeficiente de assimetria amostral.

$a_4$ : Coeficiente de curtose amostral.

**AAR:** Aumento Absoluto do Risco ( $p_1 - p_2$ ).

**ABORDAGEM BAYESIANA:** Uma das duas abordagens utilizadas em estatística, baseadas no teorema de Bayes como forma de atualizar a informação existente sobre alguma quantidade de interesse. Em inglês Bayesian approach.

**ABORDAGEM CLÁSSICA:** Uma das duas abordagens existentes em estatística, que utiliza, para a estimação de quantidades de interesse, a informação amostral obtida de um experimento. Em inglês Classical approach.

**ABORDAGEM CONJUNTA BASEADA EM ESCOLHA:** Forma alternativa de coleta de respostas e estimação do modelo conjunto. A diferença principal é que os respondentes selecionam um único estímulo de perfil completo a partir de um conjunto de estímulos, conhecido como conjunto de escolhas, em vez de avaliar ou ordenar cada estímulo separadamente.

**ABORDAGEM DE CASO COMPLETO:** Tratamento para lidar com dados perdidos que computa valores com base em dados somente de casos completos, ou seja, casos em que dados perdidos ocorrem em ensaios com dados balanceados.

**ABORDAGEM DE DISPONIBILIDADE:** Método de atribuição para dados perdidos que computa valores com base em todas as observações válidas disponíveis.

**ABORDAGEM DE EXTREMOS POLARES:** Método para construir uma variável dependente categórica a partir de uma variável métrica. Primeiro, a variável métrica é dividida em três categorias. Em seguida, as categorias extremas são usadas na análise discriminante ou na regressão logística, e a categoria do meio não é incluída na análise.

**ABORDAGEM FREQUENTISTA:** Uma das duas abordagens em estatística, baseada na utilização de informação proveniente de um experimento, deixando de lado qualquer subjetividade existente. Em inglês Frequentist approach.

**ABSCISSA EQUIVALENTE:** Corresponde a uma probabilidade ou proporção  $p$ , relativa à distribuição de função de probabilidade acumulada  $f.f. (\theta)$  é raiz t da equação

$$\int_{-\infty}^t f(\theta) d\theta = p$$

Também se diz afastamento equivalente. Cf. Afastamento normal equivalente, Próbite, Lógite.

**ABSCISSA (ABSCISSA):** A linha horizontal, ou eixo dos  $x$ 's, em um gráfico. O oposto de ordenada.

**ACURÁCIA (ACCURACY):** O grau em que a medida de um objeto se aproxima do seu valor real. Comparar com precisão.

**AGRUPAMENTO AGLOMERATIVO (AGGLOMERATIVE CLUSTERING):** O processo de agrupar objetos pegando grupos menores e os agrupando em grupos maiores, com base em sua similaridade. Ver também análise de agrupamento. Comparar com agrupamento divisivo.

**AGRUPAMENTO DIVISIVO (DIVISIVE CLUSTERING):** O processo de agrupar objetos a partir da divisão de um grande grupo, com base na dissimilaridade, em outros sucessivamente menores. Ver também análise de agrupamentos. Comparar com agrupamento aglomerativo.

**AGRUPAMENTO HIERÁRQUICO (HIERARCHICAL CLUSTERING):** Um método para agrupar objetos que se baseia na imposição de um sistema de referência externo. Em agrupamentos hierárquicos, se  $a$  e  $b$  estão em um grupo e as observações  $c$  e  $d$  estão em outro um grupo maior que inclui  $a$  e  $c$ , também deve incluir  $b$  e  $d$ . A taxonomia linneana é um exemplo de um sistema de agrupamento hierárquico.

**AGRUPAMENTO NÃO HIERÁRQUICO (NON HIERARCHICAL CLUSTERING):** Um método para agrupar objetos que não é fundamentado em um sistema de referência externo. Comparar com agrupamento hierárquico.

**AGRUPAMENTO POR K-MÉDIAS (K-MEANS CLUSTERING):** Um método não hierárquico de agrupar objetos que minimiza a soma dos quadrados das distâncias de cada objeto até o centroide de seu grupo.

**AJUSTE (FIT):** Quão consistentes as predições são em relação ao observado. Usado para descrever quão bem os dados são preditos por um dado modelo, ou quão bem os modelos são preditos pelos dados disponíveis.

**ACASALAMENTO AO ACASO:** Sistema de acasalamento em que cada fêmea tem chance igual de se acasalar com qualquer reprodutor durante uma estação de monta. Acasalamento ao acaso é necessário para testes de progênie precisos.

**ACASALAMENTO ASSOCIATIVO:** Situação em que o acasalamento entre machos e fêmeas não é aleatório, e envolve uma tendência de machos de um particular tipo acasalarem-se preferencialmente com fêmeas de um determinado tipo.

**ACASO:** Tudo o que não puder ser predito ou previsto. Não determinado. Em inglês random.

**ACASO (ESTATÍSTICA):** Termo usado para descrever os resultados de um processo estocástico, isto é, um processo no qual a probabilidade de ocorrer qualquer evento é conhecida ou pode ser determinada. Diz-se do resultado da soma de um complexo de numerosas causas cujas atuações individuais desconhecemos. Ao acaso: não significa, em estatística, a esmo, sem reflexão, inadvertidamente, mas o contrário: significa processo construído para que cada resultado possível esteja associado a uma probabilidade conhecida.

**ACASO (GERAL):** Acontecimento incerto ou imprevisível; casualidade, eventualidade. Fortuito. Destino, fortuna, sorte. Ao acaso: a esmo, sem reflexão, inadvertidamente.

**ACONTECIMENTOS INDEPENDENTES:** Os acontecimentos  $A$  e  $B$  dizem-se independentes se  $P\{A \cap B\} = P\{A\} P\{B\}$ .

**ACUMULAÇÃO:** É uma nova variável resultante da multiplicação do teor pela espessura. É comumente aplicada em cálculo de reservas em depósitos estratiformes, onde o teor médio é obtido pela divisão da acumulação pela espessura média.

**ACURÁCIA:** É a capacidade de um teste obter a medida correta, em média.

**ACURÁCIA:** No sentido estatístico geral, acurácia denota a proximidade dos cálculos ou estimações com seus valores reais. Em inglês Accuracy.

**ACURÁCIA:** O mesmo que acuidade ou validade de um teste diagnóstico.

**ACURÁCIA (DE SELEÇÃO):** Correlação entre o valor genético real (desconhecido) de um animal e um valor genético calculado.

**AÇÃO INDEPENDENTE:** Dois estímulos,  $e_1$  e  $e_2$ , de reações esperadas  $P_1$  e  $P_2$ , são ditos terem ação independente, quando, misturados, seus efeitos são aditivos. Nesse caso, o efeito  $P$  da mistura será dado por  $P = P_1 + P_2 (1 - P_1) = 1 - (1 - P_1) (1 - P_2)$ . O conceito, evidentemente, pode ser generalizado a mais estímulos.

**ALFA (ALPHA):** A primeira letra ( $\alpha$ ) do alfabeto grego. Na estatística, ela é usada para denotar uma probabilidade aceitável de cometer um erro Tipo I.

**AMOSTRA (SAMPLE):** Um subconjunto da população de interesse. Como é praticamente impossível observar ou manipular todos os indivíduos em uma população, as análises baseiam-se em amostras da população. A probabilidade e a estatística são utilizadas para extrapolar os resultados a partir de uma amostra pequena para uma população maior.

**ANÁLISE DE AGRUPAMENTO (CLUSTERANALYSIS):** Um método para agrupar objetos com base em distâncias multivariadas entre eles.

**ANÁLISE DE CAMINHOS (PATH ANALYSIS):** Um método para atribuir causas e efeitos múltiplos e sobrepostos em um tipo especial de modelo de regressão múltipla.

**ANÁLISE DE COMPONENTES PRINCIPAIS (PCA-PRINCIPAL COMPONENT ANALYSIS):** Um método de ordenação que reduz dados multivariados a um número menor de variáveis pela

criação de combinações lineares das variáveis originais. Foi originalmente desenvolvida para atribuir identidade racial a indivíduos com base em múltiplas medidas biométricas.

**ANÁLISE DE COORDENADAS PRINCIPAIS (PCOAPRINCIPAL COORDINATE ANALYSIS):** Um método de ordenação genérico que pode usar qualquer medida de distância. A análise de componentes principais e a análise fatorial, são casos especiais de uma análise de coordenadas principais.

**ANÁLISE DE CORRESPONDÊNCIA (CORRESPONDENCE ANALYSIS):** Um método de ordenação usado para relacionar a distribuição de espécies a variáveis ambientais. Também chamada de análise de gradiente indireta.

**ANÁLISE DE CORRESPONDÊNCIA CANÔNICA (CCA-CANONICAL CORRESPONDENCE ANALYSIS):** Um método de ordenação, tipicamente usado para relacionar a abundância ou as características de espécies com variáveis ambientais ou do habitat. Esse método ordena linearmente os preditores que são as variáveis ambientais ou do habitat e as respostas que são os dados das espécies para produzir eixos unimodais, ou de corcova, relativos aos preditores. Ela também é um tipo de análise de regressão multivariada. Comparar com a análise de redundância.

**ADEQUAÇÃO DO AJUSTE:** Em geral, a concordância entre valores observados e um conjunto de valores teóricos que depende de alguma hipótese. O termo é frequentemente usado no ajuste de uma distribuição teórica a um conjunto de observações.

**AJUSTE EXCESSIVO:** Adição de mais parâmetros ao modelo que o necessário.

**ALARME FALSO:** Um sinal proveniente de um gráfico de controle, quando nenhuma causa atribuída está presente.

**ABSCISSA:** É o eixo horizontal, eixo dos X's ou eixo das abscissas numa representação gráfica; ou ainda é qualquer valor da variável independente assinalado neste eixo.

**A-HISTÓRICO:** Termo que se refere ou é usado na teoria segundo o qual os fatos atuais não sofreram influência dos fatos passados. Ponto de vista que ignora a variável tempo nas pesquisas em ciências sociais.

**ALEATÓRIO:** Termo que se refere a escolhido ao acaso, por meio de sorteio, ou casualizado.

**ALFA:** Primeira letra do alfabeto grego (A,  $\alpha$ ), símbolo utilizado na estatística para representar designar o nível de significância ou probabilidade de se cometer o erro do tipo I ou de primeira espécie num teste de hipótese ou de significância estatística.

**ALFA DE CROMBACH:** Medida de fidedignidade ou consistência interna de um teste. O coeficiente alfa de Crombach varia de zero até 1,0 e indica o quanto os itens de um teste ou de uma pergunta num questionário, por exemplo, referem-se ou mensuram uma mesma variável. Quanto mais alto o valor do coeficiente, mais ele indica que os itens numa lista estão mensurando uma mesma grandeza.

**ALGÉBRA DE MATRIZES:** Todo cálculo envolvendo matrizes, vetores, determinantes, autovalores, autovetores, tais como adição, subtração, multiplicação, inversa, inversa generalizada. As matrizes são hoje em dia indispensáveis em várias aplicações, da matemática a biologia, especialmente em estatística. Alguns exemplos de aplicações são o seu uso na teoria dos grafos, na dinâmica de populações, na ecologia, na genética, na cadeia de Markov e em estatística como na análise estatística multivariada, projeto de

experimentos e na análise de variância, análise de covariância e na análise de regressão e correlação. Para mostrar uma ideia da aplicação em estatística, será considerado um modelo de regressão. Suponha que seja dado um diagrama de dispersão que consiste de pontos  $(x_i, y_i)$ ,  $i = 1, 2, 3, 4, \dots, n$ , onde  $x_i$  e  $y_i$  são medidas. Considere também que os pontos estejam próximos de certa curva e que se queira ajustar uma curva do tipo parábola quadrática. A equação é dada por:  $y = a + bx + cx^2$ , onde as constantes a, b e c são desconhecidas. As coordenadas  $x_i, y_i$  dos pontos não satisfazem exatamente uma equação do tipo  $y = a + bx + cx^2$ . Portanto pode-se escrever  $y = a + bx_i + cx_i^2 + e_i$  ( $i = 1, 2, 3, 4, \dots, n$ ), onde  $e_i$  é um erro. O sistema de equações  $y = a + bx_i + cx_i^2 + e_i$  ( $i = 1, 2, 3, 4, \dots, n$ ) pode ser escrito com a notação de matriz na forma  $y = Xb + e$  introduzindo as matrizes,

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, X = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{pmatrix}, b = \begin{pmatrix} a \\ b \\ c \end{pmatrix}, e = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}. \text{ As equações } y = a + bx_i + cx_i^2 + e_i$$

( $i = 1, 2, 3, 4, \dots, n$ ) são lineares nos coeficientes desconhecidos a, b, c. Consequentemente, chamamos  $y = a + bx_i + cx_i^2 + e_i$  ( $i = 1, 2, 3, 4, \dots, n$ ) ou  $y = Xb + e$  um modelo linear. Para encontrar coeficientes adequados aplica-se um método dos mínimos quadrados. Minimizando a soma dos quadrados dos erros  $e_1^2 + e_2^2 + e_3^2 + \dots + e_n^2 = e'e$ . Em virtude da fórmula  $y = Xb + e$  isto significa que os coeficientes desconhecidos a, b e c são determinados de tal maneira que  $e'e = (y - Xb)'(y - Xb)$  assume o menor valor possível. A solução pode ser encontrada por métodos da álgebra de matrizes, ou pode-se aplicar o cálculo diferencial igualando a zero a derivada da equação de regressão em relação ao coeficiente a, b e c, e através do sistema de equações normais encontrar os estimadores dos coeficientes a, b e c do modelo ajustado.

**AMOSTRA ALEATÓRIA SIMPLES:** Amostra selecionada de tal maneira que a escolha de um elemento da população não afeta a probabilidade de seleção de qualquer outro membro. Ou seja, cada membro da população possui probabilidade igual (equiprobabilidade) de ser selecionado para a amostra.

**AMOSTRA DE BOLA DE NEVE:** Tipo de amostra não probabilística de natureza intencional na qual um sujeito indica outro sujeito para integrar a amostra. Outra extensão deste tipo de amostra ocorre quando os próprios sujeitos são solicitados a repassarem um questionário o qual é o instrumento da pesquisa para outros, que eles próprios indicaram.

**ANÁLISE DE SOBREVIVÊNCIA:** A Análise de Sobrevivência lida com a morte de organismos biológicos e falha de sistemas mecânicos. Envolve a modelagem do tempo da ocorrência de um evento, onde a morte ou falha é considerado um evento.

**ANÁLISE DE CORRESPONDÊNCIA:** Um dos objetivos da análise correspondência é descrever o relacionamento entre duas variáveis nominais em um espaço dimensional mais baixo, enquanto descreve simultaneamente o relacionamento entre as categorias para cada variável.

**ANÁLISE DE FATORES:** A Análise de Fatores, incluindo variações como a Análise de Componentes e a Análise de Fatores Comuns objetiva encontrar uma forma de condensar a informação contida em um determinado número de variáveis originais em um conjunto menor com perda mínima de informação.



**ANÁLISE LOGLINEAR:** Dados categóricos podem ser expressos na forma de um modelo linear onde são utilizados valores de logaritmos que é a razão de a técnica receber essa denominação.

**ANÁLISE CONJUNTA:** A análise conjunta é uma técnica de avaliação de objetos tais como produtos, serviços ou idéias utilizada para entender como os respondentes desenvolvem preferências por produtos ou serviços.

**ANÁLISE DE CONFIABILIDADE:** Permite o estudo das propriedades de escalas de medidas e de seus itens. O procedimento determina medidas de confiabilidade de escalas e também fornece informações sobre relacionamentos entre itens individuais na escala.

**ANÁLISE DE FATORES:** A Análise de Fatores, incluindo variações como a Análise de Componentes e a Análise de Fatores Comuns objetiva encontrar uma forma de condensar a informação contida em um determinado número de variáveis originais em um conjunto menor com perda mínima de informação.

**ANÁLISE DE CORRESPONDÊNCIA DESTENDENCIADA (EM INGLÊS SIGNIFICA DETRENDED CORRESPONDENCE ANALYSIS):** Uma variação da análise de correspondência usada para melhorar ou remover o efeito do arco.

**ANÁLISE DE COVARIÂNCIA (ANCOVA-ANALYSIS OF COVARIANCE):** Um modelo estatístico híbrido, algo entre regressão e ANOVA, que inclui uma variável preditora contínua (a covariável) em conjunto a uma variável preditora categórica de um modelo de ANOVA.

**ANÁLISE DE DADOS GRÁFICA EXPLORATÓRIA (EDA-GRAPHICAL EXPLORATORY DATA ANALYSIS):** O uso de estruturas visuais, como gráficos, para detectar padrões, dados discrepantes e erros nos dados.

**ANÁLISE DE INTERVENÇÃO RANDOMIZADA (RANDOMIZED INTERVENTION ANALYSIS):** Uma técnica de Monte Carlo para a análise de delineamentos ADCL.

**ANÁLISE DE MONTE CARLO (MONTE CARLO ANALYSIS):** Uma das três grandes estruturas de análises estatísticas. Ela usa métodos de Monte Carlo para estimar valores de P. Ver métodos de Monte Carlo. Comparar com inferência bayesiana e análises paramétricas.

**ANÁLISE DE REDUNDÂNCIA (RDA-REDUNDANCE ANALYSIS):** Um tipo de regressão múltipla usado quando tanto a variável preditora quanto a variável resposta são multivariadas. Diferente da semelhante análise de correspondência canônica, a análise de redundância não faz nenhuma premissa acerca da forma dos eixos preditores ou resposta.

**ANÁLISE DE TABELA DE CONTINGÊNCIA (CONTINGENCY TABLE ANALYSIS):** Os métodos estatísticos usados para analisar conjuntos de dados em que tanto a variável resposta quanto a preditora são categóricas. As tabelas de contingência são compostas de linhas e colunas (h x k), e assim pode-se simbolizar uma tabela 2x3 por exemplo, como aquela que possui duas linhas e três colunas, mas estas tabelas podem ser compostas de duas, três, quatro ou mais linhas ou colunas. Contingência significa dependência, sendo assim uma tabela de contingência é aquela que mostra como duas ou mais variáveis qualitativas ou categóricas estão dependentes uma da outra ou não.

**ANÁLISE DE RESÍDUOS:** Consiste em um conjunto de técnicas para investigar a adequabilidade do modelo com base nos resíduos  $e_i = \hat{\varepsilon}_i = Y_i - \hat{Y}_i$ . A ideia básica é que se o modelo linear simples é

apropriado, os resíduos devem refletir as suposições tais como independência e variância constante para diferentes níveis de X e distribuição Normal.

**ANÁLISE IDIOGRÁFICA:** Termo oriundo do grego (idios: singular, peculiar), que se refere a uma análise de tipo qualitativa, subjetiva, hermenêutica. A análise ideográfica trata dos fatos individuais, buscando a singularidade de cada fenômeno e não a generalização.

**ANÁLISE NOMOTÉTICA:** Termo de origem do alfabeto grego denominado nomothetein, que significa baixar uma lei. Análise que busca a determinação de leis ou princípios gerais com base em fatos anteriores.

**ANÁLISE DE COVARIÂNCIA:** Em uma Análise de Covariância, ANCOVA, o pesquisador inclui uma variável contínua dentro da análise, o qual pressupõe que influencia a variável dependente. A covariável é inserida na ANOVA através de um modelo de regressão. Com isso, nós poderemos analisar a variável dependente, retirando a influência da covariável. Para diminuir a variância residual: Controlando a covariável p pode-se diminuir parte da variância residual (QMR). Com isso, esta análise melhora o modelo estatístico. Também existe o confundimento pois existe uma fonte de variação que afeta a variável dependente, mas não foi controlada durante o estudo

**ANÁLISE DE COVARIÂNCIA:** A análise de covariância pode ser entendida como um elo de ligação entre a análise de variância e a análise de regressão uma vez que um dos seus principais objetivos é avaliar o efeito de um ou mais fatores explicativos de natureza nominal numa dada variável resposta uma vez removida a influência que um ou mais fatores quantitativos podem também exercer nessa variável. Em alguns estudos demonstra-se que através da especificação de um modelo de regressão com regressores dummy é possível atingir de uma forma muito eficiente este objetivo particular da análise de covariância.

**ANÁLISE DE VARIÂNCIA (ANOVA-ANALYSIS OF VARIANCE):** Um método de partição da soma dos quadrados, creditada a Fisher. É usada primariamente para testar a hipótese nula estatística de que tratamentos distintos não possuem efeitos sobre uma variável resposta medida.

**ANÁLISE DE VARIÂNCIA MULTIVARIADA (MANOVA-MULTIVARIATE ANALYSIS OF VARIANCE):** Um método para testar a hipótese nula de que os vetores médios de múltiplos grupos são todos iguais uns aos outros. Uma ANOVA para dados multivariados.

**ANÁLISE DISCRIMINANTE (DISCRIMINANT ANALYSIS):** Um método para classificar observações multivariadas quando os grupos foram definidos com antecedência. A análise discriminante também pode ser usada como um teste post hoc para detectar diferenças entre grupos quando análise de variância multivariada (MANOVA) tenha produzido resultados significativos.

**ANÁLISE FATORIAL (FACTOR ANALYSIS):** Um método de ordenação que reescala cada observação original como uma combinação linear de um número menor de parâmetros, chamados de fatores comuns. Foi desenvolvida no começo do século 20 como forma de calcular a inteligência geral a partir de uma bateria de testes com pontuação.

**ANÁLISE PARAMÉTRICA (PARAMETRIC ANALYSIS):** Uma das três principais estruturas de análises estatísticas. Ela recai sobre o requisito de que os dados foram tirados de uma variável aleatória com distribuições de probabilidade definidas por parâmetros fixos. Também é chamada de análise frequentista ou estatística frequentista. Comparar com inferência bayesiana e análises de Monte Carlo.

**ANCOVA:** Ver análise de covariância.

**ANOVA:** Ver análise de variância.

**ARIMA:** Ver autorregressivo integrado de médias móveis.

**ARGUMENTO:** Raciocínio ou tipo especial de pensamento do qual, a partir de premissas, chega-se a uma conclusão.

**ALEATÓRIO:** Não-determinístico, ocorrendo puramente ao caso ou independente da ocorrência de outros eventos.

**ALEATORIZAÇÃO:** Tratamentos atribuídos aleatoriamente a unidades ou condições experimentais em um experimento. Isso é feito para reduzir a oportunidade para um tratamento ser favorecido ou desfavorecido (tendenciosidade) pelas condições de teste.

**AMOSTRA:** Qualquer subconjunto de elementos de uma população.

**AMOSTRA ALEATÓRIA:** Uma amostra é dita aleatória se ela for selecionada de tal maneira que cada amostra possível tenha a mesma probabilidade de ser selecionada.

**ALGORITMO:** Sequência de procedimentos e regras claras e específicas para a consecução de um objetivo. Lista passo a passo de instruções que, se executadas, levarão ao objetivo almejado. Sequência de instruções dos programas de computador.

**AMOSTRA:** Grupo de elementos ou sujeitos selecionados a partir de um grupo maior que é a população. Subconjunto da população em estudo. Quando a amostra é representativa dessa população, supõe-se que os dados obtidos por meio da amostra possam ser generalizados para a população da qual a mesma se originou. As amostras podem ser probabilísticas a qual é aquela em que todos os membros da população têm uma probabilidade conhecida de serem selecionados para a amostra. Estas podem ser amostras aleatórias simples, estratificadas, sistemáticas, por conglomerados ou por múltiplos estágios. Já as amostras não probabilísticas que são outros critérios, que não emprega a teoria da probabilidade, são utilizados para a determinação dos sujeitos ou elementos da amostra. Podem ser amostras intencionais, bola de neve, por conveniência ou por quotas. As amostragens probabilísticas possuem a vantagem de permitir a estimação da representatividade da amostra e uma melhor identificação de tendências da população. A amostragem não probabilística é menos precisa, contudo pode ser útil quando os custos e o trabalho envolvido no desenvolvimento de uma estrutura de amostragem são inconvenientes. Por exemplo, em estudos exploratórios, pré-testagem de questionários ou quando se trata de uma população sabidamente homogênea. Ver também amostragem.

**AMOSTRA DE MÚLTIPLOS ESTÁGIOS:** Tipo de amostra probabilística na qual várias etapas são realizadas até que se obtenha a amostra final. Digamos que a amostra necessária seja composta por integrantes de determinada região, por exemplo o Estado de Minas Gerais. Num primeiro estágio, desenvolver-se-á uma amostra de municípios na qual cada cidade possui determinada probabilidade, proporcional ao tamanho da sua população, de constar da amostra, chamada até aqui de amostra por conglomerados. Assim, os municípios maiores, como Belo Horizonte por exemplo, terão maiores chances de compor a amostra do que municípios menores. Num segundo estágio, obtém-se uma amostra por conglomerados dos bairros ou distritos de cada município. No terceiro estágio, seleciona-se uma amostra sistemática das

residências desses bairros e, finalmente, num quarto estágio, uma amostra aleatória de indivíduos obtida a partir da amostra anterior.

**AMOSTRA ESTRATIFICADA:** Quando ocorre a possibilidade de os sujeitos de determinada população serem subdivididos em estratos ou subclasses distintas, pode-se constituir uma amostra para cada uma das subclasses. Esse procedimento melhora a eficiência amostral, na medida em que as amostra estratificadas tendem a refletir melhor a realidade da população estudada, sob determinado ponto de vista. Por exemplo: deseja-se averiguar a opinião dos estudantes de determinada faculdade acerca do mercado de trabalho. Porém, esta faculdade possui dois cursos com características muito distintas: Agronomia e Medicina. Assim, decide-se por constituir duas amostras aleatórias simples separadas, a dos estudantes de Agronomia e a dos estudantes de Medicina. A amostra estratificada pode ser proporcional quando o número de sujeitos selecionados em cada grupo é proporcional ao número de integrantes da população ou não proporcional quando o número de sujeitos selecionados em cada grupo não é proporcional ao número de integrantes da população.

**AMOSTRAGEM:** Procedimento estatístico por meio do qual se constitui uma amostra a partir de uma população. Na maior parte das pesquisas é absolutamente inviável estudar todos os sujeitos de uma população: ao invés disso, o pesquisador estuda apenas uma amostra, constituída por um número de sujeitos bem menor que a população. Ver amostra para uma visão geral dos diversos procedimentos de amostragem.

**AMOSTRA INTENCIONAL:** Amostra na qual os sujeitos foram selecionados não probabilisticamente, mas de acordo com os critérios de um perito. Exemplo: numa pesquisa de marketing, um estilista utiliza seu julgamento para identificar os sujeitos representativos que devem compor a amostra, dentre os fregueses de uma loja de roupas. Pressupõe-se, portanto, que o pesquisador conheça de antemão as características relevantes que um sujeito deve possuir para integrar-se na amostra. O risco deste tipo de amostragem é o possível excesso de interferência do viés subjetivo do pesquisador.

**AMOSTRA POR CONGLOMERADOS:** Tipo de amostra probabilística constituída em duas ou mais etapas. É utilizada tipicamente quando o pesquisador não tem acesso a uma lista completa dos integrantes de uma população, mas sim de subgrupos chamados conglomerados dessa mesma população. Exemplo: deseja-se pesquisar as características dos estudantes do ensino médio particular na cidade de São Paulo. A lista desses estudantes não está disponível, mas, por outro lado, é possível obter-se uma lista das escolas particulares de ensino médio que são conglomerados de estudantes. Numa primeira etapa são sorteadas cinco escolas para participar do estudo. Conseguindo-se a lista de alunos dessas escolas, são sorteados então 100 estudantes de cada uma para comporem a amostra final, a qual é a segunda etapa da amostragem. Esta última leva o nome de amostra por conglomerados.

**AMOSTRA POR CONVENIÊNCIA:** Tipo de amostra não probabilística na qual os sujeitos são escolhidos para comporem a amostra de acordo com a conveniência ou facilidade do pesquisador. Este tipo de amostra deve ser utilizado com muita reserva, devido às suas deficiências inferenciais. Pode ser utilizada na pré-testagem de questionários.

**AMOSTRA POR QUOTAS:** Tipo de amostra não probabilística intencional na qual estão presentes restrições ou parâmetros preestabelecidos para o número de sujeitos de cada amostra. Por exemplo, no planejamento da amostragem, pode-se definir que uma amostra deverá conter apenas homens com idade entre 25 e 30 anos, outra amostra com mulheres acima de 20 anos dentre outros critérios. Como toda amostra não probabilística, apresenta deficiências inferenciais.

**AMOSTRA PROPOSITAL:** Ver amostra intencional.

**AMOSTRA RANDÔMICA:** Ver amostra aleatória simples.

**AMOSTRA SISTEMÁTICA:** Tipo de amostra probabilística na qual a expansão da mesma se dá sistematicamente por meio de algum critério posicional aplicado à lista original de integrantes da população. Exemplo: numa população de 1.000 funcionários de uma empresa, deseja-se extrair 100 sujeitos para a amostra. Esta seleção é realizada gerando-se uma lista aleatória dos funcionários da empresa e selecionando-se cada décimo indivíduo da lista. Alternativamente, pode-se arranjar a lista populacional de acordo com algum critério previamente estabelecido e que tenha relação com as variáveis pesquisadas, como, por exemplo, receita gerada por cada funcionário, salário, nível de escolaridade dentre outros.

**AMOSTRA:** É um conjunto menor de unidades retiradas da população, ou seja, um subconjunto da população. O desejável é que amostra represente bem a população pois, a partir dela a estatística inferencial tira conclusões sobre como deve ser a população. Entretanto, nem sempre é fácil garantir esta representatividade por uma série de fatores. O ideal é que a amostra seja casual(aleatória) para garantir que as unidades tenham igual chance de serem selecionadas e evitar possíveis vícios nas inferências que serão feitas, mas em muitas situações não é possível encontrar uma amostra casual.

**ÁREA ÚTIL:** É a porção da unidade experimental efetivamente utilizada na avaliação do tratamento.

**ATIVIDADES (PROCEDIMENTOS) PARA MINIMIZAÇÃO DOS EFEITOS DO ERRO EXPERIMENTAL:** Conjuntos de técnicas de refinamento usados na pesquisa científica que deve ser empregada pelo pesquisador para diminuir os erros mas nunca eliminá-los, os quais são: i) uniformização das parcelas experimentais; ii) parcela experimental de tamanho adequado; iii) uso de bordaduras; iv) utilização de um número adequado de repetições e, de preferência, igual para todos os tratamentos; v) manejo das unidades experimentais de forma homogênea, no que diz respeito a todos os fatores não envolvidos no estudo e vi) uso do delineamento estatístico adequado para as condições de realização do experimento, de forma a obter o melhor aproveitamento dos resultados.

**ARQUIVOS SIMPLES (FLAT FILES):** Arquivos de computador que contêm dados organizados em forma de tabela com apenas 2 dimensões que são as linhas e colunas. Cada dado é colocado em uma célula diferente.

**ÁRVORE DE CLASSIFICAÇÃO (CLASSIFICATION TREE):** Um modelo ou exibição visual de dados categóricos que é construído dividindo o conjunto de dados em grupos sucessivamente menores. A divisão continua até que nenhuma melhora no ajuste do modelo seja obtida. Uma árvore de classificação é análoga a uma chave taxonômica, exceto que as divisões em uma árvore de classificação refletem probabilidades e não identidades. Ver também diagrama de agrupamento, impureza e árvore de regressão.

**ÁRVORE DE REGRESSÃO (REGRESSION TREE):** Um modelo, ou representação visual, de dados contínuos que é construído por meio da divisão do conjunto de dados em grupos sucessivamente menores. A divisão continua até que nenhuma melhora no ajuste do modelo seja obtida. Uma árvore de regressão é análoga a uma chave taxonômica, exceto que as divisões em uma árvore de regressão refletem probabilidades, e não identidades. Ver também árvore de classificação; diagrama de agrupamento.

**ÁRVORE LÓGICA (LOGIC TREE):** Um exemplo do método hipotético-dedutivo. O investigado trabalho ao longo de uma árvore lógica escolhendo entre uma de duas alternativas a cada ponto de decisão. A árvore

vai do geral, por exemplo, animal ou vegetal, para o específico, por exemplo, 4 dedos ou 3?. Quando não restar mais escolhas, a solução foi obtida. Uma chave dicotômica é um exemplo de árvore lógica.

**ASSIMETRIA (SKEWNESS):** A descrição de quão longe uma distribuição está de ser simétrica. Abreviada como  $g_1$  e calculada a partir do terceiro momento central. Ver também assimetria à direita e assimetria à esquerda.

**ASSIMETRIA À DIREITA (RIGHT-SKEWED):** Uma distribuição que tem a maioria de suas observações inferiores à média aritmética, mas uma longa cauda de observações superiores à média aritmética. A assimetria de uma distribuição com assimetria à direita é maior que 0. Comparar com assimetria à esquerda.

**AMPLITUDE:** Ver dispersão.

**ANÁLISE:** Decomposição de um todo em suas subpartes. Estudo acerca de como as partes interagem para formar um todo; interpretação das partes para explicar o todo.

**ANÁLISE DE CONTEÚDO:** Conjunto de técnicas de investigação científicas utilizadas em ciências humanas, caracterizadas pela análise de dados linguísticos: tudo o que é dito ou escrito é suscetível de ser submetido a uma análise de conteúdo Normalmente, neste tipo de análise, os elementos fundamentais da comunicação são identificados, numerados e categorizados. Posteriormente, as categorias encontradas são analisadas face a uma teoria específica.

**ANÁLISE DE CORRELAÇÃO:** Ver correlação e coeficiente de correlação.

**ANÁLISE DE COVARIÂNCIA:** Técnica estatística na qual os efeitos de uma ou mais variáveis exógenas e independentes, são eliminados para que se possa estudar a variável dependente com maior precisão. Trata-se de um caso particular da análise de variância. A ancova é utilizada com muita frequência quando a variável dependente é muito complexa, como por exemplo, uma variável sociológica indicativa de qualidade de vida.

**ANÁLISE DE REGRESSÃO:** Ver regressão.

**ANÁLISES DE AGRUPAMENTOS:** A Análise de Agrupamento consiste na formação de grupos de indivíduos ou objetos tomando como base medidas de similaridade ou distância entre casos, que, por sua vez, são calculadas com base em uma lista de variáveis fornecidas.

**ASSIMETRIA À ESQUERDA (LEFT-SKEWED):** Uma distribuição que tem a maioria de suas observações superiores à média aritmética, mas uma longa cauda de observações inferiores à média aritmética. A assimetria de uma distribuição com assimetria à esquerda é inferior a 0. Comparar com assimetria à direita.

**ASSOCIAÇÃO (ASSOCIATIVE):** Em matemática e probabilidade é a lei, propriedade ou regra de que  $(a + b) + c = a + (b + c)$ .

**AUTORREGRESSIVO (AUTOREGRESSIVE):** Um modelo estatístico em que a variável resposta (Y) depende de um ou mais valores prévios. Muitos modelos do crescimento populacional são autorregressivos, como, por exemplo, aqueles nos quais o tamanho da população no tempo  $t + 1$  é dependente do da população no tempo  $t$ .

**AUTORREGRESSIVO INTEGRADO DE MÉDIAS MÓVEIS (ARIMA- AUTOREGRESSIVE INTEGRATED MOVING AVERAGE):** Um modelo estatístico usado para analisar dados de séries temporais



que incorpora tanto a dependência temporal nos dados quanto os parâmetros que estimam as mudanças devido aos tratamentos experimentais. Ver autorregressivo.

**AUTOVALOR (EIGENVALUE):** A solução ou as soluções para um conjunto de equações lineares de forma  $Ax = \lambda x$ . Nesta equação,  $A$  é uma matriz quadrada e  $x$  é um vetor coluna. O autovalor é a escalar, ou um único número,  $\lambda$ . Comparar com autovetor.

**AUTOVETOR (EIGENVECTOR):** O vetor coluna  $x$  que resolve o sistema de equações lineares  $Ax = \lambda x$ . Comparar com autovalor.

**AXIOMA (AXIOM):** Um princípio matemático estabelecido que é universalmente aceito, verdadeiro por autoevidência e não necessita de provas formais.

**AGRUPAMENTO PELA VARIÂNCIA MÍNIMA:** A análise de agrupamento consiste na formação de grupos de indivíduos ou objetos, tomando como base medidas de similaridade ou distância entre casos, que, por sua vez, são calculadas com base em uma lista de variáveis fornecidas. Inicialmente, cada caso constitui um grupo independente, e no agrupamento aglomerativo ou todos os casos formam inicialmente um único grupo, no agrupamento divisivo.

**ANÁLISE DE CONFIABILIDADE:** A análise de confiabilidade é verificação da capacidade de um produto desenvolver satisfatoriamente sua função predefinida sob as condições de funcionamento e sob um período de tempo predefinido. Portanto para o uso da análise de confiabilidade é necessário a existência de uma variável que represente o tempo e, pelo menos, uma variável que represente a taxa ou quantidade de falhas.

**ANÁLISE DE DEVIANCE OU DESVIO (ANADEV):** Tipo de quadro de análise para modelos mistos, onde os efeitos aleatórios são testados através do teste da razão de verossimilhança (LTR), em que um quadro semelhante ao quadro da análise de variância pode ser elaborado. Este quadro pode ser chamado de análise de deviance (ANADEV) e é estabelecido conforme os seguintes passos: i) obtenção do ponto de máximo do logaritmo da função de verossimilhança residual (LOG L) para modelos com e sem o efeito a ser testado; ii) obtenção da deviance  $D = -2 \log L$  para modelos com e sem o efeito a ser testado; iii) fazer a diferença entre as deviances para modelos sem e com o efeito a ser testado, obtendo a razão de verossimilhança (LR); iv) testar, via LTR, a significância dessa diferença usando o teste do qui-quadrado com 1 grau de liberdade, pois estas estimativas tendem a ter este tipo de distribuição teórica de probabilidade. Vale lembrar que na análise de modelos mistos, os efeitos do modelo não testados via teste F, assim como se faz no método da análise de variância (ANAVA).

**ANÁLISE FATORIAL:** A análise fatorial tem como objetivo principal explicar a correlação ou covariância, entre um conjunto de variáveis, em termos de um número limitado de variáveis não-observáveis. Essas variáveis não-observáveis ou fatores são calculados pela combinação linear das variáveis originais. Frequentemente um pequeno número de fatores pode ser usado, em lugar das variáveis originais, nas análises de regressões, análises de agrupamentos dentre outras. Os fatores são extraídos na ordem do mais explicativo para o menos explicativo. Teoricamente, o número de fatores é sempre igual ao número de variáveis. Entretanto, alguns poucos fatores são responsáveis por grande parte da explicação total. Na análise de fatores procura-se por aptidão das observações.

**ANÁLISE DE SOBREVIVÊNCIA:** Tipo de análise estatística em pesquisa que visa determinar quanto tempo os sujeitos de uma amostra permanecem em determinado estado ou condição. A análise de

sobrevivência é muito utilizada nas pesquisas da área médica que buscam avaliar a duração das doenças para comparar, por exemplo, a eficácia de dois tratamentos.

**ANÁLISE DE VARIÂNCIA (ANOVA):** Técnica estatística que tem por objetivo testar a hipótese de que as médias amostrais são diferentes, mediante determinado nível de significância. É utilizada quando existe três ou mais grupos amostrais a serem comparados. Por exemplo: uma pesquisa na área de treinamento de pessoal deseja determinar o grau de eficiência de duas diferentes modalidades de ensino: presencial e à distância. Formam-se três grupos, com 50 sujeitos em cada um. Realiza-se uma prova de conhecimentos sobre o assunto, antes de os sujeitos receberem os treinamentos que é o pré-teste. O primeiro grupo o qual é o grupo experimental 1 submeter-se-á ao treinamento presencial tradicional; o segundo grupo receberá o mesmo treinamento à distância que é o grupo experimental 2 e o terceiro grupo não se submete a nenhum treinamento, ou seja, grupo controle. Ao final dos treinamentos, que duram digamos, três semanas, todos os sujeitos se submetem novamente a uma prova de conhecimento o qual denomina-se pós-teste. O pesquisador então efetua a média das notas dos grupos nos pré e pós-testes.

**ANÁLISE DISCRIMINANTE:** Técnica estatística de análise que tem por objetivo a classificação de dados. Trata-se de uma forma específica de regressão, na qual a variável dependente possui nível de mensuração nominal e as variáveis independentes, nível intervalar ou racional. Por exemplo, em determinada pesquisa, deseja-se discriminar os pacientes de um ambulatório de saúde em relação à probabilidade de que venham a desenvolver determinada doença cuja variável dependente é nominal: adoecimento eminente [sim/não]), em relação a certos fatores tais como idade, pressão arterial, taxa de triglicérides na corrente sanguínea que são variáveis independentes, dentre outros.

**ANÁLISE DO DISCURSO:** Tipo específico de análise de conteúdo cujos procedimentos têm por objetivo inferir acerca da estrutura profunda como por exemplo, aspectos psicológicos do autor do discurso a partir dos elementos explícito-discursivos de um texto. Para outros autores, a análise do discurso é distinta da análise de conteúdo na medida em que aquela primeira envolve a reflexão acerca das condições sócio-históricas de produção dos textos analisados, visando à análise crítica da articulação da língua com a ideologia. Desta forma, a análise do discurso baseia-se no pressuposto de que o sentido de um texto sempre explicita posições sociais ou ideológicas, que devem ser reveladas por meio da própria análise.

**ANÁLISE FATORIAL:** Termo genérico que designa um conjunto de técnicas estatísticas que têm por objetivo identificar algumas poucas dimensões ou fatores que expliquem determinado fenômeno. Por exemplo, pode-se realizar uma análise fatorial de um questionário com 100 perguntas sobre hábitos de consumo e concluir-se que apenas seis grandes fatores podem ser utilizados para a análise do fenômeno do comportamento do consumidor. Desta forma, a análise fatorial tornar-se uma ferramenta importante para simplificar e sumarizar dados.

**ANÁLISE DE SOBREVIVÊNCIA:** A análise de sobrevivência é constituída de um conjunto de técnicas usadas na avaliação do comportamento de variáveis aleatórias positivas como: tempo de vida, efeito de doenças e de remédios, tempo de duração de produtos dentre outras. Quando se trata de avaliação de qualidade de produtos industriais o mais apropriado é identificar a análise como sendo de qualidade, confiabilidade dentre outras.

**ANCOVA:** Siga em inglês de Analysis of Covariance. Ver análise de covariância.

**ANOVA:** Sigla em inglês de Analysis of Variance. Ver análise de variância.

**A POSTERIORI:** Termo latino que significa que vem depois. Diz-se das conclusões derivadas de dados empíricos, ou seja, a partir de observações.

**A PRIORI:** Termo latino que significa que vem antes, antes de, em princípio. Utilizado na área da pesquisa para referir-se a conclusões alcançadas tomando-se por base pressuposições e não dados empíricos. Isso também ocorre com relação à distribuição.

**AXIOMA:** Princípio ou proposição fundamental necessária, auto-evidente e não demonstrável, largamente aceita como verdadeira e a partir da qual se constrói toda uma cadeia de raciocínio. Frequente na Matemática, por exemplo, todo número possui um sucessor ou então por dois pontos distintos passa uma e somente uma reta, não é muito comum nas Ciências Sociais. Os axiomas não podem ser considerados necessariamente verdadeiros, mas apenas pressupostos convenientes para a construção de um sistema teórico.

**AMPLITUDE:** O maior valor menos o menor valor de um conjunto de dados. A amplitude é uma medida simples de variabilidade e é largamente usada em controle de qualidade.

**AMPLITUDE DA AMOSTRA:** Veja amplitude.

**AMPLITUDE DE INTERQUARTIL:** A diferença entre o terceiro e o primeiro quartis de uma amostra de dados. A amplitude interquartil é menos sensível a valores extremos que a amplitude usual da amostra.

**AMPLITUDE MÓVEL:** O valor absoluto da diferença entre observações sucessivas nos dados ordenados no tempo. Usada para estimar a variação da chance em um gráfico individual de controle.

**AMPLITUDE NA FORMA DE STUDENT:** A amplitude de uma amostra dividida pelo desvio-padrão da amostra, a qual é uma estatística adimensional.

**ANÁLISE DE VARIÂNCIA (ANOVA):** Um método de decompor a variabilidade total de um conjunto de observações, medida como a soma dos quadrados das diferenças dessas observações em relação à sua média, em uma soma dos quadrados dos componentes que estão associados com fontes específicas e definidas de variação.

**ANÁLISE (OU GRÁFICOS) RESIDUAL:** Qualquer técnica que usa os resíduos, geralmente para investigar a adequação do modelo que foi usado para gerar os resíduos, mas também pontos discrepantes e distribucional.

**APROXIMAÇÃO NORMAL:** Um método para aproximar probabilidades para variáveis aleatórias binomial e de Poisson.

**ASSIMETRIA:** Um termo geralmente empregado com relação a um histograma de dados ou a uma distribuição de probabilidades.

**ATRIBUTO:** Uma característica qualitativa de um item ou unidade, que aparece geralmente em controle de qualidade. Por exemplo, classificar unidades de produção como defeituosas ou não-defeituosas resulta em dados de atributos.

**AXIOMAS DA PROBABILIDADE:** Um conjunto de regras as quais devem ser seguidas pelas probabilidades definidas em um espaço amostral. Veja probabilidade.

**AÇÃO SEMELHANTE:** De dois estímulos, quando misturados, é a que se caracteriza pelo fato de o efeito da mistura poder ser previsto a partir da ação dos componentes isolados, em termos de doses equivalentes.

**AÇÃO SINÉRGICA:** É dita a de dois ou mais estímulos se, misturados, a reação resultante é maior ou menor do que a soma das reações dos estímulos aplicados separadamente.

**ADESÃO (OU ADERÊNCIA):** Em conformidade com as instruções. Por exemplo, um passo intermediário na avaliação de qualquer intervenção consiste em comprovar se realmente a intervenção foi executada nos moldes propostos: ou seja, se os participantes seguiram as recomendações na forma como foram prescritas.

**ADIÇÃO FORWARD:** Método de seleção de variáveis para inclusão no modelo de regressão, que começa sem qualquer variável no modelo e então acrescenta variáveis com base em sua contribuição na previsão. Poderíamos traduzir como inclusão em avanço, mas a expressão em inglês é usual em nosso idioma.

**ADITIVIDADE:** Propriedade usada em estatística pela qual diversos fatores podem ser somados, ou são considerados por meio de soma. Em inglês Additivity.

**AFASTAMENTO:** É o mesmo que desvio.

**AGREGAÇÃO:** Palavra usada para denotar a mistura de dados originais em agregados, usualmente com o propósito de expressá-los de forma resumida. Do inglês Aggregation.

**AGRUPAMENTO SUBJETIVO:** Ver dados de confusão.

**AGRUPAMENTO DE AMOSTRAS:** Refere-se ao resultado de uma amostragem em que existe uma concentração de amostras em uma determinada região em relação área total amostrada, devido a problemas inerentes da amostragem, como por exemplo, a existência de vias de acesso, afloramentos, dentre outros.

**AIC (CRITÉRIO DE INFORMAÇÃO DE AKAIKE):** Critério para seleção de modelos baseado na utilização da razão de verossimilhança penalizada pelos graus de liberdade. Do inglês AIC (Akaike Information Criteria).

**AI:** Algoritmo de informação média.

**AI-REML:** REML obtido via algoritmo AI.

**AJUSTAMENTO DE CURVAS:** Teste de regressão destinado a verificar o modelo de dependência entre duas variáveis: linear, exponencial, logarítmica ou geométrica.

**AJUSTE DE BONFERRONI PARA ALFA:** Método para ajustar o valor de alfa quando múltiplas hipóteses estão sendo testados. Para que o risco de um achado falso-positivo, no estudo como um todo, não ultrapasse o valor de alfa o qual é geralmente igual a 0,05, o nível de alfa escolhido para rejeitar a hipótese nula é tornado mais restrito por meio da divisão do valor alfa pelo número de hipóteses que estão sendo testadas.

**AJUSTE DE CURVA:** Expressão usada em dois sentidos diferentes: primeiro para denotar o ajuste a uma curva de frequência especificada matematicamente a uma distribuição de frequência; e segundo para denotar o ajuste de uma curva matemática a qualquer conjunto de dados possíveis de serem plotadas contra o tempo ou espaço. Em inglês Curve fitting.

**AJUSTE DE UMA RETA:** Por regressão linear simples, consiste em encontrar a equação de uma reta que descreve a dependência entre uma variável aleatória  $Y$  em função de outra variável aleatória  $X$ . Geralmente a variável  $Y$  é denominada variável dependente e a variável  $X$  independente. A equação da reta tem a forma:  $Y = a + bX$ , onde  $a$  e  $b$  são os coeficiente da reta, determinados pelo método dos mínimos quadrados. Vide regressão linear simples.

**AJUSTE EXPONENCIAL:** Método de análise de tendência em série temporal.

**AJUSTE SAZONAL:** Processo de ajuste de séries temporais para compensar a variação sazonal, para tornar as variações não-sazonais mais visíveis nos dados.

**ÁLEA:** Sorte, risco ou acaso. Em inglês alea.

**ALEATÓRIO:** Diz-se que um fenômeno ou acontecimento é aleatório, casual ou estocástico se for não determinístico, isto é, se os fatores responsáveis por sua ocorrência não são determinados ou são numerosos a ponto de não se poder identificar o principal responsável. Em inglês random.

**ALEATÓRIO:** Escolhido ao acaso; o mesmo que casual; diz-se da escolha, ao acaso, das unidades que deverão compor uma amostra.

**ALEATÓRIO:** Termo usado na estatística o qual significa o evento que ocorre ao acaso, ou seja, diz-se da variável que assume valores segundo uma determinada lei de probabilidades. Por exemplo, os resultados de um jogo de dados são aleatórios. Quando é determinado por um complexo de numerosas causas somadas, mas cujas atuações individuais desconhecemos. Por exemplo, erro aleatório. Diz-se do processo construído para que cada resultado possível esteja associado a uma probabilidade conhecida. Por exemplo, em um experimento, os tratamentos são designados aos pacientes por processo aleatório.

**ALEATORIEDADE:** A designação aleatória de grupos de indivíduos para condições de tratamento a fim de assegurar representação igual de características em todos os grupos.

**ALEATORIZAÇÃO:** Uma série de objetos é considerado aleatório quando arranjados em uma ordem aleatória, e, estendendo o conceito, uma série de tratamentos aplicados para uma série de unidades é considerado aleatório quando o tratamento aplicado para qualquer unidade dada é escolhido aleatoriamente daqueles disponíveis ou ainda não alocados. Em inglês Randomization.

**ALEATORIZAÇÃO:** O mesmo que casualização e randomização.

**ALFA ( $\alpha$ ):** É a probabilidade de rejeitar  $H_0$ , a qual é a hipótese básica, nula ou de nulidade, quando ela é verdadeira. Neste caso, seria a probabilidade de afirmar que as atitudes em relação à Estatística, quando na realidade não existe nenhuma relação. Este erro é controlado pelo pesquisador, e é ele que define a margem de erro que está disposto a correr. Existem vários fatores que influenciam na escolha do nível de significância. Em pesquisas, como nas ciências exatas, biológicas, agrônomicas, onde as variáveis são mais fáceis de mensurar, onde os instrumentos de medida são confiáveis, onde o controle de fatores intervenientes é razoável, o conhecimento da área é maior, a gravidade das consequências do erro menor, entre outros, permitem um maior rigor e, portanto, pode-se ser mais exigente, diminuindo o nível de significância. Contudo, em pesquisas, nas ciências humanas, que lida com pessoas, com construtos polêmicos, instrumentos ainda não testados, as consequências do erro não são tão graves, entre outros, pode-se ser mais flexível. Via de regra, usa-se o nível de 5%.

**ALFA ( $\alpha$ ):** Nível de significância associado ao teste estatístico das diferenças entre dois ou mais grupos. Normalmente, valores pequenos, como 0,05 ou 0,01, são especificados para minimizar a possibilidade de se cometer um erro tipo I, também conhecido como erro de primeira espécie.

**ALFA:** É a probabilidade de se cometer erro do tipo I, ou seja, rejeitar a hipótese nula quando ela é verdadeira no teste de hipóteses. Em inglês alpha.

**ALFA:** Nível de significância estatística. Ver erro do tipo I.

**ALFA:** É a probabilidade de rejeitar  $H_0$ , quando ela é verdadeira. Neste caso, seria a probabilidade de afirmar que as atitudes em relação à matemática interferem na formação das atitudes em relação à estatística, quando na realidade não existe nenhuma relação.

**ALFA ( $\alpha$ ):** Símbolo usado para representar a probabilidade de um erro tipo I.

**ALFA ( $\alpha$ ):** Ver erro tipo I.

**ALFA DE CRONBACH:** Medida de confiabilidade que varia de 0 a 1, sendo os valores de 0,60 a 0,70 considerados o limite inferior de aceitabilidade.

**ALFA DE CRONBACH:** Medida comumente usada de confiabilidade para um conjunto de dois ou mais indicadores de construto. Os valores variam entre 0 e 1, com as medidas mais altas indicando maior confiabilidade entre os indicadores.

**ALGARISMOS SIGNIFICATIVOS:** Algarismos numa medida que representam as posições do número que pode ser uma unidade, dezena, centena, milhar, dentre outros, que se conhece com certeza absoluta, isto é, sem possibilidade de variação subjetiva. Ao trabalharmos com medidas é importante manter em mente que uma mesma medida deve ser expressa com igual precisão independentemente do múltiplo ou fração da unidade escolhida para expressá-la. Não se deve fazer uma medida com mais algarismos significativos que a precisão do instrumento de mensuração ou além da quantidade de algarismos significativos que o processo de mensuração pode gerar com confiabilidade.

**ALGARISMOS SIGNIFICATIVOS:** Indicam a precisão, pelo modo de ler e escrever um resultado, como segue: i) o algarismo não nulo, mais à esquerda, é o algarismo mais significativo; não havendo vírgula, o último algarismo significativo, à direita, é o algarismo menos significativo; ii) se houver vírgula, o último algarismo à direita é o algarismo menos significativo, mesmo que seja zero; iii) todos os algarismos entre o mais significativo e o menos significativo são contados como algarismos significativos.

**ÁLGEBRA:** Parte da matemática que não trata especificamente com geometria. Esta é uma definição antiga. Hoje a álgebra pode ser dividida em clássica e moderna. A álgebra clássica é a parte que trata de manipulação simbólica e da solução de equações algébricas. A álgebra moderna ou abstrata lida com um ramo da matemática conhecido como estruturas discretas que são corpos, grupos, anéis, dentre outros. Em inglês algebra.

**ÁLGEBRA DE BOOLE:** tanto na ciência matemática quanto na ciência computacional uma álgebra booleana ou de boole ou ainda um reticulado ou lattice booleano é uma estrutura algébrica que lida com as operações lógicas e, ou e não de mesma forma que lida com as operações sobre conjuntos correspondentes união, intersecção e complementação. esta estrutura foi nomeada em homenagem ao matemático



inglês george boole. a álgebra booleana é uma tentativa de se utilizar técnicas algébricas para lidar com expressões do cálculo proposicional. Em inglês boolean algebra.

**ALGORITMO:** Um algoritmo é um conjunto definido de operações e passos ou procedimentos que objetivam levar a um particular resultado. Por exemplo, com algumas exceções, os programas computacionais, as fórmulas matemáticas e de forma ideal receitas médicas e culinárias são algoritmos. Em inglês algorithm.

**ALGORITMO:** Termo definido há algum tempo atrás como sendo uma palavra equivalente à fórmula. Recentemente, significa uma sequência de passos necessários para atingir um objetivo. Do inglês algorithm.

**ALGORITMO:** Conjunto de regras ou procedimentos; é semelhante a uma equação.

**ALGORITMO:** Procedimento computacional aplicado à solução de um problema numérico. sequência de passos de uma operação que leva à solução de uma equação discreta.

**ALGORITMO DE BACK-FITTING:** Algoritmo de ajuste de modelo usado quando existem várias covariáveis, começando pelo modelo completo e eliminando covariáveis até conseguir o melhor ajuste. Utilizado para ajustar um modelo linear generalizado. Do inglês back-fitting algorithm.

**ALGORITMO DE METROPOLIS-HASTINGS:** Algoritmo de Monte Carlo em cadeia de Markov (MCMC) que baseia-se na utilização de uma distribuição auxiliar. A idéia é gerar amostras da distribuição auxiliar, chamada distribuição proposta e aceitá-las como sendo da distribuição de interesse com uma certa probabilidade. Em inglês metropolis-hastings algorithm.

**ALGORITMO DE NEWTON E RAPHSON:** Algoritmo iterativo usado para obter as raízes de uma função diferencial. Este algoritmo pode ser adaptado para a obtenção de estimadores de máxima verossimilhança ou a moda de uma distribuição, entre outras utilizações estatísticas. Em inglês Newton-Raphson algorithm.

**ALGORITMO DO AMOSTRADOR DE GIBBS:** Método de simulação estocástica que gera amostras de um vetor de parâmetros a partir das distribuições marginais completas. Em inglês Gibbs sampler algorithm.

**ALGORITMO EM:** Método geral útil para obter estimadores de máxima verossimilhança quando existem dados incompletos ou quando é mais simples maximizar a verossimilhança de um problema mais geral que envolve quantidades não observáveis. O algoritmo é iterativo e em cada iteração se alternam as operações de esperança e maximização. Em inglês EM algorithm.

**ALGORITMO GENÉTICO:** Modelos de aprendizado desenvolvidos sobre os princípios da evolução. Soluções parciais para um problema competir com outro e então as melhores soluções são selecionadas e combinadas para fornecer a base para posteriores soluções de outros problemas.

**ALIZAMENTO:** O processo de remover flutuações em uma série ordenada para que o resultado seja suave no sentido em que as primeiras diferenças são regulares e diferenças de ordem mais elevadas, menores. Em inglês smoothing.

**ALOCAÇÃO:** Processo de alocar ou designar um tratamento a uma unidade experimental.

**ALOCAÇÃO DE PARCELAS (AMOSTRAGEM ESTRATIFICADA):** Refere-se à distribuição das unidades amostrais nos diferentes estratos de forma a garantir a precisão da amostragem. Pode ser feita de diferentes formas: i) alocação proporcional, em que as unidades amostrais são distribuídas proporcionalmente ao tamanho dos estratos; ii) alocação uniforme, em que utiliza-se o mesmo tamanho de amostra em todos

os estratos; iii) alocação ótima de Neyman, em que as parcelas são distribuídas de forma a minimizar o custo, para uma precisão fixa, ou maximizar a precisão, para um custo fixo.

**ALOCAÇÃO DE PARCELAS NO CAMPO:** Ver locação de parcelas.

**ALOCAÇÃO DESPROPORCIONAL OU IDEAL:** Uma amostragem na qual o número de elementos extraídos de determinado extrato é proporcional ao tamanho relativo do extrato e ao desvio-padrão da característica sob consideração.

**ALOCAÇÃO ESTRATIFICADA:** Designação dos pacientes para diferentes grupos de risco baseada em variáveis como a gravidade da doença como por exemplo, estágios de um câncer)e idade.

**ALOCAÇÃO PROPORCIONAL:** Amostragem na qual o número de elementos selecionados de um extrato é diretamente proporcional ao tamanho.

**ALTURA DO PEITO:** Altura, convencionalmente em 1,30 metros, para se medir o diâmetro de árvores (DAP). As razões para a escolha dessa altura foram: padronização de modo que não dependa da altura do observador; a essa altura a posição de medição é cômoda; nessa altura o tronco geralmente não apresenta mais influência de dilatações na base ou sapopemas. Embora seja uma altura convencional, em países que não adotam o sistema métrico a altura fica um pouco diferente por exemplo, Estados Unidos da América 1,35 m; Inglaterra 1,29 m; Japão 1,25 m, podendo acarretar pequenas diferenças no cálculo de área transversal, dificultando a comparação entre trabalhos. Ver também diâmetro à altura do peito (DAP).

**ÂMBITO:** O âmbito é a diferença entre os dois quantis ou entre máximo e o mínimo.

**AMOSTRA:** Uma parte finita e não vazia extraída de uma população. Em inglês sample.

**AMOSTRA:** Conjunto de observações.

**AMOSTRA:** Subconjunto ou parte da população constituída de indivíduos que são analisados a fim de se fazer uma inferência sobre a população estatística.

**AMOSTRA:** Subconjunto da população de interesse.

**AMOSTRA:** É um subconjunto finito da população que se supõe representativo desta.

**AMOSTRA:** É uma parcela representativa da população que é examinada com o propósito de tirarmos conclusões sobre a essa população.

**AMOSTRA:** Qualquer conjunto cujas características ou propriedade são estudadas com o objetivo de estendê-las a outro conjunto, do qual o primeiro conjunto é considerado parte.

**AMOSTRA:** Conjunto de respondentes selecionados para estudo de forma a garantir que o que for aprendido sobre os integrantes da amostra também é verdade para a população da qual foram selecionados.

**AMOSTRA:** É todo conjunto cujas propriedades se estudam com o fim de generalizá-las a outro conjunto de que aquele é considerado parte.

**AMOSTRA:** Subconjunto de uma população a partir do qual são estimadas as propriedades e características dessa população.

**AMOSTRA:** i) é uma quantidade representativa do universo que se deseja amostrar. O método de retirada da amostra deve garantir que ela seja representativa desse universo, no que diz respeito ao(s) parâmetro(s) de interesse. Normalmente uma amostra deve ser composta pelo maior número possível de incrementos. A representatividade está ligada ao conceito de heterogeneidade ou homogeneidade, já que em um corpo totalmente homogêneo, qualquer fragmento é representativo do conjunto. Ao contrário, num corpo heterogêneo, um fragmento não representa nem seu vizinho; ii) estatisticamente é um número de observações tomado de um conjunto de todas as observações possíveis ou população.

**AMOSTRA:** Subconjunto da população. Pode ser de conveniência não-aleatória ou aleatória.

**AMOSTRA:** É um conjunto de um certo número  $n$  de elementos que é tamanho da amostra de variáveis aleatórias independentes  $X_i$  com uma densidade de distribuição comum  $F$ .

**AMOSTRA:** Parte da população que constitui um conjunto finito de valores.

**AMOSTRA:** Qualquer subconjunto de elementos de uma população finita ou infinita. Em inglês *sample*.

**AMOSTRA:** Conjunto de elementos extraídos de um conjunto maior, chamado população. Se constituição da amostra obedecer a determinadas condições, a análise das características da amostra pode servir para que se façam inferências sobre elementos da população.

**AMOSTRA ALEATÓRIA:** Amostra dos elementos de uma população extraída de tal forma que cada elemento tenha a mesma probabilidade de figurar na amostra. Sinônimo de amostra ao acaso, randômica, não viesada, equiprovável.

**AMOSTRA CONTROLADA:** Amostra extraída de tal maneira que certas características existam na mesma proporção tanto na amostra quanto população.

**AMOSTRA REPRESENTATIVA:** Amostra não tendenciosa e suficientemente ampla de modo a possuir as mesmas características da população de que foi extraída.

**AMOSTRA TENDENCIOSA:** Amostra viciada por diferenças que impedem que seja considerada representativa da população de que foi extraída, no que se refere a uma ou mais características.

**AMOSTRAGEM:** Operação pela qual se constitui um grupo de elementos cuja análise ou estudo dará informações sobre as características da população em que elementos foram retirados.

**AMOSTRAGEM CASUAL SIMPLES:** Consiste em enumerar os elementos de uma população e escolher os  $n$  elementos dessa sequência, que compõem a amostra por meio de um dispositivo aleatório qualquer, como a tabela de Tipet ou a tabela de números aleatórios.

**AMOSTRAGEM MÚLTIPLA:** É aquela em que a amostra é retirada em diversas etapas sucessivas. Em função dos resultados obtidos em cada etapa, pode-se saber se são necessárias outras etapas ou se elas são dispensadas.

**AMOSTRAGEM POLIETÁPICA:** É aquela feita por etapas, sob a forma de degraus decrescentes escolhidos ao acaso da população. Tal sistema, embora econômico e na maioria das vezes eficiente, pode apresentar pouca representatividade, uma vez que os degraus são escolhidos ao acaso.

**AMOSTRAGEM RADIAL:** Consiste em escolher, ao acaso, elementos da população, contudo efetuando a pesquisa de todos os elementos contíguos àqueles sorteados.

**AMOSTRAGEM SISTEMÁTICA:** Representa uma abreviação do processo Casual Simples. É normalmente usada quando os elementos da população apresentam-se ordenados e a retirada dos elementos da amostra é feita periodicamente. Tal processo é mais vantajoso que o Casual Simples no sentido de que há uma facilidade relativamente maior na determinação dos elementos da amostra. Contudo, se a variável que se deseja estudar sofrer variações cíclicas e o período dos ciclos coincidir com o período da retirada dos elementos da amostra, seu uso torna-se restritivo, porquanto tal método introduz vício de amostragem.

**AMPLITUDE:** Diferença entre o mais alto e o mais baixo dos valores de uma distribuição. A amplitude é uma medida de variabilidade ou de dispersão.

**AMPLITUDE TOTAL:** Expressão usada como sinônimo de amplitude quando se quer distingui-la da amplitude semi-interquartil, que, ao contrário da amplitude total, não leva em conta todos os valores.

**AMPLITUDE SEMI-INTERQUARTIL:** Metade da diferença entre o terceiro e o primeiro quartil de uma distribuição de frequências. É uma medida de variabilidade ou dispersão.

**ARRANJOS:** Arranjos simples de classe  $p$  de  $m$  elementos, sendo  $p \leq m$ , ou arranjos simples de  $m$  elementos  $p$  a  $p$ , são todos os agrupamentos de  $p$  elementos distintos tirados dentro os  $m$  elementos dados, de modo que cada agrupamento se diferencie de outro, seja pela natureza, seja pela ordem de seus elementos.

**AMPLITUDE TOTAL:** Corresponde à diferença entre o maior e o menor valor observado em um conjunto de dados. Notação dada por  $A$ ,  $A_t$  ou  $R$ .

**ASSIMETRIA:** Características de gráficos ou de curvas em que a maioria dos valores não se concentra no meio (como acontece com a curva normal), mas numa extremidade.

**ASSIMETRIA POSITIVA:** Característica de gráficos ou de curvas em que a maioria dos valores se encontram à esquerda.

**ASSIMETRIA NEGATIVA:** Característica de gráficos ou de curvas em que a maioria dos valores se encontram à direita.

**AUSÊNCIA DE VÍCIO:** Refere-se à tendência de um conjunto de medidas ser igual ao valor verdadeiro.

**AMOSTRA:** É uma parcela significativa do universo pesquisado ou de coleta de dados.

**AMOSTRA:** Uma amostra é um subconjunto de indivíduos da população alvo. Existem dois tipos de amostras, as probabilísticas, baseadas nas leis de probabilidades, e as amostras não probabilísticas, que tentam reproduzir o mais fielmente possível a população alvo. Entretanto, somente as amostras probabilísticas podem, por definição, originar uma generalização estatística, apoiada no cálculo de probabilidades e permitir a utilização da potente ferramenta que é a inferência estatística.

**AMOSTRA:** É um grupo de sujeitos selecionados de um grupo maior incluindo menos que todos os sujeitos naquele grupo maior. Um subconjunto da população em estudo; a população à qual o pesquisador pretende generalizar os resultados.



**AMOSTRA:** É um subconjunto selecionado da população. Em geral, a amostra permite estimar parâmetros para toda a população sem necessidade de analisar todos os elementos dessa população.

**AMOSTRA:** Conjunto finito de elementos extraídos de um conjunto maior, chamado população. Se a constituição da amostra obedecer a determinadas condições, a análise das características da amostra pode servir para que façam inferências sobre elementos da população. Em inglês *sample*.

**AMOSTRA (MÉDIA):** A média de uma amostra, obtida aleatoriamente de uma população, equivale a uma média de amostra de uma distribuição de médias de amostras. Interpreta-se tal média de amostra como uma estimativa da verdadeira média da população.

**AMOSTRA ACIDENTAL:** É aquele tipo que é obtida por amostragem accidental. Também se diz amostra casual, aleatória ou estocástica. O mesmo que amostra simples.

**AMOSTRA ADEQUADA:** É a amostra casual que corresponde a prévias exigências de precisão das estimativas a que ela pode dar lugar, e de capacidade de representar, sem viés, a população cujas características se pretende estimar.

**AMOSTRA A ESMO:** É a que resulta de amostragem a esmo.

**AMOSTRA:** Um subgrupo obtido de uma população.

**AMOSTRA ALEATÓRIA:** Uma seleção feita a partir da população, de modo que cada indivíduo da amostra tem a mesma chance de ser escolhido e a escolha de um indivíduo não influencia a seleção de indivíduos subsequentes. Também conhecida como amostra casual simples.

**AMOSTRA DE MÚLTIPLOS ESTÁGIOS:** Uma amostra aleatória simples de conglomerados é escolhida de uma população de conglomerados. Na amostragem em dois estágios, uma amostra aleatória simples de unidades é escolhida para ser observada, a partir dos conglomerados escolhidos. Esse processo pode se expandir e envolver mais estágios.

**AMOSTRAGEM ALEATÓRIA SIMPLES:** Um método de amostragem em que todas as unidades da população têm a mesma chance de serem selecionadas.

**AMOSTRAGEM ESTRATIFICADA:** A população é distribuída em estratos e escolhe-se uma amostra aleatória simples de unidades de cada estrato.

**AMOSTRAGEM DE GIBBS (GS):** Tipo de método de amostragem usado em genética quantitativa, para implementação prática da análise Bayesiana, no sentido de resolver as dificuldades técnicas da marginalização. Isto se justifica por que a obtenção de distribuições marginais por processos analíticos é praticamente impossível. Assim, a obtenção da distribuição marginal a posteriori, ou seja, marginalização da distribuição conjunta a posteriori, tem sido obtida pelo método da amostragem de Gibbs (GS) através da amostragem e atualização das distribuições condicionais. Este método de amostragem pertence à classe de métodos de simulação estocástica, denominados de Monte Carlo-Cadeias de Markov, a qual é sustentada em propriedades das cadeias de Markov. O nome Gibbs advém da distribuição de Gibbs, que é muito utilizada na área de física estatística ou mecânica estatística. O amostrador de Gibbs explorando as distribuições condicionais completas através de algoritmo iterativo foi proposto inicialmente por Geman e Geman (1984) para aplicação na área de processamento de imagens.

**ANÁLISE ESTATÍSTICA BAYESIANA:** Tipo de análise estatística a qual baseia-se no conhecimento da distribuição a posteriori dos parâmetros e possibilita a construção de intervalos de confiança os quais são melhor definidos como intervalos de probabilidade ou intervalos de confiança bayesianos exatos para as estimativas dos parâmetros. Assim esta análise propicia uma descrição mais completa sobre a confiabilidade dos parâmetros do que outros métodos. A inferência estatística bayesiana baseia-se na distribuição condicional do parâmetro ( $\theta$ ) dado o vetor de dados ( $y$ ), ou seja, na distribuição a posteriori do parâmetro dadas as observações, a qual equivale:

$$f(\theta/y) = \frac{f(\theta/y)f(\theta)}{\int_R f(y/\theta)f(\theta)d\theta}, \text{ em que } f(y/\theta) \text{ é a função densidade de probabilidade da distribuição}$$

condicional de um vetor de observações ( $y$ ) dado  $\theta$  denominada função de verossimilhança ou modelo para os dados.  $f(\theta)$  é a função densidade de probabilidade da distribuição a priori, que é também a densidade marginal de  $\theta$ . Esta função denota o grau de conhecimento acumulado sobre  $\theta$ , antes da observação de  $y$ . A integral  $\int_R f(y/\theta)f(\theta)d\theta$  é a distribuição marginal ou preditiva de  $y$  com respeito à

$\theta$ , em que  $R$  é a amplitude da distribuição de  $\theta$ . Verifica-se, então, que a distribuição a posteriori é proporcional à verossimilhança x priori, ou seja, a função de verossimilhança conecta a priori à posteriori usando para isto os dados do experimento (observações). Dessa forma, a distribuição a posteriori contempla o grau de conhecimento prévio sobre o parâmetro ( $\theta$ ) e também as informações adicionais propiciadas pelo experimento ( $y$ ).

**AMOSTRAGEM POR CONGLOMERADO:** Uma forma de amostragem aleatória, na qual se identificam subdivisões ou conglomerados na população; uma amostra aleatória simples de conglomerados é escolhida e todas as unidades dentro dos conglomerados escolhidos são estudadas.

**AMOSTRAGEM POR DISTÂNCIA:** Um método de estimativa de parâmetros de interesse utilizando procedimentos de amostragem que obtêm estimativas mediante o registro do número e da frequência das espécies de uma linha de corte. O método baseia-se na proposição de que a detecção de sujeitos distribuídos aleatoriamente diminui com a distância da linha de corte.

**AMOSTRAGEM PROBABILÍSTICA:** Consultar amostra aleatória.

**AMOSTRAGEM SISTEMÁTICA:** Um método de seleção de uma amostra de elementos de uma população. Define-se um ponto inicial aleatório na disposição que é o elemento  $k$ -ésimo e, então, escolhe-se cada elemento  $k$ -ésimo subsequente; exceto a escolha do elemento inicial, esta não é uma amostragem aleatória.

**AMOSTRAGEM NÃO PROBABILÍSTICA OU SEM SORTEIO DO TIPO ACESSIBILIDADE OU INACESSIBILIDADE A TODA A POPULAÇÃO:** É aquela técnica onde a amostra atinge apenas a parte acessível da população. Em um vagão com minério de ferro, por exemplo, a amostra pode ser feita em alguns casos, apenas nos 20 cm superiores, por dificuldade de se atingir todos os pontos do vagão ou às porções inferiores. Nesse tipo surge uma diferenciação entre população-objeto e população amostrada, ou seja, a população-objeto é aquela que tem-se em mente ao realizar o trabalho estatístico, mas apenas uma parte dessa população, porém, está acessível para que dela retire-se a amostra, e essa parte é a população amostrada. Se as características da variável de interesse forem as mesmas na população objeto e na população amostrada, então esse tipo de amostragem equivalerá a uma amostragem probabilística.

**AMOSTRA TOMADA A ESMO, SEM NORMA OU ACIDENTAL:** É o tipo de amostragem em que o pesquisador, para simplificar o processo de seleção, procura ser aleatório sem na realidade ter realizado propriamente o sorteio usando algum mecanismo aleatório confiável. A amostra é constituída pelos elementos que se consegue tomar de uma população. Num galpão de aves, ou em gaiolas com camundongos de mesma linhagem, por exemplo, a amostra pode ser constituída das aves ou ratos que forem tomadas no instante da coleta dos dados, sem, entretanto ter havido um sorteio prévio. Os resultados da amostragem a esmo são, em geral, equivalentes aos de uma amostragem probabilística se a população é homogênea e se não existe a possibilidade de o pesquisador ser inconscientemente influenciado por alguma característica dos elementos da população.

**AMOSTRA INTENCIONAL OU POR JULGAMENTO:** Nessa técnica o pesquisador está interessado na opinião, ação, intenção, etc. de determinados elementos da população, mas não representativos da mesma. Por exemplo, conduzir um estudo para se desejar saber como pensam os líderes de opinião de determinada comunidade, ou em pesquisa de opinião dos diretores de faculdade de comunicação social sobre a extinção da obrigatoriedade do diploma universitário para o exercício da profissão de jornalistas, a opinião de reitores de universidades federais sobre o sistema de quotas para o concurso vestibular. O pesquisador escolhe deliberadamente certos elementos para formar a amostra, por julgar tais elementos bem representativos da população, isto é baseado num pré-julgamento. O pesquisador escolhe a seu juízo os elementos da população que julgar representativos, para constituírem a sua amostra, mas sem fazer sorteio. O perigo desse tipo de amostragem é obviamente grande, pois o amostrador pode facilmente se equivocar em seu pré-julgamento. Apesar disso o uso de amostragens intencionais ou parcialmente intencionais, é bastante frequente, ocorrendo em várias situações reais.

**AMOSTRA DE VOLUNTÁRIOS OU POR CONVENIÊNCIA:** O pesquisador escolhe deliberadamente determinados elementos para fazer parte da amostra, ou seja, a amostra por conveniência é empregada quando se deseja obter informações de maneira rápida e barata, uma vez que esse procedimento consiste em simplesmente contatar unidades convenientes da amostragem, é possível recrutar respondentes tais como estudantes em sala de aula, mulheres no shopping, alguns amigos e vizinhos, entre outros. Este método também pode ser empregado em pré-testes de questionários. Como exemplo pode-se citar para ilustrar esse método no uso de pesquisa com amostras por conveniência nos seguintes casos: i) solicitar a pessoas que voluntariamente testem um produto e que em seguida respondam a uma entrevista; ii) parar pessoas no supermercado e colher suas opiniões; iii) colocar linhas de telefone adaptadas para que durante um programa de televisão os telespectadores possam dar suas opiniões.

**AMOSTRAGEM EM POPULAÇÃO FORMADA DE MATERIAL CONTÍNUO, COMO GASES E LÍQUIDOS:** Tipo de técnica de amostragem onde é impossível realizar amostragem probabilística, devido ser impraticável a realização de um sorteio rigoroso. Se a população for líquida ou gasosa, o que comumente se faz com resultado satisfatório, é homogeneizar e retirar a amostra a esmo. Esse procedimento pode às vezes, também ser usado no caso de material sólido. Outro procedimento que pode ser empregado nesses casos, especialmente quando a homogeneização não é praticável, é a enquadração, a qual consiste em subdividir a população em diversas partes. A origem do nome enquadração pressupõe a divisão em quatro partes, sorteando-se assim uma ou mais delas para constituir a amostra ou para delas se selecionar a amostra. Por outro lado as vezes se torna impraticável realizar a aleatorização ou sorteio quando a unidade amostral sofre processo de homogeneização no ato da sua retirada. É o caso de amostras no meio líquido ou gasoso. Mais especificamente, amostras do fitoplâncton são normalmente homogeneizadas quando

chegam à superfície em uma garrafa de coleta que é a unidade amostral. Analisar apenas uma parcela ou subamostra ou analisar toda unidade amostral é indiferente para efeito de análises estatísticas. Exemplo: coleta de subamostras de leite em diversos baldes com o líquido, deve-se homogeneizar e retirar a amostra a esmo. Verificação de poluição de praias e rios.

**AMOSTRA POR TIPICIDADE OU PROPOSITAL:** Tipo de técnica de amostragem onde o pesquisador escolhe um subgrupo que seja típico em relação a população como um todo. Como exemplo para ilustrar esse caso suponha uma pesquisa de um grupo típico de drogados visando obter dados sobre os fatores que levam os consumidores de drogas injetáveis a fazerem permutas de seringas e agulhas, apesar da elevada possibilidade de serem contaminados pelo vírus da AIDS. A seleção de amostras intencionais ou por julgamento são realizadas de acordo com o julgamento do pesquisador. Se for adotado um critério razoável de julgamento, pode-se chegar a resultados favoráveis. A hipótese básica que é adotada na amostragem proposital é a de que, com um bom julgamento e uma estratégia apropriada, pode-se escolher com cuidado os casos que devem ser incluídos na amostra e, deste modo, desenvolver amostras que são satisfatórias de acordo com as necessidades do estudo. Uma estratégia comum de amostragem proposital é tomar casos que são considerados típicos da população pela qual se está interessado, admitindo que os erros de julgamento cometidos na seleção tenderão a contrabalançar-se. As amostras propositais selecionadas têm sido usadas em tentativas de prognosticar eleições nacionais, por exemplo, como o detalhado a seguir: Para cada Estado, escolha um pequeno número de distritos ou municípios eleitorais para os quais o resultado eleitoral em anos anteriores se aproximou do resultado global do estado, em seguida entreviste todos os eleitores qualificados nesses distritos ou municípios sobre as suas intenções eleitorais, e tenha esperança de que os distritos selecionados ainda são típicos dos seus respectivos estados. O problema que existe com esse método de amostragem é que, quando não há mudanças significativas na opinião política dos eleitores, pode-se provavelmente obter resultado semelhante através dos prognósticos devolvidos dos anos anteriores sem fazer nenhum tipo de entrevista, mas se houve mudanças ou se elas estiverem ocorrendo é preciso saber como essas mudanças estão afetando os distritos ou municípios selecionados em comparação com outras unidades eleitorais. É comum a escolha de experts que são profissionais especializados quando se trata de amostras por julgamento. A escolha de experts é uma forma de amostragem por julgamento ou intencional usada para escolher elementos típicos e representativos para uma amostra. Em algumas situações é comum pesquisadores da área animal eleger animais para seleção amostral baseado no fenótipo exterior desses animais, por achar que eles são típicos da raça a qual pertence, como por exemplo, selecionar animais suínos ou bovinos para fazer parte de um banco de sêmen em uma determinada granja, no entanto o perigo aqui é o animal selecionado trazer consigo alguma característica genética ou gen indesejável, e assim comprometer o estudo conduzido. A abordagem da amostragem por julgamento pode ser útil quando é necessário incluir um pequeno número de unidades na amostra. O método de julgamento é muito utilizado para a escolha de uma localidade representativa de um país na qual serão realizadas outras pesquisas, sendo algumas vezes até preferida em relação à seleção de uma localidade por métodos aleatórios. A amostra por julgamento pode ser, nestes casos, mais fidedigna e representativa que uma amostra probabilística. Um exemplo de uso frequente desse tipo de prática é a escolha de uma cidade típica para representar o universo urbano e rural do país. A amostragem por julgamento também é útil quando é preciso obter uma amostra deliberadamente enviesada. Como exemplo pode-se citar o seguinte caso: quando se quer avaliar uma modificação em um produto ou serviço, podem-se identificar grupos específicos que estariam dispostos a dar sua opinião em relação à modificação. Se o pesquisador avaliar que este grupo não gostou da modificação, então ele poderia assumir que o resto da população

também teria uma percepção negativa em relação à mudança. Se o grupo gostou da modificação, então mais pesquisas poderiam ser requeridas para se chegar a uma conclusão a respeito do assunto. Como essa técnica de amostragem por tipicidade sofre das mesmas restrições aplicáveis a outras técnicas não probabilísticas, isto é, os erros de amostragem e desvios não podem ser computados e avaliados, tal técnica deve restringir-se às situações em: i) os possíveis erros não apresentam significativa influência de vício nos resultados e ii) é praticamente impossível à utilização de amostragem probabilística.

**AMOSTRA POR QUOTAS OU PROPORCIONAIS OU REPRESENTATIVAS:** Tipo de técnica de amostragem não probabilística, onde as amostras são obtidas dividindo-se a população por categorias ou estratos e selecionado certo número que são denominado quota de elementos de cada categoria de forma não aleatória, e pressupõem três etapas: i) classificar a população mediante o uso de propriedades pertinentes; ii) determinar a porcentagem da amostra a ser selecionada de cada classe e iii) fixar quotas para cada pesquisador. Na amostragem por quotas definem-se as variáveis de controle tais como sexo, idade, rendimento, etc., cujas distribuições são conhecidas e escolhe-se a amostra de forma a satisfazer essas distribuições. Por exemplo, se as mulheres de idade superior a 40 anos representarem 70 % da população em estudo, então a amostra também deverá incluir 70 % de mulheres dessa faixa etária. Normalmente, os responsáveis pelo levantamento amostral indicam aos entrevistadores a distribuição, mas a seleção final é deixada aos cuidados desses entrevistadores que podem tentar recorrer a quem estiver mais acessível, escolhendo pessoas que estejam relacionadas e, dessa forma, não distorcerem os resultados. A amostragem por quotas, algumas vezes denominada de forma equivocada como amostragem representativa, traz a segurança de garantir condições de inclusão na amostra de diversos elementos da população e assegurar-se que esses diversos elementos são tomados nas proporções em que ocorrem na população. Veja a ilustração dessa técnica através de um exemplo de um caso extremo, isto é, suponha que se esteja extraindo uma amostra de uma população com igual número de homens e mulheres e que há uma clara diferença entre os dois sexos na característica que se deseja mensurar. Se não for entrevistada nenhuma mulher, os resultados do levantamento seriam quase certamente um quadro extremamente enganoso da população total. Na verdade, mulheres e membros de grupos minoritários são, frequentemente, sub-representados em amostras acidentais. Antecipando estas possíveis diferenças entre subgrupos, aqueles pesquisadores que extrai amostras por quotas procuram garantir a inclusão na amostra de elementos suficientes de cada estrato. Como é normalmente descrito, o objetivo básico da amostragem por quotas é a seleção de uma amostra que seja a réplica da população para a qual se quer generalizar ou inferir os resultados do levantamento, por isso a noção de que ela representa aquela população. É sabido que a população tem igual número de homens e de mulheres, os entrevistadores são instruídos a entrevistarem igual número de homens e mulheres. Se for sabido que 20 % da população sob estudo estão dentro de uma faixa etária, instruem-se os entrevistadores a garantirem que 20 % da amostra estejam dentro daquela faixa etária. É fácil perceber que a amostragem por quotas é uma espécie de amostragem por estratos não aleatória. A amostra por quotas constitui um tipo especial de amostra intencional em que o pesquisador procura obter uma amostra que seja similar à população sob algum aspecto. A seleção de amostra por quotas é a forma mais usual de amostragem não probabilística. Neste caso, são consideradas várias características da população, como sexo, idade e tipo de trabalho sendo que as variáveis mais comuns são áreas geográficas, sexo, idade, raça e uma medida qualquer de nível econômico. Nesse caso com essa amostra pretende-se incluir proporções similares de pessoas com as mesmas características. A idéia de amostragem por quotas sugere que se as pessoas são representativas em termos de características, elas também poderão ser representativas em termos da informação procurada pela pesquisa. Depois de serem identificadas as

proporções de cada tipo a ser incluído na amostra, o pesquisador estabelece um número ou quota de pessoas que possuem as características determinadas e que serão contatadas pela pesquisa. O entrevistador recebe instruções para continuar a amostragem até que a quota necessária tenha sido atingida em cada estrato. Uma pesquisa com amostragem por quotas poderá ser utilizada e trazer bons resultados quando as características relevantes para controle e delineamento da amostra forem conhecidas, estiverem disponíveis ao pesquisador, estiverem relacionadas ao objeto de estudo e se constituírem em poucas categorias. As amostras por quotas são bastante úteis em estágios preliminares da pesquisa, e se for realizada com muito cuidado pode fornecer respostas definitivas. As amostras por quotas são muito usadas em pesquisa de opinião eleitoral e pesquisas de mercado. O processo de quotas produz amostras com tendências, embora esteja frequentemente de acordo com as amostras pelas probabilidades quando se trata de questões de opinião e pesquisa. Em relação à amostragem por julgamento, a amostra por quotas tem a vantagem de fornecer maior extensão, sendo também mais provável de representar, por exemplo, a atitude das pessoas jovens que uma amostra probabilística realizada com estudantes universitários. Uma amostra probabilística com 15 a 25 % de taxa de não-respostas também pode apresentar muitos vieses. A amostragem por quotas ou outras abordagens de amostragem por julgamento, que são mais rápidas e mais baratas, não deveriam ser descartadas em seu uso. Para a realização de amostragem por quotas é necessário estabelecer variáveis de controle. À medida que o número de características e categorias sob controle for sendo elevado, pode-se chegar a uma situação tal que o método não poderá ser empregado, ou pela não disponibilidade das proporções na população, ou pelo exagerado número de células a que se chega, o que permitirá um número elevado de elementos no total da amostra. A maioria das amostras por quotas tem alguns aspectos básicos em comum. Na grande parte dos casos são estimados tamanhos de subclasses na população e as quotas são divididas entre os entrevistadores para que eles encontrem pessoas que preencham os requisitos exigidos, obedecendo às restrições do controle de quotas. Uma limitação atrelada a isso é que a seleção final dos indivíduos cabe ao entrevistador o que pode gerar vieses. Outra dificuldade tem haver com a localização de tipos de indivíduos requeridos para preencher as categorias de quotas determinadas pelos pesquisadores. Não se leva em consideração que há níveis diferentes de dificuldade para encontrar tipos de indivíduos no início e no final da pesquisa, sendo que ao final da pesquisa, os entrevistadores podem perder muito tempo procurando determinado tipo de indivíduo. Muitas vezes, as instruções não estão perfeitamente claras e não se sabe se houve rejeição para responder ao questionário por razões de não preenchimento das categorias de quotas ou por outras razões pertinentes ao indivíduo. Além do mais, os entrevistadores podem exceder o número de suas quotas, o que acarretaria em um desperdício de tempo e recursos. Em uma pesquisa com amostra por quotas, é importante que as características nas quais as quotas estejam baseadas sejam facilmente identificadas, ou no mínimo estimadas pelo entrevistador, ou muito tempo será desperdiçado na identificação das pessoas que farão parte da pesquisa. O controle de quotas desafia a habilidade dos pesquisadores. As referências utilizadas como ferramentas de controle devem estar disponíveis e ser razoavelmente recentes, devendo ser factíveis de aplicação em todas as áreas em que serão feitas as entrevistas. Diversos autores concordam que a classificação dos respondentes que os entrevistadores realizam para avaliar suas características deve ser razoavelmente tipo de amostragem não probabilística é uma técnica utilizada principalmente quando se desejam obter informações detalhadas durante certo período de tempo, sobre questões particulares. Como exemplo para ilustração desse método podem ser citados estudos realizados por órgãos oficiais, principalmente sobre orçamento familiar, bolsa família, programa fome zero, etc. ou levantamentos sobre programas de rádio e televisão para aferição de audiência, estudo para mensurar a correlação entre orçamento familiar e hábitos alimentares, estudo sobre a utilização diária dos aposentos da residência, levantamentos sobre



comportamento das crianças em relação aos animais domésticos e outros animais selvagens criados em casa e apartamentos.

**AMOSTRAGEM ALEATÓRIA SIMPLES (AAS):** É o método mais simples e mais importante para a seleção de uma amostra. Além de servir como um plano próprio, o seu procedimento é usado de modo repetido em procedimentos de múltiplos estágios. É indicado quando a população for homogênea. Uma amostra é chamada aleatória se cada elemento da população tiver a mesma probabilidade de pertencer à amostra. Se a população for pequena, uma forma de se conseguir obter uma amostra aleatória é atribuindo a cada elemento um número, escrevê-lo em um pedaço de papel, misturar todos os papéis numa caixa ou urna e selecionar tantos quanto se queira para constituir a amostra. Se o número de elementos da população for grande este procedimento será impraticável. Neste último caso pode-se usar um método simples para se constituir uma amostra aleatória, o qual é baseado numa tabela de números aleatórios. Esta tabela é uma lista de dígitos decimais em ordem casual que podem ser pensados como uma sequência de observações independentes de uma variável aleatória  $D$ , com uma distribuição uniforme sobre os números inteiros 0, 1, 2, 3, ..., 9, isto é,  $P[D = d] = \frac{1}{10}$ ,  $d = 0, 1, 2, 3, \dots, 9$ . Se a população é um conjunto de indivíduos,

elementos ou objetos para os quais podem se observar valores de uma ou mais variáveis e se uma característica de interesse dessa população for a distribuição das medidas dos pesos em quilogramas ao nascer de bezerros machos da raça bovina gir, dizemos assim que a distribuição de  $X$  é a distribuição da população, ou seja, o termo população tanto indicará o conjunto dos indivíduos como a distribuição da variável aleatória  $X$ . Sendo assim chama-se de amostra aleatória  $n$  da população  $N$ , ou ainda da distribuição de  $X$  ou simplesmente de  $X$ , ao conjunto de  $n$  variáveis aleatórias independentes  $X_1, X_2, X_3, \dots, X_n$ , e cada uma tendo a mesma distribuição de probabilidade de  $X$ . Em outras palavras, pode-se dizer que uma amostra aleatória da distribuição de uma variável aleatória  $X$  corresponde a considerar  $n$  medidas repetidas de  $X$ , sob condições essencialmente idênticas. Ele pode ser caracterizado através da definição operacional, como mostrado a seguir: De posse uma lista completa com  $N$  unidades elementares, sorteiam-se com igual probabilidade  $\left(\frac{1}{N}\right)$ ,  $n$  unidades amostrais, através de um dispositivo aleatório qualquer como, por exemplo,

um computador programado para acumular números, uma calculadora, pedaços de papel numerados e colocados numa urna ou uma tabela de números ou dígitos aleatórios. A principal característica para o uso do plano denominado amostra aleatória simples (AAS) é a existência de um sistema de referências completo, descrevendo cada uma das unidades elementares. Deste modo tem-se bem listado o universo  $u = \{1, 2, \dots, N\}$ . O plano amostral é descrito do seguinte modo: i) utilizando um procedimento aleatório qualquer tais como uma tabela de números aleatórios, pedaços de papel colocados numa urna, um computador programado para acumular números, a função Random (Run#) de uma calculadora etc., sorteia-se com igual probabilidade  $\left(p = \frac{1}{N}\right)$  um elemento da população  $U$ ; ii) repete-se o processo anterior até que

sejam sorteadas  $n$  unidades. Tendo sido este número pré-fixado anteriormente; iii) caso seja permitido o sorteio de uma unidade mais de uma vez, tem-se o processo AAS com reposição. Quando o elemento sorteado é removido de  $U$  antes do sorteio do próximo, tem-se o plano AAS sem reposição. Do ponto de vista prático o plano AAS sem reposição é muito mais interessante, pois vai de encontro aos princípios intuitivo de que não se ganha mais informações se uma mesma unidade aparece mais de uma vez na amostra. Além dos estimadores obtidos apresentarem maior precisão. Por outro lado o plano AAS com reposição introduz vantagens matemáticas e estatísticas, principalmente a independência entre as unidades

sorteadas, o que facilita em muito a determinação das propriedades dos estimadores das quantidades populacionais de interesse que são os parâmetros. Neste tipo de técnica pode ser usado amostra com reposição e sem reposição. O número total de possíveis amostras aleatórias simples com reposição de tamanho  $n$  que podem ser extraídos de uma população de tamanho  $N$  é dado por  $N^n$ , e a probabilidades de se obter cada amostra é definida como  $P(\text{amostra}) = \frac{1}{N^n}$ . O número total de possíveis amostras alea-

tórias simples sem reposição de tamanho  $n$  que podem ser selecionadas de uma população de tamanho  $N$  é dado pela combinação  $C_n^N = \frac{N!}{n!(N-n)!}$ , cuja probabilidade de se obter uma dessas amostras é dada

$$\text{por } P(\text{amostra}) = \frac{1}{\binom{N!}{n!(N-n)!}}.$$

**AMOSTRAGEM ESTRATIFICADA:** Essa técnica é usada quando a população é constituída de subpopulações naturais que são os estratos e que são homogêneos internamente, podendo ser heterogêneos de estrato para estrato. Dessa forma, a amostragem deve ser realizada fazendo com que todos os estratos populacionais sejam representados na amostra final obtida. Para especificar o número de elementos por estratos que irá compor a amostra, são considerados três métodos: uniforme, proporcional e ótimo. Estes dependem basicamente do tamanho dos estratos populacionais e de sua variabilidade. Exemplos de estratos: i) na escala de uma praia, por exemplo, temos a umidade do sedimento do solo e podemos classificá-los em zonas infralitoral, medilitoral e sublitoral; ii) na escala da baía de paranaguá temos por exemplo 5 estratos de acordo com a salinidade e podemos dividi-los em regiões euhalina, mesohalina e oligohalina. Estratificar as propriedades agrícolas de uma região através do tamanho da área cultivada em hectares. Quando a população for heterogênea não se deve usar a amostra aleatória simples devido à baixa precisão das estimativas obtidas. Nesta situação, deve-se subdividir a população em subpopulações de forma que dentro das subpopulações haja homogeneidade. Este processo chama-se estratificação da população, sendo cada subpopulação um estrato. A amostra obtida nesse caso chama-se amostra aleatória estratificada. Na prática a população pode já se apresentar estratificada naturalmente, ou então depender da estratificação a ser realizada utilizando-se critérios baseados nos conhecimentos que o pesquisador tem sobre a população. Ao se executar a estratificação de uma população com relação a uma variável, é importante que esta esteja correlacionada com outras variáveis que serão estudadas no levantamento a ser executado. A amostragem estratificada consiste na divisão de uma população em grupos ou estratos homogêneos, segundo alguns (as) característica (s) conhecida (s) na população sob estudo, e de cada um desses estratos são selecionadas amostras em proporções convenientes. A estratificação é usada como melhoria da precisão das estimativas, produzir estimativas para a população toda e subpopulações, questões administrativas e outras. A distribuição das  $n$  unidades da amostra pelos estratos chama-se alocação da amostra. Essa distribuição é muito importante, pois ela é que irá garantir a precisão do procedimento amostral. Estas podem ser do tipo: i) alocação proporcional, sendo o tipo de amostragem realizado quando os tamanhos dos estratos populacionais são distintos. O estrato  $h$  fornece uma quantidade  $n_h$  de elementos, proporcional ao tamanho  $N_h$  populacional do respectivo estrato, para formar a amostra de tamanho  $n$ . Cada elemento da população tem a mesma probabilidade de pertencer a amostra. Ou seja, neste tipo de procedimento a amostra de tamanho  $n$  é distribuída proporcionalmente ao tamanho do estrato; ii) outro tipo é a alocação uniforme a qual costuma ser usada em situações em que o maior interesse é obter estimativas separadas para cada estrato, ou, ainda, quando se deseja comparar os diversos estratos. De  $K$

estratos retiram-se amostras do mesmo tamanho, ou seja, seleciona-se a mesma quantidade de elementos em cada estrato. Usada quando os estratos populacionais possuem o mesmo tamanho. Neste caso atribui-se o mesmo tamanho de amostra para cada estrato. É o procedimento indicado quando pretende-se apresentar estimativas separadas para cada estrato e iii) por último tem-se o tipo denominado de alocação ótima de Neyman, sendo o tipo de amostragem onde são considerados o tamanho e a variabilidade de cada estrato populacional para a extração da amostra. De cada estrato, retira-se uma quantidade  $n_h$  de elementos, a qual é proporcional ao tamanho ( $N_h$ ) e ao desvio padrão populacional do respectivo estrato ( $s_h$ ). Neste caso o problema é minimizar a variância do estrato para uma função custo fixada, ou minimizar o custo para uma variância do estrato fixada. As principais características deste método são: i) a obtenção da informação sobre a população é otimizada, pois quanto menor a variação nos estratos, menor é a quantidade de elementos necessária; ii) E a necessidade de conhecer o desvio padrão populacional em cada estrato representa a principal dificuldade. Podem-se realizar campanhas ou amostras piloto para pesquisar a variação dentro de cada estrato. Como exemplo pode ser citado o caso de se estratificar a área plantada com milho, no município de Mossoró, RN nos anos 2014/2015, usando como critério de estratificação o tamanho das áreas cultivadas em hectare.

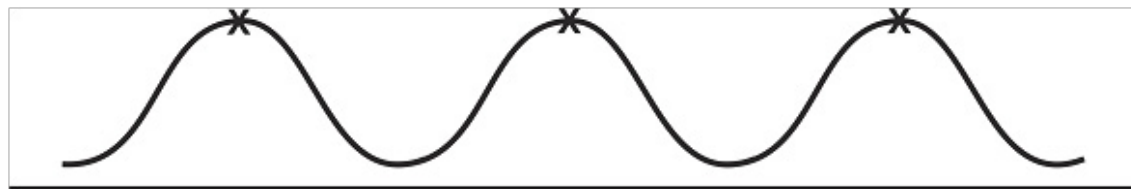
**AMOSTRAGEM SISTEMÁTICA:** É uma simplificação do processo de amostra aleatória simples (AAS). Neste caso, o primeiro elemento é sorteado e os demais são retirados em uma progressão aritmética, com uma determinada razão  $r$ , até completar o total de elementos da amostra ( $n$ ). Para isso, a população deve estar disposta de tal maneira que a sistematização do processo de obtenção das unidades amostrais seja possível. A homogeneidade da população alvo é igualmente importante. É uma variação da amostra aleatória simples. Sua aplicação exige que a população esteja devidamente ordenada de tal forma que cada um de seus elementos possa ser unicamente identificado pela colocação ou por outro critério qualquer. Isto ocorre, por exemplo, quando todos os elementos da população estão anotados numa listagem, quando um grupo de árvores está numa fileira, etc.. Neste tipo de amostragem cada elemento da população tem igual probabilidade de ser selecionado, mas cada uma das amostras possíveis não têm a mesma probabilidade de ser selecionada, e algumas das amostras possíveis têm probabilidade nula de serem escolhidas, como por exemplo, as amostras que incluem elementos consecutivos. A determinação do número de elementos nesta técnica pode ser feita assim: considere uma população com  $N$  elementos onde  $N=kn$ , e onde  $k$  é um número inteiro. Considere também que a população está ordenada de 1 a  $N$ , formando o sistema de referências, e  $n$  o tamanho da amostra desejada, define-se a quantidade  $K = \frac{N}{n}$ , a que se dá

o nome de intervalo de amostragem, passo ou intervalo de seleção e admitindo-se que  $K$  seja um número inteiro, faz-se então um sorteio ente os números 1, 2, 3, ...,  $K$ , podendo ser obtido, por exemplo, o valor  $i$ , que será chamado início casual. Nestas condições com apenas este sorteio, toda a amostra fica selecionada, sendo composta das unidades amostrais que tenham recebido os seguintes números ou ocupam a posição correspondente mostrada a seguir:  $i; i + K; i + 2K; i + 3K; \dots; i + (n-1)K$ . Sendo assim tem-se que uma unidade  $x$  é então selecionada aleatoriamente através de uma amostra aleatória simples (AAS) entre as  $k$  primeiras unidades do sistema de referências. As unidades seguintes que farão parte da amostra são obtidas a partir da primeira unidade selecionada em intervalos de comprimento  $k$ , isto é  $n$  é dado por  $n = \{x; x + k; x + 2k; x + 3k; \dots; x + (n-1)k\}$ . Por exemplo, suponha que para determinada população com mil elementos, pretende-se retirar uma amostra sistemática de 200 componentes, ou seja,  $N = 1000$  e  $n = 200$ . Portanto  $k = \frac{N}{n} = \frac{1000}{200} = 5$ . Isto, é, a população está dividida em 200 grupos de 5 unidades popula-

cionais onde um elemento será selecionado em cada grupo. Uma unidade é selecionada aleatoriamente entre as primeiras unidades. Suponha que a unidade 3 tenha sido selecionada. Então em cada um dos 199 grupos restantes, será selecionada sempre a terceira unidade, completando assim a amostra sistemática de 200 unidades elementares populacionais, e ela é obtida conforme a configuração ou ordem a seguir:  $n=\{0003; 0006; 0009; \dots; 0995\}$ . Neste caso pode-se definir população em ordem aleatória. Algumas vezes a amostragem sistemática é usada para seu melhor rendimento em populações em que a numeração das unidades é efetivamente aleatória. Assim acontece na amostragem de um fichário arrumado, alfabeticamente, por sobrenomes, se a característica que está sendo medida não tem relação com o sobrenome dos indivíduos. Neste caso não haverá tendência à estratificação em  $y_i$ , à medida que percorre-se o fichário, nem correlação entre valores vizinhos. Nessa situação, é de se esperar que a amostragem sistemática seja em essência aleatória simples e que tenham a mesma variância do estimador. A vantagem principal da amostragem sistemática (as) é a facilidade de sua execução. Também, é bem menos sujeita a erros do entrevistador que os outros esquemas de amostragem. Por outro lado, quanto a sua precisão, existem situações em que ela é mais precisa que a amostra aleatória simples (AAS). Mas na maioria das populações tem-se que sua eficiência é próxima da AAS. Estas populações são as populações que estão na chamada ordem aleatória. Em outros casos quando existem tendências do tipo linear ou existem periodicidade na população, então sua precisão pode ser bem diferente do planejamento da AAS. Isto é a precisão é menor. Um grande problema na utilização da amostra sistemática (AS) é a estimação da variância do estimador obtido. No caso em que a população está em ordem aleatória, não existem muitos problemas em se estimar a variância do estimador obtido através da amostra sistemática pela estimativa da variância do estimador  $\bar{y}$  da AAS, pois nestes casos, AAS e AS apresentam resultados similares. Por outro lado, nos casos em que a população apresenta tendências ou periodicidade pode-se, ao utilizar tal procedimento, superestimar ou subestimar a variância do estimador obtido a partir da amostragem sistemática. Por exemplo, tomar uma amostra sistemática do tamanho  $n = 100$  árvores numa população de tamanho  $N = 1000$  árvores de pinus (*Pinus elliottii*), onde mede-se o DAP que é diâmetro à altura do peito em centímetros sendo este mensurado a uma altura de 1,60 metros do solo, ou seja altura de um homem de estatura mediana, a cada décima árvore de uma fileira.  $N = 1000$ ,  $n = 100$ ,  $k = \frac{N}{n} = \frac{1000}{100} = 10$ , sendo,  $k = 10$ , e  $x = 5$  assim tem-se

que  $n=\{0001, 0002, \dots, 1000\}$ . Isto ocorre se o primeiro elemento sorteado for o 5, isto é  $[ \_\_\_\_ x = 5 \_\_\_\_ ]$ , então a amostra sistemática de tamanho 100 é dada por:  $n=\{0005; 0015; 0025; \dots; 0995\}$ . As vantagens da amostragem sistemática em relação a amostragem aleatória simples são as seguintes: maior simplicidade, economia e facilidade no processo de seleção dos elementos que comporão a amostra. O sorteio do primeiro elemento define os  $(n-1)$  outros elementos no rol dos dados que comporão a amostra. Isto facilita a supervisão da coleta dos dados controlando facilmente os erros. A amostragem sistemática é muito útil quando se trata de amostrar uma população que vai se completando ao longo do tempo e que, portanto, em um determinado momento ainda não tem todos os seus elementos ou unidades amostrais. Por exemplo, quando se quer fazer um estudo sobre óbitos registrados durante o ano no município de Mossoró no Rio Grande do Norte, não é necessário esperar que termine o ano para começar a pesquisar. Entretanto, se o pesquisador dispor de uma boa estimativa do número total de óbitos que devem ser registrados naquele ano, pode-se calcular o valor de  $K$  para o tamanho desejado da amostra e iniciar o sorteio na amostra sistemática, a partir do primeiro mês (janeiro) do ano, dando prosseguimento mensalmente à obtenção da amostra. A vantagem desta estratégia é economia de tempo, no sentido de que a pesquisa pode começar imediatamente. Outra vantagem nesse tipo de estudo reside no fato de que, quando a pesquisa envolve perguntas às famílias dos falecidos, ter que aguardar que o ano termine pode espaçar

demasiado as datas do óbito e da entrevista, introduzindo no estudo vícios devidos a problemas de memória. A amostra sistemática se distribui mais uniformemente na população podendo levar a uma maior representatividade. A população fica dividida em  $n$  estratos de tamanho  $k$ , sendo tomado um elemento de cada estrato. A precisão das estimativas quase sempre é maior. É menos sujeita a erros do entrevistador que os outros esquemas de amostragem. Por outro lado esta técnica apresenta desvantagens em relação a amostragem aleatória simples (AAS), tais como: pode haver a possibilidade de existência de ciclos de variação da variável de interesse ou seja, periodicidade, e o perigo, principalmente se o pico desses ciclos coincidirem com o período ou o momento de retirada dos elementos da amostra, como mostrado na figura a seguir:



**Figura:** Ciclos de variação da variável de interesse na amostragem sistemática.

Sendo assim essa técnica não é aconselhável quando os elementos da lista possam estar correlacionados de uma forma sequencial e irregular que possa afetar a representatividade da amostra, ou seja, com certa característica se repetindo em intervalos iguais. Por outro lado, se na população houver estratos e estes forem ordenados, a amostragem sistemática conduzirá, automaticamente, a uma partilha proporcional. Por exemplo, se o estudo tiver por objetivo a estimação da produção média de lixo urbano por habitante em uma grande cidade, pode-se ao coletar lixo nas segundas-feiras superestimar a produção diária devido ao fato do lixo coletado às segundas-feiras esteja provavelmente o lixo acumulado do final de semana como sábado e domingo, e assim este dia apresenta um valor muito mais elevado que os outros dias da semana, fazendo com que se tenta uma estimativa viciada para a produção média do lixo, bem como uma subestimativa ou superestimativa da variância desta média.

**AMOSTRAGEM POR CONGLOMERADOS OU GRUPOS:** Tipo de técnica que é usada quando a população apresenta uma subdivisão em pequenos grupos, chamados de conglomerados, então se faz a amostragem por meio desses conglomerados, a qual consiste em sortear um número suficiente de conglomerados, cujos elementos constituirão a amostra, ou seja, as unidades de amostragem, sobre as quais é feito o sorteio, são os conglomerados e não mais os elementos individuais da população. Esse tipo de amostragem é às vezes adotado por motivos de operacionalidade prática e motivos econômicos. Nessa técnica ao contrário da amostragem estratificadas onde os extratos são homogêneos a variabilidade de cada conglomerado é semelhante à da população de que o conglomerado se supõe representativo. Assim, usa-se a amostragem por conglomerados quando os grupos são muitos semelhantes entre si e semelhantes à população. Utilizada quando a população é muito dispersa, ou quando não é possível ou se tem o custo demasiadamente elevado para se arranjar uma lista de todos os elementos da população, tornando ou inviabilizando a organização de um rol com todos os elementos. Exemplo: estimar o número de cabeças de gado de uma certa região. Sorteiam-se alguns municípios dessa região e dentro deles algumas propriedades para compor a amostra. Na amostragem por conglomerado, a população é dividida primeira em grupos de elementos separados chamados de conglomerados. Cada elemento da população pertence a um e somente um conglomerado. Alguns destes conglomerados são sorteados para constituir

a amostra. Este mecanismo é o mesmo que uma amostra aleatória simples dos conglomeradas. Todos os elementos dentro de cada conglomerado formam a amostra. A amostragem por conglomerados tende a fornecer os melhores resultados quando os elementos são heterogêneos ou não parecidos. No caso ideal, cada conglomerado é uma versão em pequena escala representativa da população inteira. O valor da amostragem por conglomerado depende da representatividade que ele tem da população inteira. Se todos os conglomerados são parecidos a esse respeito, amostrar um pequeno número de conglomerados fornecerá boas estimativas dos parâmetros da população. Uma das aplicações primárias da amostragem por conglomerados é a amostragem de área onde os conglomerados são blocos de cidade como bairros ou outras áreas bem definidas. O que caracteriza bem o planejamento amostral de conglomerados é que a unidade amostral contém mais de um elemento populacional. A amostragem por conglomerado, geralmente exige um tamanho maior de amostra total do que a amostragem aleatória simples e a amostragem aleatória estratificada. No entanto, pode resultar em economia de custo por causa do fato de que quando um entrevistador é enviado para um conglomerado amostrado, por exemplo, um bairro de uma cidade, muitas observações da amostra podem ser obtidas em um tempo relativamente curto. Por isso, um tamanho maior de amostra pode ser obtido com um custo total significativamente mais baixo. O problema desta técnica reside na representatividade dos conglomerados. É, no mínimo, duvidoso que qualquer dos conglomerados possa ser representativo da população em estudo. Assim, para evitar o erro sistemático que possa estar associado a cada conglomerado, é necessário considerar o número elevado de conglomerados. Por isso, esse método, ao contrário da amostragem estratificada, pode necessitar de maior número de observações do que o método de amostragem aleatória simples para se produzir estimativas com a mesma qualidade. Esse tipo de amostragem pode ser obtido em estágio único, em dois estágios, em três estágios, etc., os quais são conhecidos como subamostragem. Um exemplo pode ser, uma pesquisa a realizada numa cidade que tem 10.000 residências e as famílias apresentam tamanho médio de 5 pessoas. Uma amostra de  $n = 600$  pessoas poderá ser obtida dividindo-se a cidade em bairros com aproximadamente 15 casas. Sorteando-se oito bairros e entrevistando todas as pessoas das residências seriam encontradas:  $8 \text{ bairros} \times 15 \text{ residências} \times 5 \text{ pessoas} = 600 \text{ pessoas}$  que constituiriam a amostra. Deve deixar claro que detalhes de ordem prática devem ser considerados, pois em geral nas pesquisas não há necessidade de se entrevistar todas as pessoas da família, sem contar ainda que as famílias apresentem em muitos casos crianças, menores ou mesmo pessoas idosas que muitas vezes não atenderiam os objetivos da amostragem.

**AMPLITUDE:** A diferença entre a maior e a menor observação em um conjunto de dados.

**AMPLITUDE DE REFERÊNCIA OU INTERVALO DE REFERÊNCIA:** A amplitude de valores de uma variável que define uma população saudável, normalmente calculada como sendo o intervalo que contém 95% das observações centrais dessa população; caso os dados apresentem distribuição Normal, é expressa por  $\text{média} \pm 1,96\text{DP}$ . Às vezes, é denominada amplitude normal.

**AMPLITUDE INTERQUARTIL:** A amplitude de valores que englobam 50% das observações centrais, caso as observações estejam disponibilizadas de modo classificatório.

**AMPLITUDE NORMAL:** Consultar amplitude ou intervalo de referência.

**ANÁLISE DE BERKSON-GAGE:** Um procedimento utilizado na análise quando os tempos de sobrevivência são agrupados em intervalos; isso serve como base para os cálculos dos métodos envolvidos nas tabelas de vida atuariais.



**ANÁLISE DE COVARIÂNCIA (ANCOVA):** Uma extensão da análise de variância, que leva em conta os valores de uma ou mais variáveis subsidiárias que são as covariáveis.

**ANÁLISE DE REGRESSÃO DE POISSON:** Uma forma específica de modelo linear generalizado com uma função de ligação do tipo log utilizada quando o resultado de interesse é a taxa que deveria ser constante durante o período de interesse; os indivíduos do estudo podem ter diferentes tempos de acompanhamento.

**ANÁLISE DE REGRESSÃO LINEAR:** Um processo formal de estimativa de coeficientes da equação de regressão linear, fazendo inferências a partir dela.

**ANÁLISE DE SOBREVIVÊNCIA:** A análise do tempo para um evento crítico, por exemplo, morte, em um grupo de animais no qual pode haver dados censurados.

**ANÁLISE DE VARIÂNCIA (ANOVA):** Consiste na decomposição dos graus de liberdade e da variância total de um material heterogêneo em partes atribuídas a causas conhecidas e independentes e a uma porção residual de origem desconhecida e de natureza aleatória.

**ANÁLISE DE VARIÂNCIA (ANOVA):** Um conjunto poderoso de procedimentos estatísticos paramétricos para a análise dos dados que compara, essencialmente, as médias de vários grupos de dados. Isso confia na separação da variação total de uma variável em seus componentes que estão associados com fontes de variação definidas.

**ANÁLISE MULTIVARIADA:** Um termo geral tradicionalmente utilizado para descrever um número de técnicas que avaliam diversas variáveis-resposta ou dependentes, simultaneamente, quando cada indivíduo assume um valor para cada uma das variáveis.

**ANÁLISES INTERINAS:** Toma-se uma decisão antes que se inicie a investigação do tamanho da amostra fixa, para realizar a análise dos dados em um período predeterminado, antes do final da investigação. Desse modo, surgiu o termo teste de significância repetido.

**ANÁLISES UNIVARIADAS:** A análise dos dados que compreende uma única variável dependente ou resposta.

**ANCOVA:** Consultar análise de covariância (ANCOVA).

**ANOVA:** Consultar análise de variância (ANOVA).

**ANOVA BIDIRECIONAL:** Uma análise dos dados que analisa o efeito de dois fatores em uma variável resposta, quando cada um desses fatores tem dois ou mais níveis.

**ANOVA DIRECIONAL DE FRIEDMAN:** Tipo de análise equivalente não-paramétrico à análise de variância (ANOVA) bidirecional.

**ANOVA UNIDIRECIONAL:** Uma extensão do teste  $t$  com duas amostras, utilizado quando pretendemos comparar as médias de mais de dois grupos de observações independentes.

**ANOVA UNIDIRECIONAL COM MEDIDAS REPETIDAS:** Pode ser considerado uma extensão do teste  $t$  pareado quando as médias são comparadas em três ou mais grupos de observações relacionadas.

**ANOVA UNIDIRECIONAL DE KRUSKAL-WALLIS:** Um procedimento não-paramétrico equivalente à análise de variância (ANOVA) unidirecional, utilizado para comparar grupos de observações independentes.

**AUTOPAREAMENTO:** Tipo de tratamento onde o animal é o seu próprio controle em um teste clínico e recebe ambos os tratamentos em ordem aleatória.

**AMOSTRA AGRUPADA:** Este método usa grupos de elementos, em vez de elementos individuais, como unidades de amostras. A partir de listas completas de agrupamentos da população, obtém-se os grupos amostra.

**AMOSTRA ALEATÓRIA:** Uma amostra selecionada de tal maneira que a seleção de um membro da população de nenhuma forma afeta a probabilidade de seleção de qualquer outro membro.

**AMOSTRA ALEATÓRIA:** Amostra selecionada de tal maneira que cada elemento da população tenha a mesma chance de ser escolhido.

**AMOSTRA ALEATÓRIA:** Processo ou técnica que envolve a seleção de um pequeno grupo para estudo retirado de um grupo muito maior de potenciais indivíduos de estudo.

**AMOSTRA ALEATÓRIA:** É aquela em que as probabilidades de seleção são as mesmas para cada elemento selecionado

**AMOSTRA ALEATÓRIA:** Amostra coletada por um método de seleção aleatória. Em inglês random sample.

**AMOSTRA ALEATÓRIA:** Amostra escolhida por um método tal que cada amostra possível tenha a mesma chance de ser escolhida; também, uma coleção de variáveis aleatórias independentes escolhidas todas elas da mesma distribuição.

**AMOSTRA ALEATÓRIA:** Tipo de amostra composta ao acaso sujeitas às leis do acaso onde as probabilidades operam em tais circunstâncias para torná-la representativa da população. Tipos principais dessa modalidade de amostra: aleatória simples, sistemática, estratificada e por conglomerados.

**AMOSTRA ALEATÓRIA:** Amostra dos elementos de uma população ou universo, extraída de tal forma que cada elemento da população tenha a mesma probabilidade de figurar na amostra. Sinônimo de amostra ao acaso, randômica, não viesada, equiprovável. Em inglês random sample.

**AMOSTRA ALEATÓRIA, CASUAL, RANDÔMICA:** Amostra escolhida sem a interferência do pesquisador, obtida por sorteio, pelos números aleatórios ou por procedimento computadorizado denominados de pseudo-randômico, constituindo, do ponto de vista matemático, amostra probabilística.

**AMOSTRA ALEATÓRIA iid:** É uma amostra em que as observações são variáveis aleatórias independentes e têm a mesma distribuição de probabilidade da população, por isso é denominada de amostra iid que significa com variáveis aleatórias independentes e igualmente distribuídas.

**AMOSTRA ALEATÓRIA SIMPLES:** Uma amostra selecionada de tal maneira que todos os membros da população têm uma igual probabilidade de serem escolhidos. No caso de amostragem sem reposição, de uma população finita, todas as possíveis amostras de um dado tamanho têm a mesma probabilidade de serem selecionadas.

**AMOSTRA ALEATÓRIA SIMPLES:** Amostra de determinado tamanho, selecionada de modo que todas as amostras possíveis de mesmo tamanho tenham a mesma chance de ser escolhidas.

**AMOSTRA ALEATÓRIA SIMPLES:** Tipo de amostra probabilística na qual as unidades que formam uma população recebem números; gera-se então um conjunto de números aleatórios, e as unidades com estes números são incluídas na amostra. Apesar de a teoria da probabilidade e os cálculos que ela prevê suporem este método básico de amostragem, por motivos práticos ele é raramente usado. Uma alternativa equivalente é a amostra sistemática, com início aleatório.

**AMOSTRA ALEATÓRIA SIMPLES:** Amostragem probabilística na qual a amostra é selecionada de tal forma que todos os elementos da população têm uma probabilidade igual e conhecida de inclusão na amostra.

**AMOSTRA AMODAL:** É uma amostra que não tem moda.

**AMOSTRA AUTOCORRIGIDA:** É a que foi obtida com um sistema de referência sabidamente incompleto, mas em que a técnica de seleção foi tal que as unidades amostrais não incluídas no sistema de referência puderam fazer parte da amostra com a mesma probabilidade das incluídas.

**AMOSTRA AUTOPONDERADA:** O mesmo que amostra proporcional.

**AMOSTRA BIMODAL:** É uma amostra que tem duas modas.

**AMOSTRA BOLA DE NEVE:** Metodologia de amostragem não-probabilística frequentemente usada em pesquisa de campo. Pede-se a cada participante sugerir outras pessoas para serem entrevistadas.

**AMOSTRA BOLA DE NEVE:** Uma amostra não probabilística na qual a seleção de respondentes adicionais é baseada nas referências fornecidas pelos entrevistados iniciais.

**AMOSTRA CASADA:** Plano experimental constituído por um conjunto de  $n$  pares  $[U_{1i}, U_{2i}]$  de unidades, tais que existe o máximo possível de homogeneidade entre as unidades experimentais do mesmo par, no tocante aos fatores estranhos ao que está sendo estudado. Usa-se com o fim de aplicar ao conjunto  $U_{1i}$  o tratamento  $T_1$  e, ao dos  $U_{2i}$ , o tratamento  $T_2$ , para, finalmente, comparar os resultados.

**AMOSTRA CASUAL:** O mesmo que amostra acidental, simples ou randômica.

**AMOSTRA CENSURADA:** É a amostra na qual a informação sobre a variável dependente não está disponível para algumas observações, embora a informação sobre os regressores ou variável independente  $X$  possa estar disponível.

**AMOSTRA CONTROLADA:** Amostra extraída de tal sorte que certas características existam na mesma proporção tanto na amostra quanto na população. Por exemplo, se na população houver 40% de mulheres e 30% de estrangeiros, a amostra deverá guardar as mesmas proporções: uma amostra de 100 indivíduos terá 40 mulheres e 30 estrangeiros. Em inglês controlled sample.

**AMOSTRA DEPENDENTE:** Amostra cujos valores estão relacionados com os valores de outra amostra.

**AMOSTRA DETERMINÍSTICA OU AMOSTRA NÃO ALEATÓRIA:** Amostra selecionada mediante um procedimento não aleatório. Estes tipos de amostras não podem ser usadas para fazer inferência. Em inglês nonrandom sample.

**AMOSTRA DE ANÁLISE:** Grupo usado para estimar a(s) função (ões) discriminante(s) ou o modelo de regressão logística. Quando se constroem matrizes de classificação, a amostra original é dividida

aleatoriamente em dois grupos, um para estimação do modelo que é a amostra de análise e o outro para validação a qual é a amostra de teste.

**AMOSTRA DE CONVENIÊNCIA:** Amostra não probabilística usada basicamente porque os dados são fáceis de serem coletados.

**AMOSTRA DE DISTRIBUIÇÃO:** Distribuição de frequência de todos os elementos de uma amostra individual.

**AMOSTRA DE JULGAMENTO:** Amostra não probabilística na qual os critérios de seleção são baseados no julgamento pessoal de que os elementos sejam representativos da população sob estudo.

**AMOSTRA DE PROBABILIDADE DE ÁREA:** Forma de amostra por conglomerado em múltiplas etapas, na qual áreas geográficas, como bairros, servem como unidade de amostragem na primeira etapa. Unidades selecionadas na primeira etapa da amostragem são então listadas todas as residências em cada quarteirão selecionadas são anotadas após visita ao quarteirão, e estas listas são sub-amostradas.

**AMOSTRA DE ROCHA DESMONTADA:** Coletada sobre a pilha de desmonte, procura representar um ponto do depósito mineral e destina-se ao controle de teores do minério a ser enviado à usina de beneficiamento. Este tipo de amostragem deve ser feito com cuidado, pois o material desmontado apresentará granulações diferentes dependendo da resistência de cada um dos componentes do minério. Uma amostra representativa da pilha de desmonte, pode ser composta a partir de pequenas porções de minério coletadas sobre os nós de um arranjo regular. Para evitar erros de amostragem devido a diferentes granulações, recomenda-se coletar toda a porção extraída de pequenos poços abertos nos pontos predeterminados.

**AMOSTRA DE TESTE:** Grupo de objetos não usados quando a(s) função(ões) discriminante(s) ou o modelo de regressão logística é computado. Esse grupo é então usado para validar a função discriminante ou o modelo de regressão logística em uma amostra separada de respondentes. É também chamada de amostra de validação.

**AMOSTRA DE TREINAMENTO:** Observações usadas na calibração de uma rede neural. Deve conter valores reais para a interconexão de saída, tal que erros na previsão do valor de saída podem ser determinados e usados no processo de aprendizado.

**AMOSTRA DE VALIDAÇÃO:** Ver amostra de teste.

**AMOSTRA EQUILIBRADA:** É a amostra acidental em que algumas unidades amostrais que fazem parte da amostra foram substituídas por outras, também escolhidas ao acaso da mesma população, a fim de que a média aritmética, da amostra fique igual, ou aproximadamente igual, a média aritmética da população originária.

**AMOSTRA ESTABILIZADA:** O mesmo que amostra equilibrada.

**AMOSTRA ESTRATIFICADA:** Reunião,  $S_N$ , das amostras independentes  $S_{a_1}, S_{a_2}, \dots, S_{a_m}$ , onde  $n$  denota o tamanho e  $\sum_{i=1}^m n_i = N$ , tais que  $S_{a_i}$  ( $i=1, 2, \dots, m$ ) é oriunda do  $i$ -ésimo estrato de uma dada população.

**AMOSTRA ESTRATIFICADA:** Quando os sujeitos num estudo de população podem ser agrupados em subclasses ou estratos e a proporção em cada subclasse é já conhecida, cada estrato é amostrado



independentemente. Quando a amostragem é do tipo proporcional, cada estrato é amostrado aleatoriamente ou aleatoriamente, independentemente de outros estratos, tirando casos suficientes de um dado estrato, de forma a fazer seu tamanho na amostra total proporcional a porcentagem de sujeitos incluídos naquele estrato na população.

**AMOSTRA ESTRATIFICADA:** Amostra retirada de camadas da população denominadas estratos não-superpostos, caracterizando-se, de um lado, pela maior homogeneidade da variável investigada dentro cada estrato e, de outro, pela maior heterogeneidade entre as camadas, como, por exemplo, nível socioeconômico de população urbana: baixo, médio e elevado. De cada camada retiram-se amostras aleatórias de tamanho proporcional a cada estrato.

**AMOSTRA ESTRATIFICADA:** Amostra probabilística que seleciona elementos de subconjuntos relevantes da população para serem mais representativos de toda a população.

**AMOSTRA ESTRATIFICADA DE CONGLOMERADOS:** É a amostra estratificada em que as unidades de amostragem são conglomerados. Ver também amostragem em estágios múltiplos.

**AMOSTRA ESTRATIFICADA SIMPLES:** É a amostra estratificada obtida por amostragem acidental simples.

**AMOSTRA FALHA:** É dita aquela amostra a que, ao fim do processo de seleção, contém um número de elementos menor que o previsto como necessário à precisão requerida.

**AMOSTRA IMPARCIAL:** É uma amostra em que todos os elementos tiveram igual oportunidade de fazer parte da mesma.

**AMOSTRA INDEPENDENTE:** Amostra cujos valores não estão relacionados com os valores de outra amostra.

**AMOSTRA INTENCIONAL:** É aquela cujos elementos foram todos eles escolhidos apenas por amostragem intencional.

**AMOSTRA INTENCIONAL:** Ver amostra por julgamento.

**AMOSTRA INTERNA IRRESTRITA:** Pesquisa montada na internet e acessível a qualquer pessoa que deseje completá-la.

**AMOSTRA MIPS:** Método de igual probabilidade de seleção. Desenho de amostra em cada membro da população tem a mesma chance de ser selecionado na amostra.

**AMOSTRA MULTIMODAL:** É uma amostra que tem mais do que duas modas.

**AMOSTRA NÃO-ALEATÓRIA (POR VEZES DITA SELECIONADA, ACIDENTAL OU VICIADA):** Tipo de plano amostral escolhido por conveniência ou por julgamento de valor; as leis do acaso não operam em tais circunstâncias.

**AMOSTRA NÃO-PROBABILÍSTICA:** Amostra selecionada de forma não sugerida pela teoria da probabilidade. Exemplos incluem amostras por julgamento que são as intencionais, por quota e bola de neve.

**AMOSTRA NÃO-PROBABILÍSTICA:** Amostra que inclui elementos da população, selecionados de maneira não aleatória.

**AMOSTRA PERMANENTE:** É a que se conserva fixa, em repetições da mesma pesquisa, em épocas sucessivas, a fim de se verificarem as variações sofridas pelo atributo em causa.

**AMOSTRA POR AGRUPAMENTO:** Amostra probabilística extraída de áreas geográficas.

**AMOSTRA POR CONGLOMERADOS EM MÚLTIPLAS ETAPAS:** Amostra em múltiplas etapas em que grupos naturais ou clusters são inicialmente amostrados, com os membros de cada grupo selecionado sendo em seguida sub-amostrados. Exemplo: selecionar num guia escolar uma amostra de faculdades e universidades, conseguir listas dos alunos em todas as escolas selecionadas e, em seguida, amostrar alunos de cada uma. Ver ainda amostra probabilística por área.

**AMOSTRA POR CONGLOMERADOS:** Amostra probabilística cujas unidades simples são obtidas de modo randômico de unidades coletivas, pressupondo-se que estas últimas apresentem homogeneidade entre si. Colméias, cardumes, blocos residenciais, são exemplos de unidades coletivas de onde são retiradas amostras randômicas.

**AMOSTRA POR COTA:** Tipo de amostra não-probabilística em que unidades são selecionadas para a amostra com base em características pré-especificadas, de modo que a amostra total tenha a mesma distribuição de características que se supõe existir na população estudada.

**AMOSTRA POR COTAS:** Amostra não probabilística na qual um subgrupo da população é classificado na base do julgamento do pesquisador.

**AMOSTRA POR JULGAMENTO:** Tipo de amostra não-probabilística em que você seleciona as unidades a serem observadas com base no seu próprio julgamento sobre quais delas serão mais úteis ou representativas.

**AMOSTRA PRINCIPAL:** É uma amostra representativa em relação a certas características populacionais, que serve como população para a seleção de outras amostras.

**AMOSTRA PROBABILÍSTICA:** Termo genérico para amostra selecionada conforme a teoria da probabilidade, tipicamente com algum mecanismo de seleção aleatória, onde cada elemento da população possui uma probabilidade conhecida e diferente de zero de fazer parte da amostra. Tipos específicos de amostras probabilísticas são amostras probabilísticas por área (APA), amostra com método de igual probabilidade de seleção (MIPS), método de amostragem de probabilidade proporcional ao tamanho (PPT) e amostra sistemática (AS).

**AMOSTRA PROBABILÍSTICA:** É a amostra com as duas características seguintes: i) as probabilidades de seleção são iguais para todos os membros da população útil em todos os estágios do processo de seleção, e ii) a amostragem é realizada com os elementos da amostra selecionados independentemente uns dos outros. As amostras probabilísticas são, muitas vezes, chamadas de amostras randômicas.

**AMOSTRA PROBABILÍSTICA:** Amostra na qual todos os elementos da população tem uma probabilidade de seleção conhecida e diferente de zero.

**AMOSTRA PROPORCIONAL:** É a amostra estratificada de tamanho  $n$ , extraída de uma população de tamanho  $N$ , de modo que, de cada estrato de tamanho  $N_i$  se extraem  $n_i = n \left( \frac{N_i}{N} \right) = np_i$  elementos. Também se diz amostra representativa.

**AMOSTRA RANDÔMICA:** É uma amostra obtida de tal forma que cada e todo elemento tem igual chance de ser escolhido em cada vez que um elemento é retirado da população.

**AMOSTRA TRUNCADA:** É aquela na qual a informação sobre os regressores (variável independente X) está disponível somente se o regressando é observado.

**AMOSTRA RECRUTADA DA INTERNET:** Levantamento controlado estabelecido na internet que é usado para populações-alvo.

**AMOSTRA REPRESENTATIVA:** É aquela que deve conter em proporção todas as características qualitativas e quantitativas da população.

**AMOSTRA REPRESENTATIVA:** É a obtida por um processo isento de viés. É aquela cujos elementos típicos apresentam valores muito próximos dos correspondentes parâmetros.

**AMOSTRA REPRESENTATIVA:** Amostra não tendenciosa e suficientemente ampla de modo a possuir as mesmas características da população de que foi extraída. Em inglês representative sample.

**AMOSTRA SELECIONADA DA INTERNET:** Uma pesquisa montada na internet que restringe os respondentes ao impor cotas baseadas em algumas características da amostra.

**AMOSTRA SIMPLES:** É a obtida por amostragem simples.

**AMOSTRA SISTEMÁTICA:** É aquela obtida por meio de amostragem sistemática.

**AMOSTRA SISTEMÁTICA:** Tipo de amostra probabilística no qual cada unidade  $k^a$  numa lista é selecionada para inclusão em amostra, por exemplo, cada vigésimo quinto aluno na listada escola. Computa-se o  $k$  dividindo-se o tamanho da população pelo tamanho desejado da amostra, e  $k$  é o intervalo de amostragem. Dentro de certas limitações, a amostragem sistemática é o equivalente funcional da amostragem aleatória simples e normalmente mais fácil de fazer. Tipicamente, a primeira unidade é selecionada aleatoriamente.

**AMOSTRA SISTEMÁTICA:** Amostra probabilística cuja primeira unidade é obtida ao acaso e as demais, a partir da primeira, escolhidas a cada  $k$  intervalo sistemático, definido pela razão entre o tamanho da população e o tamanho da amostra.

**AMOSTRA SISTEMÁTICA:** Nesta modificação, sujeitos ou unidades amostrais são selecionados sequencialmente de uma lista, em intervalo fixo, como, por exemplo, cada sujeito para um emprego será o quinto cartão de um fichário. Determine-se o ponto de partida por meio de escolha de um número ao acaso.

**AMOSTRA SISTEMÁTICA:** Amostragem probabilística na qual toda a população é numerada e os elementos extraídos pelo uso de intervalos seqüenciais.

**AMOSTRA, TAMANHO:** É a fórmula para computar ou dimensionar o tamanho de uma amostra para estimar a média de uma população é a seguinte:  $n = \frac{Z^2 \cdot \sigma_p^2}{E^2}$  onde  $n$  = tamanho da amostra;  $z$  = probabilidade crítica em termos de desvio padrão normal;  $\sigma_p$  = desvio padrão da população (estimada);  $E$  = erro permitido ( $\frac{1}{2}$  do intervalo de confiança estimado).

bilidade crítica em termos de desvio padrão normal;  $\sigma_p$  = desvio padrão da população (estimada);  $E$  = erro permitido ( $\frac{1}{2}$  do intervalo de confiança estimado).

**AMOSTRA TENDENCIOSA:** Amostra viciada por diferenças sistemáticas que impedem seja considerada representativa da população de que foi retirada, no que se refere a uma ou mais características. Em inglês biased sample.

**AMOSTRA VOLUMÉTRICA:** Refere-se a amostra que é representada por todo o material retirado durante uma escavação (trincheiras, galerias, chaminés, dentre outros). Este tipo de amostra é indicado para tratamentos em usina piloto para definição de parâmetros de beneficiamento, ou então para determinação da quantidade exata de um elemento, cuja concentração é muito pequena (ouro, platina, diamante), com pequena variância amostral, tendo em vista a relação volume x variância mantida constante.

**AMOSTRAS INDEPENDENTES:** Amostras nas quais a mensuração de uma variável em uma população não tem efeito sobre a mensuração da variável em outra.

**AMOSTRAS RELACIONADAS:** Amostras nas quais a mensuração de uma variável em uma população pode influenciar a mensuração da variável em outra.

**AMOSTRAGEM:** É o método de retirada de amostras de uma população. Em inglês sampling.

**AMOSTRAGEM:** É o ato ou processo de seleção de amostra para ser analisada como representativa de um todo e compreende todas as etapas que vão desde a coleta, o manuseio, a preparação até a sua determinação analítica.

**AMOSTRAGEM:** Procedimento de estudo de uma parte do universo. Sinônimo: Sondagem.

**AMOSTRAGEM:** Operação pela qual se constitui um grupo de elementos cuja a análise ou estudo dará informações sobre as características da população em que os elementos foram colhidos. Em inglês sampling.

**AMOSTRAGEM:** Processo de seleção de unidades que deverão compor a amostra; pode ser feita, por exemplo, por meio da tabela de números aleatórios.

**AMOSTRAGEM:** Processo ou método de conceber um número finito de indivíduos ou casos de uma população ou universo, para produzir um grupo representativo. Usado em circunstâncias em que é difícil obter informações de todos os membros da população, como, por exemplo, análises biológicas, controle de qualidade industrial e levantamento de dados sociais. É um método indutivo de conhecimento de todo o universo estatístico, por meio de um número representativo de amostras aleatórias desse universo.

**AMOSTRAGEM (QUESTIONÁRIO) (NÃO-RESPONDENTES):** As pessoas que respondem a um questionário de qualquer tipo são obviamente diferentes das pessoas que não respondem, quer porque não foram contatadas, quer porque se recusaram a responder. Em geral há três formas de cuidar dos não-respondentes, para reduzir os efeitos de tendência na amostragem. i) fazer repetidos contatos com todos que não responderam na primeira vez; ii) obter uma sub-amostra dos não-respondentes. A fim de facilitar as respostas dos não-respondentes amostrados, costuma-se reduzir o questionário às perguntas essenciais; e iii) se depois da sub-amostragem ainda houver nela alguns não-respondentes duros de dobrar, por não serem encontrados ou por simples recusa, o pesquisador usa a técnica de assumir os extremos. Pressupõe-se que alguns desses não-respondentes reagiriam de uma forma e outros de forma contrária.

**AMOSTRAGEM ACIDENTAL:** Trata-se de uma amostra formada por aqueles elementos que vão aparecendo, que são possíveis de se obter até completar o número de elementos da amostra. Geralmente utilizada em pesquisas de opinião, em que os entrevistados são acidentalmente escolhidos.

**AMOSTRAGEM ACIDENTAL:** É aquela em que a seleção de um elemento da população depende de dada lei probabilística. Distingue de amostragem intencional e de amostragem a esmo, e é o mesmo que amostragem probabilística. É o mesmo que amostragem equiprobabilística.

**AMOSTRAGEM ACIDENTAL IRRESTRITA:** É aquela que se faz diretamente da população total. Usa-se para marcar a distinção para com os processos em que existem passos intermediários, intencionais ou probabilísticos, tais como estratificação, fases, dentre outros.

**AMOSTRAGEM ACIDENTAL SIMPLES:** Na terminologia de alguns autores, é o mesmo que amostragem acidental irrestrita.

**AMOSTRAGEM ALEATÓRIA ESTRATIFICADA:** É designada para amostragem de depósitos minerais que são constituídos por dois ou mais tipos de rochas ou camadas com características e teores diferentes de mineralização que são os estratos. A amostragem deve ser planejada para ser aleatória em cada um desses estratos, o que tende a reduzir o erro de estimativa do teor médio do depósito, com pequena variância dentro dos estratos. A amostragem aleatória estratificada fornece um estimador mais eficiente que a amostragem aleatória simples se a variância dentro dos estratos for menor que a variância populacional.

**AMOSTRAGEM ADAPTATIVA:** Procedimento de seleção de amostra em que o plano de amostragem vai sendo construído durante o levantamento, em resposta aos padrões observados na população. Por exemplo, num levantamento para estimar a abundância de uma espécie, parcelas adicionais podem ser acrescentadas próximas de parcelas que apresentaram alta abundância da espécie de interesse. As aplicações mais comuns do método têm sido em levantamentos de populações de animais, plantas e minerais, e com potencial de uso em epidemiologia e controle de qualidade.

**AMOSTRAGEM A ESMO:** É o processo de seleção que, não sendo sujeito a uma disciplina mecânica, fica entregue ao critério do trabalhador de campo, este lança mão de diversos recursos de bom senso, nem sempre isentos de opinião pessoal a fim de obter uma amostra cuja estrutura se aproxime da de amostra acidental. Ou seja ele procura ser aleatório sem utilizar um dispositivo aleatório propriamente dito para realizar o sorteio.

**AMOSTRAGEM ALEATÓRIA:** A amostragem é dita aleatória, probabilística ou ao acaso se todos os elementos da população tiveram probabilidade conhecida e diferente de zero de pertencerem a amostra. Em inglês random sample.

**AMOSTRAGEM ALEATÓRIA SIMPLES:** Delineamento básico de amostragem probabilística, em que as unidades amostrais devem ser selecionadas completamente ao acaso, com  $n$  unidades diferentes selecionadas de  $N$  unidades na população, de forma que cada possível combinação das  $n$  unidades amostrais (ou amostra) tenha a mesma probabilidade de ser selecionada.

**AMOSTRAGEM ALEATÓRIA SIMPLES:** A amostragem é dita aleatória simples se todos os elementos da população possuírem a mesma probabilidade de pertencerem a amostra. Em inglês simple random sampling.

**AMOSTRAGEM ALEATÓRIA SIMPLES:** Consiste na seleção de uma amostra de tamanho  $n$ , retirada de uma população, de tal modo que cada unidade básica de amostragem tenha a mesma probabilidade de ser selecionada. Caso esta condição seja satisfeita, esta amostragem proporciona estimadores não enviesados

e, portanto, os limites de confiança podem ser calculados para a média e, se a variância populacional é conhecida, o tamanho  $n$  da amostra pode ser calculado antes da amostragem.

**AMOSTRAGEM AO ACASO:** O mesmo que amostragem acidental.

**AMOSTRAGEM CASUAL OU ALEATÓRIA SIMPLES:** É aquela em que qualquer elemento da população tem a mesma probabilidade de ser escolhido.

**AMOSTRAGEM BOLA DE NEVE:** Um tipo de amostra não probabilística na qual o pesquisador identifica alguns entrevistados e lhes pede para identificar outros que também possam vir a ser entrevistados

**AMOSTRAGEM CASUAL OU ALEATÓRIA SIMPLES:** É o processo mais elementar e freqüentemente utilizado. É equivalente a um sorteio lotérico. Pode ser realizada numerando-se a população de 1 a  $n$  e sorteando-se, a seguir, por meio de um dispositivo aleatório qualquer,  $x$  números dessa sequência, os quais corresponderão aos elementos pertencentes à amostra.

**AMOSTRAGEM COM PROBABILIDADE PROPORCIONAL AO TAMANHO:** Delineamento de amostragem em que a probabilidade de seleção não é a mesma para todas as unidades amostrais, sendo proporcional ao tamanho de cada unidade. Por exemplo, na amostragem pontual horizontal as maiores árvores têm maior probabilidade de seleção, tornando a amostragem proporcional à área basal das árvores. Além disso, esse método pode estar relacionado à seleção das unidades primárias, proporcional ao tamanho das unidades amostrais, quando são utilizados métodos de amostragem em múltiplos estágios. Ver também: amostragem em múltiplos estágios.

**AMOSTRAGEM COM REPOSIÇÃO:** Quando uma unidade de amostragem é extraída de uma população finita e é retornada para aquela mesma população, depois das suas características serem registradas, a amostragem é chamada de com reposição. Caso contrário, a amostragem será sem reposição. Em inglês sampling with replacement.

**AMOSTRAGEM COM REPOSIÇÃO:** Processo de seleção em que cada unidade amostral escolhida é, após sua observação, devolvida à população originária, antes de se proceder a nova extração.

**AMOSTRAGEM COM REPOSIÇÃO:** Método de escolha de uma amostra em que o elemento escolhido é devolvido à população, tendo, assim, chance de ser escolhido novamente.

**AMOSTRAGEM COM SUBSTITUIÇÃO PARCIAL:** Processo de seleção em que, numa sequência de repetições da mesma pesquisa, parte das unidades amostrais é conservada, e parte substituída por outras, de repetição para repetição.

**AMOSTRAGEM CONFIGURACIONAL:** É a amostragem de áreas por conglomerados de células básicas de uma graticula (figura de uma rede de pautas, geralmente ortogonais, que se constrói sobre o papel de desenho com o fim de facilitar a construção e a leitura de um gráfico), que obedece a uma configuração pré-estabelecida.

**AMOSTRAGEM CONTÍNUA:** Amostragem realizada sem interrupções, realizada ao longo de toda uma operação e por um tempo pré-determinado.

**AMOSTRAGEM DE ACEITAÇÃO:** Um procedimento estatístico no qual o número de itens defeituosos encontrados em uma amostra é usado para determinar se um lote deve ser aceito ou rejeitado.



**AMOSTRAGEM DE ÁREAS:** É aquela em que as unidades amostrais são áreas. Também se diz amostragem na base de áreas.

**AMOSTRAGEM DE CONVENIÊNCIA:** Tipo de amostra não probabilística na qual os entrevistados são selecionados de acordo com sua semelhança presumida com a população útil e sua pronta disponibilidade.

**AMOSTRAGEM DE CONVENIÊNCIA:** É aquela em que da respectiva população de interesse, se escolhe as unidades mais convenientes, no sentido de mais cômodo.

**AMOSTRAGEM DE CONVENIÊNCIA:** Amostragem em que dados são selecionados em função da facilidade de acesso aos mesmos.

**AMOSTRAGEM DE CONVENIÊNCIA:** Método de amostragem em que a amostra é composta de elementos cuja escolha é mais conveniente; não é aconselhável basear a análise estatística em amostras selecionadas por este processo.

**AMOSTRAGEM DE CONVINIÊNCIA:** Obtém-se este tipo de amostragem quando o pesquisador escolhe ou é forçado a escolher, por questões práticas, os sujeitos ou elementos que lhe são disponíveis. Embora às vezes não haja outra alternativa, tal amostragem raramente é representativa da população. Em tal situação, o pesquisador precisa dar detalhada descrição de sua amostra.

**AMOSTRAGEM DE JULGAMENTO:** Um método de não-probabilidade de amostragem em que os elementos são selecionados para a amostra com base no julgamento da pessoa que faz o estudo.

**AMOSTRAGEM DE LISTAS:** É aquela em se faz extrações por meio de uma lista, ou fichário, de endereços, nomes, dentre outros. é uma técnica comumente sistemática, com início casual.

**AMOSTRAGEM DE MASSA:** É aquela que é executada sobre uma população que se apresenta em massa, tal como um carregamento de trigo, feijão ou milho não ensacado, de carvão a granel, de petróleo fluindo num oleoduto, dentre outros.

**AMOSTRAGEM DE PONTOS:** É aquela que caracteriza-se pelo seguinte processo de escolha: se, de uma área total  $A$ , se deseja escolher um contorno de área  $a < A$ , começa-se por escolher um ponto no interior do contorno de área  $A$ , de coordenadas casuais e, em seguida, por meio de uma convenção, toma-se esse ponto para localizar  $a$ .

**AMOSTRAGEM DE ZONAS ELEITORAIS SELECIONADAS:** Técnica de amostragem não-probabilística às vezes usada em pesquisa de opinião política. Baseada em padrões eleitorais passados, a pesquisa estabelece um conjunto de zonas eleitorais que, no todo, parecem representativas dos eleitores em toda a jurisdição estudada. O risco é que as zonas amostradas e/ou toda a jurisdição podem ter mudado o bastante para que a amostra já não seja representativa do todo.

**AMOSTRAGEM DIRETA:** Diferentemente do que a expressão sugere, é o processo em que a observação do atributo em causa se faz diretamente sobre os seus portadores, e não sobre uma fonte secundária.

**AMOSTRAGEM DUPLA:** Delineamento de amostragem em que, numa primeira fase, é medida uma variável auxiliar necessária para a estimação, normalmente de medição mais rápida e a um custo menor que a variável de interesse. Somente na segunda fase é que são medidas simultaneamente a variável auxiliar e a variável de interesse.

**AMOSTRAGEM DUPLA:** É aquela que se processa em duas fases. Ver amostragem multifásica.

**AMOSTRAGEM DUPLICADA:** É aquela em que a mesma amostra é observada por dois conjuntos de pessoas.

**AMOSTRAGEM EM ESTÁGIOS MÚLTIPLOS:** É aquela que se desenvolve numa sucessão de passos, ou estágios, de acordo com o seguinte esquema: no 1º estágio, de uma população constituída por unidades amostrais chamadas de 1º estágio, ou primárias, extrai-se uma amostra acidental; no segundo estágio, define-se uma nova unidade amostral, chamada de 2º estágio, ou secundária, e definem-se tantas populações de unidades secundárias quantas foram as unidades primárias; extrai-se, em seguida, uma amostra acidental, dentre outras. Exemplo: querendo-se estimar a população de uma cidade, podemos no 1º estágio, considerar como população o conjunto dos quarteirões da cidade, e extrair uma amostra acidental de quarteirões. No 2º estágio, definimos como unidade amostral secundária os domicílios, tomando tantas populações de domicílios quantos foram os quarteirões que entraram para a amostra de 1º estágio; a seguir, escolhemos uma amostra acidental de domicílios, de cada população assim definida. Também se diz amostragem por etapas múltiplas, ou polietápica.

**AMOSTRAGEM EM ESTÁGIOS MÚLTIPLOS SIMPLES:** É a amostragem em estágios múltiplos em que a seleção é feita irrestritamente de toda a população. Distingue de amostragem em estágios múltiplos estratificada.

**AMOSTRAGEM EM FASES MÚLTIPLAS:** O mesmo que amostragem multifásica.

**AMOSTRAGEM EM MÚLTIPLOS ESTÁGIOS:** Delineamento utilizado em grandes áreas, considerando-se diversas escalas de mensuração com o objetivo de obter estimativas mais confiáveis para a maior escala. Para isso, a população de interesse é dividida em unidades primárias, que por sua vez são subdivididas em unidades secundárias e assim por diante, dependendo do número de estágios. A amostragem é feita pela seleção de um conjunto de unidades primárias, dentro das quais é selecionado e medido um conjunto de unidades secundárias e assim sucessivamente.

**AMOSTRAGEM EQUIPROBABILÍSTICA:** É aquela que dá a todas as unidades amostrais, ainda na população, ou sistema de referência, a mesma probabilidade de virem a fazer parte da amostra.

**AMOSTRAGEM ESTATÍSTICA:** A composição da amostra é definida pela seleção de unidades amostrais baseada em procedimentos probabilísticos, que garantem a representatividade da amostra e a possibilidade de desenvolver estimadores para os parâmetros de interesse e de conhecer suas propriedades estatísticas.

**AMOSTRAGEM ESTRATIFICADA:** Delineamento amostral em que a população é dividida em estratos relativamente homogêneos e cada estrato é então amostrado separadamente, utilizando-se amostras aleatórias. Em geral, as variáveis consideradas para estratificar uma floresta são a idade, classe de produtividade, qualidade de sítio, composição de espécies, topografia dentre outras. A amostragem estratificada em geral produzirá estimativas com menor erro padrão, em relação ao que seria obtido sem a estratificação. Uma vantagem inerente ao método é a obtenção de estimativas para cada estrato e para a população como um todo.

**AMOSTRAGEM ESTRATIFICADA:** Método de amostragem na qual os elementos são retirados de subpopulações agrupadas por algum critério. Em inglês stratified sampling.

**AMOSTRAGEM ESTRATIFICADA:** Um procedimento de amostragem por meio do qual a população é dividida em duas ou mais subpopulações chamadas estratos. Os elementos da amostra são aleatoriamente retirados de cada estrato.

**AMOSTRAGEM ESTRATIFICADA:** É aquela em que a população está dividida em estratos ou grupos diferenciados.

**AMOSTRAGEM ESTRATIFICADA:** Método de amostragem em que a população é dividida em camadas, ou estratos, tão semelhantes entre si quanto possível.

**AMOSTRAGEM ESTRATIFICADA:** Processo de obtenção da amostra estratificada, é precedido pela decomposição da população, ou sistema de referência, em estratos.

**AMOSTRAGEM ESTRATIFICADA:** Processo de amostragem em que as amostras são extraídas de estratos (classes).

**AMOSTRAGEM ESTRATIFICADA EM ESTÁGIOS MÚLTIPLOS:** É aquela que resulta numa amostra estratificada de conglomerados. Distingue de amostragem em estágios múltiplos simples.

**AMOSTRAGEM EXTENSA:** É aquela que cobre uma área muito grande, sem muitos pormenores de observação. Refere-se ainda a amostragem em que cada unidade é observada em relação a muitos atributos, porém, superficialmente. Também é conhecida como aquela que se estende por um período de tempo muito longo. Em cada caso mencionado anteriormente, esse tipo se opõe a amostragem intensa.

**AMOSTRAGEM INDIRETA:** É aquela que incide não sobre os portadores dos atributos a serem observados, mas sobre uma fonte secundária, tal como o arquivo das observações já feitas sobre a totalidade das unidades amostrais.

**AMOSTRAGEM, INFERÊNCIA:** Um grande problema de qualquer pesquisa é a inferência estatística. Tal problema surge porque o pesquisador quer generalizar para o universo os fatos que ele descobriu na sua amostra. Quando dizemos que uma pessoa está generalizando indevidamente queremos dizer que ela está aplicando os resultados de suas observações a um universo diferente daquele do qual ela obteve a amostra. Daí ser importante que o pesquisador defina o seu universo.

**AMOSTRAGEM INTENCIONAL:** Tipo de amostragem não probabilística na qual o pesquisador usa seu julgamento na seleção de entrevistados considerados bem informados nas áreas ligadas à pesquisa.

**AMOSTRAGEM INTENCIONAL:** Neste tipo de amostragem, o pesquisador tenta obter representatividade fazendo com que a amostra tenha as características do universo. Pressupõem-se que, sendo a amostra representativa de características conhecidas da população, ela também deverá ser representativa com relação a características desconhecidas que estão sendo pesquisadas. Um bom exemplo deste tipo de amostragem é a amostra por quota, isto é, obtém-se proporcionalidade das categorias já existentes na população.

**AMOSTRAGEM INTENCIONAL:** De acordo com determinado critério, é escolhido intencionalmente um grupo de elementos que irão compor a amostra. O investigador se dirige intencionalmente a grupos de elementos dos quais deseja saber a opinião.

**AMOSTRAGEM INTENCIONAL:** Processo de escolha no qual uma unidade amostral é selecionada para fazer parte da amostra em razão da alternativa por ela apresentada do atributo em causa.

**AMOSTRAGEM INVERSA:** Nome que se dá de forma inadequado à amostragem da qual, ao invés de prefixar-se o tamanho da amostra, o que se fixa de antemão é o número absoluto de indivíduos que apresentam uma dada alternativa de dado atributo, prosseguindo-se na extração até que esse número seja obtido.

**AMOSTRAGEM INTENSA:** Define-se por oposição a amostragem extensa.

**AMOSTRAGEM IRRESTRITAMENTE ACIDENTAL:** Ver amostragem acidental irrestrita.

**AMOSTRAGEM MISTA:** É aquela que comportando vários passos (estágios, fases, dentre outros), usa mais de um processo de seleção, nos sucessivos passos.

**AMOSTRAGEM MULTIFÁSICA:** É o processo de seleção em que há uma sequência de amostras  $S^1, S^2, S^3, \dots$  (que constituem as fases) tais que  $S^{i+1}$  é uma sub-amostra de  $S^i$ , todas elas com as mesmas unidades amostrais, sendo, porém, que os atributos estudados diferem de fase para fase. Exemplo: se o investigador pretende pesquisar consumo, pode-se, em primeiro lugar, escolher uma amostra baseada em rendimentos; dessa se extrai uma sub-amostra e nela se estuda o consumo. Também se diz amostragem por fases múltiplas.

**AMOSTRAGEM NA BASE DE ÁREAS:** O mesmo que amostragem de áreas.

**AMOSTRAGEM NÃO PROBABILÍSTICA:** Um método de seleção de amostras no qual a probabilidade de selecionar qualquer pessoa ou elemento para inclusão na amostra não é conhecida.

**AMOSTRAGEM, PLANEJAMENTO:** Ao planejar a nossa amostra devemos considerar três critérios: i) o pesquisador deve ser capaz de determinar a precisão da estimativa de sua amostra, a partir das observações que ele faz; ii) o plano de amostra escolhido deve ser administrativamente viável na situação em que ela será feita; iii) satisfeitos esses dois critérios, o melhor plano será aquele que maximiza o custo unitário, mesmo havendo uma pequena perda de precisão.

**AMOSTRAGEM POLIETÁPICA:** Ver amostragem em estágios múltiplos, que é o mesmo.

**AMOSTRAGEM PONTUAL HORIZONTAL:** Ver amostragem por enumeração angular.

**AMOSTRAGEM POR AGRUPAMENTO (MULTIESTÁGIO):** O processo de selecionar ao acaso uma amostra numa série hierárquica de estágios representados por grupos cada vez mais restritos da população útil.

**AMOSTRAGEM POR COTAS:** Tipo de amostra não probabilística na qual o pesquisador seleciona deliberadamente uma amostra para refletir a população em geral com relação a uma ou mais variáveis específicas consideradas importantes para o estudo.

**AMOSTRAGEM POR CAPTURA E RECAPTURA:** Para a estimação do teor numérico  $X$  de uma população animal. Consiste em marcar e soltar  $Y$  animais, pertencentes ou não a  $X$  e, mais tarde, recapturar uma amostra de tamanho  $x$ , anotando-se o número  $y$  dos que nela estão marcados. Então, na pressuposição, entre outras de que os  $Y$  animais se difundiram homoganeamente no meio dos  $X$  animais, será  $X = Y + \left(\frac{y}{x}\right)$ , onde  $\left(\frac{y}{x}\right)$  é chamado de índice de Lincoln.

**AMOSTRAGEM POR CONGLOMERADOS:** É aquela que usa conglomerados como unidades amostrais, as quais são grupos formados ou já cadastrados da população, sendo por natureza heterogêneos internamente, e todas as unidades amostrais dos conglomerados sorteados serão investigadas.

**AMOSTRAGEM POR CONGLOMERADO:** Método de amostragem em que a população é dividida em conglomerados; escolhem-se arbitrariamente alguns conglomerados e, em cada um deles, selecionam-se, também aleatoriamente, alguns elementos que irão compor a amostra.

**AMOSTRAGEM POR CONGLOMERADOS:** Método de amostragem onde o que é sorteado é o aglomerado e não a unidade individual. Em inglês cluster sampling.

**AMOSTRAGEM POR CONGLOMERADOS:** Ver amostragem por conglomerado em múltiplas etapas.

**AMOSTRAGEM POR CONGLOMERADOS (OU AGRUPAMENTOS):** Algumas populações não permitem, ou tornam extremamente difícil que se identifiquem seus elementos. Não obstante, isso pode ser relativamente fácil identificar alguns subgrupos da população. Em tais casos, uma amostra aleatória simples desses subgrupos (conglomerados) pode ser colhida, e uma contagem completa deve ser feita para o conglomerado sorteado.

**AMOSTRAGEM POR CONGLOMERADOS (OU AGRUPAMENTOS):** Dividir a área populacional em seções (ou conglomerados), selecionar aleatoriamente algumas dessas seções e escolher então todos os elementos das seções selecionadas.

**AMOSTRAGEM POR ENUMERAÇÃO ANGULAR:** Método de amostragem florestal em que a probabilidade de seleção de um indivíduo (árvore) é proporcional ao tamanho desse indivíduo, expresso pela sua área basal. A unidade amostral é um ponto, locado aleatoriamente, a partir do qual são selecionados os indivíduos que compõem a amostra.

**AMOSTRAGEM POR ETAPAS MÚLTIPLAS:** O mesmo que amostragem em estágios múltiplos.

**AMOSTRAGEM POR FASES MÚLTIPLAS:** O mesmo que amostragem multifásica.

**AMOSTRAGEM POR GRUPOS (CLUSTER SAMPLING):** Seleção de grupos de elementos (clusters) em vez de elementos singulares. Todos os elementos de cada grupo estão inclusos na amostra, e os grupos são selecionados aleatoriamente de uma população de grupos.

**AMOSTRAGEM POR IMPORTÂNCIA:** Este método de amostragem, usado na análise de Monte Carlo, consiste em concentrar a distribuição de pontos de amostragem naquelas partes do intervalo da variável que são de grande interesse. Em inglês importance sampling.

**AMOSTRAGEM POR INTERCEPTAÇÃO DE LINHA:** Método de amostragem florestal em que a unidade amostral é um segmento de reta ou linha e os elementos de medição tais como árvores, troncos, galhos dentre outros são selecionados se forem interceptados pela linha.

**AMOSTRAGEM POR JULGAMENTO:** Neste tipo de amostragem os elementos são escolhidos subjetivamente pelo pesquisador. O objetivo é atingir maior representatividade da população usando seus conhecimentos sobre este universo estatístico. Entretanto, diferentes pesquisadores provavelmente escolheriam diferentes elementos buscando a máxima representatividade, sob o ponto de vista científico, esse tipo de amostra não é muito adequado, porque não se pode estimar o erro em cada escolha. Em geral, as

únicas circunstâncias em que uma amostragem por julgamento pode dar bons resultados são quando: i) todos os elementos os eja, o universo inteiro podem ser vistos ao mesmo tempo; ii) o pesquisador que faz o julgamento é capaz de ser preciso sobre as características em questão; iii) o tamanho da amostra é muito pequeno menos de 10 elementos; iv) a população é homogênea; e v) a população já foi estudada.

**AMOSTRAGEM POR LINHAS:** Processo de amostragem de áreas em que são escolhidas as cortadas por linhas traçadas com determinados critério. A expressão também é usada quando uma rede de linhas é empregada a fim de determinar as áreas elementares, isto é, as que vão constituir as unidades de amostragem.

**AMOSTRAGEM POR LOTERIA:** É aquela que usa uma representação concreta do sistema de referência, tal como um conjunto de bolas numeradas contidas num recipiente onde elas são remexidas e emite uma de cada vez, tal como se faz nas extrações de loteria.

**AMOSTRAGEM POR QUOTAS:** Processo no qual os trabalhadores de campo recebem tarefas específicas quanto ao número de unidades amostrais a serem escolhidas, mas a seleção, ela própria, é feita a esmo.

**AMOSTRAGEM POR QUOTAS:** Um dos métodos de amostragem mais comumente usados em levantamentos de mercado e em prévias eleitorais. Abrange as fases de classificação da população em termos de propriedades que se sabe, ou presume serem relevantes para a característica a ser estudada; determinação da proporção da população para cada característica, com base na constituição conhecida, presumida ou estimada, da população; fixação de quotas para cada entrevistador a quem tocará a responsabilidade de selecionar entrevistados, de modo que a amostra total observada ou entrevistada contenha a proporção e cada classe.

**AMOSTRAGEM POR REMEDIÇÃO COMPLETA:** Sistema de amostragem em que as medições são repetidas nas mesmas unidades amostrais (unidades amostrais permanentes) em tempos diferentes.

**AMOSTRAGEM POR SUBSTITUIÇÃO COMPLETA:** Sistema de amostragem em que as medições são feitas na mesma população, em tempos diferentes, porém utilizando-se unidades amostrais temporárias e independentes.

**AMOSTRAGEM POR SUBSTITUIÇÃO PARCIAL:** Sistema de amostragem em que parte das unidades amostrais são remedidas em tempos diferentes (unidades amostrais permanentes), porém acrescenta-se a elas novas unidades temporárias.

**AMOSTRAGEM PROBABILÍSTICA:** O mesmo que amostragem acidental. Ou ainda é aquela onde a seleção é aleatória por meio de um sorteio e onde cada elemento da população possui uma probabilidade conhecida e diferente de zero de fazer parte da amostra.

**AMOSTRAGEM PROPORCIONAL AO TAMANHO:** É o processo de seleção probabilística em que as probabilidades associadas às diversas unidades amostrais são proporcionais ao tamanho delas. Como tamanho podem-se adotar as intensidades do próprio atributo em causa, conhecidas por uma pesquisa anterior, ou, ainda, as intensidades de qualquer outro atributo altamente correlacionado com o visado.

**AMOSTRAGEM PROPORCIONAL ESTRATIFICADA:** Quando a população se divide em estratos (sub-populações), convém que o sorteio dos elementos da amostra leve em consideração tais estratos, daí obtemos os elementos da amostra proporcional ao número de elementos desses estratos.



**AMOSTRAGEM PSEUDO-ACIDENTAL:** Assim se exprimem alguns autores com referência à amostragem sistemática de uma lista das unidades amostrais em que estas se sucedem por ordem que possa considerar-se acidental.

**AMOSTRAGEM RANDÔMICA ESTRATIFICADA:** Separação da população útil em grupos mutuamente exclusivos (estratos); a seguir, amostras randômicas são tiradas de cada estrato.

**AMOSTRAGEM RANDÔMICA SIMPLES:** A seleção randômica de membros da população útil para inclusão na amostra final.

**AMOSTRAGEM RANDÔMICA SISTEMÁTICA:** Adaptação do processo de amostragem randômica que consiste na seleção de membros da amostra por meio de uma lista de intervalos fixos a partir de um ponto inicial escolhido ao acaso.

**AMOSTRAGEM REPLICADA:** É a que resulta na obtenção de mais de uma amostra (interpenetrante), a fim de aumentar a precisão das estimativas.

**AMOSTRAGEM RETICULAR:** Amostragem por estratificação múltipla em que as unidades finais são escolhidas de modo análogo ao que se usa nos delineamentos reticulares. Nesse caso, os substratos de que vão ser extraídos os indivíduos podem, obedecer ao critério de construção de um quadrado latino.

**AMOSTRAGEM SEM REPOSIÇÃO:** Processo de seleção em que uma mesma unidade amostral não pode figurar mais de uma vez na amostra.

**AMOSTRAGEM SEM REPOSIÇÃO:** Método de escolha de uma amostra em que o elemento escolhido não é devolvido à população, não podendo, assim, ser escolhido novamente.

**AMOSTRAGEM SEM SUBSTITUIÇÃO:** Uma vez que um elemento tenha sido incluído na amostra, ele é retornado à população. Um elemento selecionado previamente pode ser selecionado novamente e, em consequência disso, pode aparecer na amostra mais de uma vez.

**AMOSTRAGEM SEM SUBSTITUIÇÃO:** Uma vez que um elemento tenha sido incluído na amostra, ele é removido da população e não pode ser selecionado uma segunda vez.

**AMOSTRAGEM SEQÜENCIAL:** É a própria à aplicação de uma prova seqüencial.

**AMOSTRAGEM SIMPLES:** O mesmo que amostragem equiprobabilística. Processo de seleção em que a probabilidade de uma unidade amostral ser escolhida é a mesma que para qualquer outra, não só antes de iniciado, como até completar-se o processo, e com o mesmo valor.

**AMOSTRAGEM SISTEMÁTICA:** Método de amostragem na qual o primeiro elemento de uma lista da população é selecionado ao acaso e após cada k-ésimo elemento é selecionado. Em inglês systematic sampling.

**AMOSTRAGEM SISTEMÁTICA:** Consiste na retirada de unidades de amostragem em posições sucessivamente predefinidas ou intervalos regulares, como por exemplo, sobre uma malha regular para amostragem em superfície, ou então furos de sondagem localizados sobre os nós de uma malha regular com testemunhagem regular na vertical segundo profundidades predefinidas. Na amostragem de depósitos minerais, a amostragem sistemática pode ou não produzir dados aleatórios, dependendo da presença de elementos não aleatórios, como tendência ou mudança do teor no minério, correlação entre teores de

pontos adjacentes e, assim por diante. A aleatoriedade implica ou na falta de ordem ou de conformidade das medidas quando elas são listadas na ordem adquirida e sua propriedade implica na inclusão dos atributos de medida na proporção direta a sua importância relativa. Nesse sentido, uma amostra aleatória deveria mostrar independência entre sucessivas unidades de amostragem, o que efetivamente não ocorre, pois teores de pontos adjacentes apresentar-se-ão invariavelmente correlacionados, diminuindo à medida que os pontos se distanciam entre si. Esta correlação ou influência entre amostras sucessivas, resultante das características localizadas da mineralização, é um elemento não aleatório que afeta a aleatoriedade inerente ou independência entre teores que podem estar presentes num conjunto de teores.

**AMOSTRAGEM SISTEMÁTICA:** É o processo que, aplicado a uma sucessão de n unidades amostrais, consiste em escolher, por amostragem acidental, um elemento entre os  $k \leq n$  primeiros, e todo subsequente k-ésimo a partir do que foi escolhido. Como, em geral, n não é múltiplo de k, várias das k possíveis amostras poderão diferir, em tamanho, por um elemento.

**AMOSTRAGEM SISTEMÁTICA:** Um procedimento de seleção, por meio do qual todos os elementos da amostra são determinados após a seleção do primeiro elemento, uma vez que cada elemento de uma lista de seleção é separado do primeiro elemento por um múltiplo do intervalo de seleção.

**AMOSTRAGEM SISTEMÁTICA:** Quando os elementos da população já se acham ordenados, não há necessidade de construir o sistema de referência.

**AMOSTRAGEM SISTEMÁTICA:** É aquela em que os elementos são escolhidos a partir de uma regra previamente estabelecida.

**AMOSTRAGEM SISTEMÁTICA:** Amostragem em que é selecionado cada k<sup>mo</sup> elementos.

**AMOSTRAGEM SISTEMÁTICA:** Delineamento amostral em que o critério probabilístico é a aleatorização da primeira unidade amostral, e a partir dela todas as demais unidades são localizadas sistematicamente, seguindo uma grade retangular ou quadrada.

**AMOSTRAGEM ÚNICA:** Expressão usada na inspeção por amostragem, quando a aceitação ou não de uma partida é resolvida pelo exame de uma única amostra dela oriunda, emprega-se, em geral, para marcar a oposição para com amostragem seqüencial.

**AMOSTRAGEM UNITÁRIA:** Processo em que os elementos que vão constituir a amostra a estudar são extraídas diretamente da população, sem fases, estágios, ou qualquer outro passo intermediário.

**AMOSTRAS CONCORDANTES, DISCORDANTES E NEUTRAS:** Seja a população das  $N = \binom{r}{n}$  partições de r valores em dois grupos de n e m = r-n valores, respectivamente; seja  $\xi$  o valor absoluto da diferença entre as médias aritméticas dos dois grupos. Seja  $M (< N)$  natural. Nesses termos, em face do nível de significância  $\frac{M}{N}$ , os resultados de uma dada participação aleatória, para a qual  $\xi = x$ , serão ditos constituírem duas amostras discordantes se o número de partições para as quais  $\xi \geq x$  for, no máximo, igual a M ou mais partições para as quais  $\xi$  é maior que x, as amostras serão ditas concordantes; neutras, quando nem concordantes, nem discordantes.

**AMOSTRAS EMPARELHADAS:** Um par, ou conjunto de amostras emparelhadas são aquelas em que cada membro de uma amostra está emparelhado, pareado ou relacionado com um membro em cada amostra por meio de uma qualidade diferente das estudadas. Em inglês matched samples.

**AMOSTRAS EMPARELHADAS:** Duas amostras que são dependentes no sentido de que os valores dos dados se correspondem ao pares.

**AMOSTRAS INDEPENDENTES:** Dois ou mais subconjuntos randômicos da população retirados de modo independente e de tamanhos iguais ou desiguais.

**AMOSTRAS INTERPENETRANTES:** São amostras oriundas da mesma população, e obtidas pelo mesmo processo de seleção.

**AMOSTRAS LIGADAS:** Assim se dizem duas amostras correspondentes  $\{x_i\}$  e  $\{y_i\}$ , tais que, ou a) a escolha de  $x_i$  determina univocamente a do correspondente  $y_i$ , ou b) dado um critério de correspondência entre as duas amostras, a cada  $x_i$  corresponde um número finito de  $y_i$ 's, podendo, dentre esses, ser escolhido um ao acaso.

**AMOSTRAS NÃO-PROBABILÍSTICAS:** Amostras escolhidas por critérios pessoais, seja por conveniência, por julgamento ou por quotas.

**AMOSTRAS PARALELAS:** Técnica de desenho de pesquisa em que amostras de populações relacionadas são selecionadas para análise contextual. Por exemplo, poderíamos selecionar para estudo uma amostra de estudantes e também seus professores.

**AMOSTRAS PAREADAS:** Subconjuntos randômicos da população em que a variável estudada é mensurada antes e depois de certo procedimento. Exemplo: medição do pulso radial antes e após esforço físico controlado. As amostras são necessariamente do mesmo tamanho.

**AMOSTRAS VIESADA:** É obtida por um processo de seleção, ou a partir de um sistema de referência, que contém erro sistemático, de modo que ela não poderá corresponder à sua definição.

**AMOVA:** Técnica estatística de análise em biologia molecular a qual consiste no estudo da variação molecular dentro de uma espécie ou entre espécies. Como exemplo desta análise está o uso de informação de divergência de haplótipos de DNA para incorporar o formato de análise de variância, derivado a partir de uma matriz de quadrados de distâncias entre os pares de haplótipos. Essa metodologia pode ser também aplicada a sequências de vírus HIV quando, por exemplo, um pesquisador está interessado na comparação de taxas de mutação entre e dentre subtipos de sequências de HIV.

**AMPLITUDE:** De um ciclo, é a metade da diferença entre os valores absolutos do máximo e do mínimo das ordenadas desse ciclo, uma vez eliminada a tendência secular da respectiva série.

**AMPLITUDE:** Medida de dispersão composta dos valores mais alto e mais baixo de uma variável num conjunto de observações. Numa turma da disciplina de estatística, por exemplo, a amplitude etária pode ser de 17 a 37.

**AMPLITUDE:** É a distância observada no variograma, a partir da qual as amostras passam a ser independentes. Nesse sentido, conforme Matheron (1971), o variograma dá um significado preciso da noção tradicional da zona de influência. A amplitude (a) no variograma, é a distância que separa o campo estruturado, amostras correlacionadas, do campo aleatório, amostras independentes.

**AMPLITUDE:** Uma medida de associação para duas variáveis, as quais são ordinais ou de intervalos.

**AMPLITUDE:** É a diferença entre os extremos de um conjunto, isto é, a distância entre o máximo e o mínimo do conjunto. O valor maior menos o valor menor de uma série de variáveis. A amplitude é uma medida de dispersão, mas em termos de média de amplitude em amostras repetidas, pode disponibilizar uma estimativa razoável do desvio padrão populacional.

**AMPLITUDE:** É a diferença entre os extremos de um conjunto, isto é, a distância entre o máximo e o mínimo do conjunto. Em inglês range.

**AMPLITUDE:** Diferença entre o mais alto e o mais baixo dos valores de uma distribuição. A amplitude é uma medida de dispersão ou variabilidade. Em inglês range.

**AMPLITUDE:** O valor maior menos o valor menor de uma série de variáveis. A amplitude é uma medida de dispersão, mas em termos de média de amplitude em amostras repetidas, pode disponibilizar uma estimativa razoável do desvio padrão populacional. Em inglês range.

**AMPLITUDE AMOSTRAL:** É a diferença entre o maior e o menor valor de uma amostra. Em inglês sample range.

**AMPLITUDE DECIL:** Oscilação de um intervalo decil.

**AMPLITUDE DE CLASSES:** Diferenças entre dois limites inferiores de classes consecutivos em uma tabela de frequências.

**AMPLITUDE DE CLASSE:** De uma classe de uma distribuição de frequência, é o valor da diferença entre os seus limites, reais ou aparentes. Também se diz intervalo unitário e intervalo de classe.

**AMPLITUDE INDEFINIDA:** É na ordem de classificação de uma distribuição de frequência, a de uma classe para a qual não é dado, ou o limite inferior que é o extremo inferior ou o limite superior que é extremo superior.

**AMPLITUDE INTERDECIL:** É, no conceito de alguns autores, o valor absoluto da diferença entre o primeiro e o nono decis.

**AMPLITUDE MÉDIA:** De um conjunto de amostras do mesmo tipo e tamanho, é a média aritmética de suas amplitudes totais.

**AMPLITUDE MÍNIMA DE INTERVALO DE CONFIANÇA (IC):** Seja  $f(x)$  a função densidade de probabilidade (f.d.p.) de uma distribuição unimodal, sendo  $\ddot{x}$  a moda. Se o intervalo  $[a, b]$  satisfaz as três (3) seguintes condições ( $0 < \gamma < 1$ ):

$$i) \int_a^b f(x) dx = \gamma,$$

$$ii) f(a) = f(b) > 0,$$

$$iii) a \leq \ddot{x} \leq b,$$

Então, entre todos os intervalos que satisfazem a condição (i), o intervalo de confiança  $[a, b]$  tem amplitude mínima.

**AMPLITUDE PERCENTIL:** Oscilação de um intervalo percentil.

**AMPLITUDE PERCENTIL:** A diferença aritmética entre o 10º e o 90º percentis.



**AMPLITUDE DE PERCENTIS 10-90:** Diferença entre o 10º e o 90º percentil.

**AMPLITUDE DE UM CONJUNTO DE DADOS:** É a diferença entre o maior valor e o menor valor desse conjunto. Se os dados estiverem agrupados em classes, a amplitude é a diferença entre o limite superior da última classe e o limite inferior da primeira. Para um intervalo do conjunto de dados de  $[a,b]$ , onde  $x_1 = a$  e  $x_n = b$ , tem-se:  $At = b-a$ .

**AMPLITUDE DO INTERVALO DE CLASSE:** É obtida por meio de diferença entre o limite superior e inferior da classe e é simbolizada por  $hi = Li-li$ . Observação: Na distribuição de frequência com classe o  $hi$  será igual em todas as classes.

**AMPLITUDE INTER-QUARTIL:** Medida da variabilidade da amostra, correspondente à diferença entre os valores do terceiro e do primeiro quartis, dando-nos informação sobre a amplitude do intervalo em que se encontram 50 % das observações centrais.

**AMPLITUDE QUARTIL:** De um conjunto de números, é o valor absoluto da diferença entre o terceiro e o primeiro quartis desse conjunto. Oscilação do intervalo quartil. Também se diz amplitude interquartil.

**AMPLITUDE SEMI-INTERQUARTIL:** Metade da diferença entre o terceiro e o primeiro quartil de uma distribuição de frequências. É uma medida de variabilidade ou dispersão. Em inglês semi-interquartile range.

**AMPLITUDE SEMIQUARTIL:** A diferença aritmética entre o 25º e o 75º percentis.

**AMPLITUDE TOTAL:** De um conjunto de dados é a diferença entre o maior e o menor deles.

**AMPLITUDE TOTAL:** A diferença aritmética entre os valores mais alto e mais baixo nos dados.

**AMPLITUDE TOTAL:** É a única medida de dispersão que não tem na média o ponto de referência. Quando os dados não estão agrupados a amplitude total é a diferença entre o maior e o menor valor observado:  $AT = X \text{ máximo} - X \text{ mínimo}$ .

**AMPLITUDE TOTAL:** Diferença entre o maior e o menor escore de uma amostra ou universo.

**AMPLITUDE TOTAL:** Expressão usada como sinônimo de amplitude quando se quer distingui-la da amplitude semi-interquartil, que, ao contrário da amplitude total, não leva em conta todos os escores. Em inglês total range.

**AMPLITUDE TOTAL:** Em um conjunto de dados é a diferença entre o maior e o menor valores.

**AMPLITUDE TOTAL DA AMOSTRA (ROL):** É a diferença entre o valor máximo e o valor mínimo da amostra (ROL). Onde  $AA = X_{\text{max}} - X_{\text{min}}$ . Observação:  $AT$  sempre será maior que  $AA$ .

**AMPLITUDE TOTAL DA DISTRIBUIÇÃO:** É a diferença entre o limite superior da última classe e o limite inferior da primeira classe.  $AT = L(\text{max}) - l(\text{min})$ .

**AMPLITUDE TOTAL ESTUDENTIZADA OU ESTUDANTIZADA:** É a amplitude total dividida por uma estimativa do desvio padrão.

**ANÁLISE:** Parte da matemática que lida com a aproximação de objetos matemáticos tais como números e funções e por outros que são mais fáceis de entender e manejar. Em inglês analysis.

**ANÁLISE:** É o trabalho de avaliação dos dados recolhidos. Sem ela não há relatório de pesquisa.

**ANÁLISE (COMO MÉTODO DE PESQUISA):** É um método de pesquisa que abrange uma grande variedade de técnicas, orientadas para dissecação de fenômenos em suas partes constituintes, como meios que favorecem maior penetração em suas naturezas básicas.

**ANÁLISE, UNIDADE DE:** O que ou quem é estudado.

**ANÁLISE AGREGADA:** Abordagem para análises de ordenação MDS (Multi-Dimensional Scaling), que são técnicas de visualização de dados multidimensionais, na qual um mapa perceptual é gerado para as avaliações de um grupo de respondentes quanto a objetos. Esse mapa perceptual composto pode ser criado por um programa de computador ou pelo pesquisador para achar alguns poucos sujeitos médios ou representativos.

**ANÁLISE BIVARIADA:** Análise simultânea de duas variáveis para determinar a relação empírica entre elas. A montagem de uma tabela percentual simples e a computação de um coeficiente de correlação simples são exemplos de análises bivariadas.

**ANÁLISE BIVARIADA:** Análise da relação existente entre uma variável independente e uma variável dependente.

**ANÁLISE BIVARIADA (RELACIONAMENTO):** A análise bivariada ou de duas variáveis é realizada não necessariamente para determinar qual é a causa de qual, mas para determinar se elas são ou não componentes do mesmo fenômeno. Existem três significados básicos para os relacionamentos bivariados: i) simétrico: relacionamento em que nenhuma das variáveis influencia a outra; ii) recíproco: relacionamento em que ambas as variáveis se influenciam e iii) assimétrico: relacionamento em que uma variável influencia a outra.

**ANÁLISE CONFIRMATÓRIA:** Uso de uma técnica multivariada para testar (confirmar) uma relação pré-especificada. Por exemplo, suponha que teorizemos que apenas duas variáveis deveriam ser preditoras de uma variável dependente. Se testarmos empiricamente a significância desses dois preditores e a não significância de todos os outros, esse teste é uma análise confirmatória. É o oposto de análise exploratória.

**ANÁLISE CONJUNTA:** É uma técnica emergente de dependência que tem trazido nova sofisticação para a avaliação de objetos, como produtos novos, serviços ou ideias. A aplicação mais direta é no desenvolvimento de novos produtos ou serviços, viabilizando a avaliação de produtos complexos e mantendo um contexto realista de decisão para o respondente. O pesquisador de mercado, por exemplo, é capaz de avaliar a importância de atributos, bem como os níveis de cada atributo, enquanto consumidores avaliam apenas uns poucos perfis do produto, os quais são combinações de níveis de produto. Por exemplo, considere que um dado produto tenha três atributos tais como preço, qualidade e cor, cada um com três níveis possíveis por exemplo, vermelho, amarelo e azul. Em vez de avaliar as 27 (3x3x3) combinações possíveis, um subconjunto como 9 ou mais pode ser avaliado por seu apelo perante consumidores, e o pesquisador sabe não apenas o quão importante cada atributo é, mas também a importância de cada nível como a atratividade de vermelho versus amarelo versus azul. Além disso, quando as avaliações do consumidor são concluídas, os resultados das análise conjunta podem igualmente ser usados em simuladores de planejamento do produto, os quais mostram a aceitação do consumidor a qualquer número de formulações do produto e ajudam no planejamento do produto ótimo.

**ANÁLISE CONJUNTA ADAPTATIVA:** Metodologia para conduzir uma análise conjunta que conta com informações de respondentes por exemplo, importância de atributos para adaptar o delineamento

conjunto, de modo a realizar a tarefa de maneira mais simples. Exemplos são os modelos auto-explicados e adaptativos ou híbridos.

**ANÁLISE CONJUNTA TRADICIONAL:** Metodologia que emprega os princípios clássicos da análise conjunta, usando um modelo aditivo da preferência de consumidor e métodos de apresentação de comparação aos pares ou de perfil completo.

**ANÁLISE DA COVARIÂNCIA (ANCOVA):** É um método para testar a significância que se baseia na razão que a variância entre os grupos mantém com a variância dentro dos grupos. Esse método é utilizado nas análises multivariáveis quando a variável dependente é contínua, algumas das variáveis independentes são categóricas isto é, nominais, dicotômicas ou ordinais e algumas das variáveis dependentes são contínuas.

**ANÁLISE DA REGRESSÃO LINEAR:** Teste estatístico referente à força da relação linear entre uma variável independente e uma variável dependente, devendo ambas serem contínuas.

**ANÁLISE DA TÁBUA DE VIDA:** Análise estatística de sobrevivência ou outro desfecho dicotômico no qual a sobrevivência proporcional é avaliada repetidamente ao longo dos intervalos de observação. Isso revela o padrão de mortalidade, bem como os coeficientes globais de mortalidade e de sobrevivência. Ver também análise da tábua de vida pelo método atuarial e análise da tábua de vida pelo método de Kaplan-Meier que são os dois métodos comumente utilizados.

**ANÁLISE DA TÁBUA DE VIDA PELO MÉTODO ATUARIAL:** É um método de análise da tábua de vida no qual a proporção de sobreviventes é avaliada em intervalos fixos, por exemplo, um certo número de meses, que foram estabelecidos antes da acumulação dos dados. Ver também análise da tábua de vida.

**ANÁLISE DA TÁBUA DE VIDA PELO MÉTODO DO PRODUTO-LIMITE:** Mesmo que análise da tábua de vida pelo método de Kaplan-Meier.

**ANÁLISE DA VARIÂNCIA:** Levando-se em conta a teoria de testes de significância multivariados, é algumas vezes conhecido como análise de variância multivariada. A variação total de um conjunto de dados pode ser separada em componentes associados a fontes de variação usadas como critério de classificação das observações.

**ANÁLISE DA VARIÂNCIA (ANOVA):** É um método para testar a significância que se baseia na razão que a variância entre os grupos mantém com a variância dentro dos grupos. Esse método é utilizado nas análises estatísticas quando a variável dependente é contínua e a variável ou as variáveis independentes são todas categóricas, isto é, nominais, dicotômicas ou ordinais. Se houver apenas uma variável independente, o método é denominado análise de variância (ANOVA) de um critério. Se houver mais de uma variável independente, o método é denominado análise de variância (ANOVA) de N (múltiplos) critérios, com o N representando o número de variáveis independentes.

**ANÁLISE DA VARIÂNCIA:** Método de análise da variância de uma população a fim de fazer inferências sobre a mesma.

**ANÁLISE DA VARIÂNCIA:** Técnica estatística que subdivide a variabilidade total de um conjunto de dados em seus componentes. Pode estabelecer, por exemplo, que as médias de vários grupo são estatisticamente diferentes.

**ANÁLISE DA VARIÂNCIA DE DOIS CRITÉRIOS:** Análise da variância envolvendo dados classificados segundo dois fatores diferentes.

**ANÁLISE DA VARIÂNCIA DE UM CRITÉRIO:** Análise da variância envolvendo dados classificados em grupos segundo um único critério.

**ANÁLISE DA VARIÂNCIA DE UM FATOR:** Veja análise da variância de um critério.

**ANÁLISE DE AGRUPAMENTOS:** A análise de agrupamentos é uma técnica analítica para desenvolver subgrupos significativos de indivíduos ou objetos. Especificamente, o objetivo é classificar uma amostra de entidades tais como indivíduos ou objetos em um pequeno número de grupos mutuamente excludentes, com base nas similaridades entre as entidades. Na análise de agrupamentos, diferentemente da análise discriminante, os grupos não são predefinidos. Ao invés disso, a técnica é usada para identificar os grupos. A análise de agrupamentos geralmente envolve pelo menos três passos. O primeiro é a medida de alguma forma de similaridade ou associação entre as entidades para determinar quantos grupos realmente existem na amostra. O segundo é o próprio processo de agrupamento, nas quais entidades são particionadas em grupos (agrupamentos). O último passo é estabelecer o perfil das pessoas ou variáveis para determinar sua composição. Muitas vezes, isso é possível pela aplicação da análise discriminante aos grupos identificados pela técnica de agrupamento.

**ANÁLISE DE CAMINHOS:** Método que emprega correlações bivariadas simples para estimar as relações em um sistema de equações estruturais. O método é baseado na especificação das relações em uma série de equações do tipo regressão retratadas graficamente em um diagrama de caminhos que podem então ser estimadas pela determinação da quantia de correlação atribuível a cada efeito em cada equação simultaneamente. Quando empregada com múltiplas relações entre construtos latentes e um modelo de mensuração, ela é então chamada de modelagem de equações estruturais.

**ANÁLISE DE COMPONENTES PRINCIPAIS:** O principal objetivo deste teste é indicar a significância relativa de variáveis preditivas, eis que, quando o número delas em um conjunto de dados é muito grande, podem surgir, pela análise de regressão múltipla, coeficientes irracionais, dificultando uma acurada identificação da importância daquelas variáveis. Trata-se, portanto, de um teste interpretativo, o qual consiste em examinar um grupo de k variáveis correlacionadas, transformando-as em outro conjunto de variáveis não correlacionadas e independentes, dispostas em combinações lineares e em ordem decrescente de importância por meio de índices designados de  $Z_p$ , de tal maneira que a variância  $\text{var}(Z_1) \geq \text{variância } \text{var}(Z_2) \geq \text{variância } \text{var}(Z_3) \geq \dots \geq \text{variância } \text{var}(Z_p)$ . Os primeiros índices são chamados de Componentes Principais, devendo-se dar maior ênfase àqueles que descrevem cerca de 80 % ou mais da variação, os quais, em alguns casos, estão representados pelo primeiro e segundo componentes. Os tamanhos das amostras podem ser iguais ou desiguais. Em inglês Principal component analysis.

**ANÁLISE DE CORRESPONDÊNCIA:** Abordagem composicional para mapeamento perceptual que relaciona categorias de uma tabela de contingência. A maioria das aplicações envolve um conjunto de objetos e atributos, em que os resultados retratam objetos e atributos em um mapa perceptual comum.

**ANÁLISE DE CORRESPONDÊNCIA:** É uma técnica de interdependência recentemente desenvolvida que facilita tanto a redução dimensional da classificação de objetos, como por exemplo, produtos, pessoas, em um conjunto de atributos quanto o mapeamento perceptual de objetos relativo a esses atributos. Os pesquisadores são constantemente defrontados com a necessidade de quantificar os dados qualitativos

encontrados em variáveis nominais. Em sua forma mais básica, a análise de correspondência emprega uma tabela de contingência, que é a tabulação cruzada de duas variáveis categóricas. Ela então transforma os dados não métricos em um nível métrico e faz redução dimensional que é análoga à análise fatorial e mapeamento perceptual semelhante a análise multidimensional. Por exemplo, preferências de respondentes a marcas podem ser tabuladas no cruzamento com variáveis demográficas, como por exemplo, sexo, categorias de renda, ocupação, indicando quantas pessoas que preferem cada marca recaem em cada categoria das variáveis demográficas. Por meio de análise de correspondência, a associação ou correspondência de marcas e diferentes características daqueles que preferem cada marca é então mostrada em um mapa bi ou tridimensional de marcas e características dos respondentes. Marcas que são percebidas como semelhantes são colocadas próximas umas das outras. Do mesmo modo as características mais eminentes de respondentes que preferem cada marca também são determinadas pela proximidade das categorias de variáveis demográficas às posições das marcas. A análise de correspondência fornece uma representação multivariada de interdependência para dados não métricos que não é possível com outros métodos.

**ANÁLISE DE CORRESPONDÊNCIA MÚLTIPLA:** Forma de análise de correspondência que envolve três ou mais variáveis categóricas relacionadas em um espaço perceptual comum.

**ANÁLISE DE CONTEÚDO:** Uma técnica que busca caracterizar os significados num dado corpo do discurso de maneira sistemática e quantitativa. É mais um método de análise, é um método de observação. Em vez de se observar o comportamento de alguém diretamente, entrevistando-o ou pedindo-lhe para usar uma escala o pesquisador pega uma comunicação produzida pela pessoa e faz perguntas em cima dela. Um domínio heterogêneo de técnicas que se foca na descrição sistemática, objetiva e quantitativa de uma comunicação ou série de comunicações.

**ANÁLISE DE COVARIÂNCIA:** É um procedimento pelo qual se fazem ajustamentos estatísticos na variável dependente. Esses ajustamentos se baseiam na correlação entre variável dependente e outra variável, chamada covariada. Um método de controle estatístico por meio do qual se ajustam os escores da variável dependente, em acordo com os escores de uma variável relacionada.

**ANÁLISE DE COVARIÂNCIA (PRESSUPOSTOS):** O primeiro pressuposto especifica que a variável a ser controlada seja independente de quaisquer efeitos do tratamento. Se existir a possibilidade de que os tratamentos possam ter afetado os dados sobre a variável de controle, a análise de covariância não é apropriada. O segundo pressuposto é o de regressão linear. Os desvios de regressão, isto é, os escores residuais ou ajustados, devem ser normal e independentemente distribuídos na população, com médias de zero e variância homogêneas. O terceiro pressuposto especifica que haja homogeneidade dos coeficientes de regressão para os diferentes tratamentos das populações.

**ANÁLISE DE CUSTO-BENEFÍCIO:** Medida e comparação de custos e dos benefícios de uma linha de ação proposta em termos de uma mesma unidade, geralmente em unidades monetárias, como dólares ou reais.

**ANÁLISE DE CUSTO-EFETIVIDADE:** Fornece um meio de comparação entre diferentes métodos propostos para alcançar um determinado fim em termos das unidades de medida mais apropriadas. Avalia os meios menos dispendiosos para que se atinja uma meta estabelecida.

**ANÁLISE DE ITENS:** Procedimento para aumentar a fidedignidade e validade de um teste, avaliando separadamente cada item, a fim de determinar se o item discrimina ou não da mesma forma que se espera do teste total. Por exemplo, dado um teste, seriam os alunos com os melhores escores nele também os que

mais acertam um dado item e os alunos abaixo da média os que mais erram o item? Se assim for, o item tem poder discriminador e contribui para a fidedignidade e validade globais do teste.

**ANÁLISE DE REGRESSÃO:** Associada com modelos preditivos, a análise de regressão trata da estimativa, fundada em dados reais, de algum relacionamento matemático (linear ou não linear) entre a variável dependente e um conjunto de variáveis independentes.

**ANÁLISE DE REGRESSÃO (FUNÇÃO):** A função básica de regressão é encontrar uma expressão algébrica que representa o relacionamento funcional entre duas (ou mais) características, dadas observações de algumas características em especificados valores de outras.

**ANÁLISE DE RESÍDUOS:** De acordo com Bussab e Morettin (2003) é um tipo de análise usada para verificar se um modelo é adequado, neste caso temos que investigar se as suposições ou pressuposições feitas para o desenvolvimento do modelo de regressão estão satisfeitas. Para tanto, estudamos o comportamento do modelo usando o conjunto de dados observados, notadamente as discrepâncias entre os valores observados e os valores ajustados pelo modelo, ou seja, quando se realiza uma análise dos resíduos, dos erros ou dos afastamentos. Esse teste de validade pode ser feito do seguinte modo: i) calculam-se os resíduos para cada valor de Y. Os resíduos representam a diferença entre aquilo que foi realmente observado e o que foi predito pelo modelo de regressão, isto é: Resíduo =  $e = (Y_i - \hat{Y}_i)$ , ii) desenha-se um gráfico no qual os resíduos são colocados no eixo vertical e os valores esperados de  $Y(\hat{Y}_i)$ , no eixo horizontal. Os pontos devem ficar distribuídos de forma equilibrada acima e abaixo de uma linha imaginária paralela ao eixo dos X'S na altura do resíduo zero, formando uma faixa aproximadamente retangular. A violação do pressuposto de homocedasticidade que são variâncias iguais produz faixas em forma de cone, enquanto violações do pressuposto de linearidade onde o fenômeno pode ser modelado por uma equação de uma reta) produz faixas curvas. Veja a seguir com maior detalhe esse tipo de análise. O i-ésimo resíduo é dado por  $\hat{e}_i = (Y_i - \hat{Y}_i)$ ,  $i = 1, 2, L, n$ . É bom lembrar que já foi utilizado estes resíduos para obter medidas da qualidade e dos estimadores dos parâmetros do modelo. Agora neste caso vamos estudar o comportamento individual e conjunto destes resíduos, comparando com as pressuposições feitas sobre os verdadeiros erros  $e_i$ . Existem na literatura especializada várias técnicas formais para conduzir esse tipo de análise, mas neste trabalho iremos ressaltar basicamente métodos gráficos.

Uma representação bastante útil é obtida plotando-se pares de valores  $(X_i - \hat{e}_i)$ ,  $i = 1, 2, L, n$ . Outras vezes, é mais importante fazer a representação gráfica dos chamados resíduos padronizados, dados pela seguinte equação.

$$\hat{z}_i = \frac{(Y_i - \hat{Y}_i)}{S_e} = \frac{\hat{e}_i}{S_e}$$

Plotando-se os pares de valores  $(x_i, \hat{z}_i)$ , é importante observar que a forma dos dois gráficos será semelhante, havendo apenas uma mudança de escala das ordenadas nos dois casos. Por isso, iremos usar a primeira representação, indicando no gráfico a posição do valor  $S_e$ .

Outro resíduo usado neste tipo de análise é o chamado resíduo estudentizado, definido pela seguinte equação:  $\hat{r}_i = \frac{\hat{e}_i}{S_e \sqrt{1 - V_{ii}}}$ , onde  $V_{ii} = \frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$ . O denominador da equação  $\hat{z}_i = \frac{(Y_i - \hat{Y}_i)}{S_e} = \frac{\hat{e}_i}{S_e}$  é o desvio

padrão de  $\hat{e}_i$ . Obtido o gráfico dos resíduos, é necessário saber como identificar possíveis inadequações.



**ANÁLISE DE SUPERFÍCIE DE TENDÊNCIA:** Tipo de análise também conhecida como análise de superfícies de tendência consiste no estudo do comportamento espacial das variáveis de interesse amostradas sobre uma região. Este estudo é feito por meio de superfícies de tendência, obtidas pelo ajuste de polinômios ou harmônicos da série dupla de Fourier. A superfície ideal estatisticamente significativa é determinada por meio de uma análise de variância e teste F, onde separa-se a variação total em variação explicada pela superfície e variação não explicada devida a resíduos aleatórios.

**ANÁLISE DE VARIÂNCIA:** Uma técnica estatística pela qual é possível separar a variância numa distribuição de escores, segundo fontes ou fatores separados; embora a variância seja distribuída, o teste estatístico testa diferença entre médias. Basicamente, este tipo de análise é usado para identificar fontes de variação na variável dependente, isto é, identificar tantas razões quanto possível do porque indivíduos diferem em suas respostas ou escores da variável dependente. Análise de variância se baseia no pressuposto de que qualquer escore na medida dependente, em qualquer ponto no tempo, determina-se pelas peculiaridades do sujeito, que está sendo medido mais a soma dos fatores relevantes que o estão influenciando no momento da mensuração.

**ANÁLISE DE VARIÂNCIA (ANOVA):** Técnica estatística usada para determinar se as amostras de dois ou mais grupos surgem de populações com médias iguais. A análise de variância emprega uma medida dependente, ao passo que a análise multivariada de variância compara amostras com base em duas ou mais variáveis dependentes.

**ANÁLISE DE AGRUPAMENTOS:** Técnica multivariada com o objetivo de agrupar respondentes ou casos com perfis similares em um dado conjunto de características. Semelhante à análise fatorial Q.

**ANÁLISE DE COMPONENTES:** Modelo fatorial no qual os fatores são baseados na variância total. Na análise de componentes, unidades (1s) são usadas na diagonal da matriz de correlação; esse procedimento implica computacionalmente que toda variância é comum ou compartilhada.

**ANÁLISE DE CONGLOMERADOS (CLUSTER ANALYSIS):** método eficiente para a classificação biológica de vários organismos em grupos relativamente semelhantes e, uma vez agrupados, as características de cada grupo podem ser analisadas para se verificar se pertencem ou não à mesma espécie. A demonstração gráfica dos conglomerados é efetuada, sobretudo, por meio de dendograma.

**ANÁLISE DE CONTEÚDO:** Método de pesquisa baseado na codificação de comunicações registradas, como livros, jornais, discursos, comerciais de televisão (TV), dentre outros.

**ANÁLISE DE CONTEÚDO:** Técnica usada para analisar material escrito geralmente em textos de anúncios ao desmembrá-lo em unidades significativas pelo uso de regras cuidadosamente aplicadas.

**ANÁLISE DE CORRELAÇÃO:** Análise do grau em que as mudanças em uma variável estão associadas a mudanças em outra.

**ANÁLISE DE COVARIÂNCIA:** É uma extensão da análise de variância para levar em conta o caso em que membros de uma classe compartilham valores de mais de uma variável. Do inglês analysis of covariance.

**ANÁLISE DE CUSTO-BENEFÍCIO:** Método de avaliação econômica que tem por objetivo identificar, medir e valorizar os custos e benefícios sociais de determinadas ações. Distingue-se dos outros métodos de avaliação econômica por valorizar tanto os custos como os resultados em termos monetários.

**ANÁLISE DE CUSTO-EFETIVIDADE:** Método de avaliação econômica teoricamente menos ambicioso do que a análise de custo-benefício. Distinguem-se desta última pelo fato de os resultados ou conseqüências dos programas de saúde serem medidos em unidades físicas (por exemplo, anos de vida ganhos ou casos detectados). Este método justifica-se quando os benefícios são dificilmente monetarizáveis.

**ANÁLISE DE CUSTO-MINIMIZAÇÃO:** Ver análise de minimização de custos.

**ANÁLISE DE CUSTO-UTILIDADE:** Forma de avaliação econômica semelhante à análise custo-efetividade, mas na qual as conseqüências dos programas de saúde são medidas numa unidade física combinada com elementos qualitativos. O termo utilizado neste caso representa o valor para o indivíduo ou para a sociedade de determinado nível de saúde. Os resultados das análises de custo por dia saudável livre de doença ou custo por ano de vida ajustado pela qualidade.

**ANÁLISE DE FATORES COMUNS:** Modelo fatorial no qual os fatores são baseados em uma matriz de correlação reduzida. Ou seja, comunalidades são inseridas na diagonal da matriz de correlação e os fatores extraídos são baseados apenas na variância comum, com as variâncias específicas e de erro excluídas.

**ANÁLISE DE MINIMIZAÇÃO DE CUSTOS:** Forma limitada de avaliação econômica em que comparam os custos de dois ou mais procedimentos alternativos para alcançar um objetivo determinado, cujas conseqüências se supõem equivalentes, o que ajuda a simplificar a análise. Sinônimo: análise de custo-minimização.

**ANÁLISE DE REGRESSÃO:** Como a análise de regressão tem por objetivo estabelecer relações entre variáveis relacionadas por leis estatísticas, existe sempre uma variável aleatória, variável função  $Y_i$ , cuja distribuição depende do valor assumido por outras variáveis  $(X_1; X_2; \dots; X_k)$  nas  $i = 1, 2, 3, \dots, n$  observações. Se os valores de  $Y_i$  dependem dos valores assumidos pelas K variáveis independentes, é razoável supor que o valor médio de  $Y_i$  para cada i também dependa dos valores assumidos pelo conjunto formado pelas K variáveis independentes. A essa dependência média, e que se indica por,  $E[Y_i / X_1; X_2; \dots; X_k] = F(X_1; X_2; \dots; X_k)$ , e que quer dizer que a esperança matemática (valor médio) de  $Y_i$  condicionada aos valores assumidos pelas variáveis explicativas ou independentes, e que é uma função F dessas  $X_1; X_2; \dots; X_k$  variáveis explicativas, damos o nome de Regressão de  $Y_i$  contra  $X_1; X_2; \dots; X_k$ , e que essa dependência é linear, ou seja,  $E[Y_i / X_i] = F(X_i) = \alpha + \beta X_i$ . A técnica desenvolvida para resolver a função F, e que como veremos consiste na resolução dos problemas de especificação, estimação e prova de hipóteses, é conhecida pelo nome de análise de regressão. Quando as variáveis independentes, ou explicativas, forem em número maior do que um ( $K > 1$ ), a técnica é conhecida por análise de regressão múltipla. Quando tivermos apenas uma variável explicativa ( $K = 1$ ), o instrumental é chamado análise de regressão simples.

**ANÁLISE DESAGREGADA:** Abordagem para MDS que são técnicas de visualização de dados multidimensionais, conhecido como escalonamento multidimensional ou MDS na qual o pesquisador gera mapas perceptuais em uma base respondente por respondente. Os resultados podem ser difíceis de generalizar para os respondentes. Portanto, o pesquisador às vezes tenta criar menos mapas por algum processo de análise agregada, na qual os resultados de respondentes são combinados.

**ANÁLISE DISCRIMINANTE:** A tarefa da análise discriminante é encontrar a melhor função discriminante linear de um conjunto de variáveis que reproduza, tanto quanto possível, um agrupamento a priori de casos considerados. Em inglês discriminant analysis.

**ANÁLISE DISCRIMINANTE:** Técnicas multivariadas que dizem respeito à separação de conjuntos distintos de objetos ou observações e à alocação de novos objetos ou observações a grupos previamente definidos.

**ANÁLISE DISCRIMINANTE CANÔNICO:** É uma técnica para redução de dimensão, relacionada à análise de componentes principais e correlação canônica. A partir de uma ou muitas variáveis quantitativas e uma variável de classificação, este procedimento calcula as variáveis canônicas que são combinações lineares das variáveis quantitativas, as quais sumarizam a variação dentro de grupo como na análise de componentes principais. Estas combinações lineares são aquelas que sumarizam as diferenças entre grupos ou classes da variável classificatória.

**ANÁLISE DISCRIMINANTE MÚLTIPLA:** A análise discriminante múltipla (MDA) é uma técnica multivariada adequada quando a única variável dependente é dicotômica, como por exemplo, o sexo masculino ou feminino ou ainda quando for multicotômica, como por exemplo, a estatura de um indivíduo, que pode ser alto, médio ou baixo, e, portanto, não – métrica. Como na regressão múltipla, pressupõe-se que as variáveis independentes sejam métricas. A análise discriminante é aplicável em situações nas quais a amostra total pode ser dividida em grupos baseados em uma variável dependente não-métrica que caracteriza diversas classes conhecidas. Os objetivos primários da análise discriminante múltipla são entender diferenças de grupos e prever a probabilidade de que uma entidade (indivíduo ou objeto) pertencerá a uma classe ou grupo em particular com base em diversas variáveis independentes métricas. Por exemplo, a análise discriminante poderia ser empregada para diferenciar inovadores de não inovadores de acordo com seus perfis demográficos e psicográficos. Em inglês multiple discriminant analysis-MDA.

**ANÁLISE EXPLORATÓRIA:** Análise que define possíveis relações apenas na forma mais geral e então permite que a técnica multivariada estime relações. O oposto da análise confirmatória, o pesquisador não busca confirmar quaisquer relações especificadas anteriormente à análise, mas, ao invés disso, deixa o método e os dados definirem a natureza das relações. Um exemplo é a regressão múltipla stepwise, na qual o método acrescenta variáveis preditoras até que algum critério seja satisfeito.

**ANÁLISE EXPLORATÓRIA:** Termo usado para denotar os métodos de descrição de dados estatísticos, em contraste com a estatística teórica que, ao tratar dados práticos, geralmente envolve algum processo de inferência em probabilidade para sua interpretação. Em inglês exploratory analysis.

**ANÁLISE DE AGRUPAMENTOS:** É uma técnica para encontrar subgrupos significativos de indivíduos ou objetos. O objetivo é classificar uma amostra de indivíduos ou objetos em um número pequeno de grupos mutuamente exclusivos. Diferentemente da análise discriminante, os grupos não são pré-definidos.

**ANÁLISE DISCRIMINANTE:** Se na regressão múltipla a única variável dependente for categórica a técnica apropriada é a análise discriminante. O principal objetivo da análise discriminante é entender diferenças entre grupos e prever a probabilidade de que uma entidade, indivíduo ou objeto pertença a uma classe em particular ou grupo baseado nas várias variáveis independentes.

**ANÁLISE FATORIAL:** Tipo de análise multivariada que inclui análise de componentes principais e análise dos fatores comuns, é uma abordagem estatística que pode ser usada para analisar inter-relações entre um grande número de variáveis e explicar essas variáveis em termos de suas dimensões inerentes comuns que são os fatores. O objetivo é encontrar um meio de condensar a informação contida em um número de variáveis originais em um conjunto menor de variáveis estatísticas ou fatores com uma perda mínima de informação.

**ANÁLISE FATORIAL:** Parte da análise multivariada na qual as variáveis observadas  $x_i, i=1, \dots, p$  são expressas como função de um número  $m < p$  de fatores  $f_j$  junto com elementos residuais. Em inglês factorial analysis.

**ANÁLISE FATORIAL:** Método algébrico complexo para determinar as dimensões ou fatores gerais existentes num conjunto de observações concretas.

**ANÁLISE FATORIAL Q:** Forma grupos de respondentes ou casos com base em sua similaridade em um conjunto de características.

**ANÁLISE FATORIAL R:** Analisa relações entre variáveis para identificar grupos de variáveis que formam dimensões latentes ou fatores.

**ANÁLISE DE FOURIER:** Teoria de representação de funções de uma variável  $t$  como a soma de séries de termos que contém senos e cossenos do tipo  $a_j \cos(2\pi j/\lambda_j), j=0,1, \dots$ . Em inglês Fourier analysis.

**ANÁLISE DE REGRESSÃO:** Método estatístico para investigar relações existentes entre diversas variáveis. Em inglês regression analysis.

**ANÁLISE DE REGRESSÃO:** Método de análise de dados em que as relações entre variáveis são representadas na forma de uma equação chamada equação de regressão.

**ANÁLISE DE REGRESSÃO BIVARIADA:** Análise da força do relacionamento linear entre duas variáveis quando uma delas é considerada a variável independente e a outra a variável dependente.

**ANÁLISE DE RESÍDUOS:** Tipo de análise auxiliar na interpretação das caselas em tabelas de contingência  $l \times c$ .

**ANÁLISE DE SOBREVIVÊNCIA:** Análise estatística do tempo de ocorrência de determinado evento, o qual, na área médica, pode ser o óbito, a recidiva de uma doença, a resposta terapêutica a uma droga, dentre outros. As instituições securitárias utilizam esse modelo de análise. São exemplos de programas de análise de sobrevivência: Atuarial e Cox (Risco Proporcional), Cox-Mantel, Gehan (Wilcoxon generalizado), Kaplan-meier e Log-Rank test. Em inglês survival analysis.

**ANÁLISE DE TOM DE VOZ:** O estudo das mudanças na frequência relativa de vibração da voz humana para medir emoções.

**ANÁLISE DE VARIÂNCIA:** Método para testar a hipótese de que vários grupos diferentes têm todos a mesma média.

**ANÁLISE DE VARIÂNCIA DE DOIS CRITÉRIOS:** Processo de teste que pode ser aplicado a uma tabela de números a fim de testar duas hipóteses: i) não há diferença significativa entre as linhas; ii) não há diferença significativa entre as colunas.

**ANÁLISE DO PROBLEMA:** Refere-se à maneira pela qual o pesquisador procede da afirmativa geral da pergunta experimental às variáveis realmente empregadas e operacionalmente definidas. De grande importância na análise do problema é a escolha da variável dependente.

**ANÁLISE DISCRIMINANTE:** Técnica analítica que adapta a lógica da análise de regressão para variáveis de nível nominal.

**ANÁLISE DISCRIMINANTE:** Dado um conjunto de observações multivariadas de uma amostra, provenientes com certeza de duas ou mais populações, o problema é definir uma regra para alocar outros indivíduos na população de origem correta com probabilidade mínima de uma classificação errada. Em inglês *discriminatory analysis*.

**ANÁLISE DISCRIMINANTE:** separação de dois ou mais grupos de indivíduos em função das medidas de suas variáveis por eles apresentadas.

**ANÁLISE FUNCIONAL:** Parte da matemática que estuda espaços vetoriais de infinitas dimensões e bijeções entre eles. Os elementos destes espaços são muitas vezes funções, como, por exemplo, o espaço das funções contínuas sobre um intervalo. Em inglês *functional analysis*.

**ANÁLISE DE TRAJETÓRIA:** Forma de análise multivariada na qual as relações causais entre variáveis são apresentadas em formato gráfico.

**ANÁLISE DE VARIÂNCIA:** Forma de análise de dados na qual a variância de uma variável dependente é examinada para toda a amostra e para subgrupos separados criados com base em alguma(s) variável(is) independente(s).

**ANÁLISE DE VARIÂNCIA:** Levando-se em conta a teoria de testes de significância multivariados, é algumas vezes conhecido como análise de variância multivariada. A variação total de um conjunto de dados pode ser separada em componentes associados a fontes de variação usadas como critério de classificação das observações. Do inglês *analysis of variance*.

**ANÁLISE DE VARIÂNCIA:** Técnica estatística cujo objetivo é testar a igualdade entre três ou mais médias. Ela permite testar se a variabilidade dentro dos grupos é maior que a existente entre grupos. A técnica supõe independência e normalidade das observações, e igualdade entre as variâncias dos grupos.

**ANÁLISE E INTERPRETAÇÃO DOS DADOS:** A última fase do trabalho estatístico é a mais importante e delicada. Está ligada essencialmente ao cálculo de medidas e coeficientes, cuja finalidade principal é descrever o fenômeno através da estatística descritiva. Na estatística indutiva a interpretação dos dados se fundamenta na teoria da probabilidade.

**ANÁLISE EXPLORATÓRIA DE DADOS (EDA):** Ramo da estatística que enfatiza a investigação de dados. Conjunto de técnicas composto de ferramentas tabulares, gráficas e de medidas resumo, tais como média, variância, etc.

**ANÁLISE FATORIAL (FUNÇÃO):** A análise fatorial reduz um grande número de variáveis associadas em um número menor de fatores que são responsáveis pela associação entre variáveis.

**ANÁLISE HARMÔNICA:** Veja teoria do potencial. Em inglês *harmonic analysis*.

**ANÁLISE LISREL:** Modelagem de equações estruturais, comumente chamada de LISREL. (o nome de um dos programas de computador mais conhecidos cujo nome é Linear Structural RELations), é uma técnica que permite separar relações para cada conjunto de variáveis dependentes. Em seu sentido mais simples, a modelagem de equações estruturais fornece a técnica de estimação apropriada e mais eficiente para uma série de equações de regressão múltiplas separadas estimadas simultaneamente. É caracterizada por dois componentes básicos: o modelo estrutural e o modelo de mensuração. O modelo estrutural é o modelo de caminhos, que relaciona variáveis independentes com dependentes. Em tais situações, teoria,

experiência prévia ou outras orientações permitem ao pesquisador distinguir quais variáveis independentes preveem cada variável dependente.

**ANÁLISE LOGIT:** Ver regressão logística.

**ANÁLISE MULTIVARIADA:** Parte da estatística que cuida da análise de duas ou mais variáveis aleatórias simultaneamente.

**ANÁLISE MULTIVARIADA:** Conjunto de técnicas usadas para investigar, simultaneamente, o efeito de duas ou mais variáveis independentes sobre uma variável dependente: por exemplo, colesterol sérico e idade sobre a mortalidade por doença coronariana. Em estatística, o termo designa os métodos analíticos para estudo simultâneo de duas ou mais variáveis dependentes e várias independentes.

**ANÁLISE MULTIVARIADA:** Em inglês *multivariate analysis*.

**ANÁLISE MULTIVARIADA:** Compreende testes estatísticos nos quais são consideradas de maneira simultânea  $n$  variáveis de  $k$  amostras, destacando-se nesta versão de BioEstat os programas: Componente principal, Distância Multivariada (Euclidiana, Penrose e Mahalanobis), teste de Hotelling, teste de Bartlett, Regressão Linear Múltipla e Regressão Linear Logística Múltipla.

**ANÁLISE MULTIVARIADA:** Análise de relações simultâneas entre diversas variáveis. Examinar simultaneamente os efeitos de idade, sexo e classe social sobre religiosidade seria um exemplo de análise multivariada.

**ANÁLISE MULTIVARIADA:** Parte de estatística que cuida da análise de duas ou mais variáveis aleatórias simultaneamente.

**ANÁLISE MULTIVARIADA:** Análise de múltiplas variáveis em um único relacionamento ou conjunto de relações.

**ANÁLISE MULTIVARIADA:** Esta expressão é usada para denotar a análise de dados multivariados no sentido de cada indivíduo ter  $p$  variáveis observadas. Em inglês *multivariate analysis*.

**ANÁLISE MULTIVARIADA:** Termo que é frequentemente usado de modo incorreto. Refere-se a métodos para analisar mais do que uma variável dependente, bem como mais de uma variável independente.

**ANÁLISE MULTIVARIADA DE VARIÂNCIA (MANOVA):** A análise multivariada de variância, MANOVA-multivariate analysis of variance em inglês, é uma técnica estatística que pode ser usada para explorar simultaneamente as relações entre diversas variáveis independentes categóricas geralmente chamadas de tratamentos e duas ou mais variáveis dependentes métricas. Como tal, representa uma extensão da análise univariada de variância (ANOVA-univariate analysis of variance). A MANOVA é útil quando o pesquisador planeja uma situação experimental como a manipulação de várias variáveis não-métricas que representam tratamento para testar hipóteses referentes à variância, em respostas nos grupos sobre duas ou mais variáveis dependentes métricas.

**ANÁLISE MULTIVARIÁVEL:** Análise de relação existente entre mais de uma variável independente e apenas uma variável dependente.

**ANÁLISE PARA A DECISÃO:** Tipo de análise cuja intenção é melhorar as decisões clínicas feitas em condições de incerteza.



**ANÁLISE PROBIT:** Este tipo de análise de regressão é apropriado para conjuntos de dados em que a variável dependente é medida em unidades do tipo Sim ou Não, morreu ou Não morreu dentre outras. Neste caso a resposta é representada pela percentagem de sobreviventes, nascidos dentre outras. Em inglês probit analysis.

**ANÁLISE QUANTITATIVA:** Representação e manipulação numérica de observações para descrever e explicar os fenômenos que tais observações refletem.

**ANÁLISE RESIDUAL:** A análise de resíduos usada para determinar se as suposições feitas sobre o modelo de regressão parecem ser válidas.

**ANÁLISE DE SOBREVIVÊNCIA OU DE SOBREVIDA:** A análise de sobrevivência, também chamada de análise de sobrevida, é o tipo de análise estatística que será utilizada quando o tempo for o objeto de interesse, seja este interpretado como o tempo até a ocorrência de um evento ou o risco de ocorrência de um evento por unidade de tempo.

**ANÁLISE DE SOBREVIVÊNCIA:** Análise de Sobrevivência é a expressão utilizada para designar a análise estatística de dados quando a variável em estudo representa o tempo desde um instante inicial bem definido até à ocorrência de determinado acontecimento de interesse. Assim sendo, a variável aleatória em estudo é não negativa e pode representar o tempo até à falha de determinada componente elétrica no âmbito da fiabilidade industrial, a duração de uma greve ou período de desemprego no contexto econômico, o tempo de resposta a um inquérito no âmbito da psicologia ou o tempo até à morte de um indivíduo com determinada doença em medicina. As observações resultantes são chamadas de tempos de vida. A origem da análise de sobrevivência remonta há vários séculos, sendo possível referir algumas datas decisivas para o desenvolvimento deste tipo de análise estatística. A análise de sobrevivência permite estudar tempos de vida, também designados por tempos de sobrevivência, ultrapassando as dificuldades inerentes a este tipo de dados. A característica fundamental é a existência de censura, ou seja, para alguns indivíduos pode não ser possível observar o acontecimento de interesse durante o período em que estiveram em observação. Outro aspecto da análise de tempos de vida é que não se pode assumir a normalidade da distribuição subjacente, pois geralmente este tipo de dados apresenta uma distribuição assimétrica positiva. Assim sendo, não podem ser utilizados os métodos estatísticos usuais.

**ANÁLISE SEQUENCIAL:** Método de análise estatística que possibilita terminar um experimento logo que uma resposta, com um certo grau de precisão, seja alcançada.

**ANÁLISE STEPDOWN:** Teste para o poder discriminatório incremental de uma variável dependente depois que os efeitos de outras variáveis dependentes foram levados em conta. Semelhante à regressão stepwise ou análise discriminante, esse procedimento, que se baseia em uma ordem especificada de entrada, determina o quanto uma variável dependente adicional acrescenta à explicação das diferenças entre os grupos na análise de variância multidimensional ou multivariada (MANOVA).

**ANÁLISE SECUNDÁRIA:** Forma de pesquisa na qual os dados coletados e processados por um pesquisador são reanalisados por outro, muitas vezes para uma finalidade diferente. Isto é especialmente apropriado nos casos de dados de survey. Arquivos de dados são repositórios ou bibliotecas para armazenamento e distribuição de dados para análise secundária.

**ANÁLISE UNIVARIADA:** Análise de uma só variável para fins descritivos. Distribuições de frequências, médias, e medidas de dispersão são exemplos de análise univariada, diferente de análise bivariada e multivariada.

**ANÁLISE UNIVARIADA:** Análise de apenas uma variável. Por exemplo, distribuição da mortalidade por sexo.

**ANÁLISE UNIVARIADA DE VARIÂNCIA (ANOVA):** Técnica estatística para determinar, com base em uma medida dependente, se várias amostras são oriundas de populações com médias iguais.

**ANÁLISES NONPARAMÉTRICAS:** Estas análises não exigem mensuração em escala intervalar; dados em escala ordinal ou nominal podem ser analisados. Além disso, para a maioria das análises nonparamétricas, não se faz necessário satisfazer os pressupostos referentes ao formato da distribuição. Por isso, elas são usadas, quando se dispõe apenas de pequenas amostras.

**ANAMÓRFICAS (CURVAS):** Diz-se das curvas de índice de sítio, quando as curvas possuem uma relação de proporcionalidade constante entre si. Ver também: Índice de Sítio.

**ANAMORFOSE:** Transformação de um conjunto de dados para reduzi-los à uma função linear ou equação da reta. Usam-se frequentemente as transformações  $\log \frac{y}{x}$ ,  $\frac{y}{\log x}$ , e  $\frac{\log y}{\log x}$ .

**ANCOVA:** Análise de covariância univariável.

**ÂNGULOS:**

Ângulo					
Degrees	0°	30°	45°	60°	90°
Radiano	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$
Sen	0	$\frac{1}{2}$	$\frac{1}{\sqrt{2}}$	$\frac{\sqrt{3}}{2}$	1
Cos	1	$\frac{\sqrt{3}}{2}$	$\frac{1}{\sqrt{2}}$	$\frac{1}{2}$	0
Tan	0	$\frac{1}{\sqrt{3}}$	1	$\sqrt{3}$	$\infty$

Considere uma circunferência de centro O e um arco AB nessa circunferência. Se o arco AB tem comprimento igual ao raio, dizemos que ele mede 1 radiano. Portanto, radiano é a medida de um arco cujo comprimento é igual ao raio da circunferência que contém o referido arco. Como ao arco está associado um ângulo central, também podemos dizer que radiano é a medida do ângulo central que determina na

circunferência um arco cujo comprimento é igual ao raio. Uma vez que uma circunferência qualquer tem comprimento  $C = (2\pi \cdot r)$  umc (unidades de medida de comprimento), o arco de uma volta tem medida igual a  $2\pi$  radianos. De modo geral, dado um arco cujo comprimento é  $L$  umc, dizemos que sua medida, em radianos, é igual a  $\frac{L}{r}$ . Assim, se a circunferência do arco considerado tem raio unitário, a medida do

arco, em radianos, é numericamente igual ao comprimento do arco. Já a medida expressa em degree, é mensurada em graus.

**ANISOTROPIA:** É dada pela variabilidade diferencial ao longo das direções pesquisadas. A anisotropia pode ser geométrica ou zonal.

**ANISOTROPIA GEOMÉTRICA:** Caracteriza-se por variogramas com amplitudes diferentes ao longo das direções, porém sob um patamar constante.

**ANISOTROPIA ZONAL:** Caracteriza-se por variogramas diferentes, com amplitudes e patamares que variam ao longo das direções. Comum em depósitos estratificados, em que cada estrato, apresenta uma média e variância distinta das demais.

**ANO-BASE:** Ano arbitrário em relação ao qual um índice de preços mede o nível médio dos preços.

**ANO EPIDEMIOLÓGICO:** Transcorre do mês em que a incidência de uma determinada condição é mais baixa até o mesmo mês do ano seguinte.

**ANOS DE VIDAS AJUSTADOS PELA QUALIDADE (QUALITY-ADJUSTED LIFE YARS- QALY):** Indicador do estado de saúde que incorpora tanto a expectativa de vida como a percepção do impacto que a enfermidade e a incapacidade provocam na qualidade de vida.

**ANOVA:** Teste estatístico cujo modelo de distribuição de probabilidades é o da variância para  $k$  amostras ou tratamentos independentes. As amostras podem ser do mesmo tamanho ou desigual, no caso de  $k$  tratamentos que é apenas um critério, e devem ser iguais quando for o caso de  $k$  tratamentos e  $r$  blocos que são dois critérios. Este teste é também conhecido como teste  $F$ , em homenagem a Ronald Aylmer Fisher, e destina-se a comparar diferenças entre médias por meio de variâncias, cujos escores amostrais devem ser mensurados a nível intervalar ou de razões.

**ANOVA:** Teste de hipótese que objetiva comparar mais de duas médias. É isto mesmo, a análise de variância é um teste para comparar médias, que é realizado por meio de variâncias dentro e entre os conjuntos envolvidos. É uma extensão do teste  $t$  para duas médias. Em inglês Analysis of Variance.

**ANOVA:** Veja a Análise da variância.

**ANOVA:** Termo do inglês analysis of variance.

**ANOVA:** Análise de variância univariável.

**ANOVA:** Técnica estatística cujo objetivo é testar a igualdade entre três ou mais médias. Ela permite testar se a variabilidade dentro dos grupos é maior que a existente entre os grupos. A técnica supõe independência e normalidade das observações, e igualdade entre as variâncias dos grupos.

**ANOVA:** A Análise de Variância é utilizada para testar se várias médias são iguais. É uma extensão do teste  $t$  de duas amostras independentes.

**ANCOVA:** A análise de covariância é um Modelo Linear Geral (MLG) com uma variável explanatória contínua e um ou mais fatores. Ela testa se determinados fatores tem algum efeito após a remoção da variância dos previsores quantitativos ou covariáveis.

**ANOVA DE MÚLTIPLOS CRITÉRIOS:** Ver análise da variância (ANOVA ou ANAVA).

**ANOVA DE UM CRITÉRIO:** Ver análise da variância (ANOVA).

**ANOVA DE UMA VIA DE KRUSKAL-WALLIS:** Teste não-paramétrico de significância usado para comparar três ou mais grupos de dados ordinais. É análogo à análise de variância (ANOVA) de uma vida usada para três ou mais grupos de dados contínuos.

**ANONIMATO:** Descreve um survey no qual a identidade do respondente não pode ser determinada por ninguém, nem mesmo pelo diretor do survey. Nota: Survey por entrevista raramente são anônimos, já que o entrevistados normalmente conhece a identidade do entrevistado.

**ANEL:** É um conjunto munido de duas operações normalmente denominadas de adição e multiplicação que satisfazem certas propriedades ou axiomas. Alguns anéis mais conhecidos são: dos reais, dos números complexos, dos polinômios, das matrizes. Muitos anéis são associativos, mas alguns podem não ser como os anéis de Lie. Em inglês ring.

**APLICAÇÃO DA PESQUISA:** A implementação do questionário pré-codificado final ao processo de pesquisa.

**APRENDIZADO:** Processo seqüencial de grandes amostras de observações conhecidas como amostras de treinamento no qual os erros de classificação ou previsão são usados para recalibrar os pesos para melhorar a estimação.

**APRESENTAÇÃO DE CAULE E FOLHA:** Um a técnica de análise exploratória de dados que classifica dados quantitativos simultânea e ordenadamente e fornece subsídios sobre a forma de distribuição.

**APRESENTAÇÃO DOS DADOS:** Há duas formas de apresentação, que não se excluem mutuamente. A apresentação tabular, ou seja, é uma apresentação numérica dos dados em linhas e colunas distribuídas de modo ordenado, segundo regras práticas fixadas pelo Conselho Nacional de Estatística. A apresentação gráfica dos dados numéricos constitui uma apresentação geométrica permitindo uma visão rápida e clara do fenômeno.

**APROXIMADAMENTE IGUAL A:** Símbolo matemático dado por  $\cong$  ou  $\approx$ .

**APROXIMAÇÃO DA PROBABILIDADE PELA FREQUÊNCIA RELATIVA:** Valor da probabilidade, estimado com base em observações reais.

**APROXIMAÇÃO DE LAPLACE:** Método de aproximação de cálculo de integrais que utiliza transformações de Laplace. Em inglês Laplace approximation.

**APROXIMAÇÃO DIOFANTINA:** Uma aproximação diofantina é aproximar um número real por meio de um número racional, ou seja razão entre inteiros. Em inglês Diophantine approximation.

**APROXIMAÇÃO NORMAL PARA BINOMIAL:** Seja  $X$  uma variável aleatória com distribuição binomial. Se o número de ensaios  $n$  é muito grande e a probabilidade  $p$  de sucesso é muito pequena de forma



que  $np > 5$  e  $np(1-p) > 5$ , então  $(X-np)/(np(1-p))$  pode ser aproximada pela distribuição Normal com média igual a zero e variância igual a 1. Em inglês Normal approximation to binomial.

**APURAÇÃO DOS DADOS:** Resumo dos dados por meio de sua contagem e agrupamento. É a condensação e tabulação de dados.

**APURAÇÃO DOS DADOS:** Processo de tomar os dados brutos, registrados em fichas clínicas ou cadernos de laboratório, e organizá-los de forma satisfatória para posterior tabulação e análise.

**APRENDIZADO:** Processo seqüencial de grandes amostras de observações (conhecidas como amostras de treinamento) no qual os erros de classificação ou previsão são usados para recalibrar os pesos para melhorar a estimação.

**ARMAZENAMENTO DE DADOS:** Assimilação e integração de dados a partir de fontes internas e externas, em um formato adequado para processamento analítico on-line e aplicações de sistemas de suporte de decisão.

**ARR:** Aumento do risco relativo (1- RR).

**ARRANJOS:** O número de arranjos de  $m$  elementos tomados  $p$  a  $p$ , simbolizado por  $A(m,p)$  é a expressão para seu cálculo dada por:  $A(m,p) = m(m-1)(m-2)\cdots(m-p+1)$ . Ver Permutações.

**ÁREA ÚTIL:** É a área da unidade experimental ou parcela de um experimento ou ensaio, sob a qual o pesquisador faz as observações e medidas, a qual é delimitada e definida, após o descarte das bordaduras. O descarte da bordadura diminui o tamanho da parcela e isso tende a aumentar o erro experimental, por isso, o descarte da bordadura só deve ser feito nos casos indispensáveis. A largura das bordaduras ou número de linhas de proteção a usar depende do tratamento em estudo.

**ÁREA ÚTIL:** Parte central da parcela ou da unidade experimental onde é realizada a coleta dos dados experimentais, os quais serão submetidos a análise estatística.

**ÁREA AMOSTRADA:** Área total das parcelas, área fixa ou área variável numa dada amostra. **Ver também:** parcela, unidade amostral, parcela de área fixa e monitoramento.

**ÁREA DE EXCESSO:** Em gráficos em linhas ou lineares, quando representamos, em um mesmo sistema de coordenadas, a variação de dois fenômenos, a parte interna da figura formada pelos gráficos desses fenômenos é denominada de *área de excesso*.

**ÁRVORE DE DECISÃO:** Modelo baseado em regras que consiste de interconexões. Pontos de decisão e ramos ou ligações entre as interconexões que alcança múltiplos resultados com base na passagem em duas ou mais interconexões.

**ASPECTO ALEATÓRIO:** Trata-se de variações marcadamente irregulares e imprevisíveis de um ponto a outro da função  $f(x)$ , conforme a interpretação matemática das variáveis regionalizadas.

**ASPECTO ESTRUTURADO:** É dado pela variação previsível descrita pelas características estruturais dos fenômenos regionalizados.

**ASSIMETRIA:** Se um conjunto de dados for dividido em duas partes a partir da mediana e estas duas partes não forem iguais, então ele é dito assimétrico. Outra maneira de verificar se um conjunto é assimétrico

é calcular o seu coeficiente de assimetria ou o momento de terceira ordem. Se ele for diferente de zero então o conjunto é dito assimétrico. Em inglês skewness.

**ASSIMETRIA:** Ou grau de assimetria, é o desvio da simetria de uma distribuição de freqüências. O coeficiente de assimetria pode ser calculado a partir do terceiro momento centrado na média:

$$CA = \frac{\frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^3}{S^3}$$

**A assimetria pode ser:** positiva ou à direita, simétrica e negativa ou à esquerda.

**ASSIMETRIA:** Medida de simetria de uma distribuição; na maioria dos casos, a comparação é feita com uma distribuição normal. Uma distribuição positivamente assimétrica tem relativamente poucos valores grandes e uma cauda mais alongada à direita, e uma distribuição negativamente assimétrica tem relativamente poucos valores pequenos e cauda mais alongada à esquerda. Valores assimétricos fora do intervalo -1 a +1 indicam uma distribuição substancialmente assimétrica.

**ASSIMETRIA:** Característica de gráficos ou de curvas em que a maioria dos escores não se concentra no meio ou centro da distribuição como acontece com a curva normal ou de Gauss, mas numa das extremidades. Em inglês skewness.

**ASSIMETRIA NEGATIVA:** Característica de gráfico ou de curva em que a maioria dos escores se encontra à direita. Em inglês negative skewness.

**ASSIMETRIA NEGATIVA:** Quando, ao contrário da assimetria positiva, a maior parte da população amostrada concentra-se sobre os valores altos e uma pequena porcentagem sobre valores baixos. A constatação de assimetria negativa em distribuições de freqüências indica mistura ou contaminação de tipos litológicos ou minérios de diferentes naturezas.

**ASSIMETRIA POSITIVA:** Características de gráfico ou de curva em que a maioria dos escores se encontra à esquerda. Em inglês positive skewness.

**ASSIMETRIA POSITIVA:** Quando a maior parte da população amostrada apresenta-se com valores baixos e uma pequena porcentagem com valores altos. As distribuições lognormais caracterizam-se por apresentarem assimetria positiva.

**ASSIMÉTRICO:** Não simétrico, prolongando-se mais para um lado do que para o outro.

**ASSIMÉTRICO:** Não-assimétrico.

**ASSIMÉTRICO NEGATIVAMENTE:** Assimétrico à esquerda.

**ASSIMÉTRICO POSITIVAMENTE:** Assimétrico à direita.

**ASSÍNTOTA:** Reta cuja distância em relação a uma curva diminui indefinidamente sem nunca cortar a curva. Em inglês assyntota.

**ASSOCIAÇÃO:** Em inglês association.

**ASSOCIAÇÃO CAUSAL:** Que exprime causa; existe associação causal quando a alteração na frequência (ou qualidade) de um dos eventos acarreta mudanças no outro.

**ASSOCIAÇÃO CAUSAL DIRETA:** O fator exerce seu efeito na ausência de fatores intermediários ou variáveis intervenientes.

**ASSOCIAÇÃO CAUSAL INDIRETA:** O fator exerce seu efeito via fatores intermediários.

**ASSOCIAÇÃO, COEFICIENTE DE:** Entre os atributos A e B, num conjunto de N indivíduos, é o valor algébrico de  $Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)}$  em que  $\alpha$  e  $\beta$  indicam, respectivamente, ausência dos atributos A

e B; (AB) a frequência dos indivíduos que são ao mesmo tempo A e B ... e  $N = (AB) + (A\beta) + (\alpha\beta) + (\alpha B)$ . Proposto por YULE.

**ASSOCIAÇÃO DIRETA:** Não há elos intermediários entre causa e efeito. Ver associação indireta.

**ASSOCIAÇÃO ESPÚRIA:** Relacionamento entre uma causa presumida e um efeito presumido que ocorre como resultado de uma variável ou conjunto de variáveis não examinadas. Conhecida também como disparatada ou sem sentido.

**ASSOCIAÇÃO ESPÚRIA:** Expressão usada com mais de um significado; o mesmo que artificial, fictícia ou falsa, pois imputada a erro de delineamento, viés metodológico ou chance. Relação que não é devida ao evento ao qual se atribui. Às vezes, é usada para significar qualquer relação estatística não-casual.

**ASSOCIAÇÃO ESTATÍSTICA:** Relação quantitativa entre dois eventos, que podem estar ou não relacionados em termos de causa-efeito.

**ASSOCIAÇÃO INDIRETA:** Dois sentidos são encontrados para esta expressão; diz-se haver associação indireta: i) quando há elos intermediários entre causa e efeito; ii) quando a relação é devida à presença de um terceiro fator, ou seja, a associação estatística entre dois eventos é explicada pela presença de um terceiro.

**ASSOCIAÇÃO NÃO-CAUSAL:** A relação estatística encontrada entre dois eventos é devida a fatores como chance, viés ou o efeito de outras variáveis.

**ASSOCIAÇÃO NÃO-CAUSAL:** A relação entre duas variáveis é estatisticamente significativa, mas não existe relação causal, seja porque a relação temporal é incorreta pois a suposta causa acontece antes, em vez de depois do suposto efeito, ou também seja porque outro fator é responsável por ambos, suposta causa e suposto efeito.

**ASSOCIAÇÃO NÃO-CAUSAL (ESPÚRIA):** A relação entre duas variáveis é estatisticamente significativa, mas não existe relação causal, ou porque a relação temporal é incorreta, ou seja a causa presumida vem depois, em vez de antes, do efeito presumido, ou porque outro fator é responsável tanto pela causa como pelo efeito presumido.

**ASSOCIAÇÃO NEGATIVA:** Eventos em direção oposta; inversamente relacionados: por exemplo, classe social e incidência de desnutrição primária.

**ASSOCIAÇÃO POSITIVA:** Os dois eventos têm a mesma direção; diretamente relacionados: por exemplo; renda e escolaridade.

**ATÍPICO:** É um valor de um conjunto de dados que se afasta dos demais. É um valor normalmente muito grande ou muito pequeno quando comparado com o restante do conjunto. Pode ter sido resultado de um erro de medida ou, então, pode ser um indicativo de um comportamento atípico do conjunto sob determinadas condições. Conhecido também pelo nome de dado aberrante ou Outlier. Em inglês outlier.

**ATRIBUTO:** Quando os dados estatísticos apresentam um caráter qualitativo, o levantamento e os estudos necessários ao tratamento desses dados são designados genericamente de estatística de atributo.

**ATRIBUTO:** É uma denominação alternativa para os possíveis valores assumidos por uma variável qualitativa. Em inglês attribute.

**ATRIBUTO:** Característica de pessoas ou coisas.

**ATRIBUTO (CARACTERÍSTICA):** Uma característica qualitativa de um indivíduo, geralmente empregado em distinção a uma variável ou característica quantitativa. Portanto, para seres humanos, o sexo é um atributo e a idade é uma variável. Do inglês Attribute.

**ATRIBUTOS QUALITATIVOS:** São atributos que estão relacionados com uma qualidade e apresentam-se com várias modalidades.

**ATRIBUTOS QUANTITATIVOS:** São atributos aos quais é possível atribuir uma medida e apresentam-se com diferentes intensidades ou valores.

**AUDITORIAS:** O exame e a verificação da venda de um produto.

**AUTOCORRELAÇÃO:** Em inglês autocorrelation.

**AUTOCORRELAÇÃO:** A correlação interna entre membros de uma série de observações ordenadas no tempo ou espaço. Do inglês Autocorrelation.

**AUTOCORRELAÇÃO:** Ver correlação em série.

**AUTOREGRESSÃO:** A geração de uma série de observações pela qual o valor de cada uma é parcialmente dependente nos valores daquelas que imediatamente a precederam. Do inglês Autoregression.

**AUTOVALOR:** Medida da quantia de variância contida na matriz de correlação, de modo que a soma dos autovalores é igual ao número de variáveis. Também conhecida como a raiz latente ou a raiz característica.

**AUTOVALOR:** Soma em uma coluna de cargas fatoriais ao quadrado para um fator; também conhecido como raiz latente. Representa a quantia de variância explicada por um fator.

**AUTOVALOR:** O autovalor ou raiz característica de uma matriz quadrada  $A$  é o valor  $\lambda$  que satisfaz  $|A - \lambda I| = 0$ , onde  $I$  é a matriz identidade. Em inglês Eigenvalue. Ver Characteristic root.

**AUTOVALORES:** Ver raízes canônicas.

**AVALIAÇÃO CEGA (OU MASCARADA):** O indivíduo observado, o observador e/ou o analista de dados desconhecem a que grupo o primeiro pertence. Em oftalmologia, prefere-se o termo mascarado. Avaliação mono-cega: quando um deles, o observado ou o observador, desconhece a que grupo o primeiro pertence. Avaliação duplo-cega: quando o observado e observador desconhecem a que grupo o primeiro pertence.

**Avaliação triplo-cega:** quando o observado, o observador e o analista e todos os demais que lidam com os dados desconhecem a que grupo o primeiro pertence.

**AVALIAÇÃO DO BLOCO:** i) do bloco de cubagem é feita no fim da pesquisa e destina-se à determinação de reservas do depósito em estudo; ii) do bloco de desmonte é feita utilizando os dados de pesquisa e os dados de furos de desmonte, tendo como objetivo a melhor estimativa dos teores de lavra.

**AVALIAÇÃO ECONÔMICA:** Análise comparativa de determinadas atividades, processos ou estruturas em termos dos seus custos e conseqüências. Os principais métodos de avaliação econômica utilizados no domínio da saúde são a análise custo-utilidade. Cada uma procura identificar, medir e comparar tanto os custos como os efeitos das alternativas em estudo.

**AVALIAÇÃO SUBJETIVA:** Método para determinar quantas dimensões são representadas no modelo MDS que são técnicas de visualização de dados multidimensionais, conhecido como escalonamento multidimensional. O pesquisador faz uma inspeção subjetiva dos mapas espaciais e questiona se a configuração parece razoável. O objetivo é obter o melhor ajuste com o menor número de dimensões.

**AVALIAÇÕES DOS RISCOS EM SAÚDE (ARS):** Uso de questionário ou de programas de computador para compilar e avaliar as informações referentes a indivíduos oriundos da prática médica em uma clínica ou em uma indústria. Cada pessoa avaliada recebe informação referente à sua expectativa de vida e aos tipos de intervenções que provavelmente terão um impacto positivo na saúde ou na longevidade.

**AXIOMA:** Proposição que se aceita verdadeira sem demonstração. Em inglês axioma.

**AXIOMA:** É uma proposição aceita como verdadeira sem demonstração. Um termo mais antigo com o mesmo propósito é postulado. Na probabilidade existem basicamente três axiomas ligados ao conceito axiomático de probabilidade. Axioma 1: A probabilidade é um número entre zero e um, isto é,  $0 \leq P(A) \leq 1$ . Axioma 2: A probabilidade do espaço todo é igual a um, isto é,  $P(S) = 1$  e Axioma 3: A probabilidade da reunião ou união de dois eventos mutuamente exclusivos ou excludentes ( $A \cap B = \phi$ ) é a soma das probabilidades, isto é,  $P(A \cup B) = P(A) + P(B)$ . Em inglês axiom.

**AXIOMAS DE KOLMOGOROV:** Conjunto de axiomas definidos pelo matemático russo Kolmogorov (1933) que fundamentam a teoria de probabilidade em termos de teoria de conjuntos e da medida. Em inglês Kolmogorov axioms.

**AXIOMAS DE PROBABILIDADE:** Axiomas que definem a teoria clássica de probabilidades. Eles estabelecem que para dois eventos quaisquer A e B, a probabilidade P satisfaz os seguintes postulados: i)  $0 \leq P(A) \leq 1$ , ii)  $P(\Omega) = 1$  sendo  $\Omega$  o espaço amostral, iii)  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ . Em inglês Probability axioms.

**AZAR:** Sinônimo de acaso. Em inglês odds.

**AMOSTRAGEM DE PARCELAS EXPERIMENTAIS:** Todo pesquisador deve levar em conta os seguintes pontos quando for realizar amostragens em parcelas ou unidades experimentais de ensaios agrícolas: i) o método de amostragem; ii) o tamanho da amostra, iii) a escolha de unidades amostrais. A escolha de unidades amostrais depende da facilidade com que elas podem ser obtidas da parcela e da natureza da cultura; iv) amostragem em diferentes estágios de crescimento; v) amostragem de características qualitativas: os experimentos de campo frequentemente requerem um outro tipo de amostragem, isto é, uma

amostra tomada na colheita que representará uma característica qualitativa de uma cultura, tal como o teor de umidade ou teor de proteína em forragens, teor de açúcar em beterraba ou cana-de-açúcar, etc. Pelo menos duas unidades amostrais devem ser tomadas em cada parcela, para a avaliação de tais características qualitativas.

**AUTOCORRELAÇÃO:** Conceito usado em análise de regressão para indicar que quando o termo erro ou resíduo em um dado período de tempo é positivamente correlacionado com o termo erro no período de tempo anterior, o pesquisador depara-se com o problema de autocorrelação positiva de primeira ordem. Isto é comum em análises de séries de tempo ou séries temporal e conduz a erros padrões viesados para menor ou subestimados e, assim, para testes estatísticos e intervalos de confiança incorretos. A presença de autocorrelação de primeira ordem é testada pela utilização da tabela da estatística de Durbin-Watson aos níveis de 5% ou 1% de significância para n observações, e k' variáveis explicativas. Se o valor calculado de d na equação  $d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}$  é menor do que o valor tabulado de  $d_L$  que é limite inferior, a hipótese

de autocorrelação positiva de primeira ordem é aceita. A hipótese é rejeitada se  $d > d_U$  o qual é o limite superior, e o teste é inconclusivo de  $d_L < d < d_U$ .

**AUTOCORRELAÇÃO:** A autocorrelação ou correlação serial refere-se ao caso em que o termo erro em um dado período de tempo correlacionado com o termo erro de qualquer outro período de tempo. Se o termo erro em um dado período de tempo estiver correlacionado com o termo erro de um período de tempo anterior, há autocorrelação de primeira ordem, mais do que a de segunda ou de ordem maior. Ainda que seja possível uma autocorrelação negativa, a maioria das séries de tempo em economia apresentam autocorrelação positiva. Autocorrelação positiva de primeira ordem significa que  $> 0$ , portanto, violando a quarta hipótese do método dos mínimos quadrados (MMQ). Isto é comum em séries de tempo ou temporais.

$A_r^n$  : Arranjos de n objetos r a r.





# B

**BANCO DE DADOS:** Conjunto de dados organizado de maneira lógica, ou seja, numa sequência que permite acesso rápido e simples.

**BANCO DE DADOS:** Coleção ou arquivo de dados organizados de maneira específica e só acessado por pessoal com a necessária competência, para propósito definido.

**BANCO DE DADOS INTERNO:** Banco de dados desenvolvido a partir de dados particulares de uma organização.

**BARICENTRO:** Da distribuição conjunta de  $n$  variáveis aleatórias, é o ponto que tem por coordenadas as esperanças matemáticas ou médias aritméticas das distribuições marginais unidimensionais implicadas. Também se diz centro de gravidade.

**BASE:** É o valor, num determinado momento (efetivo ou resultante da média tomada dentro de um intervalo de tempo) que serve de termo de comparação, quando se quer calcular uma sucessão de números-índices. Costuma-se pospor com qualificativo ao nome do fenômeno a que se refere: preço-base, dentre outros.

**BASE DE DADOS:** Conjuntos de dados; por exemplo; por exemplo, de uma pesquisa.

**BATERIA DE TESTES:** Qualquer conjunto de teste aplicado a um ou mais indivíduos.

**BATERIA PONDERADA:** É a bateria de testes em que a cada um se associa um peso levado em conta na expressão numérica do resultado da aplicação da bateria toda.

**BERNOULLI ( $p$ ):** Distribuição de Bernoulli.

**BEM INFERIOR:** Bem que o povo adquire menos quando sua renda aumenta.

**BETA:** ver erro tipo II.

**BETA ( $\beta$ ):** Símbolo usado para representar a probabilidade de um erro tipo II, ou de segunda espécie.

**BETA ( $\beta$ ):** Ver erro tipo II.

**BEHAVIORSCAN:** Sistema de pesquisa baseado em scanner que mantém um painel de aproximadamente 3000 domicílios para registrar as compras dos consumidores com base na manipulação do mix de marketing.

**BIAS:** Distorção sistemática; o mesmo que erro sistemático, viés, vício e tendenciosidade. Pronúncia: baias.

**BIAS ( $\hat{\Theta}$ ):** Enviesamento do estimador  $\hat{\Theta}$ .

**BIC (CRITÉRIO DE INFORMAÇÃO BAYESIANA):** Critério para seleção de modelos baseado na utilização da razão de verossimilhanças penalizada pelo tamanho da amostra. Em inglês BIC (Bayesian Information Criteria).

**BIJETORA OU BIJETIVA:** Diz-se que uma função é bijetora se todo  $x$  do domínio possui uma e somente uma imagem  $f(x)$  no contradomínio.

**BIMODAL:** Que tem duas modas.

**BINOMIAL (DISTRIBUIÇÃO):** Distribuição discreta de probabilidades aplicável a um experimento realizado  $n$  vezes, em que cada prova tem uma probabilidade  $p$  de sucesso e é independente de qualquer outra prova.

**Bin( $n$ ,  $p$ ):** Distribuição Binomial.

**BinNeg( $s$ ,  $p$ ):** Distribuição Binomial negativa.

**BIOESTATÍSTICA OU BIOMETRIA:** Estatística aplicada às ciências biológicas e da saúde.

**BASE (BASE):** A função logarítmica  $\log_b a = X$  é descrita por dois parâmetros,  $a$  e  $b$ , tal que  $b^X = a$  é equivalente a  $\log_b a = X$ . O parâmetro  $b$  é a base do logaritmo.

**BETA (BETA):** A segunda letra ( $\beta$ ) do alfabeto grego. Na estatística, é usada para representar a probabilidade aceitável de cometer um erro Tipo II. Ela também aparece, com subscritos, como símbolo para os coeficientes e parâmetros de regressão.

**BIC:** Ver critério de informação bayesiano.

**BI-PLOT (BI-PLOT):** Um único gráfico dos resultados de duas ordenações relacionadas. Comumente usado para ilustrar os resultados da análise de correspondência, análise de correspondência canônica e análise de redundância.

**BLOCO (BLOCK):** Uma área, ou um intervalo de tempo, dentro da qual fatores não manipulados experimentalmente são considerados homogêneos.

**BONDADE DO AJUSTE (GOODNESS-OF-FIT):** Grau com que um conjunto de dados é acuradamente predito por uma dada distribuição ou modelo.

**BOOTSTRAP (BOOTSTRAP):** Na língua inglesa, a palavra bootstrap significa meios para conseguir algo com seu próprio esforço, sem ajuda de mais ninguém. Na estatística, o bootstrap é um procedimento de aleatorização ou Monte Carlo, no qual as observações em um conjunto de dados são reamostradas com reposição a partir do conjunto de dados original, e o conjunto de dados reamostrado é então usado para recalculá-la estatística teste de interesse. Isto é repetido um grande número de vezes, em geral 1.000-10.000, ou mais. A estatística teste que foi calculada com o conjunto de dados original é comparada com a distribuição daquela resultante das repetidas reamostragens dos dados, para obter um valor de  $P$  exato. Ver também jackknife.

**BOX PLOT (BOX PLOT):** Uma ferramenta visual que ilustra a distribuição de um conjunto de dados univariado. Ilustra a mediana, os quartis superior e inferior, os decis superior e inferior, e qualquer discrepância ou outlier.

**BIOESTAT:** Programa de computador utilizado para cálculos estatísticos, voltado principalmente para a área médica e biológica. Trata-se de um programa de computador ou software brasileiro, desenvolvido

com o apoio do Conselho Nacional de Pesquisa (CNPq) e disponível gratuitamente aos pesquisadores que o requisitam.

**BIOESTATÍSTICA:** Estatística aplicada à área da saúde. Conjunto de métodos estatísticos utilizados no tratamento da variabilidade nas ciências médicas e biológicas. O objeto de estudo da bioestatística é o planejamento e a análise de estudos médicos e biológicos visando à tomada de decisão na avaliação de tratamentos e verificando a influência de fatores de risco no aparecimento de doenças.

**BIOMETRIA:** Ver bioestatística.

**BOX PLOT:** Ver diagramas de caixas.

**BIOMETRIA:** Ciência que trata da estatística aplicada a biologia.

**BISSERIAL (CORRELAÇÃO):** Técnicas de correlação em que uma variável está em escala intervalar e a outra é uma dicotomia artificial em escala ordinal. A dicotomia é artificial porque há uma distribuição contínua subjacente.

**BISETRIZ:** Reta ou semi-reta com origem no vértice de um ângulo e que forma com os lados ângulos congruentes. Em inglês bissetriz.

**BITTERLICH (POSTULADO DE):** Proposição apresentada por Walter Bitterlich, Engenheiro Florestal austríaco que relaciona a área basal de um povoamento com o número de árvores, que vistas de um determinado ponto fixo no povoamento, cujos diâmetros aparecem maiores que uma dada abertura angular. O enunciado é: o número de árvores ( $n$ ) de um povoamento, cujos os diâmetros à altura do peito (DAP) vistos de um ponto fixo aparecem maiores a um dado ângulo  $\alpha$  é proporcional à sua área basal por hectare ( $G$ ).

**BITTERLICH (WALTER):** Engenheiro Florestal austríaco que desenvolveu o método de amostragem por enumeração angular e relascópio. Nasceu em 1908, em Reutte no Tirol, onde seu pai era responsável por um grande distrito florestal. Defendeu sua tese de doutorado sobre a amostragem por enumeração angular em 1949 na Hochschule für Bodenkultur.

**BLOCO:** Conjunto relativamente homogêneo de unidades experimentais.

**BLOCO:** Em análise da variância, um grupo de indivíduos similares.

**BLOCO AO ACASO:** O mesmo que blocos casualizados.

**BLOCO CASUALIZADOS:** Delineamento experimental em que os tratamentos são atribuídos por sorteio às unidades experimentais de cada bloco.

**BLOCO CASUALIZADOS COM ALGUNS TRATAMENTOS COMUNS:** Delineamento formado por vários ensaios em blocos completos casualizados, tais que em todos esses ensaios aparecem  $c$  tratamentos ditos comuns ao lado de outros que são distintos chamados regulares.

**BLOCO:** É um conjunto ambiental homogêneo que contém todos os tratamentos ou parte deles. No delineamento em quadrado latino, em que se apresentam dois tipos de blocos, eles recebem as denominações de colunas e linhas. Também conhecido como controle local.

**BORDADURA:** É a parte da unidade experimental não coletado para avaliação do efeito do tratamento. É empregada para evitar efeito de competição ou de contaminação entre parcelas vizinhas. Normalmente é constituída pelo mesmo material da área útil. Em alguns casos pode ser montada com materiais distintos, como, no caso de alguns experimentos em fitopatologia, por exemplo, em que as bordaduras de blocos são semeadas antecipadamente para servirem de fonte de inoculo para o ensaio, como forma de se garantir maior uniformidade de infecção.

**BLOCO:** Em um planejamento de experimentos, um grupo de unidades experimentais ou materiais que é relativamente homogêneo. A finalidade de dividir unidades experimentais em blocos é produzir um planejamento experimental em que a variabilidade dentro dos blocos seja menor que a variabilidade entre os blocos. Isso permite que os fatores de interesse sejam comparados em um ambiente que tenha menor variabilidade que o experimento sem blocos.

**BLOCO COMPLETO:** Qualquer bloco que inclua todos os tratamentos do ensaio.

**BLOCO DE CUBAGEM:** Refere-se a um bloco de minério que deve ser avaliado, cuja geometria e dimensões são definidas pelos dados de pesquisa. Geralmente os métodos clássicos de avaliação de reservas fazem uso dos blocos de cubagem, os quais são definidos em função do princípio de interpretação utilizado.

**BLOCO DE LAVRA:** Unidade de mineração básica que atende a produção, dentro de um planejamento semanal, quinzenal ou mensal de mina. Os blocos de lavra são avaliados com os dados de pesquisa e atualizados com os dados de furos de desmonte.

**BLOCO, DESENHO (RANDOMIZAÇÃO):** Termo usado em experimentação para selecionar grupos de unidades que sejam tanto quanto possível homogêneas dentro de um grupo. Em seguida, designar randomicamente diferentes tratamentos a unidades no mesmo grupo. Os grupos são chamados blocos. Se cada bloco tiver exatamente tantas unidades quantos tratamentos distintos e cada tratamento for usado em apenas uma unidade em cada bloco, então, o desenho experimental é conhecido como Complete Block Design.

**BLOCOS ALEATORIZADOS:** Um experimento em que cada bloco contém uma replicação completa dos tratamentos que são alocados para várias unidades nos blocos de uma maneira aleatória e, consequentemente, permitindo estimações sem vício de erro. Em inglês Randomized blocks.

**BLOCOS AUMENTADOS DE FEDERER (BAF):** Este delineamento foi desenvolvido pelo estatístico Federer (1956), por ocasião de sua estada no Estado Americano do Hawai. Foi preparado para ser aplicado na cultura de cana – de – açúcar, sendo mais tarde expandido para outros cultivos de multiplicação assexuada e de plantas perenes e, posteriormente, para culturas anuais de multiplicação sexuada. O esquema consiste na formação de blocos completos com os materiais testemunhas e adicionalmente a inclusão nos blocos de linhagem novas a serem testadas, e cada linhagem aparece em um e somente um bloco, daí o seu nome. Este esquema pode também ser aplicado nos delineamentos quadrado latino, inteiramente casualizado, ou em qualquer outro delineamento. Os tratamentos testemunhas serão aqueles que permitirão a estimativa do erro.

**BLOCOS COMPLETOS CASUALIZADOS:** Delineamento em que cada bloco inclui uma e uma só vez todos os tratamentos. Há uma modalidade, raramente usada, em que cada bloco inclui n vezes todos os tratamentos.

**BLOCOS INCOMPLETO:** Qualquer bloco que não inclua todos os tratamentos do ensaio.

**BLOCOS INCOMPLETOS:** Forma básica de um planejamento de experimento introduzido por Yates (1936). Se o material é dividido em blocos e se deseja alocar certos tratamentos nas unidades dos blocos, os tratamentos podem ser numerosos para que todos apareçam em cada bloco. Quando um bloco contém menos do que uma replicação completa de tratamentos ele é chamado de bloco incompleto. Em inglês Incomplete blocks.

**BLOCOS INCOMPLETOS EQUILIBRADOS:** Delineamento em blocos incompletos em que todos os blocos têm o mesmo tamanho, todos os tratamentos têm o mesmo número de repetições, e qualquer par de tratamentos ocorre  $\lambda$  vezes, com ambos no mesmo bloco. Também conhecido como Resolvable designs, os outros são os não equilibrados ou Non-resolvable designs. São exemplos os delineamentos Alfa-látices.

**BLP:** Melhor preditor linear não viciado.

**BLUE:** Melhor estimador linear não viciado.

**BP:** Melhor preditor.

**BLUP:** Melhor preditor linear não viciado.

**BLUP:** Teoria genético-quantitativa sobre a qual se baseiam os sistemas mais avançados de avaliação genética; Melhor Predição Linear Não-Viciada (Best Linear Unbiased Prediction). Uma das propriedades mais importantes da metodologia BLUP é proporcionar os fatores de ponderação apropriados para que as diversas fontes de informação disponíveis a respeito de cada indivíduo possam ser resumidas a um único valor. Este valor pode ser definido como DEP.

**BONDADE DE AJUSTE:** Em geral, o nível de ajuste entre uma série de valores observados e uma segunda série que é derivada inteiramente ou parcialmente de uma base hipotética, ou seja, deriva do ajuste de um modelo aos dados. O termo é usado especialmente em relação ao ajuste de distribuições teóricas a dados observados e o ajuste de retas de regressão. Em inglês Goodness of fit.

**BORDADURA:** É a parte do material experimental que pertence à parcela, recebe os tratamentos, como o restante do material, mas não é considerada na análise dos resultados.

**BORDADURAS:** São as fileiras de proteção da parcela ou unidade experimental. Têm a finalidade de evitar a influência sobre a parcela dos tratamentos aplicados nas parcelas vizinhas ou adjacentes, bem como a competição entre os canteiros.

**BINÁRIO:** Termo usado para indicar o que possui apenas duas possibilidades. Por exemplo, uma variável binária de pesquisa pode assumir apenas dois valores como por exemplo, defeituoso e não defeituoso, masculino e feminino, sim e não dentre outras. Notação matemática que representa os números inteiros utilizando apenas algarismos zero (0) e um (1).

**BLOBOGRAMA:** Tipo de diagrama comumente utilizado nas meta-análises e que fornece um suporte visual adequado para a comparação de diversos autores ou estudos em relação a algum critério teórico.

**BONFERRONI:** Procedimento de comparação entre médias de vários tratamentos no sentido de verificar a significância estatística das diferenças entre essas medidas de tendência central, determinando-se,

a priori, o nível de alfa decisão. A comparação é efetuada após análise da variância e somente se o valor de F for significativo

**BOOSTING:** Tipo de Procedimento denominado boosting, que é um método geral para melhorar o desempenho de classificadores. Este procedimento é uma das mais importantes e bem sucedidas metodologias desenvolvidas nos últimos anos na literatura de classificação de padrões e tem recebido bastante atenção dos estatísticos desde seu surgimento na literatura de aprendizado. O procedimento boosting consiste em aplicar, sequencialmente, uma regra de classificação qualquer, chamada de classificador base a versões iterativamente reponderadas da amostra de treinamento. Em cada iteração o algoritmo atribui pesos maiores às observações incorretamente classificadas na iteração anterior. A regra final de classificação é obtida por meio de uma combinação linear dos classificadores construídos em cada iteração, produzindo um comitê de classificação. Para muitos métodos de classificação, esta estratégia simples resulta em uma impressionante melhora de desempenho. Uma das versões do procedimento boosting mais utilizadas e difundidas na literatura e na prática é o algoritmo Adaboost (Adaptive Boosting). em problemas de classificação com dois grupos o denominam Adaboost Discreto.

**BOOTSTRAP:** Método de reamostragem com reposição.

**BOOTSTRAP:** Procedimento que consiste em retirar de uma pequena amostra numerosas amostras, com reposição, cada uma com probabilidade  $\frac{1}{n}$  de ser obtida, sendo também conhecido como técnica de reamostragem ou de simulação.

**BOOTSTRAPPING:** Forma de reamostragem na qual os dados originais são repetidamente amostrados com substituição para estimação do modelo. Estimativas de parâmetros e erros padrão não são mais calculados com suposições estatísticas, mas, em vez disso, são baseados em observações empíricas.

**BOOTSTRAPPING:** Em inglês bootstrapping.

**BIMODAL:** Uma distribuição que tem duas modas ou grupos de modais.

**BIOMETRIA:** Um amplo procedimento numérico para estudo biológico que inclui delineamento do estudo, coleta de dados, análise, exibição e conclusões adequadas.

**BLOCO ALEATÓRIO:** Um delineamento particular da ANOVA em que cada um dos tratamentos do teste é distribuído aleatoriamente nas unidades, dentro de um bloco, o delineamento completo compreende certa quantidade desses blocos.

**BLOCOS:** Grupos de indivíduos em um delineamento experimental, os quais são observados em condições específicas, por exemplo, após a distribuição dos indivíduos em um tratamento. Os blocos geralmente são criados para isolar as fontes de variabilidade, de modo que os indivíduos dentro de um bloco apresentem menor variabilidade do que a população geral.

**BOOTSTRAPPING:** Um método interativo para estimar parâmetros de interesse com base em reamostragens.

**BOX & WISKERS:** É um tipo de representação gráfica, em que se realçam algumas características dos dados. O conjunto dos valores compreendidos entre o 1º e o 3º quartis, é representado por um retângulo ou caixa com a mediana ou a média aritmética indicada por uma barra. Consideram-se seguidamente duas linhas que unem os lados dos retângulos com os valores extremos.

**BQUE:** Melhor estimador quadrático não viciado.

**BRAQUISTÓNA:** É a curva de tempo mínimo, isto é, considerando todas as curvas que unem dois pontos ela é a que pode ser percorrida no menor tempo. Em inglês braquistona.

**BUE:** Melhor estimador não viciado.

**BUSCA DE CASOS:** Processo para detectar doenças assintomáticas ou fatores de risco nas pessoas, enquanto elas se acham em um ambiente clínico, isto é, enquanto elas se encontram sob assistência médica. A diferença entre rastreamento e busca de casos é frequentemente ignorada na literatura e na prática; essa distinção é importante porque muitos dos critérios para a realização de rastreamento em uma comunidade não precisam ser cumpridos no processo de busca de casos. Estudo de casos de controles. Os grupos de estudo são definidos com base na situação de doença ou de desfecho. A frequência do fator risco ou exposição nos casos de indivíduos doentes é comparada com a frequência do fator de risco ou exposição nos controles que são indivíduos não doentes.







**C:** Número de classes.

$C_r^n$ : Combinações de n objetos r a r.

**CADASTRO (OU MARCO) DE AMOSTRAGEM:** Rol de unidades ou indivíduos, como na relação de domicílios, utilizando para selecionar uma amostra.

**CADEIA DE MARKOV:** Essa expressão é usada com dois sentidos diferentes, ambos relacionados com o Processo de Markov. Em um deles, um processo  $\{x_t\}$  é chamado de cadeia se o parâmetro de tempo é discreto. No outro, é chamado de cadeia se os valores de  $x$  são discretos. O primeiro é preferível. Em inglês Markov Chain.

**CADEIAS DE MARKOV:** É um processo estocástico definido assim: Consideremos uma sequência de ensaios cujos resultados,  $x_1, x_2, \dots$ , satisfazem as seguintes propriedades. i) Cada resultado pertence a um conjunto finito de resultados, chamado o espaço dos estados do sistema; se o resultado da  $n$ -ésima tentativa é  $a_i$  dizemos que o sistema se encontra no estado  $a_i$  no instante  $n$ . ii) O resultado de qualquer ensaio depende no máximo do resultado do ensaio imediatamente anterior e não de qualquer outro dos precedentes; a cada par de estados  $(a_i, a_j)$  está associada a probabilidade  $P_{ij}$  de que  $a_j$  ocorre imediatamente após ter ocorrido  $a_i$ . Sendo assim o processo estocástico com as propriedades descritas anteriormente é chamado cadeia de Markov finita. Os números  $P_{ij}$ , chamados probabilidades de transição, podem ser dispostos segundo a matriz  $P$  abaixo, a qual é denominada de matriz de transição.

$$P = \begin{pmatrix} P_{11} & P_{12} & \cdots & P_{1m} \\ P_{21} & P_{22} & \cdots & P_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ P_{m1} & P_{m2} & \cdots & P_{mm} \end{pmatrix}$$

Assim a cada estado  $a_i$  corresponde a  $i$  – ésimas linha  $(p_{i1}, p_{i2}, \dots, p_{im})$  da matriz de transição  $P$ , se o sistema está no estado  $a_i$ , então esse vetor – linha representa as probabilidades de todos os possíveis resultados do próximo ensaio, de forma que é um vetor de probabilidade. Consequentemente, a matriz de transição  $P$  da cadeia de Markov é uma matriz estocástica. Ver matriz estocástica.

**CADEIA DE MARKOV ABSORVENTE:** Uma classe importante de cadeias de Markov não ergódicas; aquela cadeia que possui um estado absorvente. Absorbing Markov chain.

**CADEIA DE MARKOV FINITA:** Uma cadeia de Markov  $\{x_n\}$  é chamada finita com  $k$  estados se o número de valores possíveis da variável aleatória  $\{x_n\}$  é finito e igual a  $k$ . Em inglês Finite Markov chain.

**CADEIA DE MARKOV IRREDUTÍVEL:** Uma cadeia de Markov é chamada irredutível se todos os pares de estados da cadeia se comunicam. Em inglês Irreducible Markov chain.

**CADEIA ERGÓDICA:** Cadeia de Markov com propriedade de ergodicidade. Em inglês Ergodic chain.

**CAIXA DE BIGODES (BOXPLOT):** É um tipo de representação gráfica, em que se realçam algumas características da amostra. O conjunto dos valores da amostra compreendidos entre o 1º e o 3º quartis,  $Q_{25}$  e  $Q_{75}$  é representado por um retângulo ou caixa com a mediana indicada por uma barra. Consideram-se seguidamente duas linhas que unem os lados dos retângulos com os chamados valores adjacentes.

**CAIXA DE TEXTO:** Espaço destinado à entrada de uma determinada informação, requerida para o teste de hipótese em programas de computador, software ou pacote estatístico.

**CÁLCULO DE COMPOSIÇÕES:** É feito por meio da média ponderada do teor pelas espessuras selecionadas para o intervalo de trabalho, como mostra a equação a seguir:

$$t_c = \frac{\sum_{i=1}^n t_i e_i}{\sum_{i=1}^n e_i}$$

onde  $n$  é o número de trechos para compor o intervalo de trabalho;  $t_i$  é o teor do  $i$ -ésimo trecho;  $e_i$  é a espessura do  $i$ -ésimo trecho.

**CÁLCULO DE COVARIÂNCIAS:** Para fins de krigagem de bloco, onde se deseja estimar o teor médio do bloco ou painel e não de um único ponto no interior dos mesmos, é feito por meio da discretização, do domínio a ser estimado, em um número de pontos igualmente espaçados dentro do domínio, de tal forma que cada ponto represente o centro de um sub-bloco ou sub-painel. Feito isto, o cálculo da covariância média é simplesmente a média aritmética das covariâncias entre uma amostra e cada ponto ou então somente entre os pontos dentro do domínio, respectivamente para a covariância média entre uma amostra e o bloco/painel ou para a covariância média dentro do bloco/painel.

**CÁLCULO DO TAMANHO DA AMOSTA:** Cálculo matemático, feito geralmente quando o ensaio é planejado, que estabelece o número de pacientes que deve ser recrutado, a um dado nível de significância e um dado poder de teste.

**CAMADA OCULTA:** Série ou camada de interconexões em uma rede neural perceptron de multicamadas que estão entre as interconexões de entrada e saída. O pesquisador pode ou não controlar o número de interconexões por camada oculta ou a quantia de camadas ocultas. As camadas ocultas fornecem a capacidade de representar funções não-lineares no sistema de rede neural.

**CAMPO DE NÚMEROS ALGÉBRICOS:** Um campo algébrico é um sub-campo dos Reais (R) ou Complexos (C) que contém os números racionais (Q) bem como as raízes de algum polinômio com coeficientes racionais. Por exemplo, o campo  $Q(\sqrt{2})$  é o conjunto dos números reais  $\{a + b\sqrt{2} \mid a, b \text{ racionais}\}$ . Este campo contém as raízes da equação  $x^2 - 2 = 0$ . Em inglês algebraic number field.

**CAMPOS DE VARIÁVEL:** Espaços de computador destinados a uma variável e a ela correspondentes no questionário.

**CAMPO GEOMÉTRICO:** É a extensão máxima possível do variograma, considerando as informações disponíveis. Geralmente, para dados com uma extensão igual a  $L$ , tem sido aceito um campo geométrico igual a  $\frac{L}{2}$ .

**CÁLCULO DE PROBABILIDADE (PROBABILITY CALCULUS):** Temática usada para lidar com probabilidades, envolvendo quaisquer conceitos.

**CARGA (LOADINGS):** Em uma análise de componentes principais, os parâmetros que são multiplicados por cada variável de uma observação multivariada para gerar uma nova variável, ou escore de componente principal para cada observação. Cada um desses escores é uma combinação linear das variáveis originais.

**CARGA DOS FATORES (FACTOR LOADINGS):** Os coeficientes lineares com que os fatores comuns em uma análise fatorial são multiplicados para obter uma estimativa da variável original.

**CASUAL (HAPHAZARD):** Atribuição de indivíduos ou populações a grupos de tratamentos que não é verdadeiramente aleatória. Comparar com randomização.

**CCA:** Ver análise de correspondência canônica.

**CÉLULA (CELL):** Um campo em uma planilha. É identificada pelo número de sua linha e coluna e contém um único valor, ou dado.

**CENTROIDE (CENTROID):** O vetor médio de uma variável multivariada. Pense nele como o centro de uma nuvem de pontos no espaço multivariado, ou o ponto em que você poderia equilibrar essa nuvem de pontos na ponta de uma agulha.

**CENTRO DE GRAVIDADE (OU DE MASSA):** ermo usado na física para indicar o centro de equilíbrio de um corpo, barra ou sistema. Um corpo consiste de um grande número de partículas que estão todas sujeitas à gravidade. Para se encontrar o centro de gravidade pode-se citar o caso de simples de dois pontos  $P_1$  e  $P_2$  de mesma massa. Então, por simetria, o centro de gravidade é o ponto médio  $C$ , que por meio da álgebra de vetores, é representado através de setas com a origem em zero (0) e extremidades em  $P_1$  e  $P_2$ , dados por  $a_1$  e  $a_2$  respectivamente. Usando um paralelogramo, determina-se a soma dos vetores  $a_1$  e  $a_2$ . Já que as diagonais de um paralelogramo se bisseccionam, o vetor  $c$ , que aponta para  $C$  é dado por,

$$c = \frac{a_1 + a_2}{2}.$$

Isto é simplesmente a média aritmética dos dois vetores  $a_1$  e  $a_2$ . Sejam  $a_1 = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$  e  $a_2 = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$ .

Então  $c$  tem coordenadas  $\frac{(x_1 + x_2)}{2}$  e  $\frac{(y_1 + y_2)}{2}$ . Se forem dados  $n$  pontos de mesma massa, por vetores

$a_1, a_2, \dots, a_n$ , a mecânica mostra que o centro de gravidade é dado pelo vetor  $c = \frac{a_1 + a_2 + \dots + a_n}{n} = \frac{1}{n} \sum_{i=1}^n a_i$ . O resultado é simplesmente uma extensão da fórmula  $c = \frac{a_1 + a_2}{2}$ . A mesma

fórmula é de importância básica na análise estatística de direções, como aplicada em exemplos de aves migradoras. Uma generalização ocorre quando diferentes massas  $M_i$  estão localizadas nos pontos determinados por vetores  $a_i (i = 1, 2, 3, \dots, n)$ . Neste caso, a média aritmética usual se transforma na média aritmética ponderada, ou seja,

$$c = \frac{M_1 a_1 + M_2 a_2 + \dots + M_n a_n}{M_1 + M_2 + M_3 + \dots + M_n}$$

Os exemplos podem ser aplicados ou estendidos para casos de três massas, quatro massas, etc., mas a fórmula é válida nos espaços bi e tridimensionais. É também útil nos espaços de dimensões mais elevadas, apesar de perder seu significado físico. Uma importante aplicação da equação é feita na análise estatística

multivariada. A esperança matemática, média, valor médio, valor esperado ou esperança de variável aleatória discreta ou contínua, é interpretada como um centro de gravidade, a qual é dada para o caso discreto por

$$E(X) = \sum_{i=1}^m p_i x_i$$

a qual tem uma interessante correlação em mecânica, pois se for considerado cada valor  $x_i$  como a abscissa de um ponto em uma reta e se for interpretado  $p_i$  como uma massa ou um peso concentrados no ponto  $x_i$ , então, o ponto com abscissa  $E(X)$  é idêntico ao centro de gravidade (c.g.) do sistema.

**CIRCULARIDADE (CIRCULARITY):** A premissa da análise de variância de que a variação entre amostras dentro de subparcelas ou blocos é a mesma ao longo de todas as subparcelas ou blocos.

**CLASSIFICAÇÃO (CLASSIFICATION):** O processo de posicionar, colocar em grupos. Um termo geral para os métodos multivariados, como a análise de agrupamento ou a análise discriminante, que estão preocupados em agrupar observações. Comparar com ordenação.

**COEFICIENTE BINOMIAL (BINOMIAL COEFFICIENT):** A constante ou a combinação pela qual multiplica-se a probabilidade de uma variável aleatória de Bernoulli para obter uma variável aleatória binomial. O coeficiente binomial ajusta a probabilidade para levar em conta múltiplas combinações equivalentes de cada um dos dois resultados possíveis. Ele é escrito como:

$$\binom{n}{X}, \text{ e é lido como "n sobre X". É calculado como: } \frac{n!}{X!(n-X)!},$$

onde "!" indica o fatorial de um número ou o produto de n números inteiros.

**COEFICIENTE DE CAMINHOS OU DE CAMINHAMENTO OU DE TRILHA (PATH COEFFICIENT):** Um tipo de parâmetro de regressão parcial calculado na análise de caminhos que indica a magnitude e o sinal de efeito de uma variável sobre outra.

**CANÔNICO:** Termo usado em estatística para implicar algo que foi reduzido à sua forma mais simples.

**CAR (CONDICIONAL AUTO-REGRESSIVO):** Efeitos aleatórios estruturados, utilizados em modelos para dados espaciais. Inicialmente utilizados na restauração eletrônica de imagens e também usados em estatística por Besag (1974). Assumem, para um particular local geográfico, uma média igual à média dos vizinhos e uma variância inversamente proporcional ao número de vizinhos. Em inglês CAR (conditional autoregressive).

**CARACTERÍSTICA (ATRIBUTO):** Uma característica qualitativa de um indivíduo, geralmente empregado em distinção a uma variável ou característica quantitativa. Portanto, para seres humanos, o sexo é um atributo e a idade é uma variável.

**CARACTERÍSTICA QUALITATIVA:** Características em que há uma distinção bem definida entre fenótipos, tal como pelagem negra e branca ou mocho e chifrudo. Normalmente, um único ou poucos pares de genes estão envolvido (s) na expressão de caracteres qualitativos.

**CARACTERÍSTICA QUALITATIVA:** Característica que pode ser descrita em detalhes, mas que não é verdadeiramente mensurável.

**CARACTERÍSTICA QUANTITATIVA:** Característica que pode ser descrita por uma escala de medidas rígida e dimensional.

**CARACTERÍSTICA QUANTITATIVA:** Características em que não há nenhuma distinção bem definida entre fenótipos, com variação gradual de um fenótipo a outro. Por exemplo, altura, perímetro torácico, peso ao desmame, dentre outras. Geralmente, muitos pares de genes e influências ambientais estão envolvidos.

**CARACTERÍSTICAS ESPAÇO-TEMPORAIS:** Valores observados ao longo do tempo e em diversos locais geográficos. Em inglês Space-time attributes.

**CARDINAL:** Símbolo de contagem que indica número de elementos. #.

**CARGA ALOSTÁTICA:** O termo alostático provém do grego e significa encontrar estabilidade por meio da mudança. É o nível atual de demanda para adaptações em um indivíduo. Um nível muito alto pode contribuir de forma importante para muitas doenças crônicas. Em inglês Allostatic load.

**CARGA FATORIAL:** Em análise fatorial, medidas da correlação entre um item e um dos fatores calculados.

**CARGAS CANÔNICAS:** Medida da correlação linear simples entre as variáveis independentes e suas respectivas variáveis estatísticas canônicas, Essas podem ser interpretadas como cargas fatoriais e também são conhecidas como correlações de estrutura canônica.

**CARGAS CRUZADAS CANÔNICAS:** Correlação de cada variável independente ou dependente observada com a variável estatística canônica oposta. Por exemplo, as variáveis independentes são correlacionadas com a variável estatística canônica dependente. Elas podem ser interpretadas como cargas canônicas, mas com a variável estatística canônica oposta.

**CARGAS FATORIAIS:** Correlação entre as variáveis originais e os fatores, bem como a chave para o entendimento da natureza de um fator em particular. As cargas fatoriais ao quadrado indicam qual percentual da variância em uma variável original é explicado por um fator.

**CARGAS DISCRIMINANTES:** Medida da correlação linear simples entre cada variável independente e o escore Z discriminante para cada função discriminante; também chamadas de correlações estruturais. As cargas discriminantes são calculadas sendo ou não incluída uma variável independente na função discriminante.

**CARTA :** Uma carta de controle usada quando um processo de produção é medido em termos do valor médio de uma variável como comprimento, peso, temperatura e assim por diante.

**CARTA DE CONTROLE:** Uma ferramenta gráfica usada para ajudar a determinar se um processo está sob controle ou fora de controle.

**CARTA np:** Uma carta de controle usada para monitorar o processo de produção em termos do número de itens defeituosos.

**CARTA p:** Uma carta de controle usada quando um processo de produção é medido em termos da proporção defeituosa.

**CARTA R:** Uma carta de controle usada quando um processo de produção é medido em termos da amplitude de uma variável.



**CARTOGRAMAS:** São ilustrações relativas a cartas geográficas ou mapas. O objetivo desse gráfico é o de figurar os dados estatísticos diretamente relacionados com áreas geográficas ou políticas.

**CASO-ÍNDICE:** Caso em que a condição sob investigação foi identificada pela primeira vez.

**CASO SINGULARIZADO:** Uma investigação de um indivíduo ou grupo em que as variáveis medidas e os relacionamentos empíricos explorados são características do indivíduo ou grupo e não de uma sub – unidade dele.

**CASOS HEYWOOD:** Um tipo comum de estimativa transgressora, que ocorre quando o termo do erro estimado para um indicador se torna negativo, o que é um valor absurdo. O problema é remediado pela eliminação do indicador ou pela restrição do erro de mensuração para um valor positivo pequeno.

**CAPACIDADE DE PROCESSO:** A capacidade de um processo para produzir produtos dentro dos limites de especificação. Veja razão de capacidade de processo, estudo de capacidade de processo, PCR e PCR<sub>k</sub>.

**CAUSA ATRIBUÍDA:** A porção da variabilidade em um conjunto de observações que pode ser registrada como causas específicas, tais como operadores, materiais ou equipamentos. Também chamada de causa especial.

**CAUSA CASUAL:** A porção da variabilidade em um conjunto de observações que é devida somente a forças aleatórias e que não pode ser creditada a fontes específicas, tais como operadores, materiais ou equipamentos. Também chamada de causa comum.

**COEFICIENTE DE CONFIANÇA:** A probabilidade (1-α) associada com um intervalo de confiança expressando a probabilidade de que o intervalo verdadeiro estabelecido conterá o valor verdadeiro do parâmetro.

**COEFICIENTE DE CORRELAÇÃO:** Uma medida adimensional da associação linear entre duas variáveis, geralmente estando no intervalo de -1 a +1, com zero indicando a ausência de correlação, porém não necessariamente a independência das duas variáveis.

**COEFICIENTE DE DETERMINAÇÃO:** É o coeficiente que mede a qualidade do ajuste de um modelo matemático de regressão, pode múltiplo ou simples, e é dado pela porcentagem da variação total que é explicada pela equação de regressão estimada. Veja R<sup>2</sup>.

**COEFICIENTE(S) DE REGRESSÃO:** São o(s) parâmetro(s) em um modelo de regressão.

**COEFICIENTE DE CORRELAÇÃO PRODUTO-MOMENTO (R) (PRODUCT MOMENT CORRELATION COEFFICIENT):** Uma medida de quão bem duas variáveis são relacionadas uma com a outra. Ele pode ser calculado como a raiz quadrada do coeficiente de determinação, ou como a soma dos produtos cruzados dividida pela raiz quadrada do produto da soma dos quadrados das variáveis X e Y.

**COEFICIENTE DE DETERMINAÇÃO (R<sup>2</sup>) (COEFFICIENT OF DETERMINATION):** A quantidade de variabilidade da variável resposta, que é explicada por um modelo de regressão linear simples. É igual à soma dos quadrados da regressão dividida pela soma dos quadrados total.

**COEFICIENTE DE DISPERSÃO (COEFFICIENT OF DISPERSION):** A quantidade obtida pela divisão da variância da amostra pela média da amostra. O coeficiente de dispersão é usado para determinar se os indivíduos distribuídos com padrão espacial agregado, regular, aleatório ou hiperdisperso.

**COEFICIENTE DE VARIAÇÃO (CV) (COEFFICIENT OF VARIATION):** A quantidade obtida pela divisão do desvio padrão da amostra pela média da amostra. O coeficiente de variação é útil para comparar a variabilidade entre diferentes populações, pois foi ajustada ou padronizada para a média da população. Conhecido também como coeficiente de variação de Pearson CV<sub>p</sub>.

**CV<sub>r</sub>: COEFICIENTE DE VARIAÇÃO DE THORNDIKE:** É igual ao quociente entre o desvio padrão e a mediana da série de valores. Determina-se através das seguintes equações:

$$CV_r = \frac{S}{Md} \quad \text{ou} \quad CV_r = \left( \frac{S}{Md} \right) \times 100$$

quando o resultado é expresso em porcentagem (%). Ele varia entre 0 e 1, ou 0 e 100%.

**COM REPOSIÇÃO:** Um método de selecionar amostras, em que itens são repostos entre sucessivas seleções.

**CORRELAÇÃO BIVARIADO:** O coeficiente de correlação bivariado mede como variáveis ou postos estão relacionados. Coeficientes: Pearson, Spearman's rho e Kendall's tau-b.

**COEFICIENTE DE CORRELAÇÃO LINEAR SIMPLES DE PEARSON (r):** Conforme Figueiredo filho e Silva Júnior (2009), o coeficiente de correlação de Pearson tem esse nome devido ao estatístico inglês Karl Pearson, no entanto não se deve atribuir exclusivamente a Karl Pearson o desenvolvimento dessa estatística. Na verdade a origem desse coeficiente remonta o trabalho conjunto dos estatísticos ingleses Karl Pearson e Francis Galton, sendo que na literatura especializada se afirma que correlação é uma medida do grau de associação bivariada, força de correlação ou relação, ou ainda se refere ao grau de relacionamento entre duas variáveis quantitativas discretas ou contínuas, nunca variáveis categóricas. A correlação mensura a direção e o grau da relação linear entre duas variáveis quantitativas. Ou seja, o coeficiente de correlação de Pearson (r) é uma medida de associação linear entre variáveis. Sua fórmula é a seguinte:

$$r = \frac{1}{n-1} \sum_{i=1}^n \frac{(x_i - \bar{X})(y_i - \bar{Y})}{S_x S_y}$$

Dois conceitos são importantes para entendê-la: associação e linearidade. O significado que duas variáveis estão associadas é a seguinte. Em termos estatísticos, duas variáveis se associam quando elas guardam semelhanças na distribuição dos seus escores. Mais precisamente, elas podem se associar a partir da distribuição das frequências ou pelo compartilhamento de variância. No caso da correlação de Pearson (r) vale esse último parâmetro, ou seja, ele é uma medida da variância compartilhada entre duas variáveis. Por outro lado, o modelo linear supõe que o aumento ou decréscimo de uma unidade na variável X gera o mesmo impacto em Y. Em termos gráficos, por relação linear entende-se que a melhor forma de ilustrar o padrão de relacionamento entre duas variáveis é através de uma linha reta. Portanto, a correlação de Pearson (r) exige um compartilhamento de variância e que essa variação seja distribuída linearmente. O coeficiente de correlação Pearson (r) varia de -1 a 1. O sinal indica direção positiva ou negativa do relacionamento e o valor sugere a força da relação entre as variáveis. Uma correlação perfeita (-1 ou 1) indica que o escore de uma variável pode ser determinado exatamente ao se saber o escore da outra. No outro oposto, uma correlação de valor zero indica que não há relação linear entre as variáveis. Faz-se necessária a independência das observações, ou seja, a ocorrência de uma observação X<sub>1</sub> não influencia a ocorrência de outra observação X<sub>2</sub>. Sendo que a violação desta orientação implica risco de assumir correlações espúrias. Em termos mais técnicos, o pesquisador pode enfrentar o problema de lurking ou confounding variables. A

violação desses pressupostos pode comprometer os resultados, levando o pesquisador a cometer os erros do tipo I ou tipo II. O erro do tipo I consiste em concluir que a hipótese nula é falsa quando ela é verdadeira. Logo, não existe relação entre as variáveis sendo a hipótese de nulidade da ausência de correlação linear simples entre as variáveis na população  $H_0$  verdadeira, mas o pesquisador argumenta que  $X$  e  $Y$  são estatisticamente dependentes. Ou seja, ele não poderia ter rejeitado a hipótese nula. O erro do tipo II consiste em concluir que a hipótese nula é verdadeira quando ela é falsa. Logo, existe relação entre  $X$  e  $Y$ , ou seja, a hipótese de nulidade  $H_0$  é falsa, mas o pesquisador defende que as variáveis são estatisticamente independentes. Isto é, ele deveria ter rejeitado a hipótese nula. O último tópico concernente ao coeficiente de correlação de Pearson ( $r$ ) diz respeito ao problema das lurking ou counfounding variables já que elas podem produzir correlações espúrias. Por exemplo, ao se estimar a relação entre o número de leitos hospitalares e a taxa de mortalidade de um determinado estado, o pesquisador pode chegar à conclusão de que quanto mais camas, maior é a taxa de mortalidade. A variável omitida, nesse caso, é o número de pessoas internadas. Será utilizado como exemplo a relação entre gofar e engordar. Isso porque existe a crença de que bebê que gofa muito, ganha peso mais rápido. O coeficiente de correlação de Pearson ( $r$ ) é fortemente influenciado pela média da distribuição. Por esse motivo, um dos pressupostos centrais para que essa medida seja adequadamente utilizada é de que as observações obedecem a uma distribuição normal. Existem testes disponíveis para averiguar em que medida as observações estão normalmente distribuídas, sendo o teste de Kolmogorov-Smirnov e a observação gráfica dos dados um dos procedimentos mais comumente utilizados. No caso do teste, um resultado não significativo, isto é um valor  $p$  maior que 5% ( $p > 0,05$ ) indica normalidade. Caso o  $p$  valor assuma valores abaixo desse patamar ( $p < 0,05$ ), isso é um indicativo de que o pressuposto da normalidade foi violado. Em relação à análise gráfica, é comum a utilização de histogramas e gráficos Q-Q plots para analisar o formato da distribuição. Em relação ao histograma, o pesquisador deve observar em que medida a distribuição dos seus dados se aproxima da curva normal. A presença de outliers tende a distorcer o valor da média e, por consequência, o valor do coeficiente de correlação. Dessa forma, a presença de outliers pode comprometer fortemente as estimativas dos pesquisadores, levando inclusive a cometer erros do tipo I ou do tipo II. No artigo de Anscombe (1973), citado por Figueiredo filho e Silva Júnior (2009), o autor mostra um exemplo para explicar esse efeito. Nesse trabalho ele mostra os dados sobre como a correlação observada entre gofar e engordar pode ser explicada na medida em que elas têm a mesma causa: gulodice. Ou seja, essa última variável estava agindo como lurking variable. Ao se controlar pelo efeito da gulodice, a correlação entre as variáveis desaparece. Dessa forma, os pesquisadores, antes de apresentar suas conclusões, devem analisar cuidadosamente os seus dados e investigar em que medida uma correlação entre suas variáveis de interesse pode estar sendo afetada pela presença de lurking variables.

**CORRELAÇÃO CANÔNICA:** Com a correlação bivariada mede-se o relacionamento entre duas variáveis com a correlação canônica é medido como dois conjuntos de variáveis estão associados.

**CORRELAÇÃO PARCIAL:** A correlação parcial é um coeficiente que mede o relacionamento linear entre duas variáveis enquanto são controlados os efeitos de uma ou mais variáveis adicionais.

**COMBINAÇÃO:** Um subconjunto selecionado sem reposição de um conjunto usado para determinar o número de resultados em eventos e espaços amostrais.

**COMBINAÇÃO (POOLING):** Quando vários conjuntos de dados podem ter sido gerados do mesmo modelo, é possível cominá-los geralmente para finalidades de estimar um ou mais parâmetros. A associação

das amostras para essa finalidade é geralmente chamada de combinação e é comumente usada para estimar a variância.

**COMPONENTE DE VARIÂNCIA:** Os componentes individuais da variância total que são atribuídos a fontes específicas. Isso geralmente se refere a componentes individuais de variância surgindo de um modelo aleatório ou misturado de análise de variância. Nos modelos de análise de variância envolvendo efeitos aleatórios, um dos objetivos é determinar quanta variabilidade pode ser associada com cada uma das fontes potenciais de variabilidade definida pelos experimentalistas. É costume se definir uma variância associada com cada uma dessas fontes. Essas variâncias, em certo sentido, somam a variância total da resposta e são geralmente chamadas de componentes da variância.

**COMPONENTES DE VARIÂNCIA:** São variâncias associadas aos efeitos aleatórios de um modelo matemático.

**COMPRIMENTO MÉDIO DE CORRIDA (CMC):** O número médio de amostras tomadas na monitoração ou esquema de inspeção de um processo até que o esquema sinalize que o processo está operando em um nível diferente daquele no qual começou.

**COLINEARIDADE (COLLINEARITY):** A correlação entre duas variáveis preditoras. Ver também multicolinearidade.

**CASUALIZAÇÃO:** É a atribuição dos tratamentos às unidades experimentais de forma aleatória, sem nenhuma interferência por parte do pesquisador. Para cada tipo de delineamento experimental existe uma forma adequada de se fazer esta alocação dos tratamentos.

**CONTROLE LOCAL:** Tipo de técnica usada pelo pesquisador que está associada ao conhecimento do ambiente experimental e consiste na divisão das parcelas experimentais em subconjuntos homogêneos, quando é sabido que o total das unidades experimentais não possui a homogeneidade exigida na condução do ensaio. No geral, o conjunto de parcelas homogêneas constitui o que se convencionou chamar de bloco. Quando o bloco contém todos os tratamentos uma única vez, ele é chamado de bloco completo, e os delineamentos experimentais que apresentam esta característica são ditos em blocos completos. Se os blocos comportam apenas parte dos tratamentos, os delineamentos são chamados de delineamentos em blocos incompletos.

**COMBINAÇÃO (COMBINATION):** A organização de  $n$  objetos discretos com  $x$  objetos a cada vez. O número de combinações de um grupo de  $n$  objetos é calculado usando o coeficiente binomial. Comparar com permutação.

**COMBINAÇÃO LINEAR (LINEAR COMBINATION):** A reexpressão de uma variável como a combinação de outras, cujos parâmetros da nova expressão são lineares. Por exemplo, a equação  $Y = b_0X_0 + b_1X_1 + b_2X_2$  expressa a variável  $Y$  como uma combinação linear de  $X_0$  a  $X_2$ , pois os parâmetros  $b_0$ ,  $b_1$  e  $b_2$  são lineares. Em contraste, a equação  $Y = a_0X_0 + X_1^{a_1}$  não é uma combinação linear, pois o parâmetro  $a_1$  é exponencial, e não linear. Note que a segunda equação pode ser transformada em uma combinação linear, tirando os logaritmos de ambos os lados da equação. Nem sempre será o caso, entretanto, que combinações não lineares possam ser transformadas em lineares.

**COMPLEMENTO (COMPLEMENT):** Na teoria de conjuntos, o complemento  $A^c$ ,  $AC$ ,  $\bar{A}$ , ou  $A'$  de um conjunto  $A$  é tudo que não está incluído no conjunto  $A$ .

**COMPLETAMENTE CRUZADOS (FULLY CROSSED):** Um delineamento de análise de variância multifatorial em que todos os níveis de dois ou mais tratamentos são testados ao mesmo tempo em um único experimento.

**COMPLETO (EXHAUSTIVE):** Um conjunto de resultados é dito completo se inclui todos os resultados possíveis de um evento. Comparar com exclusivo.

**COMPONENTE DE VARIAÇÃO (COMPONENT OF VARIATION):** Quanto cada fator, em um modelo de análise de variância (ANOVA), explica da variabilidade dos dados observados. Inicialmente, a variação total nos dados pode ser fracionada em variação entre grupos e em variação dentro de grupos. Em alguns delineamentos de ANOVA, a variação pode ser fracionada em componentes adicionais.

**COMPONENTES PRINCIPAIS (PRINCIPAL COMPONENTS):** Os eixos principais ordenados pela quantidade de variabilidade que eles carregam do conjunto original de dados.

**CONFIABILIDADE:** A probabilidade de que uma missão especificada será completada. Geralmente se refere à probabilidade de que um tempo de vida de uma variável aleatória contínua excede um limite especificado de tempo.

**CAMPO:** Espaço a ser preenchido num formulário, ou numa planilha eletrônica de dados. Tipo de local onde os dados de uma pesquisa são coletados: qualquer local onde não haja controle do pesquisador.

**CÂNONE:** Princípio, regra fundamental, norma que rege determinado método. Por exemplo, os cânones do método indutivo representam suas normas e princípios fundamentais.

**CASUÍSTICA:** Amostra, conjunto de elementos ou casos que serão objeto de análise numa pesquisa.

**CAUSALIDADE:** Princípio ou axioma segundo o qual todo fenômeno tem uma causa.

**CAUSALIDADE BIDIRECIONAL:** Tipo de relação entre duas variáveis, na qual diz-se que as mudanças que ocorrem em uma levam a mudanças na outra e vice-versa.

**CAUSALIDADE UNIDIRECIONAL:** Tipo de relação entre duas variáveis, na qual diz-se que as mudanças que ocorrem em uma delas levam a mudanças na outra, embora a recíproca não seja verdadeira. Quando ocorre a causalidade unidirecional, a variável causal leva o nome de variável independente e a resposta recebe o nome de variável dependente.

**CASUALIZAÇÃO:** É o sorteio dos tratamentos às unidades experimentais. Ver também aleatorização.

**CÉLULA:** Região de intersecção entre uma linha e uma coluna em uma tabela ou planilha eletrônica de dados.

**CENSO:** Pesquisa realizada com todos os elementos integrantes de uma população e não apenas com uma amostra desta.

**CENSURA:** Termo usado em análise de sobrevivência o qual é referido na literatura, para resolver um dos problemas da análise de tempos de vida sendo os indivíduos com possibilidade de serem censurados. A censura ocorre quando não nos é possível observar o acontecimento de interesse durante o período de tempo em que o indivíduo está em observação. A censura pode-se dever aos seguintes motivos: a morte do indivíduo ocorrer fora do período de duração do estudo ou não ser possível observar o acontecimento

de interesse, porque o indivíduo foi perdido para o follow-up. Em qualquer um destes casos, o indivíduo entra no estudo no instante  $t_0$  e morre no instante  $t_0 + t$  onde  $t$  é desconhecido, seja pelo indivíduo ainda se encontrar vivo no final do estudo, seja por ter sido perdido para o follow-up. Se houver informação de que o indivíduo se encontrava vivo no instante  $t_0 + c$  em que  $c < t$ , o instante é designado por tempo de sobrevivência censurado e estamos perante um caso de censura à direita.

**CENSURA ALEATÓRIA:** O mecanismo de censura aleatória é aquele em que os tempos de censura são variáveis aleatórias mutuamente independentes e ainda independentes dos tempos de vida. Observa-se este tipo de censura quando um indivíduo é retirado do estudo por uma causa alheia ao próprio estudo.

**CETERIS PARIBUS:** Termo de origem latina que significa demais condições mantidas inalteradas ou se todas as outras condições relevantes permanecerem as mesmas. Em pesquisas experimentais costumam-se manter todas as variáveis importantes constantes, exceto aquela que estiver sendo investigada.

**CLASSE:** Coleção de elementos ou indivíduos que possui uma característica ou qualidade em comum.

**COEFICIENTE DE KUDER-RICHARDSON:** Técnica para verificar a fidedignidade ou consistência interna de um teste que possua respostas dicotômicas, por exemplo, sim e não, sadio e doente, dentre outras.

**CONCORDÂNCIA OBSERVACIONAL:** Procedimento usado na pesquisa no qual os dados coletados de diversos observadores são cotejados com a finalidade de consolidar apenas aqueles dados que não são divergentes.

**CONTAMINAÇÃO:** Influência que pode dificultar a verificação adequada das hipóteses de uma pesquisa.

**COORTE:** Grupo de sujeitos que possuem uma característica demográfica comum, e que são objeto de pesquisa longitudinal. Por exemplo, mulheres nascidos em 1972 formam um coorte.

**COROLÁRIO:** Proposição ou extensão de um postulado matemático que é necessária e deduzida a partir de um postulado ou axioma fundamental. Se o postulado se mostrar falso, todos os corolários dele decorrentes serão necessariamente falsos.

**COMUTATIVO (COMMUTATIVE):** Em matemática, é a propriedade de que  $a + b = b + a$ .

**CONFUNDIDO (CONFOUNDED):** A consequência de uma variável medida estar associada a outra variável que pode não ter sido medida ou controlada.

**CONJUNTO (SET):** Uma coleção de objetos discretos, consequências ou resultados. Os conjuntos são manipulados usando as operações de união, interseção e complemento.

**CONJUNTO VAZIO:** Um conjunto sem membros ou elementos.

**CONJUNTOS DISCRETOS (DISCRETE SETS):** Um conjunto de resultados discretos.

**CONTRASTE (CONTRAST):** A comparação entre grupos de tratamentos em um delineamento experimental onde se aplica a técnica de análise de variância (ANOVA). Os contrastes se referem às comparações decididas antes de um experimento ser realizado.

**CORRELAÇÃO (CORRELATION):** O método de análises usado para explorar relações entre duas variáveis. Na análise de correlação, nenhuma hipótese sobre relação de causa e efeito é criada. Comparar com regressão.



**CORRELATIVO (CORRELATIVE):** Um tipo de relação que é observado, mas que ainda não foi investigado experimentalmente de forma controlada. A existência de dados correlativos leva à expressão correlação não implica causa.

**COVARIÁVEL (COVARIATE):** Uma variável contínua que é medida em cada réplica em um delineamento de análise de covariância. A covariável é usada para explicar a variação residual nos dados e aumentar o poder do teste em detectar diferenças entre tratamentos.

**COVARIÂNCIA (COVARIANCE):** A soma do produto das diferenças entre cada observação e seu valor esperado, dividido pelo tamanho amostral. Se a covariância não for dividida pelo tamanho amostral, é chamada de soma dos produtos cruzados.

**CURVA DE SOBREVIVÊNCIA:** Representação gráfica utilizada na análise de sobrevivência que mostra o tempo em que um conjunto de indivíduos permanece em determinado estado.

**CURVA NORMAL:** Curva simétrica em formato de sino, de campânula ou de chapéu de Napoleão a qual é observado nas distribuições normais ou Gaussianas.

**CASUAL:** O mesmo que aleatório ou estocástico.

**CASUALIDADE:** Relação de causa e efeito.

**CASUALIDADE (CRITÉRIOS DE JULGAMENTO):** Conjunto de orientações que serve de base à aferição da causalidade. Os critérios mais usados são: i) a sequência cronológica entre exposição ao fator de risco e o aparecimento da doença; ii) a força da associação entre os eventos; iii) a relação dose-resposta; iv) a consistência da associação; v) a plausibilidade ou coerência da associação; vi) a analogia com outras situações; e vii) a especificidade da associação.

**CASUALIZAÇÃO:** Procedimento adotado nos ensaios clínicos casualizados; consiste em designar, por processo aleatório, tratamentos pré-escolhidos aos pacientes participantes da pesquisa. Veja randomização.

**CASUALIZAÇÃO:** O mesmo que aleatorização e randomização.

**CASUALIZAÇÃO:** Sorteio que se realiza, ao planejar um experimento, fazendo ao acaso a escolha do tratamento a ser aplicado a cada parcela.

**CHI-QUADRADO:** Ver qui-quadrado.

**CIÊNCIA:** Termo usado para traduzir conhecimento ou prática sistemáticos. Em sentido estrito, ciência refere-se ao sistema de adquirir conhecimento baseado no método científico bem como ao corpo organizado de conhecimento conseguido através de tais pesquisas. Forma especial de produção de conhecimento. Maneira específica de explicar a realidade, caracterizada pelo uso da razão associada à experimentação. Termo derivado do latim scientia, equivalente ao termo grego episteme, que, por sua vez, origina-se do termo scire, que significa conhecer, aprender. Há muita controvérsia acerca do significado do termo ciência, e mesmo sobre quais disciplinas e métodos poderiam ser chamados de científicos.

**CRITÉRIO DE INFORMAÇÃO (INFORMATION CRITERION):** Qualquer método para selecionar entre modelos estatísticos que leve em conta o número de parâmetros nos diferentes modelos em geral qualificados, como no critério de informação de Akaike ou critério de informação bayesiano.

**CRITÉRIO DE INFORMAÇÃO BAYESIANO (BIC-BAYESIAN INFORMATION CRITERION):** Um método para comparar as distribuições de probabilidades a posteriori de dois modelos alternativos quando a de probabilidades a priori não é informativa. Ele desconta o fator de Bayes pelo número de parâmetros em cada um dos modelos. Como com o critério de informação de Akaike, esse modelo é considerado o melhor.

**COEFICIENTE DE CORRELAÇÃO:** Medida objetiva do nível de relacionamento entre duas variáveis. Os coeficientes de correlação variam de -1,00 a + 1,00. Quando ocorre uma perfeita correlação negativa entre as variáveis A e B, por exemplo, o coeficiente será de -1,00. Quando ocorre uma perfeita correlação positiva, o coeficiente será de + 1,00. O valor zero indica a inexistência de correlação entre as duas variáveis. Há diversos coeficientes de correlação que podem ser utilizados dependendo do tipo de variável envolvida. Dentre os mais comuns, podem-se citar: coeficientes de correlação de Pearson, coeficiente de correlação de Spearman e o coeficiente de correlação de Kendall.

**COEFICIENTE DE CORRELAÇÃO DE KENDALL:** Coeficiente de correlação (t) utilizado para mensurar o grau de relação entre variáveis com nível de mensuração ordinal (e superiores), similar ao coeficiente de correlação de Spearman, porém considerado por alguns autores como superior a este último. Este coeficiente também é especialmente útil para o caso de amostras de tamanho reduzido, por exemplo, com  $n < 10$ . Pode ser calculado de acordo com a fórmula abaixo:

$$\frac{S}{\frac{1}{2}n(n-1)}, \text{ onde } n = \text{número de postos.}$$

**COEFICIENTE DE CORRELAÇÃO DE PEARSON:** Coeficiente de correlação (r) que mede o grau de relacionamento linear entre os valores emparelhados x e y de duas variáveis em uma amostra, podendo ser calculado conforme a fórmula abaixo:

$$r = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2} \sqrt{n(\sum y^2) - (\sum y)^2}}$$

onde n = número de pares (x e y) de observações. O coeficiente de correlação de Pearson é utilizado ( $r_s$ ) utilizado para mensurar o grau de relação entre as variáveis com nível de mensuração ordinal e superiores, daí ser também conhecido pelo nome de coeficiente de correlação por postos, podendo ser calculado de acordo com a fórmula abaixo:

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)} \text{ onde } n = \text{número de pares de observações; } d = \text{diferença entre a ordem de } x \text{ e } y.$$

**COEFICIENTE DE CORRELAÇÃO LINEAR:** Ver coeficiente de correlação de Pearson.

**COEFICIENTE DE CORRELAÇÃO PHI:** Coeficiente de correlação ( $\phi$ ) utilizado para variáveis dicotômicas e, portanto de nível de mensuração nominal. Por exemplo, numa pesquisa deseja-se determinar a correlação entre as variáveis sexo (masculino/feminino) e possui poupança (sim/não). O coeficiente  $\phi$  pode ser calculado mediante a fórmula abaixo:

$$\Phi = \sqrt{\frac{\chi^2}{n}} \text{ onde } \chi^2 = \text{qui-quadrado; } n = \text{número total de observações;}$$



**COEFICIENTE DE CORRELAÇÃO V DE CRAMER:** Coeficiente de correlação utilizado para variáveis de nível de mensuração nominal, ou seja, trata-se de uma técnica estatística não paramétrica. Por exemplo: numa pesquisa deseja-se verificar a correlação entre sexo (masculino/feminino) e preferências de lazer (cinema, teatro, televisão dentre outras). O V de Cramer pode ser calculado de acordo com a fórmula abaixo:

$$V = \sqrt{\frac{\chi^2}{n[\min(L, C)-1]}}$$

onde:  $\chi^2$  = qui-quadrado; n = número total de observações; L = número de

linhas da tabela de contingência; C = número de colunas da tabela de contingência;

**COLETA DE DADOS:** Operação por meio da qual se obtêm as informações ou dados a partir do fenômeno pesquisado. A coleta de dados é realizada mediante um instrumento de pesquisa, que irá variar em função da ciência e do método utilizado. Por exemplo, nas ciências sociais são três os caminhos para compreender o comportamento humano: i) observar o comportamento que ocorre naturalmente no âmbito real; ii) criar situações artificiais e observar o comportamento ante tarefas definidas para essas situações; iii) perguntas às pessoas sobre o que fazem ou fizeram e pensam ou pensaram. Cada uma das três famílias de técnicas para conduzir estudos empíricos, ou seja, observação, experimento e survey, apresenta vantagens e desvantagens distintas.

**CORRELAÇÃO:** Técnica matemática utilizada para medir a força ou grau de associação entre duas variáveis.

**CRAMER:** Ver coeficiente de correlação V de Cramer.

**CENTIL:** Valores que dividem uma distribuição de frequências em 100 intervalos de iguais frequências.

**CLASSE:** Grupo de valores numéricos situados em cada um dos intervalos em que é dividida uma variável cuja distribuição de frequências se quer determinar.

**COEFICIENTE DE CORRELAÇÃO:** Qualquer coeficiente que aponte o grau de relação ou inter-relação entre duas variáveis.

**COEFICIENTE DE PEARSON:** Representado pela letra r, varia de -1 a +1. Quando seu valor é -1, a correlação é perfeita negativa: os valores altos numa variável correspondem a valores baixos em outras. Quando seu valor é +1, a correlação é perfeita positiva: valores altos numa variável correspondem a valores altos na outra. Quando seu valor é 0, não existe correlação.

**CRITÉRIO DE INFORMAÇÃO DE AKAIKE (AIC-AKAIKE'S INFORMATION CRITERION):** Uma medida do poder explicativo de um modelo estatístico que leva em conta o número de parâmetros no modelo. Ao comparar muitos modelos para um mesmo fenômeno os modelos devem compartilhar no mínimo alguns parâmetros, o modelo com o menor valor do critério de informação de Akaike é considerado o melhor. Usado comumente para seleção de modelos em regressão e análise de caminhos.

**CURTOSE (KURTOSIS):** Uma medida de quão agregada ou dispersa uma distribuição é relativa a seu centro. É calculada usando o quarto momento central e é simbolizada por  $g_2$ . Ver também platocúrtica e leptocúrtica.

**CV:** Ver coeficiente de variação. Também conhecido como coeficiente de variação de Pearson, em homenagem ao seu criador o estatístico inglês Karl Pearson, simbolizado por CVp.

**COEFICIENTE DE VARIAÇÃO:** Trata-se de uma medida relativa de dispersão, útil para comparação do grau de concentração em torno da média de séries distintas.

**COMBINAÇÕES:** Combinações simples de classe p ou combinações simples de m elementos p a p ( $p \leq m$ ) são todos os agrupamentos de p elementos distintos tirados entre os m elementos dados, de modo que cada agrupamento se diferencie de outro pela natureza de seus elementos.

**CORRELAÇÃO:** Variação concomitante de dois fenômenos da correlação não se infere necessariamente causalidade.

**CORRELAÇÃO LINEAR:** A correlação linear perfeita, que corresponde aos valores -1 e +1, conforme seja negativa ou positiva, é medida pelo r de Pearson. Quando a correlação não é perfeita, os pontos de interseção no diagrama de dispersão distribuem-se, regra geral, de maneira a formar uma elipse.

**CORRELAÇÃO ORDINAL:** Correlação em que não se leva em conta os valores originais, mas sua classificação ou seus números de ordem.

**COVARIÂNCIA POPULACIONAL C(X,Y):** É um valor que mede o grau de dispersão simultânea de duas variáveis quantitativas em relação as suas médias. É a soma dos produtos dos desvios das variáveis em relação a sua média, dividido pelo número de observações:

$$\left[ \frac{\text{Soma } (X_i - \text{média } X) * (Y_i - \text{média } Y)}{N} \right].$$

É o numerador do coeficiente de correlação.

**CORRELAÇÃO POPULACIONAL (R):** É um valor que mede o grau de relação linear entre duas variáveis quantitativas. É igual a covariância dividida pelo desvio padrão de cada uma das variáveis.

**CAUDAS:** De uma distribuição de frequências representam as frequências nas extremidades da distribuição.

**COEFICIENTE DE CORRELAÇÃO:** O coeficiente de correlação de Pearson avalia o grau de associação linear entre duas variáveis. Consultar também coeficiente de correlação de postos de Spearman.

**COEFICIENTE DE CORRELAÇÃO DE POSTOS DE SPEARMAN:** Um teste não-paramétrico equivalente ao coeficiente de correlação momento-produto de Pearson; indica a associação que não necessariamente deve ser linear entre duas variáveis, as quais podem ser ordinais.

**COEFICIENTE DE CORRELAÇÃO INTRACLASSE, CCI:** A proporção entre a variância dos conglomerados e a variação total; indica a proporção da variância total em um conjunto de medidas conglomeradas, que é atribuída à diferença entre os conglomerados.

**COEFICIENTE DE CORRELAÇÃO MOMENTO-PRODUTO DE PEARSON:** Consultar coeficiente de correlação.

**COEFICIENTE DE CORRELAÇÃO MÚLTIPLA (R):** A raiz quadrada do coeficiente de determinação. R mede a associação entre os valores de variável dependente e os valores obtidos a partir da equação.

**COEFICIENTE DE DETERMINAÇÃO (R<sup>2</sup>):** É a proporção da variância total da variável dependente, y, em um modelo de regressão que é fundamentado pela regressão.

**COEFICIENTE DE REPETIBILIDADE DA BRITIS STANDARDS INSTITUTION:** Uma medida da repetibilidade de um método; fornece uma indicação da diferença máxima que deve ocorrer entre duas medidas.

**COEFICIENTE DE VARIAÇÃO (CV):** É o desvio-padrão expresso como uma porcentagem da média.

**COEFICIENTE KAPPA:** Consultar coeficiente kappa de Cohen.

**COEFICIENTE KAPPA DE COHEN:** Uma medida da concordância entre duas avaliações categóricas.

**COEFICIENTES DE REGRESSÃO:** Os coeficientes que correspondem a uma variável explicativa particular em uma equação de regressão. O termo constante ou o intercepto em uma regressão linear simples é também chamado de coeficiente de linear da reta, já o coeficiente de regressão é conhecido também como coeficiente angular ou taxa de declividade da curva da equação o qual mede o acréscimo ou decréscimo médio na variável dependente do modelo quando varia uma unidade na variável independente, isto também pode ser observado no coeficiente de regressão parcial em uma equação de regressão múltipla e multivariada.

**COEFICIENTES DE REGRESSÃO PARCIAL:** Os coeficientes na equação de regressão múltipla que correspondem a uma variável explicativa particular. Geralmente são diferentes do coeficiente de regressão que seria obtido por regressão da variável dependente naquela variável explicativa sozinha, omitindo completamente as outras variáveis explicativas da equação. Às vezes, são denominados coeficientes de regressão.

**COLABORAÇÃO COCHRANE:** Uma rede internacional de especialistas e consumidores que continuamente atualizam publicações sistemáticas e as tornam disponíveis.

**CURVA DE GAUSS:** Curva simétrica, unimodal, em forma de sino ou campânula: os valores da distribuição normal que a curva representa, concentram-se no meio, tanto que aproximadamente 68% dos casos acham-se entre um desvio padrão abaixo e um desvio padrão acima da média.

**CURVA NORMAL:** O mesmo que a curva de gauss ou gaussiana, curva simétrica, em forma de sino de igreja, campanular ou de chapéu de napoleão ou ainda curva mesocúrtica.

**CURTOSE:** Entende-se por curtose o grau de achatamento de uma distribuição.

**CASUALIZAÇÃO:** Sorteio que se realiza, ao planejar um experimento, fazendo-se, ao acaso, a escolha do tratamento a ser aplicado a cada parcela.

**CASUÍSTICA:** Registro pormenorizado de casos clínicos das doenças.

**CATANOVA:** Técnica estatística que consiste numa análise de variância para dados categóricos em tabelas de contingência sob um modelo multinomial envolvendo vários fatores. No contexto de sequências genômicas, esta técnica pode ser utilizada para comparar a variabilidade numa única posição entre e dentro grupos. Na análise de dados genômicos, uma única posição fornece pouca informação. Consequentemente precisamos considerar várias regiões do genoma.

**CATEGORIA DE REFERÊNCIA:** O nível omitido de uma variável não-métrica quando uma variável dicotômica é formada a partir da variável não-métrica.

**CATEGORIAS DE RESPOSTAS FIXAS-ALTERNATIVAS:** Uma lista de opções de respostas associadas a perguntas fechadas.

**CAUCHY ( $\mu, \sigma$ ):** Distribuição de Cauchy.

**CAUDA:** Extremidade superior ou inferior de uma distribuição.

**CAUSA E EFEITO:** De um ponto de vista prático, alguns eventos mostram similaridades, porém, estritamente falando, cada evento é certamente único. Certos eventos ocorrem em pares, no sentido de que se o primeiro ocorre, o segundo o segue. Às vezes se diz que o primeiro é a causa do segundo. Todavia, um evento pode ser considerado causa do outro, somente se a supressão dele também e sempre suprime o efeito.

**CAUSA NECESSÁRIA:** O fato ou causa precisa estar presente para que o efeito ocorra, como por exemplo doença. Entretanto, uma causa necessária pode está presente sem que haja ocorrência da doença.

**CAUSA NECESSÁRIA:** Precede a doença e tem com ela a seguinte relação: se a causa estiver ausente, a doença não poderá ocorrer. Se a causa estiver presente, a doença poderá ocorrer ou não.

**CAUSA SUFICIENTE:** Precede uma doença e tem com ela a seguinte relação: se a causa estiver presente, a doença sempre ocorrerá.

**CAUSA SUFICIENTE:** Se o fator ou causa está presente, o efeito sempre ocorrerá, inevitavelmente, como por exemplo doença.

**CAUSALIDADE:** Princípio pelo qual causa e efeito são estabelecidos entre duas variáveis. Ele requer que exista um grau suficiente de associação ou correlação entre as duas variáveis, que uma variável ocorra antes da outra, ou seja, que uma variável seja claramente o resultado da outra, e que não existam outras causas razoáveis para o resultado. Apesar de causalidade ser raramente encontrada em seu sentido estrito, na prática forte apoio teórico pode tornar possível a estimação empírica de causalidade.

**CAUSALIDADE:** Um fator produz ou contribui para a ocorrência de um desfecho específico. Ver também casualidade direta e casualidade indireta.

**CAUSALIDADE BIDIRECIONAL:** Mudanças em A levam a mudanças em B e, além disso, mudanças em B produzem mudanças em A. Exemplo: percepção de ameaças produz sentimentos de ansiedade e aumento de ansiedade faz crescer a percepção de ameaça.

**CAUSALIDADE DIRETA:** O fator sob consideração exerce seu efeito sem fatores intermediários.

**CAUSALIDADE INDIRETA:** Um fator influencia um ou mais fatores que são, por sua vez, diretamente casuais.

**CAUSALIDADE UNIDIRECIONAL:** Pode-se prever que mudanças em A produzirão subseqüentes mudanças em B, mas não se espera que mudanças em B influenciem A. Exemplo: aumento da temperatura na sala de aula causa impaciência nos alunos, mas o grau de impaciência não afeta a temperatura.

**CAUSAS COMUNS:** Variações normais ou naturais do processo de produção, devidas puramente ao acaso. Nenhuma ação corretiva é necessária quando variações da produção são devidas a causas comuns.

**CAUSAS DETERMINÁVEIS:** Variações na qualidade dos processos que são devidas a fatores tais como desgaste de ferramentas das máquinas, ajustes incorretos da máquina, matéria-prima de má qualidade, erro do operador e assim por diante. Ações corretivas devem ser tomadas quando causas determináveis da variação da produção são detectadas.

**CAVALIERI, FÓRMULA DE:** Ver também *fórmula de Newton*,

**CÉLULA:** Categoria utilizada para separar dados qualitativos, categóricos ou atributos.

**CÉLULA:** Cada uma das localizações em uma tabela de contingência. Conhecido também como casela.

**CÉLULAS:** Termo usado para designar valores em uma tabela quando utilizamos modelos mais complexos com duas ou mais variáveis independentes, como no caso de análise de variância (ANOVA) com dois fatores.

**CÉLULAS:** Em inglês cells.

**CENSO:** Abordagem exaustiva ou de 100% das unidades amostrais da população. A completa enumeração reproduz exatamente todas as características da população, ou seja, fornece o valor de seus parâmetros, e não estimativas.

**CENSO:** Coleção de dados referentes a cada elemento de uma população.

**CENSO:** Enumeração das características de uma população. Frequentemente, censo se assemelha a survey, a diferença sendo que o censo coleta dados de todos os integrantes da população enquanto o survey se limita a uma amostra.

**CENSO:** É um estudo estatístico que resulta da observação de todos os indivíduos da população relativamente a diferentes atributos pré-definidos.

**CENSO:** Dados obtidos sobre cada membro da população de interesse.

**CENSURA:** Uma amostra é considerada censurada quando certos valores são desconhecidos ou deliberadamente ignorados, porém, sua existência é conhecida. O termo é geralmente empregado nos casos onde, de um número  $n$  de valores existentes, somente os  $k$  menores ou os  $k$  maiores são observados ou  $k$  pode ser expresso como uma proporção de  $n$ . Em inglês Censoring.

**CENTIL:** Expressão sinônima de percentil: cada um dos 99 valores que dividem uma distribuição de frequência em 100 intervalos de igual frequência. Em inglês centile.

**CENTRÓIDE:** O ponto  $(, )$  de uma coleção de dados, onde obrigatoriamente passa a reta de regressão linear simples.

**CENTRÓIDE:** Valor médio para os escores  $Z$  discriminantes de todos os objetos, em uma dada categoria ou grupo. Por exemplo, uma análise discriminante de dois grupos tem dois centróides, um para os objetos em cada grupo.

**CENTRÓIDE:** Média ou valor médio dos objetos contidos no agrupamento sobre cada variável, seja usado na variável estatística de agrupamento ou no processo de validação.

**CENTRÓIDE DE AGRUPAMENTO:** Valor médio dos objetos contidos no agrupamento sobre todas as variáveis na variável estatística de agrupamento.

**CEC:** Capacidade específica de combinação.

**CGC:** Capacidade geral de combinação.

**CGH:** Capacidade geral de hibridação.

**CGS:** Seleção de genes candidatos.

**CHANCE:** É o quociente entre o número de casos favoráveis pelo número de casos desfavoráveis. É representada como uma razão. Assim se a chance de algo é de 2:1, significa que a probabilidade deste algo acontecer é

$$\frac{2}{(2+1)} = \frac{2}{3} = 66,67\% . \text{ Em inglês chance ou possibility.}$$

**CHANCES:** Divisão da probabilidade de sucesso e a probabilidade de insucesso para um experimento de Bernoulli. Em inglês Odds.

**CHANCES A FAVOR:** Razão da probabilidade da ocorrência para a probabilidade da não-ocorrência de um evento, usualmente expressa como a razão de dois inteiros sem fatores comuns.

**CHANCES CONTRA:** Razão da probabilidade da não-ocorrência de um evento para a da ocorrência de um evento, usualmente expressa na forma  $a:b$ , onde  $a$  e  $b$  são inteiros sem fatores comuns.

**CICLO:** Em um uso restrito, um movimento periódico em uma série temporal, isto é, um componente com a propriedade de que  $f(t+\tau)=f(t)$ , onde  $\tau$  é o período do ciclo. Em inglês Cycle.

**CIÊNCIA:** Conhecimento acumulado e aceito, que tem sido sistematizado e formulado com referência à descoberta de verdade gerais ou à operação de leis gerais. Pode ser definida como a crítica da experiência metodologicamente controlada, institucionalizada e comunicativa com o objetivo de garantir conhecimentos para possibilitar uma progressiva emancipação dos indivíduos agentes sociais.

**CIÊNCIA:** É um conjunto organizado de conhecimentos relativos a um determinado objeto conquistados por meio de métodos próprios de coleta de informação.

**CIÊNCIA (ASPECTOS IMPORTANTES):** Os seguintes itens mostram duas principais facetas interpretativas ou aspectos da Ciência: i) envolve um corpo de conhecimento sistemático, consistindo de conceitos, leis, princípios, e teorias usadas na explicação de um conjunto de fenômenos; ii) requer um processo de busca ou um procedimento para responder a questões, resolver problemas e planejar melhores procedimentos.

**CIÊNCIA (EVOLUÇÃO):** Quando uma ciência começa é quase inteiramente empírica in natura. Preocupa-se com fenômenos observáveis diretamente e com relacionamentos entre eles. Geralmente, à proporção que as ciências evoluem e se tornam mais maduras, preocupam-se cada vez mais com teorias e proporcionalmente menos com eventos observáveis diretamente.

**CIÊNCIA (LIMITAÇÕES):** i) A ciência não tem métodos perfeitos para coletar e empacotar conhecimentos ou informações; ii) Cientistas usam muitos métodos que diferem em precisão e custo; iii) Um método científico ou peneira é bom até onde ajuda o cientista a tomar decisões, reduzindo o número de causas ou suspeitas. iv) Não é sempre possível usar uma peneira fina num problema. Aceitamos e usamos uma peneira grossa, quando pode reduzir pelo menos um pouco nossa ignorância.

**CIÊNCIA (PROCEDIMENTOS):** i) Estimulação por uma idéia, uma pergunta, ou um problema; ii) Observação e/ou descrição de um fenômeno; iii) Análise e simplificação das observações e/ou descrições; iv) Formulação de uma hipótese; e v) Conclusões.



**CIÊNCIA (NÍVEIS PROBABILÍSTICOS DE PREDIÇÃO):**

<b>Alta</b>	<b>Ciências</b>	<b>Ciências</b>	<b>Ciências</b>	<b>Ciências</b>	<b>Baixa</b>
<b>predição</b>	<b>físicas</b>	<b>biológicas</b>	<b>sociais</b>	<b>humanas</b>	<b>predição</b>

**CIÊNCIA (INTERPRETAÇÕES):** Uma ciência é um corpo de conhecimento teórico que não é trivial. O cientista deve ter melhores teorias que o leigo ou ele realmente não é um cientista. Se a física é uma ciência porque é empírica e teórica, é também uma ciência porque as teorias dos físicos sobre o mundo físico funcionam melhor de que as teorias dos não-físicos. Um corpo de conhecimento, conectado empírica e teoricamente, é ciência somente quando as pessoas que conhecem as teorias sabem mais sobre o mundo real que as pessoas que não conhecem as teorias.

**CIÊNCIA EMPÍRICA:** Termo que se refere a conjecturas simples e assimétricas que dão lugar a observações críticas e sistemáticas; que por sua vez permitem testes sistemáticos e precisos de hipóteses sob condições controladas e, como consequência em algumas disciplinas, pode ocorrer o desenvolvimento de teorias incorporando os resultados de estudos isolados em estruturas significativas e na formulação de testes precisos e sistemáticos de hipóteses integrantes, emanadas daquelas teorias.

**CIÊNCIA E TECNOLOGIA:** Ciência é um termo que significa os acúmulos de entendimentos e conhecimento formulado sobre universos usando meios abstratos e logicamente convincentes. Já o conceito de tecnologia ou arte, significa práticas operacionais, e particularmente, aplicações controladas e preditas. A ciência ensina a saber e a arte a fazer. A ciência enfatiza doutrina, teoria, e experimentação controlada. Já a arte e a tecnologia enfatizam experiência e operacionalidade. Ciência acentua racionalidade, consistência e completa sistematização. Tecnologia enfatiza sensibilidade, uso do raciocínio e heurística.

**CIÊNCIAS SOCIAIS:** Entende-se por Ciências Sociais o conjunto de disciplinas inerentes às seguintes áreas de estudo e pesquisa: Psicologia, Sociologia, Antropologia, Economia, História e Ciência Política.

**CIENTISTA:** É aquele indivíduo que, além de compreender e usar métodos científicos, também publica trabalhos científicos.

**CLASSE:** São os intervalos de variação da variável e é simbolizada por  $i$  e o número total de classes simbolizada por  $k$ .

**CLASSE:** Grupo de valores numéricos situados em cada um dos intervalos em que é dividida uma variável cuja distribuição de frequência se quer determinar. Em inglês class.

**CLASSES:** Intervalos nos quais os valores da variável analisada são agrupados.

**CLASSE (CLASSIFICAÇÃO):** Uma coleção de coisas que têm alguma definida qualidade em comum. Não há duas coisas exatamente iguais em todos os aspectos. Um mesmo objeto será um membro de um grande número de classes diferentes; uma classe para cada qualidade que ele possua.

**CLASSE (DE COISAS):** Uma coleção de coisas que têm alguma qualidade definida, em comum. Lembrar que não existem duas coisas que são exatamente iguais em todos os seus aspectos, e que o mesmo objeto pode ser um membro de um grande número de classes diferentes, uma para cada qualidade que ele possua. Quanto mais precisamente se especificar a classe, tanto mais provável que as propriedades de membros individuais serão compartilhadas com outros membros da classe.

**CLASSE COM INTERVALO ABERTO OU ABORTADA:** Classe de dados agrupados em que não há limite superior e/ou inferior de uma classe.

**CLASSE MEDIANA:** É a classe, para dados classificados, que contém a mediana, neste caso considera-se como mediana o valor da variável estatística que corresponde a  $\frac{n}{2}$ , quer  $n$  seja par, quer  $n$  seja ímpar.

**CLASSE MODAL:** É a classe, para dados classificados, que aparece com maior frequência.

**CLASSE MODAL:** O valor que surge com mais frequência se os dados são discretos, ou, o intervalo de classe com maior frequência se os dados são contínuos.

**CLASSIFICAÇÃO:** Arranjo sistemático de objetos de qualquer tipo em grupos ou categorias de acordo com um conjunto de critérios preestabelecidos.

**CLASSIFICAÇÃO DE FOTOGRAFIAS:** Os respondentes classificam diversos tipos de pessoas, apontando as fotos daquelas que acham que usariam o produto ou serviço especificado.

**CLASSIFICAÇÃO DOS DEPÓSITOS MINERAIS:** É feita a partir da variabilidade natural dos depósitos, medida pelo coeficiente de variação, que é uma porcentagem do desvio padrão em relação à média. O Quadro abaixo apresenta a classificação dos depósitos minerais, extensivamente utilizada pela escola soviética de exploração mineral, segundo os valores de coeficiente de variação para espessura e teor. Para estabelecer o coeficiente de variação, o número mínimo de dados é de 18 para o grupo 1 (regular), 25 para o grupo 2 (irregular) e 40 para o grupo 3 (muito irregular). Classificação dos depósitos minerais segundo sua variabilidade.

grupo	variação	coeficiente de variação		
		espessura	teor	reservas
regular	usualmente regular	< 50%	< 50%	30%
irregular	usualmente aleatória	30-80 %	40-100%	80%
muito irregular	aleatória	50-100%	100-150%	130%
extremamente irregular	aleatória	80-150%	130-300%	200%

**CLASSIFICAÇÃO Q:** Distribuir um conjunto de objetos em pilhas de acordo com uma classificação especificada de categorias.

**CODIFICAÇÃO:** Processo por meio do qual dados brutos são transformados em forma padronizada adequada para processamento e análise em computador.

**CODIFICAÇÃO:** Conversão das categorias de uma variável não-métrica em uma série de variáveis binárias, uma para cada categoria. Semelhante à codificação para variável dicotômica sem a eliminação da categoria de referência.

**CODIFICAÇÃO:** O processo de agrupar e designar códigos numéricos às várias respostas a uma pergunta.

**CODIFICAÇÃO DE DADOS:** Consiste em desenvolver um sistema pelo qual os dados e informação identificadora são especificados e organizados em preparação para análise.

**CODIFICAÇÃO DE EFEITOS:** Método para especificar a categoria de referência para um conjunto de variáveis dicotômicas, sendo que a categoria de referência recebe um valor de menos um (-1) sobre o conjunto



de variáveis dicotômicas. Com esse tipo de codificação, os coeficientes da variável dicotômica representam desvios do grupo em relação à média de todos os grupos. Isso contrasta com codificação indicadora.

**CODIFICAÇÃO DOS EFEITOS:** Método para especificar a categoria de referência para um conjunto de variáveis dicotômicas no qual a categoria de referência recebe um valor de -1 no conjunto de variáveis dicotômicas. Com esse tipo de codificação, os coeficientes para as variáveis dicotômicas se tornam desvios de grupo das médias de todos os grupos. Isso contrasta com a codificação indicadora, na qual a categoria de referência recebe o valor zero em todas as variáveis dicotômicas e os coeficientes representam desvios de grupo do grupo de referência.

**CODIFICAÇÃO INDICADORA:** Método para especificar a categoria de referência para um conjunto de variáveis dicotômicas onde a categoria de referência recebe um valor 0 sobre o conjunto de variáveis dicotômicas. Os coeficientes das variáveis dicotômicas representam as diferenças de categoria em relação à categoria de referência. Ver também codificação de efeitos.

**COEFICIENTE:** Relação entre o número de eventos reais e os que poderiam acontecer, multiplicando-se o resultado dessa relação pela base referencial do denominador, que é potência de 10. Muito utilizado em saúde pública para indicar a relação ou quociente entre dois valores numéricos, no sentido de estimar a probabilidade da ocorrência ou de determinado evento. Sinônimo: taxa.

**COEFICIENTE:** Frequência ou número de novos eventos, que ocorram em um período de tempo definido, dividida pela média populacional em risco. Geralmente, a população do meio é usada como o número médio de pessoas em risco, neste caso vale lembrar a necessidade de ver a definição de coeficiente de incidência. Pelo fato de o coeficiente ser quase sempre menor do que 1,0, a não ser que todos morram ou tenham o evento de risco, a constante multiplicadora é usada para aumentar tanto o numerador como o denominador, para tornar o coeficiente mais fácil para se pensar e discutir.

**COEFICIENTE:** O termo coeficiente é também usado como tradução alternativa de rate ou taxa. Fator de ponderação usado em uma equação, sendo o peso baseado na importância relativa do fator na predição do desfecho. Em inglês coefficient.

**COEFICIENTE (OU TAXA):** Frequência com que um evento ocorre na população; proporção de uma população possuidora de uma determinada característica. Exemplo: coeficiente de morbidade, incidência ou prevalência e de mortalidade.

**COEFICIENTE AJUSTADO:** Coeficiente que sofreu transformações matemáticas de modo a eliminar o efeito de uma ou mais variáveis.

**COEFICIENTE AJUSTADO DE DETERMINAÇÃO:** Coeficiente múltiplo de determinação  $R^2$  modificado para levar em conta o número de variáveis e o tamanho da amostra.

**COEFICIENTE AJUSTADO DE DETERMINAÇÃO ( $R^2$  AJUSTADO):** Medida modificada do coeficiente de determinação que considera o número de variáveis independentes incluídas na equação de regressão e o tamanho da amostra. Apesar de a adição de variáveis independentes sempre fazer com que os coeficientes de determinação pode cair se as variáveis independentes acrescentadas tiverem pouco poder de explicação e/ou se os graus de liberdade se tornarem muito pequenos. Essa estatística é muito útil para comparação entre equações com diferentes números de variáveis independentes, diferentes tamanhos de amostras, ou ambos.

**COEFICIENTE ANGULAR:** Medida de inclinação de uma reta.

**COEFICIENTE ANGULAR:** Em inglês slope.

**COEFICIENTE ANGULAR:** Número que indica a orientação de uma reta; uma reta horizontal tem coeficiente angular zero, e uma reta vertical tem coeficiente angular infinito; obtém-se o coeficiente angular calculando-se a distância entre dois pontos quaisquer da reta dividindo-se pela distância horizontal entre esses dois pontos.

**COEFICIENTE ANGULAR OU COEFICIENTE DE REGRESSÃO OU INCLINAÇÃO DA RETA DE REGRESSÃO  $Y = a + b * X$ :** É a variação de Y por cada variação de X.

**COEFICIENTE BETA:** Coeficiente de regressão padronizado que permite uma comparação direta entre coeficientes e seus poderes relativos de explicação da variável dependente. Ao contrário dos coeficientes de regressão, que são expressos em termos das unidades da variável associada, o que torna as comparações inadequadas, os coeficientes beta usam dados padronizados e podem ser diretamente comparados. Consultar o termo padronização.

**COEFICIENTE DE ASSIMETRIA:** Medida de forma de distribuição dos escores de variáveis aleatórias contínuas, podendo ser simétrica, assimétrica positiva que é a curva cuja cauda está desviada para a direita e assimétrica negativa a qual é a curva cuja cauda está desviada para a esquerda.

**COEFICIENTE DE ATAQUE:** O mesmo que coeficiente de incidência.

**COEFICIENTE DE ATAQUE:** Proporção das pessoas expostas que se tornam doentes. Essa é a medida que se usa costumeiramente para estabelecer a gravidade de um surto.

**COEFICIENTE DE ATAQUE SECUNDÁRIO:** É a razão entre o número de casos novos surgidos a partir do contato com o caso-índice e o número total de contatos com o caso-índice, expressando-se o resultado em percentual.

**COEFICIENTE DE AUTOCORRELAÇÃO:** Se  $x_t$  é um processo estocástico estacionário de média  $m$  e variância  $\sigma^2$ , o coeficiente de autocorrelação de ordem  $k$  é definido como  $\rho_k = \rho_{-k} = E[x_t - m][x_{t+k} - m]$ , onde o valor esperado é calculado para a distribuição conjunta de  $x_t$  e  $x_{t+k}$ . Em inglês Autocorrelation coefficient.

**COEFICIENTE DE CONFIANÇA:** A medida de probabilidade  $\alpha$  associada com o intervalo de confiança expressando a probabilidade de que o intervalo inclua o valor verdadeiro do parâmetro. Em inglês Confidence coefficient.

**COEFICIENTE DE CONFIANÇA:** Corresponde a probabilidade de aceitar a hipótese nula quando esta é verdadeira e indica a probabilidade de decisão correta baseada na hipótese nula. É simbolizado por  $(1-\alpha)$ , onde  $\alpha$  é a probabilidade de se rejeitar a hipótese de nulidade quando na realidade ela for verdadeira.

**COEFICIENTE DE CONFIANÇA:** Veja grau de confiança.

**COEFICIENTE DE CONFIANÇA:** Corresponde a 1- probabilidade de aceitar a hipótese nula quando esta é verdadeira e indica a probabilidade de decisão correta baseada na hipótese nula.

**COEFICIENTE DE CORRELAÇÃO( $r$ ):** Coeficiente que quantifica o grau de correlação entre duas variáveis. Em inglês correlation coefficient.

**COEFICIENTE DE CORRELAÇÃO (r):** Coeficiente que indica a força da associação entre quaisquer duas variáveis métricas. O sinal (+ ou -) indica a direção da relação. O valor pode variar de -1 a +1, em que +1 indica uma perfeita relação positiva, 0 indica relação nenhuma e -1, uma perfeita relação negativa ou reversa, quando uma variável se torna maior, a outra fica menor.

**COEFICIENTE DE CORRELAÇÃO:** A medida da extensão do relacionamento entre duas variáveis. É uma medida do relacionamento entre duas variáveis. Pode assumir valores de -1,00 a +1,00 inclusive. Zero indica não haver relacionamento entre as variáveis X e Y.

**COEFICIENTE DE CORRELAÇÃO DE PEARSON:** Coeficiente que mede o grau de associação linear entre duas variáveis contínuas com distribuição normal bivariada, cujo valor está compreendido entre -1 e +1. Em inglês Pearson correlation coefficient.

**COEFICIENTE DE CORRELAÇÃO DE PEARSON:** Medida da força de correlação linear entre duas variáveis contínuas.

**COEFICIENTE DE CORRELAÇÃO DE PEARSON:** O coeficiente de correlação de Pearson é uma medida do grau de relação linear entre duas variáveis quantitativas. Este coeficiente varia entre os valores -1 e 1. O valor 0 (zero) significa que não há relação linear, o valor 1 indica uma relação linear perfeita e o valor -1 também indica uma relação linear perfeita mas inversa, ou seja quando uma das variáveis aumenta a outra diminui. Quanto mais próximo estiver de 1 ou -1, mais forte é a associação linear entre as duas variáveis. O coeficiente de correlação de Pearson é normalmente representado pela letra **r** e a sua fórmula de cálculo é:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(\sum (x_i - \bar{x})^2)(\sum (y_i - \bar{y})^2)}}$$

**COEFICIENTE DE CORRELAÇÃO DE SPEARMAN:** Coeficiente que mede o grau de associação entre duas variáveis qualitativas ordinais. Em inglês Spearman rho correlation coefficient.

**COEFICIENTE DE CORRELAÇÃO LINEAR (r):** Medida estatística que permite calcular o valor numérico correspondente ao grau de dependência entre duas variáveis, o qual varia entre -1 e 1. Com n = número de observações da amostra, tem-se:

$$r = \frac{\sum_{i=1}^n \sum_{j=1}^n X_i Y_j - \frac{(\sum_{i=1}^n X_i)(\sum_{j=1}^n Y_j)}{n}}{\sqrt{\left[ \sum_{i=1}^n X_i^2 - \frac{(\sum_{i=1}^n X_i)^2}{n} \right] \left[ \sum_{i=1}^n Y_i^2 - \frac{(\sum_{i=1}^n Y_i)^2}{n} \right]}}$$

**COEFICIENTE DE CORRELAÇÃO LINEAR:** É uma medida da dependência mútua entre duas variáveis aleatórias X e Y. O coeficiente de correlação linear é igual a covariância padronizada e, portanto, varia de -1 a +1. O coeficiente de correlação linear ou coeficiente de Pearson, pode ser determinado pela seguinte relação:

$$\rho_{XY} = \frac{Cov[X, Y]}{\sqrt{Var[X]Var[Y]}}$$

onde  $Cov[X, Y]$  é a covariância entre as variáveis aleatórias X e Y;  $Var[X]$  e  $Var[Y]$  são respectivamente as variâncias das variáveis aleatórias X e Y. Quando o coeficiente for igual a -1, existe uma correlação linear negativa perfeita e quando igual a +1, uma correlação linear positiva. Quando igual a zero ou próximo de zero, indica que não há correlação.

**COEFICIENTE DE CORRELAÇÃO DE MOMENTO-PRODUTO DE PEARSON:** Veja coeficiente de correlação linear.

**COEFICIENTE DE CORRELAÇÃO PARCIAL:** Valor que mede a força da relação entre a variável dependente ou critério e uma única variável independente quando os efeitos das demais variáveis independentes no modelo são mantidos constantes. Por exemplo,  $r'_{y, x_2, x_1}$  mede a variação em Y associada a  $X_2$  quando o efeito de  $X_1$  em  $X_2$  e Y é mantido constante. Esse valor é usado em métodos de estimação de modelo de regressão com seleção seqüencial de variável para identificar a variável independente com o maior poder preditivo incremental além das variáveis independentes já presentes no modelo de regressão.

**COEFICIENTE DE CORRELAÇÃO PARCIAL:** Medida do grau de associação entre três ou mais variáveis quantitativas, sendo uma delas considerada constante e o cálculo do estimador estatístico sendo realizado com as outras duas ou mais variáveis envolvidas no estudo. Em inglês partial correlation coefficient.

**COEFICIENTE DE CORRELAÇÃO POR POSTOS:** Medida da intensidade do relacionamento entre duas variáveis, baseada nos postos dos valores.

**COEFICIENTE DE CORRELAÇÃO POR PONTOS DE SPEARMAN:** Veja coeficiente de correlação por postos.

**COEFICIENTE DE CORRELAÇÃO THAU DE KENDALL:** Medida do grau de associação entre duas variáveis qualitativas. Em inglês Kendall's tau-b.

**COEFICIENTE DE CURTOSE:** Medida de forma de distribuição de variáveis aleatórias contínuas, podendo ser mesocúrtica ou em forma de sino, leptocúrtica ou afilada e platicúrtica ou achatada.

**COEFICIENTE DE DETERMINAÇÃO (R<sup>2</sup>):** É o grau de associação entre as predições baseadas na equação de regressão e as predições da variável dependente.

**COEFICIENTE DE DETERMINAÇÃO (R<sup>2</sup>):** É o quadrado do coeficiente de correlação (r), representando a quantidade da variação de uma variável dependente explicada pela variável preditiva.

**COEFICIENTE DE DETERMINAÇÃO (R<sup>2</sup>):** A porcentagem da variação total na variável dependente explicada pela variável independente.

**COEFICIENTE DE DETERMINAÇÃO:** Valor da porcentagem de variação em Y explicada pela reta de regressão.

**COEFICIENTE DE DETERMINAÇÃO (r<sup>2</sup>):** Dá uma estimativa da proporção da variância compartilhada entre duas medidas sendo correlacionadas. Quando as correlações são menos que perfeitas, o valor de r<sup>2</sup> indica a proporção pela qual a variações em Y seria reduzida se x fosse mantido constante, ou vice-versa.

**COEFICIENTE DE DETERMINAÇÃO:** Coeficiente que mede o poder explicativo de uma equação de regressão ajustada a pares de valores de variáveis quantitativas. Em inglês coefficient of determination.

**COEFICIENTE DE DETERMINAÇÃO:** Valor entre 0 e 1 que indica como as variações na variável independente de uma regressão explicam as variações na variável dependente. É simbolizado por  $r^2$ .

**COEFICIENTE DE DETERMINAÇÃO ( $R^2$ ):** Medida da proporção da variância da variável dependente em torno de sua média que é explicada pelas variáveis independentes ou preditoras. O coeficiente pode variar entre 0 e 1. Se o modelo de regressão é propriamente aplicado e estimado, o pesquisador pode assumir que quanto maior o valor de  $R^2$ , maior o poder de explicação da equação de regressão e, portanto, portanto, melhor a previsão da variável dependente.

**COEFICIENTE DE DETERMINAÇÃO ( $r^2$ ):** É o Quadrado do coeficiente de correlação, ou ainda é a Proporção da variação total na variável dependente  $y$  que é explicada por descrita pela variação na variável independente  $x$ . Pode ser expressa como fração da unidade ou como percentagem que é o quadrado do coeficiente de correlação multiplicado por 100.

**COEFICIENTE DE DETERMINAÇÃO ( $r^2$ ):** É o quadrado do coeficiente de correlação de Pearson e expresso em porcentagem. É o percentual explicado da variação da variável dependente pela reta de regressão ou modelo. O restante é explicado pelo erro, que pode ser devido a ausência de outras variáveis, erros de mensuração das variáveis e ao erro aleatório.

**COEFICIENTE DE FECUNDIDADE ESPÉCIFICA POR IDADE:** Indicador que relaciona o número de nascidos vivos, referidos a uma determinada idade da mãe, com o número total de mulheres, na mesma idade, expressando-se o resultado por 1.000. Justifica-se seu uso pela grande variação da fecundidade, em relação à idade da mulher.

**COEFICIENTE DE FECUNDIDADE GERAL:** Coeficiente dado pela relação entre o número de nascidos vivos em determinada área e período, e a população feminina de 15 a 49 anos, na área considerada e na metade do período referido, expressando-se o resultado por 1.000.

**COEFICIENTE DE FECUNDIDADE TOTAL:** Indicador obtido pela soma dos coeficientes de fecundidade específicos por idade, expressando o número de filhos vivos, de ambos os sexos, por mulher. Por não depender da estrutura etária da respectiva população e, por conseguinte, não ser influenciada pela distribuição etária, esse coeficiente é muito empregado em comparações populacionais de fecundidade.

**COEFICIENTE DE FIDEDIGNIDADE:** Uma medida da consistência de um teste. Consistência de um instrumento em medir o que quer que meça. Coeficiente de fidedignidade podem ter valores que vão de zero a 1,00. O instrumento fidedigno dará resultados similares para os mesmos indivíduos em ocasiões diferentes. Fidedignidade significa consistência numa série de mensurações. O coeficiente de fidedignidade nos diz que a proporção da variância do teste é variância não-erro. O coeficiente de fidedignidade aumenta com o tamanho do teste, e também com a dispersão ou variância de escores.

**COEFICIENTE DE GRAVIDADE:** Indicador para avaliar a virulência de um determinado bioagente e expressa a porcentagem dos casos considerados como graves segundo critérios preestabelecidos.

**COEFICIENTE DE INCIDÊNCIA:** Ver incidência.

**COEFICIENTE DE INCIDÊNCIA:** Coeficiente calculado como o número de casos incidentes em um período de estudo definido, dividido pela população em risco no ponto médio desse período de estudo.

Os coeficientes de ocorrência de nascimentos, mortes e de casos novos de doenças são todos formas de um coeficiente de incidência.

**COEFICIENTE DE INCIDÊNCIA:** Constitui medida de risco de doença ou agravamento, fundamentalmente nos estudos da etiologia de doenças agudas e crônicas. É a razão entre o número de casos novos de uma doença que ocorre em uma coletividade, em um intervalo de tempo determinado, e a população exposta ao risco de adquirir referida doença no mesmo período, multiplicando-se o resultado por potência de 10, que é a base referencial da população.

**COEFICIENTE DE LETALIDADE:** Coeficiente resultante da relação entre o número de óbitos decorrente se determinadas causa e o número de pessoas que foram realmente acometidas pela doença, expressando-se sempre em percentual. É um indicador útil para avaliar a virulência de um determinado bioagente.

**COEFICIENTE DE LETALIDADE (PROPORÇÃO DE CASOS FATAIS):** Proporção das pessoas clinicamente doentes que morrem por causa da condição em estudo. É um indicador de virulência. Em inglês Case fatality ratio.

**COEFICIENTE DE MORBIDADE:** Relação entre o número de casos de uma doença e a população exposta a adoecer. Discriminado em coeficiente de incidência e coeficiente de prevalência. Muito útil para o objetivo de controle de doenças ou de agravos, bem como para estudos de análise do tipo causa/efeito.

**COEFICIENTE DE MORTALIDADE:** Relação entre a frequência absoluta de óbitos e o número dos expostos ao risco de morrer. Pode ser geral, quando inclui todos os óbitos e toda a população da área em estudo, e pode ser específico por idade, sexo, ocupação, causa dentre outros.

**COEFICIENTE DE MORTALIDADE ESPECÍFICO POR IDADE:**

$$= \frac{\text{Número de óbitos de pessoas em um grupo etário em particular (local e período de tempo definido)}}{\text{População no meio do período (mesmo grupo etário local e período de tempo)}} \times 100$$

**COEFICIENTE DE MORTALIDADE FETAL TARDIA:** Ver coeficiente de natimortalidade.

**COEFICIENTE DE MORTALIDADE GERAL:** Coeficiente que se expressa dividindo-se o número de óbitos concernentes a todas as causas em um determinado ano pela população naqueles ano, circunscritos os dados a uma determinada área, e multiplicando-se por 1.000, base referencial para a população exposta.

**COEFICIENTE DE MORTALIDADE INFANTIL:** Um dos mais sensíveis indicadores de saúde e, talvez, o mais utilizado dentre os coeficientes de mortalidade. É obtido da divisão do número de óbitos de crianças menores de 1 ano de idade, em certa área e em determinado ano, pelos nascidos vivos, na área e ano considerados, multiplicando-se por 1.000 o valor encontrado. Mede o risco de morte para crianças menores de 1 ano e é diretamente influenciado pelas condições sócio-econômicas da população. Seus componentes: coeficiente de mortalidade neonatal e coeficiente de mortalidade pós-neonatal.

**COEFICIENTE DE MORTALIDADE INFANTIL:**

$$= \frac{\text{Número de óbitos de crianças com menos de 1 ano de idade (local e período de tempo definido)}}{\text{Número de nascidos vivos (mesmo local e período de tempo)}} \times 1.000$$

**COEFICIENTE DE MORTALIDADE INFANTIL PRECOCE:** Ver coeficiente de mortalidade neonatal.

**COEFICIENTE DE MORTALIDADE INFANTIL TARDIA:** Ver coeficiente de mortalidade pós-natal.



**COEFICIENTE DE MORTALIDADE MATERNA:** Relaciona os óbitos por complicações da gravidez, parto e purpúreo que são as mortes maternas obstétricas, com o número de nascidos vivos, e deve ser expresso por 100.000 nascidos vivos. As mortes maternas não-obstétricas não são incluídas no cálculo deste coeficiente. Ver mortalidade materna.

**COEFICIENTE DE MORTALIDADE MATERNA:**

$$= \frac{\text{Número de óbitos relacionado à gestação (local e período de tempo definidos)}}{\text{Número de nascidos vivos (mesmo local e período de tempo)}} \times 100.000$$

**COEFICIENTE DE MORTALIDADE MATERNA DIRETA:** Coeficiente obtido da divisão do número de mortes maternas de causas obstétricas diretas pelo total de nascidos vivos em lugar e período determinados, multiplicando-se por 1.000. Ver mortalidade materna de causa obstétrica direta.

**COEFICIENTE DE MORTALIDADE MATERNA INDIRETA:** Coeficiente derivado da divisão do número de mortes maternas de causas obstétricas indiretas pelo total de nascidos vivos em lugar e período determinados, multiplicando-se o resultado por 1.000. Ver mortalidade materna de causa obstétrica indireta.

**COEFICIENTE DE MORTALIDADE NEONATAL:** É a relação do número de óbitos de 28 dias de idade (0|28) e de 1.000 g ou mais por 1.000 recém-nascidos vivos de 1.000 g ou mais ao nascer em um intervalo de tempo e lugar determinados. A maioria dos óbitos nesse período é decorrente de causas perinatais tais como problemas de gestação, de parto, fatores maternos vários, e de anomalias congênicas.

**COEFICIENTE DE MORTALIDADE NEONATAL:**

$$= \frac{\text{Número de óbitos de crianças com menos de 28 dias de idade (local e período de tempo definido)}}{\text{Número de nascidos vivos (mesmo local e período de tempo)}} \times 1.000$$

**COEFICIENTE DE MORTALIDADE NEONATAL PRECOCE:** É a relação do número de óbitos de crianças de 1.000 g ou mais, com menos de 7 dias completos a partir do nascimento (0|—7), por 1.000 recém-nascidos vivos de 1.000 g ou mais ao nascer, em período e lugar determinados.

**COEFICIENTE DE MORTALIDADE PERINETAL:** Coeficiente resultante do número e natimortos de 1.000 g ou mais somado ao número de mortos de 1.000 g ou mais e de até 6 dias completos de vida (0|—7), dividido pelo número total de nascidos vivos e mortos de 1.000 g ou mais ao nascer, em período e lugar determinados, multiplicando-se o resultado por 1.000.

**COEFICIENTE DE MORTALIDADE PERINATAL:**

$$= \frac{\text{Número de natimortos (local e período de tempo definidos)} + \text{Número de óbitos em crianças com menos de 7 dias de idade (mesmo local e período de tempo)}}{\text{Número de natimortos (mesmo local e período de tempo)} + \text{Número de nascidos vivos (mesmo local e período de tempo)}} \times 1.000$$

Vale lembrar que diversas outras fórmulas são usadas em diferentes partes do mundo.

**COEFICIENTE DE MORTALIDADE POR CAUSA:** Coeficiente que resulta da divisão do número de óbitos ocorridos por determinada causa e a população exposta e, a seguir, multiplicando-se o resultado por 100.000. Deve-se definir precisamente qual a população exposta, a fim de não se incorrer no erro de, por exemplo, colocar mulheres no denominador quando se quer mensurar a mortalidade por câncer de próstata.

**COEFICIENTE DE MORTALIDADE POR CAUSA ESPECÍFICA:**

$$= \frac{\text{Número de óbitos devido a causa particular (local e período de tempo definido)}}{\text{População no meio do período (mesmo local e período de tempo)}} \times 1.000$$

**COEFICIENTE DE MORTALIDADE POR DOENÇAS TRANSMISSÍVEIS:** Calculado a partir do número de habitantes no denominador e do número de óbitos por doenças infecciosas e parasitárias no numerador, tomando-se como referência a Classificação Internacional de Doenças e Causas de Óbitos.

**COEFICIENTE DE MORTALIDADE PÓS-NEONATAL:** É a relação do número de óbitos ocorridos no período que vai do 28º dia de vida até o 12º mês, antes de 1 ano de idade, por 1.000 nascidos vivos em um intervalo de tempo e determinados. Os óbitos desse período são resultantes, principalmente, de causas ligadas a fatores ambientais, como doenças infecciosas e desnutrição. Sinônimo: coeficiente de mortalidade infantil tardia.

**COEFICIENTE DE MORTALIDADE PÓS-NEONATAL:**

$$= \frac{\text{Número de óbitos de crianças entre 28 e 365 dias de idade (local e período de tempo < definidos)}}{\text{Número de nascidos vivos (mesmo local e período de tempo)} - \text{Número de óbitos neonatais (mesmo local e período de tempo)}} \times 1.000$$

**COEFICIENTE DE MORTALIDADE PÓS-NEONATAL APROXIMADO:**

= Coeficiente de mortalidade – Coeficiente de mortalidade neonatal.

**COEFICIENTE DE NATIMORTALIDADE:** Natimortos com peso de 1.000 g ou mais por cada 1.000 nascimentos totais que são os natimortos mais os nascidos vivos de 1.000 g ou mais, em um período e lugar determinados. Sinônimo: coeficiente de mortalidade fetal tardia.

**COEFICIENTE DE PEARSON:** Representado pela letra r minúscula, varia de -1 a +1. Quando seu valor é -1, a correlação é perfeita negativa: os valores altos numa variável correspondem a valores baixos em outra. Quando seu valor é +1, a correlação é perfeita positiva: a valores altos numa variável correspondem valores altos na outra. Quando seu valor é zero, não existe correlação. Em inglês Pearson coefficient.

**COEFICIENTE DE PEARSON:** A correlação de Pearson requer que ambas as variáveis sejam medidas pelo menos em escala intervalar e que suas implícitas distribuições sejam contínuas, algo simétricas e homoscedásticas.

**COEFICIENTE DE PREVALÊNCIA:** Coeficiente que mede a força com que subsiste a doença na coletividade. Expressa-se como a relação entre o número de casos conhecidos de uma dada doença e a população, multiplicando-se o resultado pela base referencial da população, que é potência de 10, usualmente 1.000, 10.000 ou 100.000.

**COEFICIENTE DE PREVALÊNCIA:** Proporção de pessoas com uma doença ou condição definida no momento em que são estudadas. Essa proporção não é verdadeiramente um coeficiente, embora seja convencionalmente referida como tal.

**COEFICIENTE DE PREVALÊNCIA:** Termo que mede a proporção a qual é geralmente a porcentagem de uma população que tem doença ou condição definida em um ponto particular no tempo. Embora usualmente chamado de coeficiente, é na verdade uma proporção.

**COEFICIENTE DE PREVALÊNCIA INSTANTÂNEA:** Ver coeficiente de prevalência pontual.



**COEFICIENTE DE PREVALÊNCIA LÁPSICA:** Refere-se ao coeficiente de prevalência que abrange um lapso de tempo mais ou menos longo e que não concentra a informação em um dado ponto desse intervalo. É a medida que expressa o número total de casos de uma doença, somando-se a prevalência pontual ao começo de um período específico com todos os casos novos que ocorrem durante esse período. No numerador estão incluídos as altas, óbitos e emigrações ocorridas neste intervalo de tempo.

**COEFICIENTE DE PREVALÊNCIA MOMENTÂNEA:** Ver coeficiente de prevalência pontual.

**COEFICIENTE DE PREVALÊNCIA PONTUAL:** Refere-se ao coeficiente de prevalência em um ponto definido no tempo, seja o dia, a semana, o mês ou o ano. Isto é, mede a proporção de uma população que, a um determinado instante, apresenta doença. No numerador são incluídos os casos que estão vivos e diagnosticáveis na época de avaliação, retirando aqueles que já faleceram ou que foram curados ou que emigraram. Sinônimo: coeficiente de prevalência momentânea e coeficiente de prevalência instantânea.

**COEFICIENTE DE REGRESSÃO:** É uma constante que determina o grau de inclinação da reta de regressão, simbolizado pela letra b. Na regressão múltipla há tantos coeficientes de regressão quanto o número de variáveis independentes testadas.

**COEFICIENTE DE REGRESSÃO (INCLINAÇÃO):** É o gradiente de uma linha, i.e., a variação na variável dependente ou variável resposta (y) por unidade de variação na variável independente ou preditora (x).

**COEFICIENTE DE REGRESSÃO ( $b_n$ ):** Valor numérico da estimativa do parâmetro diretamente associado com uma variável independente; por exemplo, no modelo  $y = b_0 + b_1 X_1$  é o coeficiente de regressão para a variável  $X_1$ . O coeficiente de regressão representa o montante de variação na variável dependente em relação a uma unidade de variação na variável independente. No modelo preditor múltiplo (por exemplo,  $y = b_0 + b_1 X_1 + b_2 X_2$ ), os coeficientes de regressão são coeficientes parciais, pois cada um considera não apenas as relações entre Y e  $X_1$  e entre Y e  $X_2$ , mas também entre  $X_1$  e  $X_2$ . O coeficiente não é limitado nos valores, já que é baseado tanto no grau de associação quanto nas unidades de escala da variável independente. Por exemplo, duas variáveis com a mesma associação a Y teriam coeficientes diferentes se uma variável independente fosse medida em uma escala de 7 pontos e outra fosse baseada em uma escala de 100 pontos.

**COEFICIENTE DE REPRODUCIBILIDADE:** Medida da extensão pela qual um escore de escala permite a reconstrução exata dos dados específicos que entraram na montagem da escala.

**COEFICIENTE DE TANQUE:** A taxa de incidência a uma população específica ou a um grupo bem definido de pessoas, limitadas a uma área e tempo restritos. Muito útil para investigar surtos epidêmicos logo em sua eclosão e durante a sua vigência.

**COEFICIENTE DE TRAJETORIA:** Medida da associação entre duas variáveis numa análise de trajetória.

**COEFICIENTE DE VARIAÇÃO:** É o quociente entre o desvio padrão e a média de um conjunto de dados. É um percentual e, portanto adequado para efetuar comparações entre diferentes conjuntos de dados. Em inglês coefficient of variation.

**COEFICIENTE DE VARIAÇÃO:** Medida de variabilidade que resulta da divisão do desvio padrão e a média. Às vezes é apresentada em forma de porcentagem. Em inglês Coefficient of variation.

**COEFICIENTE DE VARIAÇÃO:** É o desvio-padrão expresso com porcentagem da média.

**COEFICIENTE DE VARIAÇÃO:** Quando temos que comparar projetos que envolvem significativas diferenças em seus valores, usamos a idéia de variação relativa para medir risco relativo. Uma dessas medidas é conhecida como coeficiente de variação (V ou CV) definido como:  $V = CV = \frac{\sigma}{\bar{X}}$ . Em geral, quanto maior o coeficiente de variação, mais arriscado o projeto.

**COEFICIENTE DE VARIAÇÃO:** Medida de variação relativa com características de que relaciona o desvio padrão com a média aritmética. Sua utilidade está em que se expressa em porcentagem (%) de número abstrato. É usado para comparar a variabilidade de duas séries, uma vez que nele não intervêm as unidades de medida. É frequentemente utilizado para comparações de variações relativas entre variáveis de diversas séries.

**COEFICIENTE DE VARIAÇÃO:** Desvio padrão dividido pela média; compara o tamanho da dispersão com a média.

**COEFICIENTE DE VARIAÇÃO:** É igual à razão entre o desvio padrão e a média e pode ser expressa em porcentagem, se multiplicada por 100.

$$CV = \frac{S}{\bar{X}}$$

Como o coeficiente de variação é adimensional, frequentemente é utilizado para comparar a dispersão relativa de valores em torno da média entre diferentes distribuições. Neste sentido, este parâmetro, é extensivamente utilizado pela escola soviética de exploração mineral para classificação de depósitos minerais para fins de pesquisa. O coeficiente de variação é também uma medida da forma da distribuição, sendo uma medida do grau de achatamento da curva, pois para uma distribuição com mesma média e valores variáveis de desvio padrão, o coeficiente de variação será tão alto quanto o grau de achatamento.

**COEFICIENTE DE VARIAÇÃO:** Razão entre o desvio padrão e a média aritmética, vezes 100(%).

**COEFICIENTE DE VARIAÇÃO:** Desvio padrão dividido pela média, multiplicado por 100.

**COEFICIENTE DE VARIAÇÃO (C.V.):** É um número abstrato, isto é, não tem unidade. Mas é uma medida de dispersão que expressa percentualmente o desvio padrão por unidade de medida.

**COEFICIENTE DE VARIAÇÃO (C.V.):** Estatística ou estimador obtido pela fórmula  $C.V. = \frac{100s}{\hat{m}}$  onde s é uma estimativa do desvio padrão e  $\hat{m}$  uma estimativa da média.

**COEFICIENTE DE VARIAÇÃO DE PEARSON ( $CV_p$ ):** É a razão entre o desvio padrão e a média referentes a dados de uma mesma série.

$$CV_p = \frac{S}{\bar{X}} \cdot 100(\%)$$

**COEFICIENTE DE VARIAÇÃO DE THORNDIKE ( $CV_t$ ):** É igual ao quociente entre o desvio padrão e a mediana.

$$CV_t = \frac{S}{Md} \cdot 100(\%)$$

**COEFICIENTE DISCRIMINANTE:** Ver peso discriminante.

**COEFICIENTE ESPECÍFICO:** Coeficiente que indica a frequência de resultado adverso de um grupo específico de população. No numerador, deverá constar o número do evento ocorrido no grupo, e, no denominador, o número total de indivíduos que compõem esse grupo, em um tempo e lugar determinados.

**COEFICIENTE GAMA:** Em inglês gamma coefficient.

**COEFICIENTE GERAL DE MORTALIDADE:**

$$= \frac{\text{Número de óbitos (local e período de tempo definido)}}{\text{População no meio do período (mesmo local e período de tempo)}} \times 1.000$$

**COEFICIENTE GERAL DE NATALIDADE:** Indicador que relaciona o número de nascidos vivos com a população total. Sinônimo: taxa bruta de natalidade.

**COEFICIENTE GERAL DE NATALIDADE:** Definido pela relação entre o número anual de nascidos vivos, e em área e período determinados, e a população dessa área, no meio do período considerado, multiplicando-se o resultado por 1.000.

**COEFICIENTE GERAL DE NATALIDADE**

$$= \frac{\text{Número de nascidos vivos (local e período de tempo definidos)}}{\text{População no meio do período (mesmo local e período de tempo)}} \times 1.000$$

**COEFICIENTE LOGÍSTICO:** Coeficiente que atua como o fator de ponderação para as variáveis independentes em relação a seu poder discriminatório. Semelhante a um peso de regressão ou um coeficiente discriminante.

**COEFICIENTE MÚLTIPLO DE DETERMINAÇÃO AJUSTADA:** Uma medida da eficiência da equação de regressão múltipla estimada, que ajusta para o número de variáveis independentes no modelo e desta forma evita superestimar o impacto de adicionar mais variáveis independentes.

**COEFICIENTE MÚLTIPLO DE DETERMINAÇÃO:** Mede quão bem uma equação de regressão múltipla se ajusta aos dados amostrais.

**COEFICIENTE PHI ( $r_\phi$ ):** Coeficiente de correlação entre duas variáveis qualitativas e dicotômicas, dispostas em tabela de contingência 2x2.

**COEFICIENTE OU TAXA:** É uma relação ou quociente entre dois valores numéricos que estimaria uma probabilidade ou determinado risco.

**COEFICIENTE OU TAXA:** Frequência ou número de eventos que ocorrem em um período definido de tempo, dividida pelo tamanho médio da população sob risco.

**COEFICIENTE ANGULAR OU INCLINAÇÃO DA RETA DE REGRESSÃO:** É o valor do beta 2, ou b no modelo  $Y = a + b * X$ , e significa a variação para mais ou para menos de b unidades em Y por cada unidade de variação de X.

**CORRELAÇÃO MONOTÔNICA:** A correlação é chamada de monotônica se uma das variáveis aumenta ou diminui sistematicamente quando a outra decresce, com associações que podem ter forma linear ou não linear.

**COEFICIENTE QUARTÍLICO DE VARIAÇÃO (CVQ):** Esse coeficiente é definido pela expressão:

$$CVQ = \frac{(Q_3 - Q_1)}{(Q_3 + Q_1)}$$

$$\text{ou } CVQ = \frac{(Q_3 - Q_1)}{(Q_3 + Q_1)} \times 100 \text{ para resultado em porcentagem (\%)}$$

**COEFICIENTES AJUSTADOS (TAXAS AJUSTADAS):** O mesmo que coeficientes padronizados ou taxas padronizadas.

**COEFICIENTES BRUTOS:** Coeficientes que dizem respeito à população como um todo, sem referência a qualquer característica dos indivíduos que a compõem.

**COEFICIENTES ESPECÍFICOS:** Coeficientes que são calculados após a população ser categorizada em grupos, por uma característica particular. Exemplos incluem coeficientes específicos por idade e coeficientes específicos por sexo. São, geralmente, necessários para comparações válidas de coeficientes.

**COEFICIENTES ESPECÍFICOS POR CAUSA:** Coeficientes que fornecem numeradores que são comparáveis com referência ao diagnóstico.

**COEFICIENTES GERAIS:** Coeficientes aplicados a uma população inteira, sem preferência a nenhuma das características dos indivíduos contidos na mesma. Os coeficientes gerais, frequentemente, não são úteis para comparações, porque as populações podem ser muito diferentes em sua composição, principalmente em relação à idade.

**COEFICIENTES PADRONIZADOS:** Coeficientes brutos que foram modificados de modo a permitir comparações válidas entre elas. A padronização é geralmente necessária para corrigir diferentes distribuições etárias em populações distintas. Ver padronização direta e padronização indireta.

**COEFICIENTES PADRONIZADOS (AJUSTADOS):** Coeficientes gerais modificados ou ajustados para controlar efeitos da idade ou de outras características e, assim, permitir comparações válidas desses coeficientes.

**CONFUNDIMENTO:** Situação que ocorre quando os efeitos de duas ou mais variáveis não podem ser distinguidos uns dos outros.

**CONFUNDIMENTO:** Confusão entre duas variáveis supostamente causais, de modo que parte do pretenso efeito de uma variável é realmente devido à outra.

**CO-INTERVENÇÃO:** Realização de procedimentos diagnosticados e terapêuticos adicionais, em um dos grupos.

**COLETA DE DADOS:** Fase operacional. É o registro sistemático de dados, com um objetivo determinado.

**COLETA DE DADOS:** É a fase da pesquisa em que se reúnem dados por meio de técnicas específicas.

**COLETA DIRETA:** Quando é obtida diretamente da fonte. Pode ser contínua como registros de nascimento, óbitos, casamentos, dentre outras, periódica tais como recenseamento demográfico, censo industrial e ocasional como por exemplo, o registro de casos de dengue.

**COLETA INDIRETA:** É feita por deduções a partir dos elementos conseguidos pela coleta direta, por analogia, por avaliação, indícios ou proporcionalização.

**COLINEARIDADE:** Condição em que algumas covariáveis utilizadas numa regressão resultam da combinação linear de outras. Em inglês Collinearity.

**COLINEARIDADE:** Relação entre duas (colinearidade) ou mais (multicolinearidade) variáveis. As variáveis exibem colinearidade completa se seu coeficiente de correlação é 1, e uma completa falta de colinearidade se seu coeficiente de correlação é 0.

**COLINEARIDADE:** Termo que expressa a relação entre duas (colinearidade) ou mais (multicolinearidade) variáveis independentes. Diz-se que duas variáveis independentes exibem colinearidade completa se seu coeficiente de correlação é 1, e completa falta de colinearidade se o coeficiente de correlação é 0. A multicolinearidade ocorre quando qualquer variável independente é altamente correlacionada com um conjunto de outras variáveis independentes. Um caso extremo de colinearidade/multicolinearidade é a singularidade, na qual uma variável independente é perfeitamente prevista, ou seja, uma correlação de valor igual a 1,0 por uma outra variável independente ou mais de uma.

**COLUNA:** Um dos tipos de bloco que ocorrem no delineamento experimental do quadrado latino.

**COMBINAÇÃO:** Na análise combinatória é o subgrupo formado pela escolha sem reposição de objetos de um grupo em que a ordem não é considerada. Em inglês combination.

**COMBINAÇÃO LINEAR:** A soma de duas ou mais variáveis, como por exemplo, a expressão  $Y = b_0 + b_1 * X_1 + b_2 * X_2 + \dots + b_k * X_k$ . Se qualquer das variáveis estiverem ligadas por um operador diferente de + ou -, a combinação não será mais linear.

**COMBINAÇÃO LINEAR:** É um processo matemático da média ponderada empregado na krigagem ordinária, na ponderação pelo inverso da potência da distância ou então em qualquer processo de médias móveis, em que o valor procurado é encontrado como resultado da combinação linear ou somatória dos produtos dos valores de dados com seus respectivos ponderadores. Ver o conceito de média ponderada.

**COMBINAÇÕES:** Número de escolhas de j objetos distintos de um grupo de n objetos, quando a ordem em que se faz a escolha não e elevada em conta.

**COMBINAÇÃO DE n ELEMENTOS TOMADOS k a k:** Chamaremos cada seleção de n elementos tomados k a k desprezando a ordem como o número total de combinações por  $C_k^n$  ou seja,

$$C_k^n = \frac{n(n-1) \dots (n-k+1)}{k!} = \frac{n!}{k!(n-k)!},$$

esta extensa fórmula é normalmente abreviada pelo símbolo  $\binom{n}{k}$ . Não existe nenhuma forma padrão

para a leitura deste símbolo. Uma maneira apropriada seria n acima de k, mas utiliza-se em geral a seguinte: combinação de n k a k. O símbolo deveria ser cuidadosamente diferenciado de n sobre k, que significa a fração  $\frac{n}{k}$ . Vejamos um exemplo de combinação: De quantas maneiras três (3) dentre cinco (5) animais

suínos da Raça Duroc podem ser selecionados? A equação  $C_k^n = \frac{n(n-1) \dots (n-k+1)}{k!}$  dá a resposta.

$$C_3^5 = \binom{5}{3} = \frac{5!}{3!(5-3)!} = \frac{5!}{1.2.3} = 10. \text{ Se for representado os cinco animais, arbitrariamente, por}$$

a, b, c, d, e, as dez combinações serão: abc, abd, abe, acd, ace, ade, bcd, bce, bde, cde. Observemos que dce ou edc não ocorrem nesta listagem já que eles são somente arranjos de cde. Da mesma forma, obtemos,

$$C_1^5 = \binom{5}{1} = \frac{5!}{1!(5-1)!} = \frac{5!}{4!} = 5$$

$$C_2^5 = \binom{5}{2} = \frac{5!}{2!(5-2)!} = \frac{5.4}{1.2} = 10$$

$$C_4^5 = \binom{5}{4} = \frac{5!}{4!(5-4)!} = \frac{5.4.3.2}{1.2.3.2} = 5$$

Deve ser observado que

$$\binom{5}{5} = 1, \text{ já que pode-se selecionar 5 objetos somente de uma maneira. Pode-se observar também que}$$

$$\binom{5}{1} = \binom{5}{4}, \binom{5}{2} = \binom{5}{3}$$

Esta propriedade tem um significado particular. Se selecionarmos um objeto dentre cinco, os quatro objetos restantes formam também uma seleção. A cada seleção de um objeto corresponde exatamente uma seleção de quatro e vice e versa. Portanto, seus números devem ser iguais. O mesmo é verdadeiro para seleções de dois e três objetos dentre cinco. Em geral, temos que  $\binom{n}{k} = \binom{n}{n-k}$  para  $k = 1, 2, \dots, n-1$ .

Deve-se observar que em  $\binom{n}{k} = \binom{n}{n-k}$  para  $k = 1, 2, \dots, n-1$ , o símbolo  $\binom{n}{n}$  não ter par. Formalmente,

ele seria  $\binom{n}{0}$ . Entretanto, não existe nada como uma combinação de zero objeto dentre n. Introduzimos

$\binom{n}{0}$  somente para a segurança da simetria. Para satisfazer  $\binom{n}{k} = \binom{n}{n-k}$  para  $k = 1, 2, \dots, n-1$ , definimos

$$\binom{n}{0} = \binom{n}{n} = 1.$$

Aqui, n é qualquer inteiro positivo. Mas, vamos um pouco além e permitimos que n seja zero. Então definimos  $\binom{0}{0} = 1$ . Os números  $\binom{n}{k}$  são frequentemente dispostos de tal forma que eles formam um

triângulo de Pascal [Blaise Pascal (1623-1662), matemático, físico e filósofo francês].

$\binom{0}{0}$	1
$\binom{1}{0} \binom{1}{1}$	1 1
$\binom{2}{0} \binom{2}{1} \binom{2}{2}$	1 2 1
$\binom{3}{0} \binom{3}{1} \binom{3}{2} \binom{3}{3}$	1 3 3 1
$\binom{4}{0} \binom{4}{1} \binom{4}{2} \binom{4}{3} \binom{4}{4}$	1 4 6 4 1
$\binom{5}{0} \binom{5}{1} \binom{5}{2} \binom{5}{3} \binom{5}{4} \binom{5}{5}$	1 5 10 10 5 1



Existe uma forma fácil de calcularmos os números do triângulo de Pascal. Observamos que cada número é a soma dos dois números mais próximos na linha imediatamente superior. Então,  $5 = 1 + 4$ ,  $10 = 4 + 6$  dentre outros. A próxima linha não representada aqui, conteria, portanto, os números  $1 = 5 = 6$ ,  $5 + 10 = 15$ ,  $10 + 10 = 20$  dentre outros.

Uma aplicação importante do coeficiente  $\binom{n}{k}$  é feita na álgebra. Consideramos potências de um binômio com inteiros positivos como expoentes dados por:

$$(a + b)^2 = a^2 + 2ab + b^2,$$

$$(a + b)^3 = a^3 + 3a^2b + 3ab^2 + b^3,$$

$$(a + b)^4 = a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4, \text{ dentre outros.}$$

Os coeficientes formam as linhas do triângulo de Pascal. Em geral, pode-se escrever o seguinte:

$$(a + b)^n = \binom{n}{0}a^n + \binom{n}{1}a^{n-1}b + \binom{n}{2}a^{n-2}b^2 + \dots + \binom{n}{n}b^n,$$

Uma forma de demonstrar esta fórmula utiliza combinação, e para explicar a idéia, vamos mostrar o caso especial em que  $n=4$ . Consideramos o produto.

$$(a_1 + b_1)(a_2 + b_2)(a_3 + b_3)(a_4 + b_4).$$

Quando fazemos as multiplicações, obtemos uma soma de muitos termos. Cada um dos termos contém exatamente quatro fatores, tais como  $(a_1b_2b_3b_4)$  ou  $(a_3a_4b_1b_2)$ . Todas as combinações dos a e b apa-

recem. Existem  $c_3^4 = \binom{4}{3}$  com três b,  $c_2^4 = \binom{4}{2}$  com dois b, dentre outros. Quando finalmente igualamos

$a_1 = a_2 = a_3 = a_4 = a$  e  $b_1 = b_2 = b_3 = b_4 = b$ , o termo  $ab^3$  aparece  $\binom{4}{3}$  vezes, o termo  $a^2b^2$  aparece

$\binom{4}{2}$  vezes dentre outros. Isto explica a fórmula para  $(a + b)^4$ . O mesmo argumento pode ser usado para

qualquer outro expoente inteiro  $n$  ( $n > 0$ ).

A fórmula  $(a + b)^n = \binom{n}{0}a^n + \binom{n}{1}a^{n-1}b + \binom{n}{2}a^{n-2}b^2 + \dots + \binom{n}{n}b^n$ , é chamada o Teorema do Binômio.

Os coeficientes  $\binom{n}{k}$  são conhecidos como coeficientes binomiais. Quando transformamos  $(a + b)^n$  em uma soma de termos  $\binom{n}{k}a^{n-k}b^k$ , também dizemos que expandimos  $(a + b)^n$ .

**COMPANHIAS DE SERVIÇO DE CAMPO:** Empresas que apenas coletam dados para clientes empresariais ou empresas de pesquisa.

**COMPARAÇÃO DE GRUPOS:** Comparação entre grupos de indivíduos, geralmente na base de valores representativos como a média de cada um. Em inglês Group comparison.

**COMPARAÇÃO DE TRATAMENTOS:** Qualquer comparação que envolva dois ou mais tratamentos ou grupos.

**COMPARAÇÕES MÚLTIPLAS:** Refere-se ao fato de que dois ou mais tratamentos devem ser comparados, sempre em relação à mesma variável, em determinado momento do ensaio em geral ou no fim.

**COMPARAÇÃO PLANEJADA:** Teste a priori para uma comparação específica de diferenças de médias de grupos. Esses testes são executados em conjunto com os testes para efeitos principal e de interação usando-se um contraste.

**COMPLEMENTO:** De um evento, conjunto de todos os resultados que não fazem parte do evento.

**COMPLEMENTO DE UM CONJUNTO:** Dado um conjunto A o complemento de A é o conjunto de todos os elementos que não estão em A, isto é,  $S-A$ . Representa-se por:  $A'$ ,  $A^c$  ou ainda  $\bar{A}$ .

**COMPLEMENTO DE UM EVENTO:** Todos os resultados em que o evento original não ocorre.

**COMPLETAMENTE PERDIDOS AO ACASO (MCAR):** Classificação de dados perdidos aplicável quando valores perdidos de Y não dependem de X. Quando os dados perdidos são MCAR, os valores observados de Y são uma amostra verdadeiramente aleatória de todos os valores de Y, sem um processo inerente que induza tendências aos dados observados.

**COMPLETITUDE DA ESTATÍSTICA:** Uma estatística  $h(x)$  é completa somente se sua família de densidades é completa. Em inglês Completeness of statistic.

**COMPONENTE DE TENDÊNCIA:** Componente de uma série temporal que aumenta ou diminui suavemente com o tempo.

**COMPONENTE DE VARIÂNCIA:** Qualquer das fontes de variação numa análise de variância que contribui com parte da soma de quadrados dos desvios das observações com relação à média. Em inglês Component of variance.

**COMPONENTE IRREGULAR:** Componente de uma série temporal que aumenta ou diminui sem qualquer padrão visível.

**COMPORTAMENTO, CIÊNCIA DO (BEHAVIORAL SCIENCE):** Quando aplicado à administração, a ciência do comportamento é um estudo de observáveis e verificáveis no comportamento humano em organizações. Possui três níveis de análise: i) comportamento do indivíduo; ii) comportamento do grupo; iii) efeitos da organização sobre o comportamento.

**COMPOSIÇÃO DE AMOSTRAGEM:** A lista de membros da população útil da qual é tirada a amostra.

**COMPOSIÇÃO DE AMOSTRAS DE FUROS DE SONDAGEM:** É o processo pelo qual as amostras dos furos de sondagem, coletadas a intervalos predeterminados são compostas para o intervalo de trabalho adequado ao tipo de depósito e objetivo da avaliação. O objetivo de se fazer composições de amostras é de obter amostras representativas de uma unidade mineralógica particular ou unidade de mineração, às quais podem ser aplicadas, por meio de alguma função de extensão, para estimar o teor ou valor de um volume muito maior da mesma unidade. Existem muitos tipos de depósitos minerais, cada um dos quais irá requerer um tratamento especial dos dados amostrados para a obtenção dos melhores intervalos de



composição para avaliação do depósito. Basicamente são três os tipos de composições possíveis em amostras de furos de sondagem para o intervalo de trabalho: 1. bancadas; 2. tipos litológicos; e 3. elevações.

**COMPOSIÇÃO DE VARIÂNCIAS:** É feita por meio da relação de aditividade de variâncias. A composição de variâncias permite o cálculo da variância do depósito, a partir das variâncias parciais dos blocos:

$$\varepsilon_D^2 = \frac{\sum_{i=1}^{nb} \varepsilon_{B_i}^2}{nb} + \frac{\sum_{i=1}^{nb} (T_{B_i} - T_D)^2}{nb}$$

**onde:** nb é o número de blocos do depósito;  $\sigma_{B_i}^2$  é a variância do i-ésimo bloco obtido pelas;  $T_{B_i}$  é o teor médio do i-ésimo bloco;  $T_D$  é o teor médio do depósito, que pode ser obtido como:

$$T_D = \frac{1}{D} \sum_{i=1}^{nb} V_i T_{B_i}$$

**onde:**  $V_i$  é o volume do i-ésimo bloco, o volume do depósito, portanto, é igual a:

$$D = \sum_{i=1}^{nb} V_i$$

As variâncias parciais dos blocos, podem ser obtidas pela seguinte equação:

$$\varepsilon_B^2 \sum_{i=1}^n \lambda_i (T_i - T_B)^2$$

onde:  $T_i$  é o teor da i-ésima amostra de furo vizinho; n é o número de amostras de furos vizinhos. Observe-se que a equação para cálculo da variância parcial do bloco não é a mesma utilizada para cálculo da variância de krigagem, pois esta não permite a composição como se apresenta aqui, visto que o Teorema da Combinação das Estimativas de Krigagem não se aplica para o depósito todo, uma vez que é impossível utilizar o mesmo conjunto de amostras para todos os blocos do depósito, além do variograma estar definido no máximo até o campo geométrico. A equação de cálculo da variância parcial do bloco é a que determina a variância de interpolação, que é uma medida da dispersão dos teores de amostras vizinhas em relação ao teor médio do bloco. O primeiro termo da equação de composição de variâncias é a variância dentro dos blocos, ou seja, a variância devida aos blocos propriamente dito, e o segundo termo é a variância entre os sub-blocos, devida à diferença de teores entre os blocos e o teor médio do depósito, que corresponde à parcela da variância pelo aumento de suporte, mantendo-se a relação volume-variância constante. Observe que a variância dentro dos blocos é eliminada, se não totalmente em grande parte, pelo método de eliminação, por meio das operações de desmonte, britagem e homogeneização, restando somente a variância entre blocos, ou seja, a dispersão de teores esperada em relação ao teor médio do depósito, que depende somente da dimensão do bloco e do método de mineração.

**COMPOSIÇÃO POR BANCADAS:** É feita pelo agrupamento das amostras dos furos de sondagem, segundo a altura das bancadas. Este tipo de composição é indicada para se fazer a avaliação de reservas em depósitos, cuja lavra se dará a céu aberto, onde o interesse está voltado para o teor médio da bancada a ser desmontada.

**COMPOSIÇÃO POR ELEVAÇÕES:** É feita por meio do agrupamento de amostras para alguns intervalos de elevações selecionados, segundo as características geológicas, mineralógicas ou estruturais do depósito.

**COMPOSIÇÃO POR TIPO LITOLÓGICO:** É obtida pelo agrupamento de amostras conforme a espessura de uma ou conjunto de litologias que suporta a mineralização, sendo indicada para depósitos estratiformes ou localizados em veios.

**COMPOSIÇÕES LINEARES:** Ver variáveis estatísticas canônicas.

**COMPRADORES MISTERIOSOS:** Pessoas contratadas para agir como consumidores e fazer compras nos concorrentes do empregador ou em suas próprias lojas para comparar preços, displays e coisas de gênero.

**CO-MORBIDADE:** Agravo à saúde de que o paciente é portador, excluído de consideração o agravo à saúde objeto de investigação.

**COMUNALIDADE:** Quantia total de variância que uma variável original compartilha com todas as outras variáveis incluídas na análise.

**CONCEITO:** Ideia ou imagem usada para resumir e representar um conjunto de objetos, experiência ou pensamentos. Por exemplo, o conceito cadeira representa um conjunto variado de objetos. Conceitos típicos das ciências sócias são classe alta e alienado.

**CONCEITO:** Pode ser definido como um nome ou símbolo dado a um grupo de objetos, pessoas, ou eventos tendo algo em comum. Exemplos: mobília é um conceito que se aplica a cadeira, mesa, armário, dentre outros. Emoções é o conceito para alegria, felicidade, afeição, ira, dentre outros.

**CONCEITO DE MARKETING:** Filosofia de negócios baseada na orientação para o consumidor, orientação de metas e orientação de sistemas.

**CONCEITO OU VARIÁVEL NÃO OBSERVADOS:** Ver construto ou variável latentes.

**CONCEITOS:** São denominações de classes que podem referir a objetos, eventos, ou propriedades de eventos.

**CONCEITUAÇÃO:** Termo que significa o processo mental por meio do qual noções vagas e imprecisas que são os conceitos se tornam mais específicas e precisas. Imagine que você quer estudar preconceito. O que quer dizer preconceito? Há tipos diferentes de preconceito? Quais são?.

**CONCEITUAÇÃO:** Refere-se ao processo de frasar um problema ou um evento que nos preocupa, dentro dos conceitos e leis de uma teoria geral já bem estabelecida. Requer algum talento para relacionar o efêmero e concreto ao duradouro e mais abstrato.

**CONCLUSÕES:** Generalizações que respondem às perguntas levantadas pelos objetivos da pesquisa ou que de outra forma satisfazem aos objetivos.

**CONCORDÂNCIA DE KENDALL (W):** Teste de associação entre k variáveis dispostas em vários conjuntos de postos e mensuradas a nível ordinal.

**CONDIÇÃO DE ORDEM:** Exigência para identificação de que os graus de liberdade do modelo devem ser maiores ou iguais a zero.

**CONDIÇÃO DE NÃO ENVIESAMENTO DA KRIGAGEM ORDINÁRIA:** Tem origem na imposição que as médias real e estimada sejam iguais:  $E[Z_v] = E[Z_v^*]$ , fazendo  $E[Z_v] = m$  (supondo estacionaridade) e tendo que:

$$E[Z_v^*] = E\left[\sum_{i=1}^n \lambda_i Z(X_i)\right] = \sum_{i=1}^n \lambda_i E[Z(X_i)] = E[Z(X_i)] \sum_{i=1}^n \lambda_i = m \sum_{i=1}^n \lambda_i,$$

como  $E[Z_v] = E[Z_v^*] = m$ , a condição de não enviesamento do estimador  $Z_v^*$  fica:  $\sum_{i=1}^n \lambda_i = 1$ .

**CONDIÇÃO DE ORDENAÇÃO:** Exigência para identificação que cada parâmetro estimado seja definido algebricamente. Muito mais trabalhosa do que a condição de ordem ou estimação empírica de identificação.

**CONDIÇÃO NECESSÁRIA, E SUFICIENTE:** Há que se fazer uma distinção entre uma condição necessária para a verdade de uma afirmativa, e uma que é suficiente. A condição de que um número termine em zero é suficiente para provar que ele é divisível por 5, mas não é necessário. Por outro lado, a condição de que um número seja par é uma necessária condição para sua divisibilidade por 6, mas não é suficiente. É necessário e suficiente que um número seja par para seu quadrado ser par.

**CO-NEGATIVIDADE:** Substituto para especificidade, quando o padrão de referência para a doença não é o diagnóstico de certeza.

**CONFEDERADOS:** Grupo de sujeitos que atuam em convivência com o pesquisador para estudar reações de um ou mais inocentes. Estudos que envolvem confederados empregam manipulações sociais. São empregados em estudos de processo perceptual. Um bom exemplo é o experimento de Asch (1948). O experimento foi construído de tal forma que, em cada tentativa a um problema, o inocente respondia após os confederados, que em vários exemplos escolhiam claramente a alternativa incorreta.

**CONFIABILIDADE:** O mesmo que variável de confundimento ou confusão de variáveis.

**CONFIABILIDADE:** Extensão em que uma variável ou um conjunto de variáveis é consistente com o que se pretende medir. Se medidas repetidas forem executadas, as medidas confiáveis serão consistentes em seus valores. É diferente de validade, por se referir não ao que deveria ser medido, mas do modo como é medido.

**CONFIABILIDADE:** Grau em que uma variável ou conjunto de variáveis é consistente com o que se pretende medir. Se múltiplas medidas são realizadas, as medidas confiáveis serão muito consistentes em seus valores. É diferente de validade, no sentido em que não se relaciona com o que deveria ser medido, mas com o modo é medido.

**CONFIABILIDADE:** Qualidade de métodos de medida que sugere que os mesmos dados teriam sido coletados cada vez em repetidas observações do mesmo fenômeno. No contexto de um survey, esperamos que a pergunta você foi a igreja na semana passada? teria mais confiabilidade que a pergunta quantas vezes na vida você já foi à igreja? Não confundir com validade.

**CONFIABILIDADE:** Grau em que um conjunto de indicadores de construtos latentes são consistentes em suas mensurações. Em termos mais formais, confiabilidade é a extensão em que dois ou mais indicadores compartilham em sua mensuração de um construto. Os indicadores de construtos altamente confiáveis são altamente inter-correlacionados, indicando que eles estão todos medindo o mesmo construto latente. Quando a confiabilidade diminui, os indicadores se tornam menos consistentes e assim são indicadores mais pobres de construto latente. Confiabilidade pode ser computada como 1,0 menos o erro de mensuração. Ver também validade.

**CONFIABILIDADE:** Este termo é usado como a probabilidade de sobrevivência após o tempo  $t_n$  ou seja,  $1 - F(t)$  onde  $F(t)$  é a função de distribuição dos tempos de vida. Outras definições também estão presentes nas áreas de ensaios biológicos e testes educacionais. Em inglês Reliability.

**CONFIABILIDADE:** Medidas consistentes de uma administração para a seguinte.

**CONFIABILIDADE (RELAÇÃO ENTRE VARIÁVEIS):** É um conceito muito menos intuitivo, mas extremamente importante. Relaciona-se à representatividade do resultado encontrado em uma amostra específica de toda a população. Em outras palavras, diz quão provável será encontrar uma relação similar se o experimento fosse feito com outras amostras retiradas da mesma população, lembrando que o maior interesse está na população.

**CONFIABILIDADE DE CONSISTÊNCIA INTERNA:** A habilidade de se produzir resultados semelhantes com o uso de amostras diferentes para medir um fenômeno durante o mesmo período de tempo.

**CONFIABILIDADE DE FORMA EQUIVALENTE:** A habilidade de se produzir resultados semelhantes usando-se dois ou mais instrumentos, os mais parecidos possíveis, para medir a mesma coisa.

**CONFIABILIDADE DO TESTE-RETESTE:** A habilidade de o mesmo instrumento produzir resultados consistentes quando usado uma segunda vez sob condições as mais semelhantes possíveis.

**CONFIANÇA:** A proporção de vezes em que dois ou mais eventos ocorrem conjuntamente. Por exemplo, se o produto A é comprado 60% das vezes em que um cliente também compra o produto B, temos uma confiança de 60% ao dizer que os produtos A e B são comprados juntos.

**CONFUNDIDOR:** O mesmo que variável de confundimento ou confusão.

**CONFUNDIMENTO:** Técnica pela qual se confunde o efeito de blocos com o de interações de pouco interesse, tendo em mira o uso de blocos menores em ensaios fatoriais.

**CONFUNDIMENTO:** Situação em que os efeitos de duas variáveis, sobre uma teoria, não estão separados, por exemplo, na avaliação exposição-doença como sedentarismo e incidência de coronariopatias, um outro fator como a idade, além da exposição em questão o sedentarismo, está determinado parte do efeito, ou seja, é responsável também por parte da incidência de coronariopatias. Em consequência, a idade funciona como variável de confundimento ou de confusão, pois perturba o estudo da associação entre sedentarismo e coronariopatias.

**CONFOUNDING:** Termo em inglês que significa confundimento ou confusão de variáveis.

**CONFUSÃO DE VARIÁVEIS:** ver confundimento.

**CONGLOMERAÇÃO:** Em epidemiologia, nome dado ao processo de agrupamento dos casos individuais de determinado agravo à saúde, de acordo com o lapso de tempo decorrido entre decorrido entre o evento supostamente causal e a manifestação mórbida em estudo, e o local em que concorreram.

**CONGLOMERADO:** Grupo de unidades amostrais com disposição espacial sistemática onde a locação aleatória é realizada para o grupo e não para as unidades amostrais individualmente.

**CONGLOMERADO [CLUSTER]:** Agrupamento natural de unidades, usado em amostragem por conglomerado em múltiplas etapas. Exemplos: nas faculdades há conglomerados de alunos; em bairros há conglomerados de residências.

**CONGLOMERADO DE CASOS OU DE ÓBITOS:** Um conjunto de casos ou de óbitos para os quais poder-se-ia hipotetizar origem idêntica, seja a ação de uma substância química, de um agente infeccioso, a retirada de um fator ambiental e, até mesmo, os modos de vida.

**CONGLOMERADO ESPACIAL DE CASOS:** Casos de doença de etiologia conhecida ou desconhecida, com doentes exibindo sintomas e sinais iguais, para os quais pode ser suspeitada ou evidenciada uma origem idêntica, ou mesmo comum, associada a algum fator ou fatores surgidos em um território circunscrito cujos limites possam ser perfeitamente definidos.

**CONGLOMERADO TEMPORAL DE CASOS:** Um grupo de casos para os quais se suspeita de um fator comum e que ocorrem dentro dos limites de intervalos de tempo significativamente iguais, medidos a partir do evento que supostamente lhes deu origem.

**CONHECIMENTO CIENTÍFICO:** É o conhecimento racional, sistemático, exato e verificável da realidade. Sua origem está nos procedimentos de verificação baseados na metodologia científica. Podemos então dizer que o conhecimento científico:

- É racional e objetivo.
- Atém-se aos fatos.
- Transcende aos fatos.
- É analítico.
- Requer exatidão e clareza.
- É comunicável.
- É verificável.
- Depende de investigação metódica.
- Busca e aplica leis.
- É explicativo.
- Pode fazer previsões.
- É aberto.
- É útil.

**CONHECIMENTO CIENTÍFICO:** Para o cientista, conhecimento é público, no sentido de que é comunicável de alguma maneira a outros: objetivo, porque abrange todos os fatos existentes; verificável, pois implica observações factíveis e concretas, sistemáticos e causais entre eventos. O conhecimento científico é também conceptual, uma característica que ele compartilha com outros tipos de conhecimentos não-científicos.

**CONHECIMENTO EMPÍRICO (OU CONHECIMENTO VULGAR):** É o conhecimento obtido ao acaso, após inúmeras tentativas, ou seja, o conhecimento adquirido por meio de ações não planejadas.

**CONHECIMENTO FILOSÓFICO:** É fruto do raciocínio e da reflexão humana. É o conhecimento especulativo sobre fenômenos, gerando conceitos subjetivos. Busca dar sentido aos fenômenos gerais do universo, ultrapassando os limites formais da ciência.

**CONHECIMENTO PRÁTICO:** Conhecimento é no máximo um fator de grande número de fatores de input em qualquer situação prática. Nenhum problema prático pode ser resolvido usando apenas conhecimento, tem-se que considerar uma imensidade de outros fatores, inclusive os econômicos, os sociais, e os motivacionais.

**CO-HOMOLOGIA:** A co-homologia calcula invariantes algébricos de espaços topológicos que são formalmente duais para a homologia. Os invariantes obtidos são em geral mais poderosos do que aqueles obtidos pela homologia e normalmente apresentam mais estruturas algébricas. Em inglês cohomology.

**CÔNICA:** Uma cônica ou então secção cônica é uma das três seguintes curvas: parábola, hipérbole ou elipse que podem ser obtidas pela interseção de um plano com um cone duplo. Em inglês conic.

**CONJUNTO:** Grupo de objetos semelhantes e considerados de acordo com alguma propriedade em comum. Em inglês set.

**CONJUNTO:** Uma coleção bem definida de objetos.

**CONJUNTO ESCOLHA:** Conjunto de estímulos de perfil completo construído por meio de princípios de delineamento experimental e usado na abordagem baseada em escolha.

**CONJUNTO FECHADO DE ESTADOS:** É dito o conjunto  $C \subset R$  de estados  $E_i$  de uma cadeia se, sendo  $P_{ij}^{(n)} = P$  a probabilidade de transição superior de  $E_i$  para  $E_j$ , tiver-se que:

$$P = 0, \text{ se } E_i \in C \text{ e } E_j \in (R - C)$$

$$P > 0, \text{ se } E_i \text{ e } E_j \in C, \text{ para algum } n \geq 0$$

**CONJUNTO VAZIO:** Conjunto desprovido de elementos, e simbolizado pela letra grega fi maiúscula, ou seja,  $\Phi$ .

**CONJUNTO UNIVERSO:** Conjunto referência e que contém outros, simbolizado pela letra U. Em inglês universal set.

**CONHECIMENTO TEOLÓGICO:** Conhecimento revelado pela fé divina ou crença religiosa. Não pode, por sua origem, ser confirmado ou negado. Depende da formação moral e das crenças de cada indivíduo.

**CONSISTÊNCIA:** Propriedade que descreve o comportamento de um estimador quando o tamanho da amostra tende ao infinito. Em inglês consistency.

**CONSISTÊNCIA DE UM TESTE:** Um teste é dito consistente para um determinada alternativa, se o seu poder tende para 1 (um), quando o tamanho da amostra tende para o infinito ( $\infty$ ).

**CONSISTÊNCIA DO VARIOGRAMA EXPERIMENTAL:** Além do campo geométrico, aceito entre metade e um terço da extensão dos dados originais, deve-se considerar a consistência dos pontos do variograma experimental, pois como se sabe à medida que a distância aumenta diminui o número de pontos e, conseqüentemente, a dispersão dos últimos pontos. Assim, deve-se considerar para fins de cálculo do variograma médio, ao menos entre 30-50 pares de pontos.

**CONSTANTE DE REGRESSÃO (INTERCESSÃO):** Ponto onde a linha de regressão intercepta o eixo dos y's quando  $x = 0$ . É medida pela constante  $a$  ou  $b_0$  de uma equação de regressão.



**CONSTRUTO:** Conceito que o pesquisador pode definir em termos teóricos mas que não pode ser diretamente medido, como por exemplo, o respondente não pode articular uma única resposta que fornecerá no todo e de forma perfeita uma medida do conceito ou medido sem erro. Construtos são a base para a formação de relações causais, uma vez que eles são a mais pura representação possível de um conceito. Um construto pode ser definido em diversos graus de especificidade, variando de conceitos muito estreitos, tais como renda familiar total, até conceitos mais complexos ou abstratos, tais como inteligência ou emoções. Não importa qual o seu nível de especificidade, porém, um construto não pode ser medido direta e perfeitamente, mas deve ser aproximadamente medido por indicadores. Ver erro de mensuração.

**CONSTRUCTO:** Nas ciências físicas, constructos incluem massa, força, energia, próton, electrón e fóton. Em psicologia e educação, exemplos de constructos incluem inteligência, criatividade, habilidades para ler, autoconceituações, ansiedade. Constructos não podem ser observados diretamente, mas podem ser inferidos por meio do comportamento. São generalizações empíricas que resumiriam observações de fatos e dos relacionamentos entre eles.

**CONSTRUTO ENDÓGENO:** Construto ou variável que é a variável dependente ou de resultado em pelo menos uma relação causal. Em termos de um diagrama de caminhos, existem uma ou mais setas ou indicações que conduzem até o construto ou variável endógena.

**CONSTRUTO EXÓGENO:** Construto ou variável que atua simplesmente como preditor ou causa para outros construtos ou variáveis no modelo. Em diagramas de caminhos, os construtos exógenos têm apenas indicações causais que partem deles e não são previstos por quaisquer outros construtos no modelo.

**CONSTRUTO OU VARIÁVEL LATENTE:** Operacionalização de um construto em modelagem de equações estruturais. Uma variável latente não pode ser diretamente medida mas pode ser representada ou medida por uma ou mais variáveis que são os indicadores. Por exemplo, a atitude de uma pessoa em relação a um produto jamais pode ser medida precisamente a ponto de não haver incerteza, mas fazendo-se várias perguntas podemos avaliar muitos aspectos da atitude dessa pessoa. Em combinação, as respostas a tais questões fornecem uma medida razoavelmente precisa do construto latente ou atitude para um indivíduo.

**CONSULTA:** Acesso direto a uma base de dados solicitando informação específica. Pode ser usada para escavar a partir de dados agregados.

**CONTADORES DE TRÁFEGO:** Máquinas usadas para medir o fluxo de veículos em determinado trecho de estrada.

**CONTAGEM-z:** Um valor encontrado dividindo-se o desvio ao redor da média pelo desvio-padrão  $s$ . Uma contagem-z é definida como um valor padronizado e denota o número de desvios-padrões  $x_i$  ao redor da média.

**CONTAMINAÇÃO:** Influências que dificultam um teste adequado da hipótese de pesquisa.

**CONTAMINAÇÃO:** A inclusão, no teste de um grupo, de respondentes que normalmente não estão lá. Por exemplo, compradores externos que veem uma propaganda dirigida somente àqueles da área de testes e que entram nessa área para comprar o produto que está sendo testado.

**CONTAMINAÇÃO:** Presença de uma agente infeccioso na superfície do corpo, roupas, alimentos dentre outros. A aplicação, grupo-controle, de procedimentos reservados ao grupo de estudo.

**CONTAR HISTÓRIAS:** Os respondentes contam histórias para descrever suas experiências.

**CONTÉM:** Símbolo matemático dado por  $\supset$ .

**CONTINGÊNCIA C:** Teste de associação entre conjuntos de variáveis mensuradas a nível nominal, dispostas em tabelas de contingência l linhas x c colunas, isto é, com quaisquer números de categorias.

**CONTINGÊNCIA, TABELA DE:** Tabela que mostra quantas observações está em cada célula; onde as células representam todas as combinações possíveis de dois fatores; no teste qui-quadrado, usa-se uma tabela de contingência para determinar se dois fatores são independentes.

**CONTINUIDADE:** O grau de regularidade da mineralização é mostrado pelo comportamento do variograma próximo a origem. Assim, quanto ao grau de continuidade, podem ser reconhecidos quatro tipos de variogramas: i) alto grau de continuidade: quando o variograma descreve uma curva parabólica na origem, como o variograma de espessuras de camada; ii) continuidade média: quando o variograma é caracterizado por um comportamento linear na origem, típico de muitos depósitos metálicos; iii) efeito pepita: quando o variograma apresenta descontinuidade na origem; iv) efeito pepita puro: quando o variograma reflete a variação espacial de um fenômeno totalmente, sendo a variabilidade constante para qualquer distância. Esse termo tem origem na mineração de ouro, onde a inclusão de uma pepita de ouro em uma pequena amostra de um testemunho de sondagem é um evento aleatório.

**CONTINUIDADE DO CORPO DE MINÉRIO:** Refere-se a extensão do corpo de minério, interpretada a partir dos dados de pesquisa e desenvolvimento disponíveis, para fins de indicação ou inferência de reservas.

**CONTRADOMÍNIO DA FUNÇÃO (IMAGEM DA FUNÇÃO):** Dada uma função  $f: A \rightarrow B$  a imagem de  $f$ , que representamos por  $\text{Im}(f)$  ou por  $f(A)$ , é o conjunto dos elementos  $b \in B$  para os quais existe  $a \in A$  satisfazendo  $f(a)=b$ . Simbolicamente tem-se o seguinte:  $\text{Im}(f) = f(A) = \{b \in B : \exists a \in A, b = f(a)\}$ . O conjunto de chegada B, também possui um sinônimo, é chamado de contradomínio.

**CONTRASTE:** Procedimento para investigar diferenças de grupos específicos de interesse em conjunção com a análise de variância (ANOVA) e a análise de variância multivariada (MANOVA), por exemplo, comparar diferenças de médias de grupos para um par específico de grupos.

**CONTROLE:** Um delineamento de pesquisa tem duas finalidades básicas: primeira, e mais óbvia, dar respostas a perguntas de pesquisa e, segunda, controlar variância ou variabilidade. Há três tipos de variância que devem ser controladas: i) maximize a variância experimental: os efeitos sistemáticos da variável associada com as hipóteses da pesquisa; ii) Controle a variância exógena: os efeitos de outras variáveis sistemáticas, porém indesejáveis que podem influenciar os resultados experimentais, mas que não são objeto de estudo; iii) Minimizar a variância ou erro: os efeitos de flutuações randômicas, incluindo os chamados erros de mensuração.

**CONTROLE:** Há duas formas principais em que este termo é usado em estatística. Se um processo produz uma série de dados sob o que são essencialmente as mesmas condições e as variações internas são aleatórias, então o processo pode ser considerado estatisticamente sob controle. Cada observação isolada é, na verdade, proveniente de uma população distribuída de acordo com alguma lei estatística estabelecida. O segundo uso preocupa-se com os testes de um novo método, processo ou fator contra um padrão aceito. Esta parte do teste que envolve a comparação do padrão é conhecida como controle. Em inglês Control.



**CONTROLE:** Quando aplicado a doenças transmissíveis e algumas não-transmissíveis, significa operações ou programas desenvolvidos com o objetivo de reduzir sua incidência e/ou prevalência, ou eliminá-las.

**CONTROLE:** Ato de regular ou fiscalizar; conjunto de ações para prevenir, reduzir ou eliminar doenças e agentes mórbidos através de programa de controle de doenças; grupo-testemunha; remoção do efeito de uma ou mais variáveis que confundem a interpretação.

**CONTROLE:** Uma série de atividades destinadas a reduzir a prevalência de uma agravo até alcançar um determinado nível que não mais constitua problema de saúde pública.

**CONTROLE (EM EXPERIMENTO):** Espécimes similares num teste, que são sujeitos tanto quanto possível aos mesmos tratamentos aplicados aos objetos do experimento, exceto a variável experimental.

**CONTROLE DE PROCESSO ESTATÍSTICO (CPE):** Utilização de processos estatísticos, como cartas de controle, para analisar um processo ou seus resultados, a fim de tomar medidas apropriadas para atingir e manter uma situação de controle estatístico e melhorar a capacidade do processo.

**CONTROLE DE PROJETO:** Uso de design experimental para controlar fatores casuais estranhos.

**CONTROLE DE CHARLIER:** O controle de Charlier tem por finalidade verificar a exatidão do cálculo efetuado da média e da variância ou do desvio padrão principalmente quando esses cálculos são realizados por meio de processos breves ou abreviados ou desvio em classes de cálculo de determinação dessas medidas em uma tabela de frequência. O controle de Charlier, para o cálculo dos momentos pelo processo abreviado ou desvios em classe consiste ou utilizam-se as seguintes identidades.

$$\sum f(d_i + 1) = \sum fd_i + \sum f = \sum fd_i + n$$

$$\sum f(d_i + 1)^2 = \sum f(d_i^2 + 2d_i + 1) = \sum fd_i^2 + 2\sum fd_i + \sum f = \sum fd_i^2 + 2\sum fd_i + n.$$

$$\sum f(d_i + 1)^3 = \sum fd_i^3 + 3\sum fd_i^2 + 3\sum fd_i + n$$

$$\sum f(d_i + 1)^4 = \sum fd_i^4 + 4\sum fd_i^3 + 6\sum fd_i^2 + 4\sum fd_i + n$$

**CONTROLE DE QUALIDADE:** Uma série de inspeções e medidas que determinam se os padrões da qualidade estão sendo atingidos.

**CONTROLE ESTATÍSTICO:** Ajuste para efeito de confundir variáveis ajustando-se estatisticamente o valor da variável dependente para cada condição de tratamento.

**CONTROLE EXTERNO (HISTÓRICO, NÃO-PARALELO, NÃO-SIMULTÂNEO, NÃO-CONCOMITANTE OU NÃO-CONCORRENTE):** Grupo de indivíduos cujos dados foram coletados, em geral, antes dos do grupo de estudo, dados estes usados como padrão para comparação dos resultados.

**CONTROLE INTERNO (PARALELO, SIMULTÂNEO, CONCOMITANTE OU CONCORRENTE):** Grupo de indivíduos formado durante o próprio desenrolar da investigação; é encontrado na investigação controlada. Exemplo, estudo de caso-controle.

**CONTROLE FÍSICO:** Manter o valor ou o nível de variáveis estranhas constante durante o curso de um experimento.

**CONTROLE NEGATIVO:** Tratamento sem qualquer efeito farmacológico ou fisiológico, isto é, placebo ou pseudoprocimento. Veja controle positivo.

**CONTROLE POSITIVO:** Normalmente o tratamento padrão, mas sempre um tratamento que envolve o uso de uma substância farmacologicamente ativa. Veja controle negativo.

**CONTROLE SOCIAL:** Participação popular no planejamento e na avaliação das ações do governo. Na área da saúde essa participação se traduz no controle pela população das ações do Sistema Único de Saúde (SUS), por meio de Conselhos de Saúde.

**CONVERGÊNCIA EM PROBABILIDADE:** Uma sequência de variáveis aleatórias  $\{X_n, n \geq 1\}$  em um espaço de probabilidade  $(\Omega, F, P)$  converge em probabilidade para  $X, X_n \xrightarrow{p} X$ , se existe uma variável aleatória (v.a.)  $X$  em  $(\Omega, F, P)$  tal que  $P_r\{|X_n - X| > \varepsilon\} \rightarrow 0, \forall \varepsilon > 0, \text{ quando } n \rightarrow \infty.$

**CONVERGÊNCIA QUASE CERTA:** Uma sequência de variáveis aleatórias  $\{X_n, n \geq 1\}$  em um espaço de probabilidade  $(\Omega, F, P)$  converge em probabilidade para  $X, X_n \xrightarrow{p} X$ , se existe uma variável aleatória (v.a.)  $X$  em  $(\Omega, F, P)$  tal que  $P_r\{|X_n - X| > \varepsilon\} \rightarrow 0, \forall \varepsilon > 0, \text{ quando } n \rightarrow \infty.$  E, se existir uma variável aleatória  $X$  em  $(\Omega, F, P)$  tal que  $P_r\{\lim_{n \rightarrow \infty} X_n = X\} = 1$  teremos convergência quase certa  $(X_n \xrightarrow{q.c.} X).$

**CONVERSÃO ALIMENTAR (EFICIÊNCIA ALIMENTAR):** Unidade de peso ganha por unidade de alimento consumida. Também pode-se considerar a produção de carne, leite, ovos, lã, dentre outras por unidade de alimento consumida.

**COOKIE:** Um arquivo de texto colocado no computador do usuário a fim de identificá-lo quando ele revisita o site.

**COORTE:** Grupo de pessoas com alguma característica em comum; por exemplo, crianças nascidas em determinado ano.

**COORTE:** Grupo de pessoas claramente definido que é estudado durante um período de tempo para determinar a incidência de morte, doença ou acidente.

**COORTE:** Grupo claramente definido de pessoas estudado ao longo do tempo.

**COORTE:** Grupos bem definidos de uma população, possuindo alguma característica em comum e cujos indivíduos permanecem no conjunto durante determinado tempo, registrando-se e avaliando-se as ocorrências havidas entre os elementos no período considerado.

**COORTE:** Qualquer grupo definido de pessoas selecionadas, tendo em vista um propósito ou estudo especial.

**COORTE (COHORT):** Subseção de uma população com um fator em comum, geralmente a idade. Exemplo: todos os indivíduos nascidos no Rio Grande do Norte em 2015.

**COORTE DE INÍCIO (INCEPTION COHORT):** Grupo de pessoas com características comuns, por exemplo, pacientes portadores de uma dada doença utilizados em investigação sobre prognóstico desta doença, reunidas em fase inicial dos transtornos clínicos, tal como na época de exposição à causa suspeita ou no momento do diagnóstico, e que são seguidas por um certo tempo de modo a conhecer os defeitos clínicos.

**CO-POSITIVIDADE:** Substituto para sensibilidade, quando o padrão de referência para doença não é o diagnóstico de certeza.

**CORPO:** Um corpo é um anel em que cada elemento não-nulo tem um inverso multiplicativo. Entre as estruturas deste tipo mais importantes estão os números reais,  $R$ , os números complexos,  $C$  e vários corpos finitos que são utilizados na teoria dos números em Combinatória. Em inglês field.

**CORREÇÃO DA AMOSTRAGEM FINITA:** É o fator dado por  $\frac{[N-n]}{[N-1]}$ , pelo qual a variância  $Var(\bar{X}) = \frac{\sigma^2}{n}$

da média aritmética da amostra de tamanho  $n$ , extraída com reposição de uma população de tamanho finito  $N$  deve ser multiplicada quando a amostragem é feita sem reposição. Não se trata, na realidade, de uma correção, mas, sim, de uma relação.

**CORREÇÃO DA ANISOTROPIA GEOMÉTRICA:** A anisotropia geométrica é uma feição frequentemente encontrada em modelos de variogramas de depósitos minerais. Portanto, encontrado um modelo de variograma que apresente anisotropia, esta deve ser corrigida para cálculo do sistema de equações de krigagem para correta estimativa da variável de interesse por meio dos ponderadores resultantes da solução do referido sistema. A correção da anisotropia geométrica se dá aplicando o ângulo do eixo de anisotropia e a razão de anisotropia à equação do cálculo da distância corrigida para a qual é determinada a semi-variância ou covariância. Para a correção da anisotropia deve-se convencionar que o ângulo de anisotropia, igual a  $q$ , é positivo quando medido no sentido anti-horário a partir do eixo das abscissas e a razão de anisotropia é  $K$ , dada pela divisão da maior amplitude pela menor (sempre  $> 1$ ). A equação para cálculo da distância corrigida da anisotropia geométrica é:

$$h' = \sqrt{[(X_1 - X_2) \cos \theta + (Y_1 - Y_2) \sin \theta]^2 + K^2 [-(X_1 - X_2) \sin \theta + (Y_1 - Y_2) \cos \theta]^2}$$

Observe-se que se não houvesse anisotropia geométrica ( $q = 0$  e  $K = 1$ ), a equação anterior resumiria-se a:

$$h = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2}, \text{ ou simplesmente a distância euclidiana entre dois pontos quaisquer } (x_1, y_1) \text{ e } (x_2, y_2).$$

**CORREÇÃO DA CONTINUIDADE DO TESTE DE QUI QUADRADO ( $\chi^2$ ):** Em uma tabela quádrupla de contingência ou  $2 \times 2$  frequência hipotética é  $\tilde{n}_{ij} = \frac{1}{N} n_i n_j$  (hipótese de independência), consiste em

subtrair 0,5 do valor absoluto,  $|n_{ij} - \tilde{n}_{ij}| = d_{ij}$ , das diferenças entre frequência observada e frequência teórica, antes de elevá-las ao quadrado, para, então, entrar na fórmula, ou seja, tomar

$$\chi^2 = \sum_{j=1}^2 \sum_{i=1}^2 \frac{(d_{ij} - 0,5)^2}{\tilde{n}_{ij}}. \text{ Aplica-se especialmente, quando existem frequências teóricas, ou hipotéticas iguais}$$

ou menos que 5.

**CORREÇÃO DE CONTINUIDADE:** Ajuste feito quando uma variável aleatória discreta é aproximada por uma variável aleatória contínua.

**CORREÇÃO POR CONTINUIDADE:** Ver Yates correction. Em inglês Continuity correction.

**CORREÇÃO POR CONTINUIDADE DE YATES:** É aquele tipo de correção que se usa quando se aproxima a distribuição de probabilidade de uma variável aleatória discreta  $X_D$  por meio de uma distribuição

teórica ou especial de probabilidade de variável aleatória contínua à qual se associa a variável aleatória  $X_C$ , e assim é necessário efetuar uma correção de continuidade de mais ou menos um meio  $\left[\pm \frac{1}{2}\right]$ , para o

cálculo de probabilidades. Efetua-se então a correção seguinte:

$$P\left(X_D \leq x\right); P\left(X_C \leq x + \frac{1}{2}\right); P\left(X_D < x\right); P\left(X_C \leq x - \frac{1}{2}\right) \text{ e } P\left(X_D = x\right); P\left(x - \frac{1}{2} \leq X_C \leq x + \frac{1}{2}\right)$$

**CORREÇÃO DE POPULAÇÃO FINITA:** Ajuste do erro padrão para levar em conta populações de pequeno tamanho.

**CORREÇÃO DE SHEPPARD:** Ajuste usado ao se calcular o desvio padrão de uma distribuição de frequência com os dados agrupados em intervalos de classes. Esse processo conduz a um certo erro ou qual denomina-se erro de agrupamento. Adota-se o ponto médio de cada classe como sendo o de todos os valores da classe, ou seja que os dados se distribuem simetricamente dentro de cada classe. Isso, segundo alguns estatísticos, pode nos conduzir a erros de agrupamentos. Os erros de agrupamentos é óbvio, aumentam o valor da variância e do desvio padrão, e serão tanto maiores quanto mais amplos forem os intervalos de classe e menor o conjunto total de observações ou dos dados. Para corrigir os possíveis erros de agrupamentos utiliza-se uma fórmula criada por Sheppard, e que é a seguinte:

$$\sigma_c = \sqrt{\sigma^2 - \frac{c^2}{12}}$$

Onde, os termos significam que.

$\sigma_c$  = ao desvio padrão corrigido;  $\sigma$  = desvio padrão; e

$c$  = amplitude do intervalo de classe.

O termo corretivo  $\frac{c^2}{12}$ , que é subtrativo, é denominado correção de Sheppard, e é usado para distribuições de variáveis contínuas, cujas extremidades tendem gradualmente para zero em ambas as direções.

Não há unanimidade entre os estatísticos sobre a conveniência ou não da aplicação da correção de Sheppard. Os pesquisadores da linha matemática recomendam aplica-la somente quando conhecemos os dados primitivos, ou seja, um exame completo da situação, o que nos permitirá um exame profundo da situação da distribuição de valores. No entanto é preferível não aplica-la, pois raramente se tem acesso aos dados primitivos, e a aplicação da correção de Sheppard pode, não somente, supercorrigir o valor da variância, como também substituir erros existentes por novos erros.

As correções de Sheppard para os momentos são as seguintes:

$$m_2 \text{ corrigido} = m_2 - \frac{1}{12} c^2;$$

$$m_4 \text{ corrigido} = m_4 - \frac{1}{2} c^2 m_2 + \frac{7}{240} c^4.$$

Os momentos de primeira e terceira ordens,  $m_1$  e  $m_3$  respectivamente não necessitam de correção.

**CORREÇÃO DE WILLIAMS:** Correção do teste G para obter melhor aproximação com o teste do Qui-Quadrado.

**CORREÇÃO DE YATES:** Correção de continuidade do teste do Qui-Quadrado, onde se subtrai 0.5 do numerador de cada termo, desde que o grau de liberdade seja igual à unidade (1).

**CORREÇÃO DE YATES PARA A CONTINUIDADE:** Ajuste do valor do qui-quadrado recomendado quando as contagens na tabela de contingência são pequenas, porque a distribuição binomial é descontínua.

**CORREÇÃO GEOMÉTRICA DE FINNEY:** O valor da média geométrica, obtida pela exponenciação da média logarítmica, tende a subestimar, dependendo principalmente da variância amostral, a média aritmética dos mesmos dados, sendo necessário proceder à correção daquela por um fator denominado correção geométrica de Finney. O fator de correção geométrica de Finney, pode ser determinado, pela aproximação da série cujos primeiros termos são:

$$\varphi_n(t) = 1 + \frac{n-1}{n}t + \frac{(n-1)^3 t^2}{n^2(n+1)2!} + \frac{(n-1)^5 t^3}{n^3(n+1)(n+3)3!} + \dots$$

onde:  $t = \frac{S^2}{\mu}$ , ou seja, a metade da variância dos logaritmos dos dados;

n é o número total de dados.

**CORREÇÃO DE BESSEL:** O mesmo que grau de liberdade.

**CORREÇÃO POR POPULAÇÃO FINITA:** Quando o tamanho da população N é conhecido, é necessário corrigir a variância do estimador de um parâmetro multiplicando-a pelo fator dado por  $\frac{N-n}{N-1}$ , que é chamado fator de correção por população finita. Em inglês Finite population correction.

**CORREGIONALIZAÇÃO:** Fenômeno ou resultados de ensaios regionalizado representado por variáveis correlacionadas entre si.

**CORRELAÇÃO:** Variação concomitante de dois fenômenos. Da correlação não se infere necessariamente causalidade. Em inglês correlation.

**CORRELAÇÃO:** Indicação do grau de associação entre duas grandezas; seu valor está sempre entre -1 e 1.

**CORRELAÇÃO:** A extensão de relacionamento entre duas variáveis.

**CORRELAÇÃO:** É a relação ou dependência entre as duas variáveis de uma distribuição bidimensional.

**CORRELAÇÃO:** Define a dependência mútua entre duas variáveis aleatórias X e Y, podendo ser medida por meio da covariância ou do coeficiente de correlação linear de Pearson.

**CORRELAÇÃO:** Relação mútua entre dois eventos; a presença de um prediz ou favorece a ocorrência do outro; a correlação simples mede a força do relacionamento de duas variáveis; a correlação múltipla envolve três ou mais variáveis.

**CORRELAÇÃO:** Teste de associação entre variáveis, medindo-se a magnitude e o sentido ou somente o grau dessa correlação, conforme o teste empregado. Não há qualquer dependência funcional de uma variável em relação à outra.

**CORRELAÇÃO:** Um termo geral utilizado para descrever o fato de que duas ou mais variáveis ou seja se os conjuntos de dados estão relacionados. Galton (1869), foi provavelmente o primeiro a utilizar o termo

com este sentido. O termo é utilizado mais precisamente para denominar relacionamento linear entre dois conjuntos de dados ou variáveis. Em inglês Correlation.

**CORRELAÇÃO:** É a relação ou dependência entre as duas variáveis de uma distribuição bidimensional.

**CORRELAÇÃO:** Método de análise usado para estudar a possível associação entre duas variáveis. Em inglês Correlation.

**CORRELAÇÃO:** Termo ambíguo usado para indicar o co-relacionamento ou correspondência entre variáveis; quando os valores de uma variável mudam, os valores da outra mudam padronizadamente. Fala-se de correlação positiva quando os valores das duas variáveis aumentam ou diminuem juntos e de correlação negativa quando os valores de uma variável aumentam enquanto os da outra diminuem. O termo é também usado com significado mais específico para se referir à correlação produto-momento r de Pearson.

**CORRELAÇÃO:** Índice do grau de covariação entre medidas pareadas (x e y) em um modo linear. É expresso como um coeficiente (r), com valores entre -1 e +1. Coeficientes de correlação não dão informação sobre relação causa efeito, e só podem ser confiáveis se a relação entre x e y for linear.

**CORRELAÇÃO:** Medida que serve para mensurar como duas características variam juntas. Uma correlação de +1,00 significa que quando uma característica aumenta, a outra também aumenta em uma relação positiva perfeita. Já uma correlação de -1,00 representa uma associação inversa perfeita entre características; enquanto uma aumenta a outra diminui. Uma correlação de 0,00 significa que não há nenhuma relação consistente entre as características; quando há aumento de uma delas, a outra pode aumentar ou diminuir. Os coeficientes de correlação podem variar entre +1,00 e -1,00.

**CORRELAÇÃO:** Medida de associação linear entre duas variáveis. É um valor que pode estar entre -1 e 1 e indica o quanto duas variáveis estão alinhadas. Representa-se pela letra grega  $\rho$  (rô). A correlação pode ser obtida dividindo-se a covariância pelo produto das variabilidades das duas variáveis consideradas. Assim  $\rho_{(x,y)} = \frac{Cov(X,Y)}{(\sigma_x \sigma_y)}$ . Em inglês correlation.

**CORRELAÇÃO AMBIENTAL:** Ver correlação interatributos.

**CORRELAÇÃO AMOSTRAL (r):** Serve para estimar a correlação linear populacional.

**CORRELAÇÃO BISSERIAL:** Medida de correlação empregada para substituir a correlação produto-momento quando uma variável medida em termos métricos é associada com urna medida binária não-métrica (0,1). Ver também correlação polisserial.

**CORRELAÇÃO CANÔNICA:** Análise multivariada que permite investigar o relacionamento entre duas séries de variáveis: um grupo de variáveis X e outro do grupo Y.

**CORRELAÇÃO CANÔNICA:** Em inglês canonical correlation.

**CORRELAÇÃO CANÔNICA:** O objetivo da correlação canônica é determinar uma combinação linear para cada grupo de variáveis dependentes e independentes que maximize a correlação entre os dois grupos. Funções adicionais não correlacionadas com a primeira podem ser calculadas, sendo limitadas pelo número de variáveis do menor grupo.



**CORRELAÇÃO CANÔNICA:** Medida da força das relações gerais entre as composições lineares que são as variáveis estatísticas canônicas para as variáveis independentes e dependentes. Com efeito, ela representa a correlação bivariada entre as duas variáveis estatísticas canônicas.

**CORRELAÇÃO CANÔNICA:** Tipo de análise que é vista como uma extensão lógica da análise de regressão múltipla. Vale lembrar que a análise de regressão múltipla envolve uma única variável dependente métrica e várias variáveis independentes métricas. Com análise canônica, o objetivo é correlacionar simultaneamente diversas variáveis dependentes métricas e diversas variáveis independentes métricas. A regressão múltipla envolve uma única variável dependente já a correlação canônica envolve múltiplas variáveis dependentes. O princípio subjacente é desenvolver uma combinação linear de cada conjunto de variáveis independentes e dependentes para maximizar a correlação entre os dois conjuntos. Em outras palavras, o procedimento envolve a obtenção de um conjunto de pesos para as variáveis dependentes e independentes que fornece a correlação simples máxima entre o conjunto de variáveis dependentes e o de variáveis independentes.

**CORRELAÇÃO CANÔNICA:** Em análise multivariada pode ser provado que as variáveis  $x_1, \dots, x_p$  e  $x_{p+1}, \dots, x_{p+q}$  podem ser transformadas linearmente em variáveis  $\lambda_1, \dots, \lambda_p$  e  $\lambda_{p+1}, \dots, \lambda_{p+q}$  de forma que (a) os membros de cada grupo não são correlacionados entre eles, (b) cada membro de um grupo é independente de todos exceto um membro do outro grupo e (c) as correlações diferentes de zero entre membros de grupos diferentes são maximizadas. Estas quantidades são chamadas correlações canônicas e os dois conjuntos de variáveis  $\lambda_1, \dots, \lambda_p$  e  $\lambda_{p+1}, \dots, \lambda_{p+q}$  são chamadas variáveis canônicas. Em inglês Canonical correlation.

**CORRELAÇÃO CURVILINEAR:** Variação concomitante de duas variáveis cuja representação típica é uma curva e não uma reta. Assim, a escores baixos na variável X correspondem escores altos na variável Y, a escores médios na variável X correspondem escores baixos na variável Y e escores altos na variável X correspondem escores altos na variável Y. O coeficiente próprio para fornecer o grau de correlação é rho e não r. Se for usado impropriamente o coeficiente r, o valor de r será uma subestimação de grau de correlação. Isto quer dizer que a correlação existente é maior do que o coeficiente possa sugerir. Em inglês Curvilinear correlation.

**CORRELAÇÃO DEFASADA:** Correlação entre duas séries, uma das quais está defasada com relação à outra. Em inglês Lag correlation.

**CORRELAÇÃO DE KENDALL:** Teste de associação entre duas variáveis mensuradas a nível ordinal ou postos, calculando-se a correlação pelo coeficiente de Kendall  $\tau$ .

**CORRELAÇÃO DE ORDEM DE CLASSIFICAÇÃO SPEARMAN:** Técnica de análise de correlação para uso com dados ordinais.

**CORRELAÇÃO DE PEARSON:** Usada para determinar o quanto o relacionamento é linear, isto é, o quanto a variação em uma medida é consistentemente acompanhada por uma variação direta ou inversa nas outras medidas.

**CORRELAÇÃO DE SPEARMAN:** Teste de associação entre duas variáveis mensuradas a nível ordinal ou postos, calculando-se a correlação pelo coeficiente de Spearman r.

**CORRELAÇÃO E ASSOCIAÇÃO (DIFERENÇA):** Correlação sugere dois fatores que crescem ou decrescem conjuntamente. Associação sugere dois fatores que ocorrem conjuntamente.

**CORRELAÇÃO EM SÉRIE:** Problema que surge na análise de regressão envolvendo dados de séries temporais, quando os valores sucessivos do termos erro aleatório não são independentes. É o mesmo que auto-correlação.

**CORRELAÇÃO ESPÚRIA (DISPARATADA OU SEM SENTIDO):** Esta é aquela correlação que apesar do valor do coeficiente obtido ser elevado, ela não faz sentido do ponto de vista prático e nem científico, como exemplo podemos citar a correlação linear simples direta e forte  $r = + 0,95$  entre o número de ovos postos por galinhas em uma granja no período de 30 dias e o consumo de gasolina em litros medidos nesse mesmo período e nessa localidade.

**CORRELAÇÃO FENOTÍPICA ENTRE CARACTERES (r):** Medida da magnitude da associação fenotípica entre caracteres em um indivíduo.

**CORRELAÇÃO FENOTÍPICA PERMANENTE ( $r_{fp}$ ):** Medida da magnitude da associação fenotípica entre caracteres em um indivíduo, a qual permanece constante de uma medição a outra, em um indivíduo.

**CORRELAÇÃO FRACA OU NULA:** Quando o diagrama de dispersão não permite o ajustamento de nenhuma reta, o que significa que  $r = 0$ . Diz-se, então, que não existe nenhuma relação entre as variáveis da distribuição bidimensional.

**CORRELAÇÃO GENÉTICA ENTRE CARACTERES ( $r_g$ ):** Medida da magnitude da associação genética entre caracteres em um indivíduo.

**CORRELAÇÃO GEOLÓGICA:** Semelhança litológica, estrutural e mineralógica de um depósito com relação a outro, ou entre partes de um mesmo depósito.

**CORRELAÇÃO INTERATRIBUTO:** Correlação entre atributos, também conhecida como correlação ambiental, que torna as combinações de atributos inacreditáveis ou redundantes. Uma correlação negativa descreve a situação na qual se considera que dois atributos naturalmente operam em diferentes direções, como potência e desempenho em termos de quilômetros percorridos por litro de gasolina. Quando um aumenta, o outro naturalmente diminui. Assim, por conta dessa correlação, todas as combinações desses dois atributos, por exemplo, alto desempenho e alta potência, são inacreditáveis. Os mesmos efeitos podem ser vistos para correlações positivas, nas quais talvez preço e qualidade são considerados como positivamente correlacionados. Pode ser inacreditável encontrar um produto de elevado preço e baixa qualidade em tal situação. A presença de fortes correlações interatributos exige que o pesquisador examine cuidadosamente os estímulos apresentados aos respondentes e evite combinações inacreditáveis que não são úteis na estimação das utilidades parciais.

**CORRELAÇÃO INTRACLASSE:** A correlação intraclasses é um aplicativo para testar replicabilidade amostral para dados contínuos, não sendo tratadas as variáveis como X e Y e sim como replicação 1 e replicação 2, portanto é uma medida de replicabilidade amostral.

**CORRELAÇÃO LINEAR:** A correlação linear perfeita, que corresponde aos valores -1 e +1, conforme seja negativa ou positiva, é medida pelo r de Pearson. Quando a correlação não é perfeita, os pontos de interseção no diagrama de dispersão se distribuem regra geral, de maneira a formar uma elipse. Em inglês linear correlation.



**CORRELAÇÃO LINEAR (PEARSON):** Teste de associação linear entre duas variáveis mensuradas a nível intervalar ou de razões, medindo-se o grau e a direção pelo coeficiente de correlação linear  $r$ .

**CORRELAÇÃO MOMENTO-PRODUTO DE PEARSON:** Técnica de análise de correlação para se utilizar com dados métricos.

**CORRELAÇÃO MÚLTIPLA:** É uma técnica de relacionamento destinada à manipulação de mais de duas variáveis. Descreve o relacionamento de uma variável e duas ou mais variáveis consideradas simultaneamente.

**CORRELAÇÃO MÚLTIPLA:** Neste caso o estudo envolve o cálculo sobre o grau de relação existente entre três ou mais variáveis.

**CORRELAÇÃO NEGATIVA FORTE:** Quando a reta de regressão, obtida a partir do diagrama de dispersão, tem declive negativo. A correlação é negativa quando  $r$  varia entre -1 e 0 e será tanto mais forte quanto  $r$  se aproxima de -1.

**CORRELAÇÃO NEGATIVA PERFEITA OU LINEAR:** Quando a reta de regressão, obtida a partir do diagrama de dispersão, tem declive negativo com  $r = -1$ .

**CORRELAÇÃO NOMINAL:** Neste caso o estudo envolve variáveis qualitativas nominais ou atributos dispostas em tabelas de contingência, e no caso da tabela de contingência ser de 2 linhas com 2 colunas (tabela 2x2) ela recebe o nome de correlação tetracórica. Por exemplo, o coeficiente de contingência ( $C$ ), a correlação de atributos ( $\gamma$ ).

**CORRELAÇÃO ORDINAL:** Correlação em que não se levam em conta os escores originais, mas sua classificação ou seus números de ordem. Em inglês rank correlation.

**CORRELAÇÃO ORDINAL:** Neste caso o estudo leva em consideração não os valores das variáveis envolvidas, mas os dados dispostos em ordem de tamanho, importância, ou outro critério semelhante. Como exemplo, podem ser citados o coeficiente de correlação ordinal de Sperman ( $r_s$ ), o coeficiente de correlação ordinal de Kendall ( $\tau$ ).

**CORRELAÇÃO PARCIAL:** É uma técnica de relacionamento destinada à manipulação de mais de duas variáveis. Permite ao estatístico descrever o relacionamento entre duas variáveis, controlando o relacionamento de uma outra variável.

**CORRELAÇÃO PARCIAL:** Teste de associação de duas variáveis quantitativas  $X$  e  $Y$  as quais são testadas juntamente com outra variável  $Z$ , a fim de se verificar se a correlação existente entre as primeiras é alterada pela presença da terceira variável introduzida.

**CORRELAÇÃO PARCIAL:** Neste caso o estudo visa verificar o grau de relacionamento entre duas variáveis, considerando uma terceira, uma quarta, ou uma quinta, sexta dentre outras. variável (is) constantes.

**CORRELAÇÃO PARCIAL BIVARIADA:** Correlação simples entre duas variáveis ou entre dois conjuntos de resíduos que são as variâncias inexplicadas que permanecem depois que a associação de outras variáveis independentes é removida.

**CORRELAÇÃO POLICÓRICA:** Medida de associação empregada como uma substituição para a correlação produto-momento quando ambas as variáveis são medidas ordinais com três ou mais categorias.

**CORRELAÇÃO POLISSERIAL:** Medida de associação usada como urna substituta para a correlação produto-momento quando uma variável é medida em uma escala ordinal e a outra variável é metricamente medida.

**CORRELAÇÃO POPULACIONAL ( $\rho$ ):** É um valor que mede o grau de relação linear entre duas variáveis quantitativas. É igual a covariância dividida pelo desvio padrão de cada uma das variáveis.

**CORRELAÇÃO POSITIVA FORTE:** Quando a reta de regressão, obtida a partir do diagrama de dispersão, tem declive positivo. A correlação é positiva quando  $r$  varia entre 0 e 1 e será tanto mais forte quanto  $r$  se aproxima de 1.

**CORRELAÇÃO POSITIVA PERFEITA OU LINEAR:** Quando a reta de regressão, obtida a partir do diagrama de dispersão, tem declividade positiva com  $r = 1$ .

**CORRELAÇÃO PRODUTO-MOMENTO:** Também conhecida como momento-produto. Ver correlação.

**CORRELAÇÃO SEMIPARCIAL:** Valor que mede a força da relação entre uma variável dependente e uma única variável independente quando os efeitos preditivos das demais variáveis independentes no modelo de regressão são removidos. O objetivo é retratar o efeito preditivo único devido a uma só variável independente em um conjunto de variáveis independentes. Difere do coeficiente de correlação parcial, que envolve efeito preditivo incremental.

**CORRELAÇÃO SEMIPARCIAL:** Ver coeficiente de correlação semiparcial.

**CORRELAÇÃO TETRACÓRICA:** Medida de associação usada para relacionar duas medidas binárias em uma tabela de contingência do tipo 2x2 com duas linhas por duas colunas. Ver também correlação policórica.

**CORRELAÇÕES ESTRUTURAIS:** Ver cargas discriminantes.

**CORRELAÇÕES FENOTÍPICAS:** Correlações entre duas características causadas por fatores genéticos e ambientais que influenciam ambas as características.

**CORRELAÇÕES GENÉTICAS:** Correlações entre duas características que surgem porque alguns genes afetam ambas as características. Quando duas características são positiva e altamente correlacionadas, por exemplo, peso à desmama e peso ao ano em bovinos de corte, a seleção para aumentar uma delas resultará em aumento na outra. Caso as características sejam negativa e altamente correlacionadas, por exemplo, peso ao nascimento e facilidade de parição, a seleção para aumentar delas resultará em diminuição na outra.

**CORRELACIONADOS POSITIVAMENTE:** Duas grandezas dizem-se correlacionadas positivamente se estão relacionadas de tal modo que, quando uma aumenta, a outra tende a aumentar; sua correlação é maior do que zero (0).

**COVARIÂNCIA:** O primeiro momento do produto de duas variáveis em torno de seus valores médios. O termo é usado também para o estimador da covariância. Simbolizada por  $Cov(X, Y)$ . Em inglês Covariance.

**COVARIÂNCIA:** i) Medida do grau de associação entre variáveis quantitativas utilizada para medir o grau de dependência mútua entre duas variáveis aleatórias e é definida como sendo a esperança matemática do produto dos desvios das variáveis aleatórias em torno de suas respectivas médias, podendo ser calculada como:  $Cov[X, Y] = E(XY) - \bar{X}\bar{Y}$ , que pode ser estimada na prática como:  $Cov[X, Y] = \frac{1}{N} \sum_{i=1}^N X_i \cdot Y_i - \bar{X}\bar{Y}$ . A

magnitude dessa medida será determinada pela grandeza das variáveis aleatórias X e Y; ii) A covariância entre pontos separados por uma distância h, pode ser calculada como:  $C(h) = E[z(x+h).z(x)] - m^2$

onde: Z(x+h) e Z(x) são os valores da variável regionalizada nos pontos (x+h) e (x), respectivamente. m é a média aritmética da variável regionalizada.

**COVARIÂNCIA:** Produto do desvio de uma observação em relação à média da variável x, multiplicado pelo desvio da mesma observação em relação à média da variável y.

**COVARIÂNCIA:** Medida semelhante ao coeficiente de correlação, uma vez que mede a relação entre um par de variáveis, porém, ao contrário deste, ele não é padronizado como coeficiente de correlação, a covariância é dividida pelo desvio padrão de x e y. Por ser a covariância não padronizada, não há limite para os valores, sendo então difícil compara-las.

**COVARIÂNCIA:** Indicação do grau de associação entre duas grandezas; é associação à correlação, mas seu valor não está restrito a permanecer no intervalo [-1, 1].

**COVARIÂNCIA:** É a variação conjunta de duas variáveis. A covariância é definida como sendo a expectativa do produto menos o produto das expectativas. Sejam X e Y duas variáveis aleatórias. A covariância entre X e Y, anotada por Cov(X, Y) é dada por:  $Cov(X, Y) = E(XY) - E(X).E(Y)$ , onde E(X) significa a expectativa ou valor esperado da variável X. Em inglês covariance.

**COVARIÂNCIA POPULACIONAL COV (X,Y):** É um valor que mede o grau de dispersão simultânea de duas variáveis quantitativas em relação as suas médias. É a soma dos produtos dos desvios das variáveis em relação a sua média, dividido pelo número de observações:  $\sum \left[ \frac{(X_i - \bar{X})(Y_i - \bar{Y})}{N} \right]$ . É o numerador do coeficiente de correlação.

**COVARIÁVEL:** Uma variável que apresenta um efeito que não se tem interesse direto. A análise da variável de interesse apresentará melhores resultados se a variação da covariável for controlada. Em inglês covariate or control variable.

**COVARIÁVEL:** Outro nome para variável independente ou regressor. Uma variável que apresenta um efeito que não se tem interesse direto. A análise da variável de interesse apresentará melhores resultados se a variação da covariável for controlada. Em inglês Covariate.

**COVARIÁVEIS OU ANÁLISE COVARIÁVEIS:** Uso de procedimentos do tipo regressão para remover variação estranha ou inconveniente nas variáveis dependentes devido a uma ou mais variáveis independentes métricas que são as covariáveis não controladas. As covariáveis são consideradas linearmente relacionadas com as variáveis dependentes. Depois de ajustar a influência de covariáveis, uma análise de variância (ANOVA) ou análise de variância multivariada (MANOVA) padrão é realizada. Esse processo de ajuste conhecido como ANCOVA ou MANCOVA geralmente permite testes mais sensíveis de efeitos de tratamento.

**COVARIÂNCIA DEFASADA:** O primeiro momento dos produtos de duas séries, uma das quais está defasada com relação à outra. Em inglês Lag covariance.

**COVARIÂNCIA ENTRE AS VARIÁVEIS ALEATÓRIAS X E Y:** Medida do grau de associação entre duas variáveis. Simbolizado por Cov (X,Y).

**COVARIOGRAMA:** É a representação gráfica da covariância em função da distância, numa determinada direção. O uso da função covariância se faz por meio da função variograma, que pode ser modelada mais facilmente que a primeira, valendo-se da relação:  $C(h) = C(0) - \gamma(h)$ ,

**onde:** C(h) é a função covariograma; C(0) é a variância a priori; g(h) é a função variograma.

**COVRATIO:** Medida da influência de uma única observação sobre o conjunto inteiro de coeficientes de regressão estimados. Um valor próximo de 1 indica pouca influência. Se o valor COVRATIO menos 1 é maior que  $\pm 3 \frac{P}{n}$ , onde p é o número de variáveis independentes + 1 e n é o tamanho da amostra. A

observação é considerada influente com base nessa medida.

**COV (X,Y):** Covariância entre as variáveis aleatórias X e Y.

**CRESCIMENTO EXPONENCIAL:** Um aumento no qual a taxa de crescimento é sempre proporcional à quantidade de material existente a cada momento.

**COMPETIÇÃO INTERPARCELAR (O EFEITO DE BORDADURAS):** Bordadura é a área que circunda as parcelas de um experimento, ou mesmo todo o experimento com a finalidade de dar proteção e condições homogêneas à unidade experimental considerada como área útil. O efeito de bordadura entre as parcelas pode influir aumentando o erro experimental, pois as plantas nas bordaduras e nos extremos das parcelas são menos afetadas, pela competição por água, luz e nutrientes do que aquelas que crescem no interior das parcelas. Os tipos são isoespecífica quando é idêntica ao material da parcela. Heteroespecífica quando o material é diferente do da parcela e mista, quando o material é idêntico ao material da parcela mais o material diferente da parcela.

**CRIATIVIDADE:** Produção de uma idéia, conceito, criação ou descoberta que é nova, original, útil, ou que satisfaça seu criador ou alguém mais, por um certo período de tempo.

**CRITÉRIO DAS CHANCES PROPORCIONAIS:** Outro critério para avaliar a razão de sucesso, no qual a probabilidade média de classificação é calculada considerando todos os tamanhos de grupos.

**CRITÉRIO DE ACEITAÇÃO:** O número máximo de itens defeituosos que pode ser encontrado na amostra e ainda permite aceitação do lote.

**CRITÉRIO DE CHANCE MÁXIMA:** Medida de precisão preditiva na matriz de classificação que compara o percentual corretamente classificado, também conhecido como a razão de sucesso com o percentual de respondentes no maior grupo. A ideia é que a melhor escolha desinformada é classificar cada observação no maior grupo.

**CRITÉRIO DE FATORIZAÇÃO DE NEYMAN:** A estatística t é suficiente para  $\theta$  se e somente se  $p(x|\theta) = f(t, \theta)g(x)$  onde f e g são funções não negativas. Em inglês Neyman's factorization criteria.

**CRITÉRIO DE VIZINHANÇA:** Em estatística espacial, critério utilizado para definir os vizinhos dentro de uma região. Em inglês Neighboring criteria .

**CRITÉRIO DE WHIPPLE:** É aquele que permite classificar uma população humana em normal, acessiva ou secessiva, conforme a frequência relativa porcentual da classe de 0-15 anos de idade contiver aproximadamente 50% do total, muito mais, ou muito menos, respectivamente.

**CRITÉRIO-PADRÃO (OU DE REFERÊNCIA):** Método, medida ou procedimento tido como o melhor, entre as alternativas; o que mais se aproxima da verdade. Um exame ou conjunto de exames diagnósticos que constitui o melhor standard para a afeição da sensibilidade, especificidade e outros atributos de um novo teste de rastreamento ou de diagnóstico.

**CRITÉRIOS:** Para se avaliar objetivamente qualquer fenômeno é necessário compará-lo a um outro fenômeno. Padrões de comparação são chamados critérios que são divididos em dois tipos: i) critérios principais visam a principal finalidade de comparação; ii) critérios modificadores visam os fatores práticos e, às vezes, cruciais, tais como: custo, conveniência, tempo envolvido, dentre outros.

**CRÍTICA:** Por crítica entende-se tanto a análise como o conhecimento e o posicionamento valorativo diante de uma realidade. A crítica pode referir-se de maneira lógica ou empírica à estrutura e validade das afirmações, ou esclarecer doutrinas e esquemas de pensamento. Ela pode contribuir de muitas maneiras para o esclarecimento da prática e para o desenvolvimento de alternativas de soluções.

**CRONOGRAMA:** É o planejamento das atividades da pesquisa, descrito na metodologia, dentro de um espaço pré-determinado de tempo. Normalmente é demonstrado por meio de um gráfico.

**CRUZ GEOESTATÍSTICA:** Consiste na realização de furos de pesquisa dispostos segundo uma cruz, daí seu nome. Este procedimento foi adotado inicialmente para determinação da malha ótima e possível anisotropia do depósito, a partir do variograma. Contudo, devido ao alto custo para realização dos furos esta prática foi abandonada, dando lugar a definição da malha de sondagem por critérios geológicos e econômicos, o que vem dando resultados satisfatórios.

**CRUZAMENTO GRÁFICO:** Uma das mais úteis apresentações gráfica para análise de dados. Pode ser usado com praticamente qualquer tipo de dados e tem o poder gráfico de apontar similaridades e diferenças em fortes contrastes. Em sua forma elementar, ele aparece na tabela contigencial 2x2 comum.

	GANHADORES	PERDEDORES
CLASSE ALTA	8	22
CLASSE MÉDIA	19	11

**CUBAGEM:** Termo que designa os diferentes métodos e técnicas que visam medir o volume de árvores, troncos e toras.

**CUBAGEM, FÓRMULAS DE:** Fórmulas definidas a partir do estudo de sólidos de revolução utilizadas para determinar, de modo aproximado, o volume de toras de madeira. As principais fórmulas de cubagem são denominadas pelos nomes de seus proponentes: Smalian, Huber e Newton.

**CUBAGEM, REGRA DE:** Método para definir o volume de toras baseado na experiência prática ou numa convenção, gerando resultados aproximados. No Brasil, as regras de cubagem em uso são a regra de Francon e a regra da alfândega de Paris. Diferem das fórmulas de cubagem por não representarem aproximações do volume de sólidos geométricos.

**CUBAGEM RIGOROSA:** Cubagem baseada na medida direta do volume do lenho, pelo método do deslocamento de água ou pelo método do empuxo, ou baseada na determinação do volume do lenho por fórmulas de cubagem que utilizam medidas do diâmetro dos troncos e ramos em diversas posições ao longo desses.

**CUBICAGEM:** Ver cubagem.

**CÚBICO:** Modelo de variograma com patamar, descrito pela seguinte equação:

$$\gamma(h) = C_0 + C \left[ 7 \left( \frac{h}{a} \right)^2 - 8,75 \left( \frac{h}{a} \right)^3 + 3,5 \left( \frac{h}{a} \right)^5 - 0,75 \left( \frac{h}{a} \right)^7 \right]. \text{ Para } h < a, \gamma(h) = C_0 + C \text{ para } h^3 a.$$

**COLINEARIDADE (MULTICOLINEARIDADE):** Quando duas ou mais variáveis explicativas ou independentes de um modelo de regressão múltipla exibem uma relação linear e são altamente correlacionadas.

**COMPARAÇÕES MÚLTIPLAS:** O processo de realizar vários testes de hipótese em um conjunto de dados; resulta no aumento do risco tipo I, a menos que se façam os ajustes.

**CONCORDÂNCIA DE MÉTODO:** Também denominado reprodutibilidade, envolve a aferição da semelhança de diferentes métodos de medidas, por exemplo, observadores diferentes utilizando a mesma técnica ou um único observador utilizando técnicas diferentes.

**CONFUSAS:** Uma variável explicativa relacionada com a resposta de interesse e com outra ou mais de uma variável explicativa, de modo que é impossível separar os efeitos das duas variáveis explicativas na resposta.

**CONTROLE ATUAL:** Os animais são designados para o grupo controle ou grupo do tratamento em teste em períodos similares dentro do período de estudo.

**CONTROLE HISTÓRICO:** Os animais não são direcionados ao(s) tratamento(s) controle(s) ou teste(s) contemporaneamente; a informação proveniente de controles históricos é obtida a partir de relatórios antigos.

**CONTROLE NEGATIVO:** O animal não recebe um tratamento ativo. Esses animais compõem o grupo controle negativo.

**CONTROLE POSITIVO:** Em testes clínicos, é o tratamento padrão que serve para comparação para novos tratamentos. Em estudos laboratoriais, é o tratamento que induz a resposta máxima.

**CORREÇÃO DE BONFERRONI:** Um método de redução do risco do erro tipo I, quando se empregam múltiplas comparações; envolve a multiplicação do valor de P obtido em qualquer teste, pelo número das várias comparações.

**CORREÇÃO DE CONTINUIDADE:** Uma correção aplicada a um teste estatístico para facilitar a aproximação de uma distribuição discreta do teste estatístico a uma distribuição contínua como a distribuição qui-quadrado.

**CORREÇÃO DE YATES:** Um ajuste aplicado ao teste estatístico qui-quadrado, quando é utilizado para testar uma hipótese em uma tabela de contingência 2x2. Torna a distribuição discreta do teste estatístico uma melhor aproximação à distribuição qui-quadrado contínua.

**CORRELAÇÃO LINEAR, COEFICIENTE:** Uma medida da associação linear entre duas variáveis.

**CORRELAÇÃO SERIADA:** Uma medida de dependência seriada em uma sequência de observações, como uma série de momentos. Sua presença implica que os desvios acerca de quaisquer tendências de longo prazo estão associados.



**CO-VARIÁVEIS:** Fatores que podem influenciar a resposta. Também denominadas variáveis explicativas ou independentes.

**CRITÉRIO DE INFORMAÇÃO DE AKAIKE (AIC):** Critério usado para seleção de modelos o qual penaliza a verossimilhança pelo número de parâmetros independentes ajustados, e por esse critério, qualquer parâmetro extra deve aumentar a verossimilhança por ao menos uma unidade para que o mesmo entre no modelo. O AIC é dado por  $AIC = -2 \log L + 2p$ , em que  $p$  é o número de parâmetros estimados. Menores valores de AIC refletem um melhor ajuste global. Assim, os valores de AIC são calculados para cada modelo e aquele com menor valor de AIC é escolhido como melhor modelo. O primeiro termo do AIC pode ser interpretado como uma medida de ajuste do modelo eo segundo termo como uma penalização. Desse modelo, no caso em que se comparam modelos com mesmo número de parâmetros, necessita-se comparar apenas o Log L. A vantagem do AIC é que as comparações não se limitam a modelos com estrutura hierárquica de fatores, fato que faz do AIC uma ferramenta genérica para seleção de modelos. Pode ser usado, por exemplo, para a comparação entre modelos com erros, apresentando diferentes distribuições, no entanto, os modelos devem ter a mesma estrutura de efeitos fixos.

**CRITÉRIO DE INFORMAÇÃO BAYSIANO (BIC):** Critério usado para seleção de modelos o qual é dado por  $BIC = -2 \log L + p \log v$ , em que  $v = N - r(x)$  é o número de graus de liberdade do resíduo. O BIC é calculado para cada modelo e aquele com menor valor é escolhido como melhor modelo. Pode ser usado quando os modelos não possuem estrutura hierárquica, no entanto, os modelos devem ter a mesma estrutura de efeitos fixos.

**CURTOSE:** Termo utilizado para descrever o pico de uma curva de frequência unimodal.

**CURVA DE SOBREVIVÊNCIA DE BERKSON-GAGE:** Uma curva de sobrevivência baseada em tabelas de vida atuariais; seus cálculos requerem conhecimento dos intervalos de tempo dentro dos quais os eventos críticos ocorrem, por exemplo, morte.

**CURVA DE SOBREVIVÊNCIA DE KAPLAN-MEIER:** É uma curva de sobrevivência baseada em tempos de sobrevivência conhecidos e que incluem dados censurados.

**CURVA RECEIVER OPERATING CHARACTERISTIC (CURVA ROC):** Um gráfico da sensibilidade de um teste oposto (1-especificidade); pode ser utilizada para escolher o valor de corte de uma variável contínua que levará ao teste diagnóstico ótimo ou teste de triagem, ou para comparar testes quanto à determinada condição.

**CURTOSE:** É uma medida de um conjunto de dados que informa o quanto este conjunto se desvia do modelo ou curva normal. É o grau de achatamento do conjunto. A curva normal teria um coeficiente de curtose igual a três, sendo que alguns autores utilizam zero e é denominada de mesocúrtica. Uma medida acima de três ou positiva caracterizaria o conjunto como leptocúrtico que é mais afinado que a curva normal. Se o coeficiente de curtose for menor do que três ou negativo então o conjunto é denominado de platicúrtico o qual é mais achatado do que a curva normal.

**CURTOSE:** É o grau de achatamento de uma distribuição em relação à distribuição normal. O coeficiente de curtose pode ser calculado a partir do quarto momento centrado na média:

$$CC = \frac{\frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^4}{S^4}$$

A curtose, de acordo com seu valor, pode ser classificada em platicúrtica, mesocúrtica e leptocúrtica.

**CURTOSE:** É uma medida de um conjunto de dados que informa o quanto este conjunto se desvia do modelo ou curva normal. É o grau de achatamento do conjunto. A curva normal teria um coeficiente de curtose igual a três alguns autores utilizam zero e é denominada de mesocúrtica. Uma medida acima de três ou positiva caracterizaria o conjunto como leptocúrtico o qual é mais afinado que a curva normal. Se o coeficiente de curtose for menor do que três ou negativo então o conjunto é denominado de platicúrtico, ou seja, mais achatado do que a curva normal. Foi proposto por Karl Pearson antes de 1905. É representado por  $\gamma_2$  e calculado por:  $\gamma_2 = \frac{\mu_4}{\mu_2^2}$ , onde  $\mu_4$  é o momento central de quarta ordem e  $\mu_2$  é

a variância ou momento de segundo ordem. Em inglês kurtosis.

**CURTOSE:** Medida da elevação ou do achatamento de uma distribuição quando comparada com uma distribuição normal. Um valor positivo indica uma distribuição relativamente elevada e um valor negativo indica uma distribuição relativamente achatada.

**CURTOSE:** Distorção vertical de uma distribuição de frequências.

**CURVA ACUMULATIVA:** i) derivada diretamente do histograma, é obtida desenhando-se uma série de retângulos contíguos com bases iguais aos intervalos de medidas e alturas proporcionais às frequências acumuladas. Os pontos da curva acumulativa são lançados nos pontos médios dos intervalos de medida, fazendo com que a curva seja representada por um número de pontos igual ao número de classes do histograma. Na verdade esta representação é conhecida como polígonos de frequências acumuladas; ii) uma outra forma de curva acumulativa é obtida lançando-se todos os pontos, ao invés de dados agrupados em intervalos, que dá uma noção exata da distribuição de frequências em estudo, pois os pontos por si só delimitam uma curva contínua, sendo que nesse caso quanto maior o número de dados melhor o desenho da curva acumulativa.

**CURVA CARACTERÍSTICA DE OPERAÇÃO:** Um gráfico mostrando a probabilidade de aceitar um lote como uma função da porcentagem defeituosa no lote. Essa curva pode ser usada para ajudar a determinar se um determinado plano de amostragem de aceitação atende ambos os requisitos do risco, do produtor e do consumidor.

**CURVA DE CRESCIMENTO:** Em geral, uma expressão que fornece o tamanho de uma população como função de uma variável do tempo, descrevendo o curso de seu crescimento. Em inglês Growth curve.

**CURVA DE DENSIDADE:** Gráfico de uma distribuição contínua de probabilidade.

**CURVA DE FREQUÊNCIA:** Representação gráfica de uma distribuição de frequências contínua ou uma variável discreta quando esta possui valores muitos dispersos e grande volume de dados, com a variável na abscissa e a frequência na ordenada. Em inglês Frequency curve.

**CURVA DE FREQUÊNCIA (CURVA POLIDA):** Representação gráfica de uma distribuição de frequências contínua, com a variável na abscissa e a frequência na ordenada. Enquanto o polígono de frequência nos dá a imagem real do fenômeno estudado, a curva de frequência nos dá a imagem tendencial.



**CURVA DE GAUSS:** Ou curva normal é a representação gráfica da função densidade de probabilidade da distribuição normal, caracteriza-se por apresentar forma de sino e ser simétrica em torno da média.

**CURVA DE GAUSS:** Curva simétrica, unimodal, em forma de sino ou campânula: os escores da distribuição normal que a curva representa concentram-se no meio, tanto que aproximadamente 68% dos casos se acham entre 1 desvio-padrão abaixo e 1 desvio-padrão acima da média. Em inglês Gauss curve.

**CURVA DE GOMPERTZ:** Ver curva de crescimento. Em inglês Gompertz curve.

**CURVA DE REGRESSÃO:** É a representação geométrica e uma equação de regressão com uma só variável independente.

**CURVA-GUIA, MÉTODO DA:** Método de construção de curvas de índice de sítio por meio de ajuste estatístico, por regressão linear ou não-linear, de um modelo da altura média das árvores dominantes em função da idade do povoamento. Um dos parâmetros do modelo é tomado como indicador de sítio expressando-o em função da idade base e da altura média das árvores dominantes nessa idade que é o índice de sítio. A substituição do parâmetro no modelo ajustado pela expressão resultante gera um modelo que expressa a altura média das dominantes em função da idade, da idade base e do índice de sítio. O método da curva-guia pode ser aplicado em dados de parcelas temporárias de inventário florestal, pois não requer que uma mesma parcela tenha sido medida em mais de uma ocasião. **Ver também:** *índice de sítio*.

**CURVA LOGÍSTICA:** Uma curva em S formada pela transformação logit que representa a probabilidade de um evento. A forma em S é não-linear porque a probabilidade de um evento deve se aproximar de 0 e 1, porém jamais ser maior. Assim, apesar de haver uma componente linear no meio do intervalo, à medida que as probabilidades se aproximam dos limites inferior e superior de probabilidade (0 e 1), elas devem se amenizar e ficar assintóticas nesses limites.

**CURVA NORMAL:** O mesmo que curva de Gauss. Em inglês normal curve.

**CURVA OPERATIVA-CARACTERÍSTICA:** Gráfico que representa as probabilidades de aceitar hipóteses alternativas quando uma hipótese nula é verdadeira nas ordenadas contra os valores da hipótese alternativa nas abscissas. Em inglês Operating characteristic curve.

**CURVA DE ROC (RECEIVER OPERANTIG CHARACTERISTIC CURVE):** É um método estatístico e gráfico para determinar o melhor ponto de corte (cutoff point) de um teste diagnóstico.

**CURVA TEOR X TONELAGEM:** É a curva que relaciona a reserva e o teor médio em função do teor de corte. O teor de corte é o parâmetro principal que define a reserva, qualquer que seja a sua categoria: geológica, lavrável ou compatibilizada. Assim, a partir da curva teor X tonelagem pode-se determinar para qualquer teor de corte a reserva, bem como o teor médio.

**CURVAS DE MORTALIDADE PROPORCIONAL:** Constituem representações gráficas dos vários índices de mortalidade proporcional, índice de Moraes, segundo grupos etários prefixados: grupo infantil (<1 ano), crianças em idade pré-escolar (1-4 anos), as crianças e adolescentes (5-19 anos), os adultos jovens (20-49) e as pessoas e meia-idade e idosas (50 e mais anos). De acordo com os tipos de curva resultantes, avalia-se o nível de saúde: muito baixo (irregular), baixo (jota invertido), regular (forma em U) e elevado (forma de jota-J).

**CUSTO DE OPORTUNIDADE:** Valor da melhor alternativa não concretizada em consequência de se utilizarem recursos escassos na produção de um dado bem ou serviço. O custo de oportunidade é o verdadeiro custo em que a sociedade incorre ao fornecer um programa de saúde à população, na medida em que os recursos humanos e materiais empregados nesse programa ficam indisponíveis para outros fins. Há que se notar ainda que, quanto maior for a escassez de recursos, maiores serão os custos de oportunidade de determinado da decisão.

**CUSTO SOCIAL:** O custo de determinada atividade para a sociedade no seu todo, e não apenas para os indivíduos ou instituições envolvidas na sua realização. O custo social de qualquer procedimento médico incluirá tanto os custos incorridos pela clínica ou hospital como aqueles suportados pelos usuários e por outros setores da sociedade. O custo social não equivale necessariamente ao somatório dos custos privados. No caso de um programa de imunização, por exemplo, o custo social terá menor do que os custos privados. No caso de um programa de imunização, por exemplo, o custo social será menor do que os custos privados, dado que outros indivíduos, além dos vacinados, virão a se beneficiar com a implementação do programa. Como é evidente, o conceito de custo social está estreitamente relacionado com o conceito de extremidade.

**CUSTOS DIRETOS:** Custos incorridos com a organização e operacionalização de determinado programa de saúde. As categorias de custo direto com maior peso são geralmente as despesas com pessoal, materiais consumíveis e energia e os gastos com capital. Para além desses custos de produção incorridos pelo serviços, incluem-se ainda, sob a rubrica de custos diretos, os gastos efetuados pelos usuários e seus familiares.

**CUSTOS INDIRETOS:** Custos associados à perda de produção econômica devido à participação do usuário em determinado programa de saúde e ainda os chamados custos psíquicos ou intangíveis, tais como ansiedade, dor e desconforto associados aos próprios tratamentos.

**CUSTOS FIXOS:** Aqueles que não são passíveis de alteração a curto prazo, pelo fato de serem independentes do volume de produção, por exemplo, rendas, gastos, gasto com capital, dentre outros.

**CUSTOS VARIÁVEIS:** São custos cuja dimensão depende do volume de produção, por exemplo, aproveitamento, pagamentos ao ato à vista, alimentação dentre outros.

**CV:** Coeficiente de variação amostral. Conhecido também como coeficiente de variação de Pearson (CVp) ou coeficiente de variação de Thornidike (CVt).

**CONTRASTE:** Função linear de médias ou de totais de tratamentos com coeficientes que totalizam zero. Um contraste é um resumo das médias dos tratamentos que é de interesse em um experimento.

**CONTROLE DE QUALIDADE:** Sistemas e procedimentos usados por uma organização para assegurar que as saídas dos processos satisfazem os consumidores.

**CONTROLE ESTATÍSTICO DE PROCESSO(CEP):** Um conjunto de ferramentas de resolução de problemas, baseado em dados, cujo objetivo é melhorar o processo.

**CONTROLE ESTATÍSTICO DE QUALIDADE:** Métodos estatísticos e de engenharia usados para medir, monitorar, controlar e melhorar a qualidade.

**CONTROLE EXCESSIVO:** Ajustes desnecessários feitos nos processos que aumentam os desvios em relação ao alvo.

**CONVOLUÇÃO:** Um método para derivar a função densidade de probabilidade da soma de duas variáveis aleatórias independentes a partir de uma integral ou soma de funções densidades de probabilidade.

**CORREÇÃO DE CONTINUIDADE:** Um fator de correção usado para melhorar a aproximação das probabilidades binomiais por uma distribuição normal.

**CORRELAÇÃO:** No uso mais geral, uma medida de interdependência entre os dados. O conceito pode incluir mais de duas variáveis. O termo é mais comumente usado em um estreito senso para expressar a relação entre variáveis quantitativas ou postos.

**COVARIÂNCIA:** Uma medida de associação entre duas variáveis aleatórias, obtida como o valor esperado do produto de duas variáveis aleatórias em torno de sua média ou esperança matemática, ou seja,  $Cov(X,Y) = E[(X-\mu_x)(Y-\mu_y)]$ .

**CURTOSE:** Uma medida do grau de achatamento de uma distribuição unimodal.

**CURVA DE NÍVEL:** Um gráfico bidimensional para uma função de densidade de probabilidade bivariada que dispõe curvas para as quais a função densidade de probabilidade é constante.

**CURVAS CARACTERÍSTICAS OPERACIONAIS (CURVAS CCO):** Um gráfico de probabilidade do erro tipo II versus alguma medida de extensão que afirma que a hipótese nula é falsa. Tipicamente, uma curva CCO é usada para representar cada tamanho de amostra de interesse.

**CUSUS TABULAR:** Um algoritmo numérico usado para detectar causas atribuídas em um gráfico de controle de soma cumulativa. Veja Máscara V.

**COMPETIÇÃO INTERPARCELAR (O EFEITO DE BORDADURAS):** Bordadura é a área que circunda as parcelas de um experimento, ou mesmo todo o experimento com a finalidade de dar proteção e condições homogêneas à unidade experimental considerada como área útil. O efeito de bordadura entre as parcelas pode influir aumentando o erro experimental, pois as plantas nas bordaduras e nos extremos das parcelas são menos afetadas, pela competição por água, luz e nutrientes do que aquelas que crescem no interior das parcelas. Estas podem ser dos tipos: i) Isoespecífica que é idêntica ao material da parcela; ii) heteroespecífica cujo material é diferente do da parcela e iii) mista cujo material idêntico ao material da parcela mais o material diferente da parcela.

**CATI:** Acrônimo para Computer Assisted Telephone Interviewing que significa Entrevista Telefônica Assistida por Computador. É a técnica de inquérito estatístico realizada através do telefone. O entrevistador é orientado durante a entrevista seguindo um algoritmo disponibilizado pelo computador. Esse software tem a capacidade de configurar a ordem e o tipo de questões baseando-se nas respostas anteriores, ou na informação conhecida sobre o participante.

**CAPI:** Acrônimo para Computer Assisted Personal Interviewing, que significa Entrevista Pessoal Assistida por Computador. Trata-se de uma técnica de inquérito que utiliza um questionário baseado no computador. Como alternativa aos questionários em papel, o CAPI permite ao entrevistador orientar a sua entrevista seguindo um algoritmo disponibilizado pelo computador. O software tem a capacidade de configurar a ordem e o tipo de questões baseando-se nas respostas anteriores, ou na informação que já dispõe sobre o participante. Desse modo, as entrevistas deixam de ser genéricas e passam a ser mais focadas no entrevistado.



# D

**DADO:** Qualquer tipo de representação que tenha um significado.

**DADO BRUTO:** O mesmo que dado primitivo.

**DADO CENSURADO:** É o dado cujo valor está além de um limite conhecido, embora não se saiba o valor exato do dado.

**DADO CENSURADO:** Dizemos que um dado é censurado quando sabemos que seu valor está além de um limite, embora não saibamos exatamente qual é esse valor. Em ensaios clínicos de longo prazo por exemplo, em que o resultado de interesse é o tempo de sobrevivência. O tempo de sobrevivência dos pacientes que continuaram vivos ao término do ensaio é um dado censurado. Nas medidas obtidas por meio de aparelhos que têm um limite mínimo para as medições. O valor real de algumas medições pode estar abaixo do nível mais baixo que o aparelho pode medir. Esses dados são censurados. Nos experimentos com animais em que o tempo de seguimento é fixo e a observação de interesse é o tempo que alguma condição específica demora a aparecer ou a desaparecer, e se nada acontece com alguns animais até o final do experimento, os dados desses animais são censurados.

**DADO ESTATÍSTICO:** É um dado numérico, considerado a matéria-prima sobre a qual iremos aplicar os métodos estatísticos.

**DADO NUMÉRICO:** são expressos por números. Veja também dado quantitativo.

**DADO ESTATÍSTICO:** É o resultado da observação de um atributo/variável qualitativa ou quantitativa.

**DADO PRIMÁRIO:** É o dado estatístico colhido pela própria entidade que o vai elaborar ou analisar. Opõe-se a dado secundário.

**DADO PRIMITIVO:** Diz-se o dado estatístico que ainda não sofreu nenhuma espécie de elaboração, achando-se, portanto, apenas colhido. Também se diz dado bruto.

**DADO SECUNDÁRIO:** É o dado estatístico colhido por entidade diversa daquela que o vai elaborar ou analisar. Opõe-se a dado primário.

**DADOS:** Informação fundamental coletada em um estudo, em geral expressa numericamente.

**DADOS:** Escores obtidos de observações ou de experimentos, podendo ser de fontes primárias ou secundárias e de amostras ou de populações.

**DADOS:** Conjunto de qualquer tipo de informação detalhada e quantificada, resultado de medições ou experiências realizadas com objetivos específicos, usado como referência para determinações, estudos e trabalhos científicos. Toda a informação factível de ser resumida em um código, uma cifra, um esquema, um plano ou uma foto. Quer dizer, informação que não requer um texto ou um comentário para ser inteligível ou utilizável.



**DADOS:** Números ou informações que descrevem alguma característica.

**DADOS:** Os números e atributos que são coletados, analisados e interpretados. Em inglês data.

**DADOS AGREGADOS:** Classe de dados estatísticos em que os valores para intervalos unitários pequenos podem ser somados para formar séries de valores relacionadas a intervalos maiores. Em inglês Integrated data.

**DADOS AGREGADOS:** Dados criados por um processo de combinação ou resumo. Eles podem assim ser armazenados nessa forma resumida em um armazém de dados.

**DADOS AGRUPADOS:** Disposição dos dados em que, em lugar de relacioná-los todos um a um, se elabora uma tabela que mostra quantos valores se enquadram em cada categoria.

**DADOS ANALÍTICOS:** Dados operacionais que foram integrados e combinados com dados analíticos existentes por meio de uma operação de armazenamento de dados para formar um registro histórico de um evento de negócios ou objeto, por exemplo, cliente, firma ou produto.

**DADOS BINÁRIOS:** Dados que apresentam dois possíveis valores, muitas vezes relacionados com variáveis qualitativas de duas categorias e que podem ser representados pelos números 0 e 1. Em inglês Binary data.

**DADOS BINÁRIOS:** Referem-se a uma variável binária, que só assume uma de dois valores possíveis, zero (0) ou um (1).

**DADOS BIVARIADOS:** Par de valores correspondente a um dado indivíduo ou resultado experimental.

**DADOS BIVARIADOS:** Dados expostos em pares.

**DADOS BRUTOS:** Medida e observações registradas em fichas clínicas ou cadernos de laboratório, mas ainda não organizadas para interpretação. Listagens de dados obtidos em computador, mas na forma como foram coletados, antes de edição, resumo e análise.

**DADOS BRUTOS:** Dados iniciais com que o pesquisador começa seu trabalho, antes de passar à análise.

**DADOS CATEGÓRICOS (DE ATRIBUTOS):** Dados que podem ser separados em diferentes categorias que se distinguem por alguma característica não-numérica.

**DADOS CATEGORIZADOS:** São aqueles distribuídos em categorias mutuamente exclusivas. Veja também dados qualitativos.

**DADOS CENSURADOS:** Uma observação é dita censurada se ela não pode ser medida de forma precisa, mas sabe-se que está além, ou aquém, de um limite. Em inglês Censored data.

**DADOS CENSURADOS:** Observações que são incompletas de uma maneira sistemática e conhecida. Um exemplo ocorre nos estudos de causas de morte em uma amostra na qual alguns indivíduos ainda estão vivos. Os dados censurados são um exemplo de dados perdidos ignoráveis.

**DADOS CLASSIFICADOS:** São valores que uma dada variável pode tomar dentro de certo intervalo. Estes dados são classificados ou agrupados em classes.

**DADOS CONTÍNUOS:** Dados quantitativos que podem tomar todos os valores numéricos, compreendidos no seu intervalo de variação.

**DADOS CONTÍNUOS:** Referem-se a uma variável contínua. São obtidos por mediação.

**DADOS CONTÍNUOS:** Dados que resultam de um número infinito de valores contínuos que podem ser associados a pontos de uma escala contínua, sem lacunas ou interrupções.

**DADOS DE ATRIBUTOS:** Dados que podem ser separados em categorias diferentes que se distinguem por alguma característica não numérica.

**DADOS DE CONFUSÃO:** Procedimento para obter percepções de respondentes sobre dados de similaridades. Os respondentes indicam as similaridades entre pares de estímulos. O par ou confusão de um estímulo com um outro é assumido para indicar similaridades. Também conhecido como agrupamento subjetivo.

**DADOS DE DENOMINADOR:** Dados que definem a população sob risco.

**DADOS DE DESEMPENHO:** Registro individual de um animal para reprodução, produção, e mérito de carcaça. Os caracteres incluídos poderiam ser pesos ao nascimento, desmama e ao ano, facilidade de parto, intervalo entre partos, produção de leite, produção de lã, ovos, dentre outros.

**DADOS DE ESCALA DE INTERVALOS:** Dados envolvendo um nível de medição que estabelece um valor exato para cada categoria da variável em termos de unidades específicas de medição.

**DADOS DE ESCALA NOMINAL:** Dados envolvendo um nível de medição que simplesmente identifica ou classifica as observações em categorias.

**DADOS DE ESCALA ORDINAL:** Dados envolvendo um nível de medição que procura classificar categorias pela extensão até onde elas representam a variável.

**DADOS BRUTOS:** Dados na forma em que foram coletados, sem nenhum tratamento, ordenamento ou qualquer tipo de organização.

**DADOS DE NUMERADOR:** Dados que definem os eventos ou a condição de interesse.

**DADOS DE PROCESSO:** **Dados dispostos segundo uma sequencia temporal que medem uma característica de bens ou serviços resultante de uma combinação de equipamento, pessoas, materiais, métodos e condições.**

**DADOS DE SEÇÃO TRANVERSAL:** Dados coletados no mesmo ou aproximadamente no mesmo ponto do tempo. Em inglês cross section data.

**DADOS DE SÉRIE HISTÓRICA:** Dados coletados em períodos de tempo sucessivos. Em inglês data.

**DADOS DE SÉRIES TEMPORAIS:** Dados referentes a observações de uma grandeza em diferentes pontos do tempo.

**DADOS DE SIMILARIDADES:** Dados usados para determinar quais objetos são os mais semelhantes entre si e quais são os mais distintos. Implícita nas medidas de similaridades está a habilidade de comparar todos os pares de objetos. Três procedimentos para obter dados de similaridades são comparações aos pares de objetos, dados de confusão e medidas derivadas.

**DADOS DISCRETOS:** Dados quantitativos que só podem assumir um número finito, ou infinito numerável, de valores distintos.

**DADOS DISCRETOS:** Referem-se a uma variável discreta. Surgem de processos de contagem.

**DADOS DEMOGRÁFICOS:** Dados usados para descrever o grupo estudado. Inclui: dados sobre idade, sexo, etnicidade, lugar de nascimento, status marital, número de filhos, nível educacional, ocupação, lugar de residência, e outros dados fatuais.

**DADOS MÉTRICOS:** Também chamados de dados quantitativos, dados intervalares ou dados proporcionais, essas medidas identificam ou descrevem indivíduos ou objetos não apenas na posse de um atributo, mas também pela quantia ou grau em que o indivíduo pode ser caracterizado pelo atributo. Por exemplo, a idade ou o peso de alguém são dados métricos.

**DADOS MULTIDIMENSIONAIS:** Elemento de dados com múltiplas dimensões ou atributos.

**DADOS NÃO-MÉTRICOS:** Também chamados de dados qualitativos, são atributos, características ou propriedades categóricas que identificam ou descrevem um indivíduo ou objeto. Diferem dos dados métricos no sentido de indicarem a presença de um atributo, mas não a quantia. Exemplos são ocupações de médicos, advogado, professor ou status do comprador como comprador e não-comprador. São também conhecidos como dados nominais ou dados ordinais.

**DADOS NUMÉRICOS:** Dados numéricos que representam contagens ou mensurações.

**DADOS NÃO-PARAMÉTRICOS:** Dados para os quais não podem ser obtidos parâmetros descritivos, como a média e o desvio-padrão, porque não há escala de medidas. Não se elabora nenhuma hipótese referente à distribuição de frequências subjacente.

**DADOS OPERACIONAIS:** Dados utilizados por sistemas operacionais, por exemplo, contabilidade, administração de balanço que refletem o status corrente da organização que é necessário para funções de negócios diários.

**DADOS PARAMÉTRICOS:** Dados para os quais podem ser obtidos parâmetros descritivos tais como, a média e o desvio-padrão.

**DADOS PERDIDOS:** Informação não disponível de um indivíduo ou caso sobre o qual outra informação está disponível. Os dados perdidos frequentemente ocorrem quando um respondente deixa de responder uma ou mais questões da pesquisa.

**DADOS PERDIDOS IGNORÁVEIS:** Processo de dados perdidos que é explicitamente identificável e/ou está sob o controle do pesquisador. Os dados perdidos ignoráveis não demandam ações corretivas, pois os dados perdidos são explicitamente tratados na técnica empregada.

**DADOS PRIMÁRIOS:** Quando são publicados pela própria pessoa ou organização que os haja recolhido.

**DADOS PRIMÁRIOS:** Novos dados coletados para ajudar a resolver problemas imediatos.

**DADOS PRIMITIVOS:** Elementos de dados mantidos em seu mais elementar nível de detalhes, como transações individuais.

**DADOS QUALITATIVOS:** Representam a informação que identifica alguma qualidade, categoria ou característica, não susceptível de medida, mas de classificação, assumindo várias modalidades.

**DADOS QUANTITATIVOS:** Representam a informação resultante de características susceptíveis de serem medidas, apresentam-se com diferentes intensidades, que podem ser de natureza discreta que se refere a dados discretos, ou contínua referente a dados contínuos.

**DADOS QUANTITATIVOS:** São aqueles distribuídos em categorias mutuamente exclusivas. Veja também dados categorizados.

**DADOS QUANTITATIVOS:** Dados numéricos que representam contagens ou medidas.

**DADOS QUANTITATIVOS:** São expressos por números. Veja também dado numérico.

**DADOS SECUNDÁRIOS:** Quando são publicados por outra organização, instituto de pesquisa ou órgão público ou privado.

**DADOS SECUNDÁRIOS:** Dados que já foram coletados anteriormente por órgãos públicos ou privados.

**DADOS SIMPLES:** São valores associados a uma dada variável e cuja representação é feita por meio de uma tabela.

**D'AGOSTINO:** Teste empregado para testar a normalidade dos dados amostrais.

**D'AGOSTINO-PEARSON:** Teste destinado a verificar a normalidade dos dados amostrais.

**DAP:** Diâmetro à Altura do Peito. Medida tomada a 1,30 metros da base do tronco de uma árvore, podendo ser obtida diretamente com suta ou compasso florestal ou indiretamente com fita métrica ou fita diamétrica, assumindo-se que a secção transversal do tronco é um círculo para fazer a conversão da medida de circunferência em diâmetro. Em alguns casos, principalmente em florestas tropicais, não é possível medir o DAP na altura convencional. Nessas situações geralmente mede-se 30 cm acima da área com deformidade ou raízes, lembrando que é importante anotar na ficha de campo a altura em que o DAP foi tomado.

**DADO:** Unidade de informação fundamental em um estudo ou pesquisa, normalmente mas não necessariamente expressa numericamente. Exemplo: i) a média das idades dos sujeitos de uma amostra; ii) Conceito mais recorrente num texto, obtido a partir de uma análise de conteúdo.

**DELINEAMENTO DE PESQUISA:** i) Plano ou esquema geral a ser utilizado numa pesquisa; ii). Planejamento dos procedimentos que serão utilizados na coleta e análise dos dados obtidos em uma pesquisa. Há quatro grandes delineamentos, vinculado ao tipo de pesquisa: levantamento, correlacional, experimental e quase-experimental; dentro destes últimos, diz-se que há também os subdelineamentos.

**DESVIO-PADRÃO:** Medida de dispersão de um conjunto de dados, comumente abreviada como  $s$  ( $s$  minúsculo) ou  $\sigma$  (letra grega sigma). Calcula-se por meio da fórmula:

$$s = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}} \quad \text{ou} \quad \sigma = \sqrt{\frac{\sum_{i=1}^n (X_i - \mu)^2}{N}}$$

onde:  $x$  = cada valor individual do conjunto de dados;

$n$  = número de valores do conjunto de dados;

= média dos valores do conjunto de dados;

Ver dispersão e depois variância.

**DIAGRAMA DE BARRAS:** Ver gráfico de barras.

**DIAGRAMA DE CAIXAS:** Diagrama de apresentação de dados cujo objetivo principal é o de facilitar a visualização da variabilidade de um conjunto de dados. Apresenta os valores centrais bem como a amplitude de determinada distribuição, por meio de cinco valores básicos: valor mínimo, quartil inferior, mediana, quartil superior e valor máximo.

**DIAGRAMA DE COLUNAS:** Ver gráfico de barras.

**DIAGRAMA DE EXTREMOS E QUARTIS:** Ver diagrama de caixas.

**DIAGRAMA DE PARETO:** Representação gráfica que permite discriminar as causas mais importantes de um problema, estabelecendo prioridades para a sua solução. Neste tipo de diagrama, os fatores causadores do problema são colocados em ordem de importância, mostrando-se assim sequência ideal para resolução do problema.

**DADOS BIVARIADOS (BIVARIATE DATA):** Dados que consistem de duas variáveis para cada observação. Em geral são ilustrados em gráficos de dispersão.

**DADOS MULTIVARIADOS (MULTIVARIATE DATA):** Os dados consistem de mais de duas variáveis para cada observação. Com frequência são ilustrados usando matrizes de gráficos de dispersão ou outras ferramentas de visualização em mais dimensões.

**DADOS UNIVARIADOS (UNIVARIATE DATA):** Os dados consistem de uma única medida para cada observação. Muitas vezes são ilustrados usando box plots, histogramas ou gráficos de caule e folhas.

**DECIL (DECILE):** Um décimo. Em geral refere-se aos decis superior e inferior, os 10% superiores e os 10% inferiores de uma distribuição.

**DECOMPOSIÇÃO (DECOMPOSITION):** Em álgebra de matrizes, é a re-expressão de qualquer matriz  $m \times n$   $A$  conforme o produto de três matrizes,  $VWU$ , onde  $V$  é uma matriz com as mesmas dimensões de  $A$ ,  $U$  e  $W$  são matrizes quadradas (dimensão  $n \times n$ );  $V$  e  $U$  são matrizes de colunas ortonormais; e  $W$  é uma matriz diagonal com valores singulares na diagonal.

**DEDUÇÃO (DEDUCTION):** O processo de raciocínio científico que progride do geral para o específico, por exemplo, todos os cisnes são brancos, esta ave, sendo um cisne, é branca. Comparar com indução.

**DISPERSÃO:** Quando observamos um conjunto de dados, verificamos que eles se distribuem ao redor de uma média. Quanto mais próximos dessa média os valores estiverem, mais homogênea será a amostra, ou seja, menor a dispersão dos dados. As duas principais medidas de dispersão de uma amostra são a Variância e o Desvio-padrão. Exemplos: Suponha a existência de dois grupos, A e B, com três indivíduos cada. A variável que estamos estudando nesses dois grupos é a variável idade.

**DISTRIBUIÇÃO:** Agrupamento dos dados de uma variável de acordo com determinado critério.

**DISTRIBUIÇÃO DE FREQUÊNCIA:** Técnica estatística que permite a organização e visualização de dados de acordo com a ocorrência de diferentes resultados observados. Uma distribuição de frequência pode ser construída na forma de uma tabela que indica basicamente o número de ocorrências de determinado dado ou valor numa variável.

**DISTRIBUIÇÃO DE PROBABILIDADES:** A distribuição de probabilidades de determinada variável é a descrição das probabilidades associadas a todos os possíveis valores dessa variável.

**DISTRIBUIÇÃO DO TEMPO DE VIDA:** Seja  $T$  uma variável aleatória, não negativa, que representa o tempo de vida de um indivíduo proveniente de uma dada população homogênea. Considerando  $T$  como sendo uma variável aleatória absolutamente contínua, cuja distribuição pode ser caracterizada por qualquer uma das seguintes funções: função densidade de probabilidade, função de distribuição, função de sobrevivência, função de risco e função de risco cumulativa

**DISTRIBUIÇÃO À PRIORI:** Tipo de distribuição usada em inferência Bayesiana para obter informação a priori a qual requer a especificação desta distribuição para a quantidade de interesse  $\theta$ . Seja  $\theta$  o parâmetro do qual se deseja fazer uma inferência. A informação já disponível sobre  $\theta$  permite associar a ele uma distribuição de probabilidade que descreverá as incertezas e crenças de um pesquisador, chamada distribuição à priori. Tome como exemplo no lançamento de uma moeda a probabilidade  $\theta$  de tirar cara. Qual é a porcentagem de caras em certa quantidade de lançamentos? Se a moeda for honesta, espera-se que esta probabilidade esteja centrada em 0,5. Porém, suponha que isso não aconteça. Pergunta-se que informação prévia o dono da moeda possui. Pode-se começar pensando em uma priori não informativa, ou seja, aquela que associa igual chance a todas as porcentagens. Essa distribuição seria uma Uniforme ou uma Beta (1,1).

**DECAIMENTO EXPONENCIAL:** Um declínio no qual a taxa de decaimento é sempre proporcional à quantidade de material existente a cada momento.

**DECIL:** Qualquer dos nove pontos que dividem uma escala de escores de dez intervalos, cada um dos quais contendo um décimo da frequência total. Em inglês decile.

**DECIL:** Cada um dos nove valores que dividem a frequência total em dez partes iguais. Em inglês Decile.

**DECIL:** Cada um dos nove valores que dividem dados ordenados em dez grupos com aproximadamente 10% dos valores em cada grupo.

**DECIS:** Medida de tendência central com características de que seu valor divide os termos da série em grupos de dez partes. O quinto decil dessa divisão coincide com a mediana. Tem sua aplicação para divisão da série em dez partes iguais permitindo sínteses para comparações.

**DECIS:** A definição dos decis obedece ao mesmo princípio dos quartis, com a modificação da porcentagem de valores que ficam aquém e além do decil que se pretende calcular. De especial interesse é o quinto decil, que divide o conjunto em duas partes iguais. Assim sendo, o quinto decil é igual ao segundo quartil, que por sua vez é igual à mediana.

**DECISÃO, TOMADA DE (CARACTERÍSTICAS):** i) Qualquer decisão deve identificar-se necessariamente com escolha entre alternativas; ii) Qualquer decisão requer algum valor que possa servir como um objetivo; iii) Uma terceira alternativa de tomada de decisão é uma estimativa de custos, acompanhando



cada alternativa; iv) Tem-se que ter algum conhecimento das probabilidades de várias conseqüências para cada alternativa.

**DECIL:** Qualquer dos nove pontos que dividem escala de valores em dez intervalos, cada um dos quais contendo um décimo da frequência total.

**DESVIO:** Diferença entre um escore e a média aritmética. É equivalente à expressão afastamento, a qual, todavia, é menos empregada.

**DESVIO-PADRÃO:** Medida de dispersão ou variabilidade de uma distribuição de frequência. É igual à raiz quadrada da soma dos desvios ao quadrado dividida pelo número de casos. Em outras palavras, é a raiz quadrada da média dos desvios elevados ao quadrado

**DIAGRAMA CARTESIANO:** Representação gráfica por meio de dois eixos perpendiculares que são denominadas de coordenadas cartesianas, sendo a abscissa correspondendo ao eixo horizontal e ordenada ao eixo vertical.

**DIAGRAMA DE DISPERSÃO:** Gráfico que, por meio das coordenadas cartesianas, é construído com o fim de mostrar a relação entre valores de dois testes obtidos por indivíduos de certo grupo. Os valores de um teste são representados ao longo da abcissa e os valores do outro ao longo da ordenada. Se os valores nas duas variáveis são altamente correlacionados, os pontos tendem a situar-se ao longo de uma reta.

**DISTRIBUIÇÃO BIMODAL:** Distribuição de frequências dotada de duas modas, ou seja, de dois valores que se distinguem dos demais por alta frequência. Graficamente, a distribuição bimodal possui duas saliências ou corcovas, que podem não apresentar a mesma altura.

**DISTRIBUIÇÃO DE FREQUÊNCIA:** Tabela que indica as frequências com que ocorrem os casos correspondentes a cada intervalo.

**DISTRIBUIÇÃO EM J:** Distribuição de frequências acentuadamente assimétrica, quer positiva quer negativamente. Sua representação gráfica assemelha-se a um jota de pé ou invertido.

**DISTRIBUIÇÃO NORMAL:** Também chamada distribuição de gauss, possui as seguintes características: é unimodal, simétrica e campanular. Muitas medidas antropométricas distribuem-se normalmente.

**DISTRIBUIÇÃO BINOMIAL:** É aquela que envolve um número finito de tentativas; os resultados das diversas tentativas são independentes, de tal modo que a probabilidade de que certo resultado seja a mesma em cada tentativa; cada tentativa admite somente dois resultados, mutuamente exclusivos, tecnicamente chamados sucesso e fracasso.

**DISTRIBUIÇÃO DE POISSON:** Trata-se do caso limite da Binomial quando o número de prova  $n$  tende para o infinito e a probabilidade  $p$  do evento em cada prova tende para zero.

**DECLINAÇÃO MAGNÉTICA:** É a diferença entre a leitura da bússola e o norte verdadeiro. Em inglês magnetic declination.

**DADO ORIGINAL:** Consultar leituras.

**DADOS CENSURADOS:** Animais em teste de sobrevivência que sobrevivem ao final do estudo ou que são perdidos na etapa de acompanhamento.

**DEFINIÇÃO DA FREQUÊNCIA DE PROBABILIDADE:** Envolve a contagem da frequência de ocorrência do evento em um grande número de repetições de testes similares.

**DELINEAMENTO:** O plano do estudo ou experimento que deve levar em conta todos os fatores que possam influenciar a variável resposta de interesse.

**DELINEAMENTO DE BLOCO INCOMPLETO:** Os blocos de um experimento não contêm o conjunto inteiro de tratamentos.

**DELINEAMENTO EM BLOCO ALEATÓRIO COMPLETO:** Cada bloco do delineamento experimental contém um conjunto completo de tratamentos.

**DELINEAMENTO SEQUENCIAL DE GRUPO:** Envolve uma análise temporária dos dados.

**DELINEAMENTO:** É o esquema adotado para distribuição dos tratamentos.

**DELINEAMENTOS COM MEDIDAS REPETIDAS:** Um delineamento experimental em que cada animal é analisado em cada nível de um fator, de modo que os efeitos de interesse sejam analisados “dentro” dos animais; por exemplo, quando cada animal recebe todos os tratamentos ou é analisado em vários momentos. Consultar também análise de variância (ANOVA) unidirecional com medidas repetidas.

**DELINEAMENTOS DE GRUPOS PARALELOS:** Cada animal, individualmente, recebe apenas um tratamento e as comparações no tratamento são feitas entre grupos de animais, ao invés de entre os animais.

**DELINEAMENTO ADCI (BACI DESIGN):** Um método experimental para avaliar os efeitos de um tratamento, frequentemente um impacto ambiental adverso em relação a uma população sem ameaças. Nos delineamentos ADCI (Antes-Depois-Controle-Impacto) com frequência há pouca replicação espacial, portanto a replicação é temporal. O efeito nas populações controle e nas populações ameaçadas são medidos tanto antes quanto depois do tratamento.

**DELINEAMENTO ADITIVO (ADDITIVE DESIGN):** Um delineamento experimental usado para estudar competição entre uma espécie de interesse alvo, ou focal, e seus competidores. Em um delineamento aditivo, a densidade da espécie-alvo é mantida constante e os diferentes tratamentos representam diferentes números de competidores. Comparar com delineamentos substitutivos e de superfície de resposta.

**DELINEAMENTO ANINHADO (NESTED DESIGN):** Qualquer delineamento de análise de variância em que há subamostragem dentro das réplicas.

**DELINEAMENTO COM PARCELAS SUBDIVIDIDAS (SPLIT-PLOT DESIGN):** Um delineamento de análise de variância multifatorial em que cada parcela experimental é subdividida em subparcelas, e cada uma recebe um tratamento diferente.

**DELINEAMENTO COM UM FATOR (SINGLE FACTOR DESIGN):** Delineamento de análise de variância que manipula apenas uma variável. Comparar com delineamento multifatorial.

**DELINEAMENTO DE ANÁLISE DE VARIÂNCIA (ANALYSIS OF VARIANCE DESIGNS):** A classe de configurações experimentais usada para explorar relações entre variáveis preditoras categóricas e variáveis resposta contínuas.



**DESLOCADA À DIREITA (DESVIO POSITIVO):** A distribuição de frequência não é simétrica e tem uma cauda do lado direito estendida.

**DESLOCADA À ESQUERDA (DESVIO NEGATIVO):** A distribuição de frequência não é simétrica e tem uma cauda do lado esquerdo estendida.

**DESVIO NORMAL PADRÃO (DNP):** Uma variável que segue a distribuição Normal, com média zero e desvio-padrão igual a um (1).

**DESVIO-PADRÃO (DP):** Uma medida de amplitude que pode ser considerada, aproximadamente, como uma média dos desvios das observações da média aritmética. É igual à raiz quadrada da variância.

**DESVIO-PADRÃO DA PROPORÇÃO:** É o desvio-padrão da amostragem de distribuição de uma proporção, normalmente denominado de erro padrão da proporção. É uma medida da precisão de p como uma estimativa de  $\pi$ .

**DIAGNÓSTICOS DE REGRESSÃO:** Os procedimentos utilizados para verificar as hipóteses nas quais se baseia um modelo de regressão e para determinar os valores anômalos ou outliers e casos que exercem influência.

**DIAGRAMA BOX-AND-WHISKER:** É um diagrama que mostra a distribuição de dados numéricos ou ordinais. Normalmente tem o formato de uma caixa cujos limites horizontais são definidos pelos quartis superior e inferior, envolvendo 50% das observações, com mediana marcada por uma linha horizontal dentro da caixa. Os whiskers são linhas verticais que se estendem na caixa, desde o 2,5º percentil até o 97,5º percentil.

**DIAGRAMA DE CAULE-E-FOLHAS (STEM-AND-LEAF DIAGRAM):** Um diagrama, geralmente gerado por computador, que mostra a distribuição de um conjunto de dados; o caule representa o valor principal das observações, por exemplo, o valor da unidade antes da casa decimal e cada folha é uma sequência de dígitos unitários ordenados, um para cada observação, que seguem o valor principal, por exemplo, a primeira casa decimal.

**DIAGRAMA DE DISPERSÃO:** Uma projeção bidimensional em que cada eixo representa a escala de medida de uma ou duas variáveis; cada ponto representa as coordenadas para um valor em ambas as escalas.

**DIAGRAMA/GRÁFICO DE PONTOS:** Utilizado para mostrar a distribuição de um conjunto de dados. Cada observação é marcada como um ponto em uma linha ajustada nas unidades de medida da variável.

**DIAGRAMAS:** Uma maneira de expor dados por meio de ilustrações.

**DIFERANÇA PADRONIZADA:** Utilizada nos cálculos do tamanho da amostra necessário para um teste de hipótese; baseia-se no efeito do tratamento significativo um que não se pretende omitir, dividido pelo desvio padrão relevante.

**DIMENSIONAMENTO AMOSTRAL:** Tamanho de amostra n, para ser usado na estimação da média populacional  $\mu$  de uma população infinita, normal, com variância ( $\sigma^2$ ) conhecida por exemplo, com um erro de estimação ou de amostragem de no máximo igual a e, e com uma confiança de  $[1 - \alpha]\%$  de probabilidade, pode-se determinar o tamanho amostral n. Partindo-se da expressão do erro de amostragem

ou de estimação e do intervalo de confiança ou fiducial, chega-se a equação para se obter o valor do tamanho da amostra n.

$$\text{I.C.}[\mu] \quad \bar{X} \pm Z\left(\frac{\alpha}{2}\right) \cdot \frac{\sigma}{\sqrt{n}} \quad \therefore \bar{X} \pm e$$

$$e = Z\left(\frac{\alpha}{2}\right) \cdot \frac{\sigma}{\sqrt{n}} \quad \therefore \sqrt{n} \cdot e = Z\left(\frac{\alpha}{2}\right) \cdot \frac{\sigma}{\sqrt{n}} \cdot \sqrt{n} \quad \therefore \sqrt{n} \cdot e = Z\left(\frac{\alpha}{2}\right) \cdot \sigma \quad \therefore \frac{\sqrt{n} \cdot e}{e} = \frac{Z\left(\frac{\alpha}{2}\right) \cdot \sigma}{e}$$

$$\sqrt{n} = \frac{Z\left(\frac{\alpha}{2}\right) \cdot \sigma}{e} \quad \therefore (\sqrt{n})^2 = \left(\frac{Z\left(\frac{\alpha}{2}\right) \cdot \sigma}{e}\right)^2 \quad \therefore n = \frac{\left[Z\left(\frac{\alpha}{2}\right)\right]^2 \cdot \sigma^2}{e^2}$$

Ou seja, esse é o menor tamanho n que deve ter a amostra para ser usada na estimação da média populacional  $\mu$ , em uma população infinita, normal, com variância ( $\sigma^2$ ) conhecida, com um erro de estimação ou de amostragem de no máximo e, e uma confiança de  $[1 - \alpha]\%$  de probabilidade.

**DELINEAMENTO DE BLOCOS ALEATORIZADOS (RANDOMIZED BLOCK DESIGN):** Um delineamento de análise de variância em que conjuntos de réplicas, e todos os níveis dos tratamentos, são dispostos em uma área física no espaço ou em um ponto fixo no tempo, um bloco por exemplo, no intuito de reduzir a variabilidade nas condições não manipuladas.

**DELINEAMENTO DE DOIS FATORES (TWO-WAYDESIGN):** Um delineamento de análise de variância (ANOVA) com dois efeitos principais em que cada um deles tem dois ou mais níveis de tratamento.

**DELINEAMENTO OU DESENHO DE MEDIDAS REPETIDAS (REPEATED MEASURES DESIGN):** Um tipo de análise de variância (ANOVA) em que múltiplas observações são feitas em uma mesma réplica individual, em diferentes pontos no tempo.

**DELINEAMENTO DE REGRESSÃO (REGRESSION DESIGN):** A classe de esboço experimental usado para explorar relações entre variáveis preditoras contínuas e variáveis resposta contínuas ou categóricas.

**DELINEAMENTO OU DESENHO DE SUPERFÍCIE DE RESPOSTA (RESPONSE SURFACE DESIGN):** Um delineamento experimental usado para estudar competição entre uma espécie de interesse alvo, ou focal, e seus competidores. Em um delineamento de superfície de resposta, a densidade da espécie alvo e de seus competidores são variadas sistematicamente. Os delineamentos de superfície de resposta também podem ser usados em vez de análise de variância (ANOVA) quando os fatores são melhor representados como variáveis contínuas. Comparar com delineamento aditivo e delineamento substitutivo.

**DELINEAMENTO FATORIAL (FACTORIAL DESIGN):** Um delineamento para análise de variância que inclui todos os níveis de dois ou mais tratamentos ou fatores de interesse.

**DELINEAMENTO OU DESENHO MULTIFATORIAL (MULTI FACTOR DESIGN):** Delineamento de análise de variância (ANOVA) que inclui dois ou mais tratamentos ou fatores.

**DELINEAMENTO OU DESENHO SUBSTITUTIVO (SUBSTITUTIVE DESIGN):** Um delineamento experimental usado para estudar a competição entre uma espécie de interesse alvo, ou focal, e seus

competidores. Em um delineamento substitutivo, a densidade total da espécie alvo e de seus competidores é mantida constante em certo nível, mas a proporção relativa da espécie alvo e seus competidores é variada sistematicamente. Comparar com delineamento aditivo e com delineamento de superfície de resposta.

**DIMENSIONAMENTO AMOSTRAL:** Tamanho de amostra  $n$  para ser usado na estimação da média populacional  $\mu$  de uma população infinita, normal, com variância ( $\sigma^2$ ) desconhecida, com um erro de estimação ou de amostragem de no máximo e uma confiança de  $[1 - \alpha]\%$  de probabilidade. Qual deve o tamanho da amostra  $n$ ? Partindo-se da expressão do erro de amostragem ou de estimação e do intervalo de confiança ou fiducial tem-se que, chega-se a equação para se obter o valor do tamanho da amostra  $n$ .

$$IC_{[\mu]}: \bar{X} \pm t_{\left[\frac{\alpha}{2}\right]} \cdot \frac{S}{\sqrt{n}} \quad \therefore \bar{X} \pm e \quad e = t_{\left[\frac{\alpha}{2}\right]} \frac{S}{\sqrt{n}} \quad \therefore \sqrt{n} \cdot e = t_{\left[\frac{\alpha}{2}\right]} \frac{S}{\sqrt{n}} \cdot \sqrt{n} \quad \therefore$$

$$\frac{\sqrt{n} \cdot e}{e} = \frac{t_{\left[\frac{\alpha}{2}\right]} \cdot S}{e} \quad \therefore \sqrt{n} = \frac{t_{\left[\frac{\alpha}{2}\right]} \cdot S}{e} \quad \therefore (\sqrt{n})^2 = \left( \frac{t_{\left[\frac{\alpha}{2}\right]} \cdot S}{e} \right)^2 \quad \therefore n = \frac{t_{\left[\frac{\alpha}{2}\right]}^2 \cdot S^2}{e^2}$$

Ou seja, esse é o menor tamanho  $n$  que deve ter a amostra para ser usada na estimação da média populacional  $\mu$  de uma população infinita, com  $[1 - \alpha]\%$  de probabilidade de confiança e um erro de amostragem ou de estimação de no máximo igual a  $e$ .

**DIMENSIONAMENTO AMOSTRAL:** Tamanho de amostra  $n$  para ser usado na estimação da verdadeira proporção populacional  $P$  de uma população infinita, normal com  $[1 - \alpha]\%$  de probabilidade de confiança e um erro de amostragem ou de estimação de no máximo igual a  $e$ . Qual deve ser o tamanho da amostra  $n$ ? Partindo-se da expressão do erro de amostragem ou de estimação e do intervalo de confiança ou fiducial tem-se que, chega-se a equação para se obter o valor do tamanho da amostra  $n$ .

$$e = Z_{\left[\frac{\alpha}{2}\right]} \cdot \sqrt{\frac{\hat{P}(1 - \hat{P})}{n}} \quad \therefore e = Z_{\left[\frac{\alpha}{2}\right]} \cdot \frac{\sqrt{\hat{P}(1 - \hat{P})}}{\sqrt{n}} \quad \therefore \sqrt{n} \cdot e = Z_{\left[\frac{\alpha}{2}\right]} \cdot \frac{\sqrt{\hat{P}(1 - \hat{P})}}{\sqrt{n}} \cdot \sqrt{n} \quad \therefore$$

$$\sqrt{n} \cdot e = Z \cdot \sqrt{\hat{P}(1 - \hat{P})} \quad \therefore \frac{\sqrt{n} \cdot e}{e} = \frac{Z \cdot \sqrt{\hat{P}(1 - \hat{P})}}{e} \quad \therefore \sqrt{n} = \frac{Z \cdot \sqrt{\hat{P}(1 - \hat{P})}}{e} \quad \therefore$$

$$(\sqrt{n})^2 = \frac{(Z)^2 \left( \sqrt{\hat{P}(1 - \hat{P})} \right)^2}{(e)^2} \quad \therefore n = \frac{(Z)^2 \hat{P}(1 - \hat{P})}{(e)^2}$$

Portanto, este é o menor tamanho que deve ter a amostra para se estimar a verdadeira proporção populacional de uma população infinita, normal, com  $[1 - \alpha]\%$  de probabilidade de confiança e um erro de amostragem ou de estimação de no máximo igual a tantos pontos percentuais  $e$ .

**DISTÂNCIA DE COOK:** Uma medida geral de influência utilizada na análise de regressão, que inclui tanto os valores leverage quanto os residuais.

**DISTORÇÃO:** Um termo utilizado para descrever a assimetria de uma distribuição de frequência.

**DADO DEMOGRÁFICO:** Tipo de dado referente a descrição de características específicas das populações humanas. Informações tais quais as taxas de natalidade, taxa de fertilidade e índice de imigração, idade, sexo, estado civil, local de nascimento, número de filhos, nível socioeconômico são exemplos de dados demográficos de uma população.

**DADO PRIMÁRIO:** Dados obtidos ou levantados e apresentados por uma pesquisa.

**DADOS EXPERIMENTAIS OU RESPOSTAS DO EXPERIMENTO:** São as medidas de características do objeto em estudo em uma pesquisa.

**DATA MINING:** Prospecção, mineração de dados ou é um termo que significa um conjunto de técnicas estatísticas que visam encontrar padrões ainda não descobertos numa massa de dados.

**DEDUÇÃO:** Tipo especial de inferência ou raciocínio que parte de pelo menos uma premissa universal e que conduz a uma conclusão particular.

**DISTRIBUIÇÕES A PRIORI:** A distribuição a priori constitui-se no único elemento novo na análise bayesiana em relação à análise estatística clássica. A determinação desta distribuição é subjetiva, entretanto dados experimentais podem ser utilizados nesta etapa. A premissa desta distribuição é que ela represente o estado atual de conhecimento sobre o parâmetro  $\theta$ , antes de serem analisados os resultados experimentais. Existem dois tipos de informação a priori, uma baseada em dados passados e outra completamente subjetiva não baseada em dados. Na presença de dados experimentais prévios ou quando baseada em fundamentos teóricos, a priori deve ser obtida de uma maneira científica por formas funcionais. A distribuição a priori tem importância maior quando a quantidade  $n$  de dados experimentais atuais é pequena. Caso contrário, a priori tende a ser dominada, pois recebe menor peso pela verossimilhança, ou seja, com  $n \rightarrow \infty$ , as abordagens frequentistas e bayesiana são coincidentes. Em outras palavras, o peso da verossimilhança é maior do que o peso da priori, se a variância dos dados observados é menor que a variância da priori. É importante ressaltar, entretanto, que atribuir probabilidade zero a priori, significa atribuir zero também a posteriori, pois, neste caso, sendo a posteriori = priori x verossimilhança, não há verossimilhança que faça a posteriori diferente de zero.

**DAFAULT:** Em informática ou computação científica, o termo refere-se a uma ação ou valor que será assumido se nenhum outro for informado, por exemplo, ao não se preencher um campo numa tela de dados, o software, programa ou pacote estatístico assume um valor padrão para aquele campo.

**DENDROGRAMA (DENDROGRAM):** Um diagrama, que vem do termo grama, de árvore que se origina do termo dendro. Um método de representar graficamente os resultados de uma análise de agrupamentos. Em princípio, é similar a uma árvore de classificação. Também é muito usado para ilustrar a relação entre espécies ou filogenias.

**DESIGUALDADE DE TCHEBYCHEFF:** Também conhecido como teorema de Tchebycheff é aplicado em situações em se conhece a média ou esperança matemática de uma variável aleatória  $E(X)$  e a sua variância  $V(X)$ , mas não se conhece a distribuição de  $X$ , isto é,  $p(x)$  ou  $f(x)$ , a desigualdade de Tchebycheff

ajuda o pesquisador a obter limitantes para certas medidas, por exemplo, para  $P\{|X - E(X)| < \varepsilon\}$ . Sejam  $c$  um número real qualquer e  $\varepsilon$  um número positivo. Pode ser mostrado que:  $P\{|X - c| \geq \varepsilon\} \leq \frac{1}{\varepsilon^2} E((X - c)^2)$  ou,  $P\{|X - c| < \varepsilon\} \geq 1 - \frac{1}{\varepsilon^2} E((X - c)^2)$ . Note que se  $c = E(X)$ , então a desigualdade pode ser escrita por:  $P\{|X - E(X)| \geq \varepsilon\} \leq \frac{V(X)}{\varepsilon^2}$ , vale lembrar que  $E((X - E(X))^2) = V(X)$ .

.Note ainda que se  $\varepsilon = k\sigma(X)$ , então:  $P\{|X - E(X)| \geq k\sigma(X)\} \leq \frac{1}{k^2}$ , lembrando que  $V(X) = \sigma(X)^2$ . Veja a prova do teorema.

$$P\{|X - c| \geq \varepsilon\} = \int_{\{x: |x-c| \geq \varepsilon\}} f(x) dx$$

$$\leq \int_{\{x: |x-c| \geq \varepsilon\}} \frac{(x-c)^2}{\varepsilon^2} f(x) dx$$

Dado que  $\frac{(x-c)^2}{\varepsilon^2} \geq 1$  e  $f(x) \geq 0$ , para  $\{x: |x-c| \geq \varepsilon\} \leq \int_{-\infty, \dots, \infty} \frac{(x-c)^2}{\varepsilon^2} f(x) dx$ , dado que  $\frac{(x-c)^2}{\varepsilon^2} \geq 1$  e  $f(x) \geq 0$ , para  $\{-\infty \leq x \leq \infty\} = \frac{1}{\varepsilon^2} E((X - c)^2)$ .

**DESIGUALDADE TRIANGULAR (TRIANGLE INEQUALITY):** Para três objetos multivariados, a, b e c, a distância entre a e b mais a entre b e c sempre é maior que, ou igual à, distância entre a e c.

**DESTENDENCIADA (UNBIASED):** A propriedade estatística de medidas, observações ou valores não serem nem consistentemente muito grandes nem muito pequenas. Ver também acurácia e precisão.

**DESVIO (DEVIATION):** Diferença entre a predição do modelo e dos valores observados.

**DESVIO PADRÃO (DP) (STANDARD DEVIATION-SD):** Uma medida de dispersão. Igual à raiz quadrada da variância.

**DESVIO PADRÃO AMOSTRAL (SAMPLE STANDARD DEVIATION):** Uma estimativa, sem viés, do desvio padrão de uma amostragem. Igual à raiz quadrada da variância amostral.

**DETERMINANTE (DETERMINANT):** Em álgebra de matrizes, é a escalar calculada como a soma de todos os produtos possíveis, contendo um elemento na linha e um na coluna de uma dada matriz.

**DIAGRAMA DE VENN (VENN DIAGRAM):** Um gráfico usado para ilustrar as operações de união, interseção e complemento de conjuntos.

**DIMENSÃO (DIMENSION):** Um vetor bidimensional especificando o número de linhas e o de colunas em uma matriz.

**DEFINIÇÃO OPERACIONAL:** Definição clara e específica que delimita o escopo de uma variável, para o uso prático ou operacional em uma pesquisa. Por exemplo, podemos dizer que a definição operacional de obeso seja: pessoa que esteja 20 % acima do seu peso ideal, de acordo com uma tabela padronizada de pessoas, alturas, idades e sexo.

**DELTA:** Quarta letra do alfabeto grego ( $\Delta, \delta$ ), frequentemente utilizada para indicar a diferença entre dois valores. Por exemplo,  $\Delta_{xy}$ , significa a diferença entre x e y.

**DEMONSTRAÇÃO:** Meio de se chegar a uma prova, evidência. Processo por meio do qual se explica o método utilizado para se chegar a uma conclusão. Por exemplo, quando se demonstra um teorema da probabilidade dos eventos independentes.

**DETECÇÃO DE CASOS:** Técnica de pesquisa utilizada principalmente na área da saúde pública ou epidemiologia, que consiste no exame de sujeitos assintomáticos ou saudáveis, com a finalidade de se estabelecer a probabilidade de terem a doença que é o objeto do estudo.

**DISTRIBUIÇÃO NORMAL:** Tipo de distribuição de dados de uma variável aleatória contínua, na qual os valores, quando descritos graficamente, são representados pela curva normal, que pode ser descrita matematicamente pela equação de Laplace-Gauss. A curva normal ou de Gauss é simétrica, unimodal e tem a forma de sino, de campânula ou de chapéu de Napoleão.

**DISTRIBUIÇÃO NORMAL PADRÃO:** Distribuição normal que possui média zero e desvio padrão um (1).

**DOXOMETRIA:** Qualquer método que tenha por finalidade a determinação das opiniões públicas por meio de sondagens ou levantamentos estatísticas, como ocorre, por exemplo, nas pesquisas de intenção de voto, pesquisas de mercado, pesquisas de opinião, dentre outras.

**DISTRIBUIÇÃO:** Consultar distribuição de frequência empírica e distribuição de probabilidade.

**DISTRIBUIÇÃO ALEATÓRIA:** A distribuição dos animais nos grupos, por exemplo, tratamento de maneira aleatória, ou seja, o procedimento baseia-se em sorteio. Também conhecida como escolha aleatória.

**DISTRIBUIÇÃO AMOSTRAL DA MÉDIA:** A distribuição das médias da amostra; é uma distribuição hipotética obtida ao considerar todas as amostras repetidas possíveis de um dado tamanho da população, calculando-se a média amostral em cada amostra. Essas médias amostrais podem ser representadas graficamente para mostrar a distribuição em uma forma diagramática.

**DISTRIBUIÇÃO A MOSTRAL DA PROPORÇÃO:** A distribuição das proporções amostrais; é uma distribuição hipotética obtida quando consideramos várias amostras repetidas de um dado tamanho da população, calculando-se a proporção amostral em cada amostra.

**DISTRIBUIÇÃO AMOSTRAL DA MÉDIA OU DAS MÉDIAS  $[\bar{X}]$ :** Conceito muito importante para inferência estatística, o qual será descrito a seguir. Inicialmente tem-se a distribuição amostral propriamente dita: Considere todas as possíveis amostras de tamanho n de uma população, isto é,  $N^n$  ou  $\binom{N}{n}$ ,

pode-se para cada amostra, calcular sua média obtendo dessa maneira, uma distribuição desses resultados dando origem a distribuição amostral ou por amostragem da média.

Seja X uma população infinita ou finita e amostragem com reposição de média  $\mu$  e variância  $\sigma^2$ , isto é  $E[X_i] = \mu_x = \mu$  e  $\text{Var}[X_i] = \sigma^2$

Quando consideramos  $E[\bar{X}]$  e  $\text{Var}[\bar{X}]$  existe alguma relação entre estes valores e a média e a variância da população objeto de estudo  $\mu$  e  $\sigma^2$ , e é essa associação entre parâmetros e estimadores que é visto a seguir.

**Tabela:** Distribuição de amostragem da média, para k amostras de tamanho n retiradas de uma população de tamanho N.

AMOSTRA	TAMANHO	ESTIMADOR
1	N	$\bar{X}_1 = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$
2	N	$\bar{X}_2$
.	.	.
.	.	.
.	.	.
K	N	$\bar{X}_k$

Seja  $(x_1, x_2, \dots, x_n)$ , uma amostra aleatória de n elementos e  $\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$ , a média amostral, a qual é uma combinação linear de variáveis aleatórias independentes,  $\bar{X}$  é, ela mesma também uma variável aleatória.  $E[\bar{X}] = \frac{1}{n} \{E[X_1] + E[X_2] + \dots + E[X_n]\}$ , cada observação  $X_i$  tem a distribuição da população com esperança

$\mu$ , de modo que:

$$E[\bar{X}] = \frac{1}{n} \{\mu + \mu + \dots + \mu\} \therefore E[\bar{X}] = \frac{1}{n} [n \cdot \mu] = \mu$$

ou ainda,

$$E[\bar{X}] = E\left[\frac{\sum_{i=1}^n x_i}{n}\right] = \frac{1}{n} \cdot \sum_{i=1}^n E[x_i] = \frac{1}{n} \cdot \sum_{i=1}^n \mu = \frac{1}{n} \cdot n \cdot \mu = \mu$$

Para a variância tem-se que:

$$\text{Var}[\bar{X}] = \text{Var}\left[\frac{\sum_{i=1}^n x_i}{n}\right] = \frac{1}{n^2} \{ \text{VAR}[X_1] + \text{VAR}[X_2] + \dots + \text{VAR}[X_n] \}$$

Devido todos os componentes de  $\bar{X}$  serem independentes. Cada observação  $X_i$  tem a distribuição de probabilidade da população, cuja variância é representada por  $\sigma^2$ , de modo que.

$$\text{Var}[\bar{X}] = \frac{1}{n^2} \{ \sigma^2 + \sigma^2 + \dots + \sigma^2 \} = \frac{1}{n^2} [n \cdot \sigma^2], \text{Var}[\bar{X}] = \frac{\sigma^2}{n} \therefore \sigma_{(\bar{X})} = \frac{\sigma}{\sqrt{n}}$$

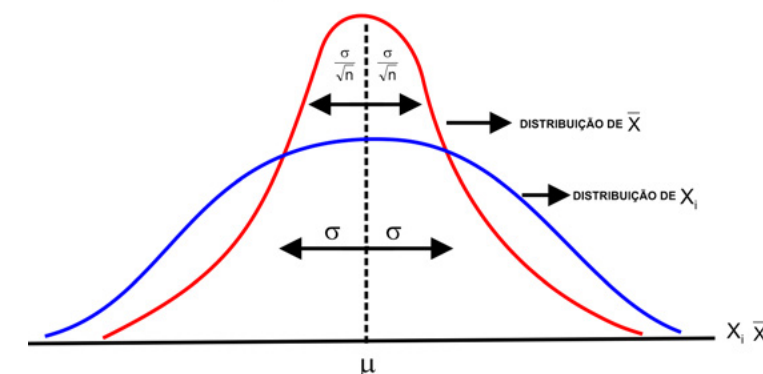
ou ainda:

$$\text{VAR}[\bar{X}] = \text{VAR}\left[\frac{\sum_{i=1}^n x_i}{n}\right] = \frac{1}{n^2} \cdot \sum_{i=1}^n \text{VAR}[x_i] = \frac{1}{n^2} \cdot \sum_{i=1}^n \sigma^2 = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}$$

$$D.P.(\bar{X}) = \sigma_{(\bar{X})} = \sqrt{\text{VAR}(\bar{X})} = \frac{\sigma}{\sqrt{n}}$$

Portanto, se  $X \cap (\mu, \sigma^2)$ , então  $\bar{X} \cap N\left(\mu, \frac{\sigma^2}{n}\right)$ , sendo  $Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$  a variável normal padronizada da variável aleatória  $\bar{X}$ .

Graficamente teremos,



**Figura:** Gráfico representativo da distribuição de uma população qualquer e da distribuição amostral da média.

No caso de amostragem sem reposição de população finita temos:  $X \cap (\mu, \sigma^2)$ , e

$$\mu_{(\bar{X})} = \mu \text{ e } \sigma_{(\bar{X})}^2 = \frac{\sigma^2}{n} \left[ \frac{N-n}{N-1} \right], \text{ sendo } Z = \frac{\bar{X} - \mu}{\frac{\sigma}{n} \left[ \frac{N-n}{N-1} \right]}, \text{ isto é: } \bar{X} \cap N\left\{ \mu; \frac{\sigma^2}{n} \left[ \frac{N-n}{N-1} \right] \right\}$$

Onde  $\left[ \frac{N-n}{N-1} \right]$  é o fator de correção para população finita. Note que esse fator tende à unidade quando

o tamanho da população tende ao infinito, sendo que N é o número de elementos da população, e n é o número de elementos da amostra.

**DISTRIBUIÇÃO AMOSTRAL DA MEDIANA [Md]:** A mediana Md de uma amostra aleatória simples é também uma variável aleatória. De uma maneira geral, para as populações parentes cuja distribuição é contínua, a esperança matemática da mediana da amostra é igual à mediana da população, ou seja:  $E[md]=MD$ . Para tamanhos suficientemente grandes, a variância correspondente é dada de maneira aproximada pela relação:

$$\text{Var}[md] \cong \frac{1}{4nf^2(MD)},$$

em que  $f[MD]$  é o valor da função densidade de probabilidade correspondente à mediana da população. Em particular, para uma população normal deduz-se que:  $E[md]=MD$ , e  $\text{Var}[md] \cong \frac{\pi \cdot \sigma^2}{2 \cdot n}$ . A variância da

mediana é, portanto, superior à variância da média, exceto no caso de amostras com 2 elementos ( $n = 2$ ), para as quais a média e a mediana se confundem. No limite, para  $n \rightarrow \infty$ , podemos escrever:

$$\frac{\text{Var}[md]}{\text{Var}[\bar{X}]} \rightarrow \frac{\pi \cdot \sigma^2}{2 \cdot n} \cdot \frac{n}{\sigma^2} = \frac{\pi}{2} = 1,57$$



Finalmente, para toda a população parente contínua, a distribuição de amostragem da mediana é assintoticamente normal. A distribuição amostral da mediana Md de amostras de uma população simétrica X, com média  $\mu$  e variância  $\sigma^2$ , segue aproximadamente uma distribuição normal com média  $E(Md)=\mu$  e  $Var(Md)=\frac{\sigma^2}{n} \cdot \frac{\pi}{2}$ , com desvio padrão dado por:

$$\sigma_{med} = \sigma \sqrt{\frac{\pi}{2n}} = \frac{1,2533\sigma}{\sqrt{n}}. \text{ Para } n \geq 30, \text{ a distribuição amostral da mediana é muito aproximada da normal.}$$

O resultado obtido somente é válido quando a população é normal ou aproximadamente normal.

**DISTRIBUIÇÃO AMOSTRAL DA DIFERENÇA ENTRE DUAS MÉDIAS**  $[\bar{X}_1 - \bar{X}_2]$ : Sejam duas populações com médias  $\mu_1$  e  $\mu_2$ , então se o interesse é na diferença entre as médias dessas duas populações, isto é  $(\mu_1, \mu_2)$ , e ainda se temos duas amostras aleatórias com médias  $\bar{X}_1$  e  $\bar{X}_2$ .

Sendo assim,  $E[\bar{X}_1 - \bar{X}_2] = E(\mu_1) - E(\mu_2) = \mu_1 - \mu_2$  e  $Var[\bar{X}_1 - \bar{X}_2] = Var[\bar{X}_1] + Var[\bar{X}_2] = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$ , e

desvio padrão igual a  $D.P.(\bar{X}_1 - \bar{X}_2) = \sigma_{(\bar{X}_1 - \bar{X}_2)} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ , sendo a variável normal padronizada Z

da variável aleatória  $[\bar{X}_1 - \bar{X}_2]$  igual a:

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

**DISTRIBUIÇÃO AMOSTRAL DA PROPORÇÃO**  $[\hat{P}]$ : Distribuição de frequência da proporção relativa, ou porcentagem de sucesso na amostra. Primeiro veja a distribuição amostral propriamente dita.

**Tabela:** Distribuição de amostragem da proporção, para k amostras de tamanho n retiradas de uma população de tamanho N.

AMOSTRA	TAMANHO	ESTIMADOR
1	n	$\hat{P}_1 = \frac{n_1}{n}$
2	n	$\hat{P}_2 = \frac{n_2}{n}$
.	.	.
.	.	.
.	.	.
K	n	$\hat{P}_k = \frac{n_k}{n}$

Considere uma variável ou população X com distribuição Binomial, isto é:  $X \in B(n, p)$ , com média  $E[X] = \mu_x = n.p$  e variância  $Var[X] = n.p.q$ , seja  $\hat{P} = \frac{X}{n} = \frac{n_i}{n}$  a proporção ou frequência relativa de sucessos na amostra. Sendo assim temos que:

$$\mu_{\hat{P}} = E[\hat{P}] = E\left[\frac{X}{n}\right] = \frac{1}{n} \cdot E[X] = \frac{1}{n} \cdot n.p = P$$

$$VAR[\hat{P}] = VAR\left[\frac{X}{n}\right] = \frac{1}{n^2} \cdot VAR(X) = \frac{1}{n^2} \cdot n.p.q = \frac{p.q}{n}$$

$$D.P.(\hat{P}) = \sigma_{(\hat{P})} = \sqrt{\frac{P.q}{n}}$$

Se  $n \geq 30$  então  $\hat{P} \in N\left[p, \frac{p.q}{n}\right]$ , sendo assim a variável normal padrão Z da variável aleatória  $\hat{P}$ , é dada

$$\text{por: } Z = \frac{\hat{P} - P}{\sqrt{\frac{p.q}{n}}}$$

**DISTRIBUIÇÃO AMOSTRAL DA DIFERENÇA ENTRE PROPORÇÕES**  $[\hat{P}_1 - \hat{P}_2]$ : Sejam duas populações binomialmente distribuídas com parâmetros  $P_1, q_1$  e  $P_2, q_2$  respectivamente. Onde  $P_1$  e  $P_2$  são as proporções de sucessos e  $q_1$  e  $q_2$ , as proporções de fracassos, sendo assim temos que:

$E[\hat{P}_1 - \hat{P}_2] = \mu_{\hat{P}_1 - \hat{P}_2} = \mu_{\hat{P}_1} - \mu_{\hat{P}_2} = P_1 - P_2$  e  $\sigma_{\hat{P}_1 - \hat{P}_2} = \sqrt{\sigma_{p_1}^2 + \sigma_{p_2}^2} = \sqrt{\frac{p_1 \cdot q_1}{n_1} + \frac{p_2 \cdot q_2}{n_2}}$ , sendo a variável

normal padronizada Z da variável aleatória  $[\hat{P}_1 - \hat{P}_2]$ , dada por:  $Z = \frac{(\hat{P}_1 - \hat{P}_2) - (P_1 - P_2)}{\sqrt{\frac{p_1 \cdot q_1}{n_1} + \frac{p_2 \cdot q_2}{n_2}}}$ .

**DISTRIBUIÇÃO AMOSTRAL DA VARIÂNCIA**  $[S^2]$ : Selecionando-se todas as amostras aleatórias possíveis, de tamanho “n”, de uma população, e calculando a variância de cada amostra, obtemos a distribuição amostral das variâncias.  $S^2 = \frac{1}{n}(X - \bar{X})^2$  isto é:  $S_1^2 = \frac{1}{n}(X_1 - \bar{X})^2$ ,  $S_2^2 = \frac{1}{n}(X_2 - \bar{X})^2$  ...

$$S_n^2 = \frac{1}{n}(X_n - \bar{X})^2$$

Estas variâncias podem ser consideradas como valores observados da mesma variável aleatória:  $S_i^2 = \frac{1}{n}(X_i - \bar{X})^2$ , a distribuição desta variável é a distribuição de amostragem da variância. A média

desta distribuição é, para toda a amostra aleatória simples dada por:

$$E[S^2] = E\left[\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2\right] = E\left[\frac{1}{n} \sum_{i=1}^n (X_i - m)^2 - (\bar{X} - m)^2\right] = \frac{1}{n} \sum_{i=1}^n E(X_i - \bar{X})^2 - E(\bar{X} - m)^2 = \sigma^2 - \frac{\sigma^2}{n}$$

Ou ainda:  $E[S^2] = \frac{(n-1) \cdot \sigma^2}{n}$ . Em média, a variância das amostras é, portanto, inferior à variância da

população, sendo a diferença tanto maior quanto menor for o tamanho da amostra. Quanto a variância da distribuição de amostragem da variância, demonstra-se que, no caso de uma população normal, ela é dada por:

$$Var[S^2] = E\left[S^2 - \frac{(n-1) \cdot \sigma^2}{n}\right]^2 = \frac{2 \cdot (n-1) \cdot \sigma^4}{n^2}$$

A distribuição da variância  $S^2$  segue uma distribuição teórica de probabilidade chamada modelo de qui-quadrado ( $\chi^2$ ) quando a população parente tem distribuição normal, ou seja se extrairmos amostras aleatórias de tamanho  $n$  de uma população normal, então a variável amostral

$$\frac{n \cdot S^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2}{\sigma^2} = \sum_{i=1}^n \left( \frac{X_i - \bar{X}}{\sigma} \right)^2 = \sum_{i=1}^n \left( \frac{X_i - \bar{X}}{\sigma} \right)^2$$

tem distribuição qui-quadrado com  $n-1$  graus de liberdade. Logo podemos escrever:

$$\chi_{n-1}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} = \frac{n-1}{\sigma^2} \cdot \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = \frac{(n-1) \cdot S_x^2}{\sigma^2}, \text{ donde resulta que } S_x^2 = \frac{\sigma^2}{n-1} \chi_{(n-1)}^2. \text{ Verifica-se,}$$

pois, que, a menos de uma constante, a estatística  $S^2$  se distribui conforme uma distribuição do tipo qui-quadrado ( $\chi^2$ ) com  $n-1$  graus de liberdade.

**DISTRIBUIÇÃO AMOSTRAL DO DESVIO PADRÃO [S]:** A média da distribuição amostral do desvio padrão é dada por:  $E[S] = \mu_{(s)} = \sqrt{\frac{n-1}{n}} \sigma$ . Os erros padrões da distribuição amostral do desvio padrão

S são dados por  $\sigma_s = \frac{\sigma}{\sqrt{2 \cdot n}}$  e  $\sigma_s = \sqrt{\frac{\mu_4 - \sigma^4}{4 \cdot n \cdot \sigma^2}}$ . Para  $n \geq 100$  a distribuição amostral de S é assintoticamente

normal ou muito aproximadamente normal.  $\sigma_s$  é dado por  $\sigma_s = \frac{\sigma}{\sqrt{2 \cdot n}}$  só se a população é normal ou

aproximadamente normal. Se a população é não normal, pode-se empregar  $\sigma_s = \sqrt{\frac{\mu_4 - \sigma^4}{4 \cdot n \cdot \sigma^2}}$ . Note-se que

o desvio padrão de S,  $\sigma_s = \sqrt{\frac{\mu_4 - \sigma^4}{4 \cdot n \cdot \sigma^2}}$  se reduz a  $\sigma_s = \frac{\sigma}{\sqrt{2 \cdot n}}$  quando  $\mu_{(4)} = 3 \cdot \sigma^4$  o que é verdadeiro para

populações normais. Para  $n \geq 100$ ,  $\mu_{(s)} = \sigma$  muito aproximadamente.

**DISTRIBUIÇÃO AMOSTRAL DA RAZÃO ENTRE VARIÂNCIAS**  $\left[ \frac{S_1^2}{S_2^2} \right]$ : Considere agora que de duas populações normais com a mesma variância  $\sigma^2$  ou o que seria equivalente, de uma mesma população normal, sejam extraídas duas amostras independentes com, respectivamente,  $n_1$  e  $n_2$  elementos e determine o quociente das variâncias dessas amostras. Utilizando a expressão  $S_x^2 = \frac{\sigma^2}{n-1} \cdot \chi_{n-1}^2$ , pode-se concluir que

a distribuição amostral desse quociente será uma distribuição  $F[(n_1-1; n_2-1)]$ , pois:

$$\frac{S_1^2}{S_2^2} = \frac{\left[ \frac{\sigma^2}{(n_1-1)} \right] \cdot \chi_{n_1-1}^2}{\left[ \frac{\sigma^2}{(n_2-1)} \right] \cdot \chi_{n_2-1}^2} = \frac{\chi_{n_1-1}^2}{\chi_{n_2-1}^2} = F_{(n_1-1), (n_2-1)}$$

**DISTRIBUIÇÃO BINOMIAL:** Uma distribuição de probabilidade discreta de uma variável representando o número de sucessos em testes nos quais só há dois resultados possíveis, sucesso com uma probabilidade fixa e falha ou fracasso.

**DISTRIBUIÇÃO DE FREQUÊNCIA:** Consultar distribuição.

**DISTRIBUIÇÃO DE FREQUÊNCIA EMPÍRICA:** Mostra a frequência de ocorrência das observações em um conjunto de dados.

**DISTRIBUIÇÃO DE FREQUÊNCIA RELATIVA CUMULATIVA:** Mostra as proporções acumuladas de indivíduos, os quais estão contidos em uma categoria ou classe e em todas as categorias.

**DISTRIBUIÇÃO DE POISSON:** Uma distribuição probabilística discreta da contagem do número de eventos que ocorrem aleatoriamente no tempo ou no espaço, em uma taxa constante na média.

**DISTRIBUIÇÃO DE PROBABILIDADE:** Uma distribuição teórica que especificamos matematicamente e utilizamos para calcular a probabilidade teórica de ocorrência de um evento.

**DISTRIBUIÇÃO DE PROBABILIDADE TEÓRICA:** Determinado a partir de um modelo matemático.

**DISTRIBUIÇÃO GAUSSIANA:** Consultar distribuição Normal ou Gaussiana.

**DISTRIBUIÇÃO LOGNORMAL:** É uma distribuição de probabilidade contínua. Caso os dados que são deslocados para a direita sejam transformados em log e a distribuição resultante seja Normal, diz-se que os dados se aproximam de uma distribuição Lognormal.

**DISTRIBUIÇÃO NORMAL OU GAUSSIANA:** Uma distribuição de probabilidade contínua. É uma distribuição com formato de sino, sendo aproximada por diversas variáveis biológicas.

**DISTRIBUIÇÃO NORMAL PADRÃO:** Uma distribuição Normal com média zero e desvio padrão um (1).

**DISTRIBUIÇÃO QUI-QUADRADO ( $\chi^2$ ):** Uma distribuição de probabilidade contínua frequentemente utilizada em testes de hipótese de proporção.

**DISTRIBUIÇÃO SIMÉTRICA:** A forma de distribuição à direita de um valor central é uma imagem espelho daquela distribuição à esquerda do valor central.

**DISTRIBUIÇÃO t DE STUDENT:** Consultar distribuição t.

**DISTRIBUIÇÃO-F:** Uma distribuição de probabilidade contínua. É utilizada para comparar variâncias.

**DISTRIBUIÇÃO-t:** Descoberta por Student, um pseudônimo de William Sealey Gosset, é uma distribuição de probabilidade contínua; a distribuição é simétrica ao redor da média, sendo caracterizada pelos graus de liberdade. À medida que aumentam os graus de liberdade, ela se torna mais parecida com a distribuição Normal.

**DISTRIBUIÇÕES DE FREQUÊNCIAS RELATIVAS:** Indicam a proporção ou porcentagem das observações em cada classe ou categoria de distribuição.

**DUMMY:** Consultar placebo.

**DUPLO-CEGO:** Em um teste clínico, nenhum dos tratadores dos animais, tampouco o assistente da resposta ao tratamento, teste ou controle, sabe qual tratamento cada animal ou paciente está recebendo.

**DECOMPOSIÇÃO:** O ato da divisão de uma série temporal em suas partes constituintes pelo uso de métodos estatísticos. Uma série temporal típica é usualmente composta de quatro partes: i) Um termo de longo prazo ou tendência; ii) oscilações de períodos ou amplitude mais ou menos regulares em torno da tendência; iii) uma componente sazonal; iv) uma componente aleatória. Em inglês Decomposition.

**DECOMPOSIÇÃO DE CHOLESKI:** Um dos métodos mais comuns para achar a matriz raiz quadrada é chamado de decomposição de Choleski. Em inglês Choleski decomposition.

**DEDUÇÃO:** Modelo lógico no qual expectativas específicas ou hipóteses são desenvolvidas com base em princípios gerais. Partindo do princípio de que todos os coordenadores são durões, pode-se prever que este coordenador não o deixará mudar de curso. Esta antecipação resulta de dedução.

**DEDUÇÃO:** Conclusão baseada em algumas proposições ou resultados de experiências.

**DEDUÇÃO:** Como método de raciocínio, é um conjunto de proposições particulares contidas em verdades universais. Vai do conhecimento geral para o particular. Quando se formulou uma hipótese, adequada aos fatos observados, torna-se possível aplicar as regras da lógica formal e deduzir várias conseqüências. É neste ponto que a lógica faz sua entrada na ciência.

**DEFASAGEM:** Um evento que acontece no tempo  $t+k$  ( $k>0$ ) é chamado defasado com relação ao tempo  $t$ , sendo a defasagem de ordem  $k$ . Em inglês Lag.

**DEFINIÇÃO:** É uma descrição mais curta e mais sumária de algum conceito.

**DEFINIÇÃO AXIOMÁTICA:** Em inglês axiomatic definition.

**DEFINIÇÃO CONCEITUAL:** Especificação da base teórica para um conceito representado por um fator.

**DEFINIÇÃO CONCEITUAL:** Em contraste à definição operacional, definição conceitual trata do relacionamento entre o conceito que tem-se em mente e outros conceitos que formam a teoria da qual ele faz parte.

**DEFINIÇÃO DO PROBLEMA:** Saber exatamente aquilo que se pretende pesquisar é o mesmo que definir corretamente o problema.

**DEFINIÇÃO DO PROBLEMA:** É a primeira fase do estudo estatístico e consiste na definição e formulação correta do problema a ser estudado.

**DEFINIÇÃO OPERACIONAL:** Definição concreta e específica de algo em termos das operações por meio das quais observações devem ser categorizadas. A definição operacional de tirar nota A neste curso poderia ser responder corretamente pelo menos 90 % das questões no exame final.

**DEFINIÇÃO OPERACIONAL:** Uma definição expressa em termos dos processos ou operações que serão usadas para medir as características em estudo. Esclarecimento de um termo, especificando as operações ou ações. Especifica o instrumento ou as operações a serem usadas para medir a variável.

**DISCREPÂNCIAS (OUTLIERS):** Dados inesperadamente grandes ou pequenos.

**DISPERSÃO (SPREAD):** Uma medida da variação dentro de um conjunto.

**DISSIMILARIDADE (DISSIMILARITY):** Quão diferentes dois objetos são. Normalmente quantificados com uma distância.

**DISTÂNCIA (DISTANCE):** Quão separados dois objetos estão. Há muitas formas para medir distâncias, a mais familiar é a distância euclidiana.

**DISTÂNCIA EUCLIDIANA (EUCLIDEAN DISTANCE):** Uma medida de distância ou dissimilaridade entre dois objetos. Se cada objeto pode ser descrito por um conjunto de coordenadas em um espaço  $n$ -dimensional. A distância euclidiana pode ser calculada como:

$$d_{i,j} = \sqrt{\sum_{k=1}^n (y_{i,k} - y_{j,k})^2}$$

**DISTRIBUIÇÃO DE PROBABILIDADES (PROBABILITY DISTRIBUTION):** A distribuição de resultados  $X_i$  de uma variável aleatória  $X$ . Para variáveis aleatórias contínuas, a distribuição de probabilidades é uma curva suave e um histograma apenas é uma aproximação da distribuição de probabilidades.

**DISTRIBUIÇÃO DE PROBABILIDADES A POSTERIORI (POSTERIOR PROBABILITY DISTRIBUTION):** A distribuição resultante da aplicação do Teorema de Bayes a um determinado conjunto de dados. Essa distribuição dá a probabilidade para o valor de qualquer hipótese de interesse conforme os dados observados.

**DISTRIBUIÇÃO DE PROBABILIDADES A PRIORI (PRIOR PROBABILITY DISTRIBUTION):** Uma de duas distribuições, a outra é a verossimilhança, usada no numerador do Teorema de Bayes para calcular a distribuição de probabilidades a posteriori. Ela define a probabilidade para o valor de qualquer hipótese de interesse antes que os dados sejam coletados. Ver também probabilidade a priori e verossimilhança.

**DISTRIBUIÇÃO UNIFORME (DISTRIBUIÇÃO UNIFORME DISCRETA EM  $n$  PONTOS)-DU( $i, j$ ):** Este é o caso mais simples de distribuição discreta, no qual cada possível valor da variável aleatória ocorre com a mesma probabilidade. A variável aleatória,  $X$ , diz-se que tem distribuição Uniforme discreta em,  $n$ , pontos quando tem função probabilidade,  $f(x_i) = P(X = x_i) = \frac{1}{n}$ ,  $i = 1, 2, \dots, n$ . A variável aleatória  $X$ ,

com distribuição uniforme discreta, com os  $k$  valores de seu domínio dados por  $x_1, x_2, \dots, x_k$ , tem função de probabilidade definida pela equação abaixo.  $P(X=x_i) = P(x_i) = f(x_i) = \frac{1}{k}$ , para  $i=1, 2, \dots, k$ . Uma variável

aleatória discreta  $X$  assumindo os valores  $x_1, x_2, \dots, x_k$ , terá distribuição uniforme discreta se todos os elementos forem equiprováveis.

**DISTRIBUIÇÃO DE BERNOULLI-BERNOULLI(P):** Muitos experimentos são tais que os resultados possíveis apresentam ou não uma determinada característica, normalmente de natureza qualitativa. i) uma moeda é lançada: o resultado ou é cara ou não é; ii) uma peça é escolhida, ao acaso, de um lote, contendo 500 peças; esta peça é defeituosa ou não defeituosa; iii) um parto é realizado numa matriz bovina, o bezerro macho ou é fêmea.

Em todos os casos, estamos interessados na ocorrência de um sucesso (ocorrência de cara, peça defeituosa, bezerro macho, etc.). Para cada experimento acima, podemos definir uma variável  $X$  que assume apenas dois valores: o valor 1, se ocorre sucesso e o valor 0, se ocorre fracasso indicamos  $p$  a probabilidade de sucesso,  $p(\text{sucesso}) = p(s) = p$ , onde  $0 < p < 1$ . Ensaios que resultam numa variável aleatória que assume apenas dois valores: sucesso [ $X = 1$ ] e fracasso [ $X = 0$ ], cujas probabilidades são respectivamente  $p$  e  $1 - p = q$  sendo que  $p + q = 1$  são denominados ensaios de Bernoulli. Dessa forma experimentos que resultam numa variável aleatória de Bernoulli são chamados ensaios de Bernoulli. A notação é descrita conforme a seguir:  $X \cap \text{Ber}(p)$  indica que a variável aleatória  $x$  possui distribuição de bernoulli com parâmetro  $p$ . A Função de probabilidade é dada por:  $P(X = x) = p^x(1 - p)^{1-x}$  para  $x = 0$  ou  $x = 1$ .

**DISTRIBUIÇÃO BINOMIAL OU SEQUÊNCIA DE BERNOULLI-Bin (n, p):** Esse modelo foi criado e deduzido em fins do século XVII pelo matemático Suíço Jacob Jacques Bernoulli (1654-1705). É a mais importante das distribuições teóricas para variáveis aleatórias discretas, e foi a primeira distribuição introduzida na estatística. podemos afirmar que a distribuição Binomial está para as distribuições discretas assim como a distribuição Normal está para as contínuas. Esta distribuição é muito aplicada em amostragem e em situações em que conhecemos o tamanho da amostras e sabemos quantas vezes é que um acontecimento ou evento ocorreu. Imagine agora que repetimos um ensaio de Bernoulli n vezes, ou, como se diz também, obtemos uma amostra de tamanho n de uma distribuição de Bernoulli. Suponha ainda que as repetições sejam independentes e, o resultado de um ensaio não tem influência nenhuma no resultado de qualquer outro ensaio. Uma amostra particular será constituída de uma seqüência de sucessos e fracassos ou se quisermos de uns e zeros.

Características de uma experiência binomial

Para cada ensaio, a variável aleatória pode assumir somente um dos dois valores sucessivo ( $X = 1$ ) ou fracasso ( $X = 0$ );

- i) Os ensaios repetidos são independentes;
- ii) O valor de P que é a probabilidade de sucesso permanece constante de ensaio para ensaio.
- iii) Um número fixo de ensaios (n) serão conduzidos;
- v) Interessa ao pesquisador o número X, quantidade de sucessos obtidos em n tentativas.

A função de probabilidade [P(x)] é obtida conforme o seguinte procedimento.

i) Seja p a probabilidade de ocorrência (sucesso) e  $q=1-p$ , a de não ocorrência (fracasso), evidentemente  $p + q = 1$ ;

ii) Admitimos que em n provas independentes ou repetições há x sucessos;

iii) A probabilidade de se obter x sucessos nas n provas é dada por:  $p.p.p \dots p.p.q.q.q \dots q = p^x q^{n-x}$ , devido a independência dos ensaios. Mas qualquer seqüência com x sucessos e n-x fracassos terá a mesma probabilidade. Multiplicando esta probabilidade pelo número de maneiras distintas de se conseguir X sucesso em n tentativas que é  $\binom{n}{x}$ , obtém-se o resultado pretendido. Portanto resta saber quantas seqüên-

cias com x sucessos e n-x fracassos podemos formar. É fácil ver que, considerando todas as n-úplas com x sucessos ou melhor como nada foi dito sobre a ordem dos sucessos, a probabilidade será obtida multiplicando-se o resultado acima por

$C(n, x) = \binom{n}{x}$ ; e a função de probabilidade apropriada é:

$P(X = x) = \binom{n}{x} \cdot p^x \cdot q^{n-x} = \frac{n!}{x!(n-x)!} \cdot p^x \cdot q^{n-x}$ , que é o termo geral do desenvolvimento do

Binômio de Newton,  $(p + q)^n = \sum_{x=0}^n \frac{n!}{x!(n-x)!} p^x q^{n-x}$  é claro que  $(p + q)^n = \sum_{x=0}^n P(x) = 1$ . Os termos

de p(x) dão as probabilidades dos vários resultados possíveis. Por exemplo, a probabilidade de nenhum sucesso ocorrer é dada por:  $P(X = 0) = P(0) = q^n$ .

**DISTRIBUIÇÃO DE POISSON-POISSON ( $\lambda$ ):** Esse modelo foi criado pelo matemático e físico francês Siméon Denis Poisson (1781-1840), em 1837. É um modelo usado para descrever as probabilidades do número de ocorrências de acontecimentos ou eventos raros, em geral num intervalo de tempo, distância, área ou volume os quais são contínuos, como por exemplo o número de ligações erradas num circuito telefônico. O número de automóveis que passam numa esquina, observe que pode-se anotar o número de automóveis que passaram que é o sucesso, porém o número de carros que deixaram de passar não poderá ser determinado o qual é o fracasso. Note-se que a unidade de medida como tempo, área ou volume é contínuo, mas a variável ou o número de ocorrências é discreta. Este modelo é denominado distribuição de eventos raros.

As características de um experimento ou processo de Poisson pode ser descrito conforme a seguir: Um processo de Poisson refere-se normalmente ao número de eventos que ocorrem num intervalo temporal ou numa região espacial, sendo assim a utilização da distribuição de Poisson, baseia-se nas seguintes hipóteses: i) A probabilidade de uma ocorrência é a mesma em todo o campo de observação; ii) A probabilidade de mais de uma ocorrência ou evento num único ponto ou intervalo muito pequeno é aproximadamente zero. Quando n cresce indefinidamente, p tende para zero; iii) O número de ocorrências ou eventos que ocorre em qualquer intervalo é independente do número de ocorrências ou eventos que ocorre em outros intervalos disjuntos, diz-se então que essa distribuição não tem memória.

A função de probabilidade desta distribuição é mostrada a seguir:

Se o resultado de uma variável aleatória discreta é descrita por uma distribuição de Poisson, então a probabilidade de se realizar um determinado número de ocorrências por unidade de medida tais como minuto, hora,  $cm^2$ ,  $cm^3$ , etc. é dado pela seguinte fórmula:

$P(x = k) = \frac{\lambda^k e^{-\lambda}}{k!}$ ,  $k = X \in \{0, 1, 2, \dots, \infty\}$ , onde,  $x = k =$  número de ocorrências; e = base do logaritmo

natural;  $e = 2,7182$ ;  $\lambda =$  taxa média por intervalo ou unidade a qual é uma constante positiva dada,  $\lambda = \mu t$ ;  $\lambda =$  número médio de eventos ocorrendo no intervalo considerado. A distribuição de Poisson é caracterizada por um único parâmetro que é a média do processo a qual é simbolizada pela letra do alfabeto grego lambda  $\lambda$ .

**DISTRIBUIÇÃO HIPERGEOMÉTRICA-HIPGEOM (N, K, n):** A distribuição Binomial é o modelo adequado para estudar as propriedades dos esquemas probabilísticos do seguinte tipo: de uma urna que contém, N, bolas, apenas diferentes na sua cor, das quais, K, são brancas e, N-K, são pretas,  $0 < p < 1$ ,  $p + q = 1$ , tiram-se, n, bolas, com reposição, isto é, devolvendo à urna cada bola logo após verificar a sua cor; qual a probabilidade de obter, x, bolas brancas? Este esquema presta-se a caracterizar um sem número de situações correntes, em que se colhem com reposição amostra de, n, elementos de um universo ou população com, N, elementos, dos quais, K, possuem determinado atributo tais como: são fumadores, clientes de uma certa marca, doadores de sangue, etc. e, N-K, não possuem esse atributo.

A função de probabilidade ou de massa de probabilidade é mostrada como segue: O modelo matemático dessa distribuição de probabilidade é dada por:

$$P(X = x) = \frac{\binom{k}{x} \binom{N-k}{n-x}}{\binom{N}{n}}$$



em que  $\text{Max}(0, n-N+k) \leq x \leq \text{min}(k, n)$ ,  $0 < p < 1$ ,  $P + q = 1$ . A notação desta variável aleatória é dada por:  $x \sim H(N, n; p)$ : Significa que a variável aleatória discreta  $X$  tem distribuição de probabilidade hipergeométrica.

**DISTRIBUIÇÃO GEOMÉTRICA OU DE PASCAL-GEOM (P):** A distribuição geométrica ou Pascal desempenha um importante papel na estatística. O nome distribuição geométrica ocorre porque os termos de sua função da probabilidade, para cada valor da variável aleatória, seguem uma série geométrica. Considerar uma seqüência de ensaios independentes de Bernoulli, cuja probabilidade do sucesso em cada ensaio é igual a  $p$ . A variável aleatória  $X$ , definida como sendo o número de ensaios requeridos até a ocorrência do primeiro sucesso, segue a distribuição geométrica. O valor de  $X(x)$  representa o número de fracassos consecutivos até a ocorrência do primeiro sucesso. O primeiro sucesso ocorrerá no ensaio  $x + 1$ , tendo  $x$  fracassos o antecedendo. Como os ensaios são independentes, os  $x$  fracassos têm probabilidade igual a  $(1-p)^x$  e o sucesso do ensaio  $x + 1$  tem probabilidade  $p$ . Então, para se obter a função de probabilidade do evento requerido, a qual está apresentada, a seguir basta multiplicar essas duas probabilidades. A distribuição geométrica é a única distribuição discreta que não tem memória, isto é, o número de tentativas que tem de se realizar, a partir de um dado momento até se conseguir o primeiro sucesso, não depende do número de tentativas já realizadas. Assim,  $P(X \leq k+c | X > k) = P(X \leq c)$ , ou de forma equivalente,  $P(X > k+c | X > k) = P(X > c)$ . A função apresentada abaixo é a função de probabilidades da distribuição geométrica. A função de probabilidade é dada pela seguinte equação:  $P(X = x) = (1-p)^x p$ , em que  $x=0, 1, 2, 3, \dots$  representa o número de falhas até a ocorrência do primeiro sucesso.

**DISTRIBUIÇÃO BINOMIAL NEGATIVA-BinNeg (r, p):** Quando se fixa o número de provas de Bernoulli, seja  $n$ , o número de “sucessos” é um variável aleatória que assume valores da sucessão finita,  $0, 1, \dots, n$ , e tem, como se viu na secção anterior, distribuição Binomial. Quando se fixa, em determinado problema, o número de sucessos, seja,  $K$ , e se considera o número de provas que têm de realizar-se até se obterem  $K$  sucessos então o número de provas, seja  $Y$ , passa a ser uma variável aleatória, assume valores da sucessão infinita,  $K, K+1, K+2, \dots$

É fácil ver que a variável aleatória  $Y$ , tem função de probabilidade ou de massa de probabilidade dada por:

$$f(y) = P(Y = y) = \binom{y-1}{y-k} p^k q^{y-k},$$

$$y = K, K+1, K+2, \dots, 0 < p < 1, q = 1 - p$$

De fato, se somente ao fim de,  $y$ , ( $y \geq K$ ), provas se completam,  $K$ , sucessos, na  $y$ -ésima prova tem de obter-se um sucesso, o  $K$ -ésimo, e nas,  $y-1$ , provas anteriores têm de obter-se,  $K-1$ , sucessos e,  $y-K$ , insucessos. Assim, considerando que as provas são independentes,  $P(Y = y) = \text{Prob}\{(K-1) \text{ sucessos em } (y-1) \text{ provas}\} \times \text{Prob}\{1 \text{ sucesso em } 1 \text{ prova}\}$ ,

$$P(Y = y) = \binom{y-1}{k-1} p^{k-1} q^{y-1-(k-1)} \cdot p$$

Em vez de,  $Y$ , é corrente empregar,  $X = Y-K$ , onde,  $X$ , exprime o número de falhas ou insucessos verificados até ocorrerem,  $K$ , sucessos. Se, por hipótese, se realiza uma prova por unidade de tempo,  $X$ , pode considerar-se o tempo de espera por,  $K$ , sucessos. A variável aleatória,  $X$ , tem função probabilidade dada por:

$$f(x) = P(X = x) = \binom{x+k-1}{x} p^k q^x, \quad x = 0, 1, 2, 3, \dots, \quad 0 < p < 1, q = 1 - p.$$

A variável aleatória com essa função de probabilidade diz-se que tem distribuição Binomial negativa. Essa família de distribuição de probabilidade é uma generalização da geométrica, ou seja, a geométrica é um caso particular da binomial negativa. A notação desta variável aleatória é dada por:  $X \sim \text{BN}(r; P)$ : significa que a variável aleatória  $X$  possui distribuição binomial negativa com parâmetros  $r$  e  $P$ .

**DISTRIBUIÇÃO MULTINOMIAL-Mult (n, p<sub>i</sub>):** Outra distribuição de probabilidade que desempenha um importante papel na estatística é a distribuição multinomial que é uma generalização da distribuição binomial. Seja  $A_1, A_2, \dots, A_k$  um conjunto exaustivo de  $k$  eventos mutuamente exclusivos, ou de atributos ou de classes e  $p_1, p_2, \dots, p_k$  com  $\sum_{i=1}^k p_i = 1$ , as correspondentes probabilidades associadas às ocorrências

desses eventos em um dado ensaio. Seja  $X_i$  a variável aleatória correspondente ao número de ocorrências no evento  $A_i$  e  $X_i$  o número de realizações do evento  $A_i$  ( $i = 1, 2, \dots, k$ ) em  $n$  independentes ensaios. Então:  $\sum_{i=1}^k x_i = n$  e  $\sum_{i=1}^k p_i = 1$ . O modelo ou equação que representa a função de probabilidade desta variável

aleatória é mostrada a seguir: A função de probabilidade ou de massa conjunta das  $k$  variáveis é denominada função de probabilidade multinomial e está apresentada abaixo:

$$P(X_1 = x_1, X_2 = x_2, \dots, X_k = x_k) = \frac{n!}{x_1! x_2! \dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}. \quad \text{Quando } k = 2, \text{ ou seja, quando há}$$

apenas duas classes ou dois atributos na população, a função de probabilidade acima se especializa na distribuição binomial.

**DISTRIBUIÇÃO BINOMIAL BINOMIAL TRUNCADA EM ZERO (0):** Uma outra importante distribuição teórica de probabilidade é a binomial truncada em 0. O modelo binomial é realista para caracterizar a distribuição da variável aleatória  $X$ , que, por alguma razão, o zero não pode ser observado. A função de probabilidade dessa distribuição é apresentada abaixo:

$$P(X = x) = \frac{\binom{n}{x} p^x q^{n-x}}{1 - q^n}; \quad x = 1, 2, \dots, n$$

**DISTRIBUIÇÃO DE POISSON TRUNCADA EM ZERO (0):** O modelo da distribuição teórica de probabilidade de Poisson truncada em zero é dada pela função de probabilidade abaixo:

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!(1 - e^{-\lambda})}; \quad x = 1, 2, 3, \dots, \infty.$$

**DISTRIBUIÇÃO BETA BINOMIAL:** Uma outra função de probabilidade, não menos importante, é a beta-binomial, muitas vezes usada para construção de planos de amostragem seqüencial ou em pesquisa de amostragem com vários amostradores ou entrevistadores envolvidos. A função de probabilidade dessa distribuição é apresentada na equação a seguir. Os parâmetros dessa distribuição são: a)  $n$ , inteiro ou não negativo, representando o tamanho da amostra; e b) os parâmetros reais  $\alpha$  e  $\beta$ , tais que  $\alpha > 0$  e  $\beta > 0$ . Essa distribuição surge naturalmente em situações em que  $n$  ensaios independentes de Bernoulli são realizados, mas cada ensaio tem probabilidade de sucesso diferente dos demais, e a distribuição beta (contínua) é usada para descrever a variabilidade de  $p$  nos diferentes ensaios.

$$P(X = x) = \binom{n}{x} \cdot \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(x + \alpha)\Gamma(n + \beta - x)}{\Gamma(n + \alpha + \beta)}; \quad x = 0, 1, 2, 3, \dots, n$$

em que  $\Gamma(m) = \int_0^\infty x^{m-1} e^{-x} dx$  é a chamada função gama. Se o argumento m dessa função é inteiro positivo, então  $\Gamma(m) = (m - 1)!$ . Se  $\alpha = \beta = 1$ , então a função de probabilidade em 20 se especializa na função de probabilidade uniforme discreta para  $x = 0, 1, 2, \dots, n$ .

**DISTRIBUIÇÃO DE PROBABILIDADE UNIFORME-U (a,b):** Uma variável aleatória contínua que admita distribuição constante em algum intervalo  $(a, b)$  e zero para valores externos é conhecida como distribuição retangular ou uniforme, a qual tem uso mais comum em primeira tentativa em casos em que apenas os limites dos dados são conhecidos. A distribuição uniforme é uma das mais simples distribuições e é usada comumente em situações em que não há razão para atribuir probabilidades diferentes a conjuntos possíveis de valores da variável aleatória em determinado intervalo, Por exemplo, o tempo de chegada de um vôo pode ser considerado distribuído uniformemente em certo intervalo de tempo; a distribuição da distância de posição de cargas em uma ponte, em relação a um pilar terminal, também pode ser representada adequadamente por uma distribuição uniforme sobre o vão da ponte. Não é demais observar que, por vezes, associamos uma distribuição uniforme a determinada variável aleatória, simplesmente por falta de informação mais precisa além do conhecimento do seu intervalo de valores. A função densidade de probabilidade  $[f(x)]$  deste modelo é definida com o segue. Uma variável aleatória possui distribuição uniforme em um dado intervalo se sua função densidade de probabilidade (f.d.p.) é dada por:

$$\int_a^b K dx = 1, \quad K(X)|_a^b = 1, \quad (b - a)K = 1, \quad K = \frac{1}{b - a}$$

$$f(x) = \begin{cases} \frac{1}{(b - a)} & \text{se } a \leq X \leq b \\ 0 & \text{caso contrário} \end{cases}$$

**DISTRIBUIÇÃO DE PROBABILIDADE NORMAL, SIMÉTRICA, CAMPANULAR OU GAUSSIANA-N ( $\mu, \sigma$ ):** Esse modelo de distribuição é conhecida também como distribuição de gauss ou curva gaussiana em honra ao maior matemático de seu tempo, o cientista alemão Karl Friedrich Gauss (1777-1855). Outras denominações são curva de Laplace-Gauss, curva normal de probabilidades, curva normal dos erros, em forma de sino de igreja, em forma campanular, em forma de chapéu de Napoleão, simétrica. O cientista Abrahan De Moivre (1667-1754) em 1733, desenvolveu a equação matemática e Gauss (1777-1855) em 1810 derivou sua equação partindo do estudo de erros em medidas repetidas de uma mesma medida na área da astronomia. Já o cientista Pierre Simon Laplace (1749-1827), usou o modelo em estudo no tratamento analítico de probabilidades. Ou seja, o primeiro aparecimento dessa distribuição se deu, de forma indireta em 1733, numa discreta publicação posteriormente incluída na segunda edição da obra *Doctrine of Chances* do matemático francês Abraham de Moivre (1667-1754), que proppunha, nessa publicação, o uso da integral de  $e^{-x^2}$ , como forma de aproximação ao binômio  $(a + b)^2$ , com n suficientemente grande. Grandes contribuições ao estudo da distribuição normal foram feitas pelo matemático francês Pierre Simon Laplace (1749-1827), através de sua obra máxima *Thèorie*

*Analytique dès probabilités*, publicada em 1812. A laplace, provavelmente, deve-se o resultado  $\int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi}$ , importantíssimo para o cálculo de áreas sob a curva normal. Como Laplace, o matemático

alemão Karl Friedrich Gauss (1777-1855) tinha grande interesse nas aplicações da matemática aos estudos de astronomia. Assim, estudando erros de medida, Gauss encontrou uma forma matemática para modelagem desse erros partindo de algumas hipóteses consideradas naturais, entte elas a de que o valor mais provável de uma variável cujos valores fossem todos dignos de confiança era a média aritmética. A Gauss é creditado o uso pioneiro da função  $f(X) = \frac{h}{\sqrt{\pi}} e^{-h^2 X^2}$ , como fórmula matemática para descrição probabilísitca dos erros de medida. Gauss ainda associava a h o valor  $\frac{1}{\sigma\sqrt{2}}$  em notação moderna, dizendo que, através dele, se poderia ter uma idéia sobre a precisão das medidas.

É a mais importante distribuição de probabilidade de variável contínua em todo o campo da estatística tanto sob o ponto de vista teórico quanto nas aplicações práticas, dela derivando todos os procedimentos conhecidos como estatística paramétrica. Seu gráfico conhecido como curva normal é uma curva de forma campanular que descreve de forma aproximada a distribuição de freqüências ou de probabilidades da maioria dos dados ou variáveis de mensuração que ocorrem nas ciências físicas, biológicas e sociais, ou na natureza, na indústria e nas pesquisas de maneira geral, por exemplo, gramas, cm, °C, etc. É um modelo constantemente utilizado para o desenvolvimento de estudos estatísticos de inferência, pis a base teórica da estatística matemática está na distribuição normal. Exemplos de variáveis aleatórias contínuas com distribuição normal: Temperatura média diária em Mossoró-RN, produção de soja no Mato Grosso, peso de bezerras ao nascer, etc. A função densidade de probabilidade  $[f(x)]$  desta variável é dada conforme a seguir: Uma variável aleatória contínua é normal ou Gaussiana, se sua função densidade de probabilidade (f. d. p.) tem a forma

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < \infty, \text{ onde, } e = 2,7183; \pi = 3,1416, X \Rightarrow \mathfrak{R}, \mu = \text{é a média da distri-}$$

buição (parâmetro de localização),  $\sigma =$  é o desvio padrão da distribuição (parâmetro de escala) ( $\sigma > 0$ ). A distribuição normal é caracterizada pelos parâmetros  $\mu, \sigma^2$  que são respectivamente sua média e variância, com  $\sigma^2 > 0$ . A notação usada é dada conforme a seguir:  $X \cap N(\mu, \sigma^2)$  ou  $X \sim N(\mu, \sigma^2)$ : isso significa que X tem distribuição normal de média  $\mu$  e variância  $\sigma^2$ . O gráfico da distribuição densidade de probabilidade normal tem um formato de sino, campânula, chapéu de Napoleão, ou pano de circo, como é mostrado na Figura abaixo.

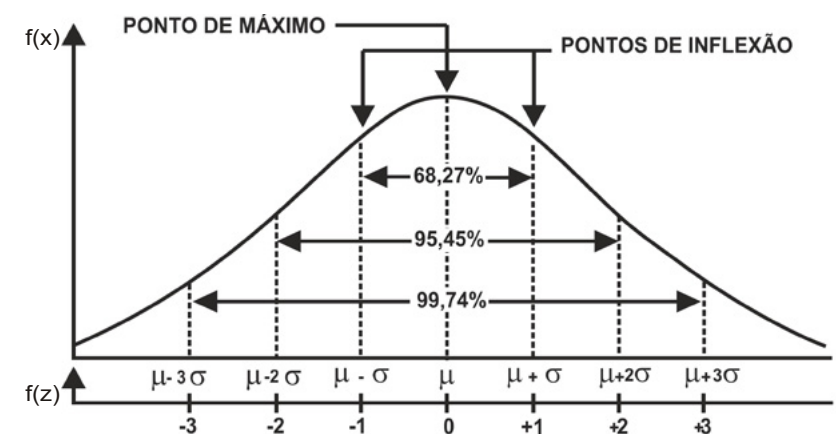


Figura: Gráfico representativo da distribuição normal  $[f(x)]$  e normal padrão  $[f(z)]$ .

**DISTRIBUIÇÃO NORMAL COMO APROXIMAÇÃO DA DISTRIBUIÇÃO BINOMIAL (APROXIMAÇÃO DE MOIVRE-LAPLACE):** Se  $n$  é grande e nem  $p$  nem  $q$  são demasiadamente próximos de zero, a distribuição binomial pode ser satisfatoriamente aproximada por uma distribuição normal com variável aleatória padronizada, dada por  $Z = \frac{x - n \cdot p}{\sqrt{n \cdot p \cdot q}}$ . Aqui  $x$  é a variável aleatória que dá o número de

sucessos em  $n$  provas de Bernoulli e  $p$  a probabilidade de sucesso. A aproximação melhora com o crescer de  $n$  e é exata no caso limite. Na prática, a aproximação é muito boa, se  $np$  e  $nq$  são ambos maiores que 5. Pode-se descrever o fato de que a distribuição teórica de probabilidade binomial tende para a distribuição normal, escrevendo

$$\lim_{n \rightarrow \infty} P\left(a \leq \frac{x - np}{\sqrt{n \cdot p \cdot q}} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{u^2}{2}} du.$$

Em palavras significa que a variável aleatória padronizada  $\frac{(x - np)}{\sqrt{n \cdot p \cdot q}}$  é assintoticamente normal. A justificativa formal de tal aproximação é dada pelo chamado teorema do limite central.

**DISTRIBUIÇÃO NORMAL COMO APROXIMAÇÃO DA DISTRIBUIÇÃO DE POISSON:** Pode-se mostrar que se  $X$  é a variável de Poisson e  $\frac{(x - \lambda)}{\sqrt{\lambda}}$  é a correspondente variável aleatória padronizada,

então  $\lim_{\lambda \rightarrow \infty} P\left(a \leq \frac{x - \lambda}{\sqrt{\lambda}} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{u^2}{2}} du$ . Isto é a distribuição de Poisson tende para a distribuição normal quando  $\lambda \rightarrow \infty$  ou seja,  $\frac{x - \lambda}{\sqrt{\lambda}}$  é assintoticamente normal. Condições a serem satisfeitas: quando

a média lambda tiver os seguintes valores:  $\lambda \geq 5$ ,  $\lambda \geq 10$ ,  $\lambda \geq 15$ ,  $\lambda = \mu \geq 10$  ou  $\lambda \rightarrow \infty$ . Sendo assim a variável aleatória padronizada ou reduzida dada por

$$Z = \frac{x - \mu}{\sigma} = \frac{x - \lambda}{\sqrt{\lambda}} \sim N(0,1), \text{ onde } \begin{cases} \mu = \lambda \\ \sigma = \sqrt{\lambda} \end{cases} \text{ segue o modelo normal de probabilidades.}$$

**DISTRIBUIÇÃO DE PROBABILIDADE EXPONENCIAL-EXP ( $\lambda$ ):** As variáveis aleatórias exponenciais surgem naturalmente no tempo decorrente entre ocorrências consecutivas de variáveis Poisson. Esse tempo decorrente para que ocorra um sucesso é modelado pela distribuição exponencial. Para ilustrar situações em que a modelagem exponencial é adequada pode-se considerar a variável  $X$  como sendo o tempo decorrente entre ocorrências sucessivas de uma doença tropical bovina rara: o tempo de vida de componentes eletrônicos; o tempo decorrente entre terremotos consecutivos em regiões com falhas geológicas; o tempo decorrido até a falha romper ou quebrar uma viga de aço; o tempo decorrido até a falha de um trator; entre outros. A distribuição exponencial é a versão contínua da distribuição geométrica, a qual é discreta. A variável aleatória com densidade exponencial é conhecida como variável aleatória contínua de tempo de espera, ou seja, ela refere-se ao tempo necessário para a ocorrência do primeiro sucesso (FERREIRA, 2005). O modelo da função densidade de probabilidade  $[f(X)]$  desta variável aleatória ou estocástica é descrita conforme a seguinte equação:

$$f(X) = \begin{cases} \lambda \cdot e^{-\lambda \cdot X} & \text{Para } X > 0 \\ 0 & \text{para outros valores de } x. \end{cases}$$

**DISTRIBUIÇÃO DE PROBABILIDADE DE QUI-QUADRADO -  $\chi^2_{(v)}$ :** Estabelecida em 1876, pelo físico e astrônomo alemão Friedrich Robert Helmert (1843-1917), e reexaminada em 1900, pelo estatístico inglês Karl Pearson (1857-1936). Esse modelo possui várias aplicações em estatística. Uma delas é a de propiciar mecanismos para a realização de inferências sobre o parâmetro de uma população normal. Outra aplicação refere-se aos testes de falta de ajuste de um modelo teórico aos dados observados em um experimento ou levantamento amostral. A definição da variável e conforme a seguir: Seja  $(Z_1, Z_2, \dots, Z_v)$  uma amostra aleatória simples, retirada de uma distribuição normal padronizada  $[N(0,1)]$ . Então a variável  $Y = \chi^2 = Z_1^2 + Z_2^2 + \dots + Z_v^2 = \sum_{i=1}^v Z_i^2$ , tem distribuição de qui-quadrado com  $V$  graus de liberdade (g.l.). A

função densidade de probabilidade (F.D.P.) desta variável aleatória é mostrada como a seguir: A variável aleatória  $Y$  possui distribuição Qui-quadrado ( $\chi^2$ ), se sua função densidade de probabilidade for da forma:

$$f(\chi^2) = \frac{1}{2^{\frac{v}{2}} \cdot \left(\frac{v-2}{2}\right)!} (\chi^2)^{\frac{v-2}{2}} \cdot e^{-\frac{\chi^2}{2}}, \quad 0 < \chi^2 < \infty.$$

A distribuição teórica de probabilidade de Qui-quadrado é caracterizada pelo parâmetro  $v$  que é o número de graus de liberdade, o qual corresponde ao número de variáveis normais independentes ( $Z_i$ ) ao quadrado que estão sendo somadas.

**DISTRIBUIÇÃO DE PROBABILIDADE t de Student-T<sub>(v)</sub>:** A distribuição t de Student é uma distribuição de probabilidade contínua e tal como distribuição de qui-quadrado só tem um parâmetro, o número de graus de liberdade, que é um número inteiro positivo, tem notável aplicação em estatística, principalmente em estudos envolvendo pequenas amostras ( $n < 30$ ), pois proporciona os elementos necessários para se realizar inferências sobre médias populacionais. Foi estabelecida em 1908 pelo estatístico inglês William Sealey Gosset (1876-1937), o qual realizou uma importante publicação a seu respeito, e que adotou o pseudônimo de Student ao publicar seus trabalhos, durante a primeira parte do século XX, Gosset publicou a maioria dos seus trabalhos contra a vontade de seus empregadores de uma cervejaria que o considerava top secret, por isso que no início do século XX ele adotou o pseudônimo de Student. Em 1925, o estatístico inglês Ronald Aylmer Fisher (1890-1962), deu rigorosa demonstração dessa distribuição.

A definição da variável aleatória de t de student é mostrada como segue: Seja  $Z$  uma variável aleatória normal padrão  $[N(0,1)]$ , e  $U$  uma variável aleatória qui-quadrado ( $\chi^2_{(v=1)}$ ), independente de  $U$  com  $v$  graus de liberdade. Então a variável aleatória  $t = \frac{Z}{\sqrt{\frac{U}{v}}}$  possui distribuição t de STUDENT, com  $v$  graus de liberdade.

A função densidade de probabilidade (F.D.P.) desta variável aleatória é mostrada como segue: Uma variável aleatória contínua, possui distribuição t de “Student” se sua função densidade de probabilidade for da forma:

$$f(t) = \frac{\left[\frac{v-1}{2}\right]!}{\sqrt{\pi v} \left[\frac{v-2}{2}\right]!} \cdot \frac{1}{\left[1 + \frac{t^2}{v}\right]^{\frac{(v+1)}{2}}}, \quad -\infty < t < +\infty$$

Essa distribuição é simétrica, leptocúrtica para pequenas amostras e aproxima-se da distribuição Normal a medida que o tamanho da amostra  $n$  aumenta, apresenta um só máximo para  $t = 0$  e admite como assíntota o eixo dos  $t$ .



**DISTRIBUIÇÃO DE PROBABILIDADE F DE SNEDECOR (1881-1974)-Fisher (1890-1962)-F<sub>(v1,v2)</sub>:** Esse modelo foi estabelecido pelo estatístico inglês Ronald Aylmer Fisher, pois Fisher estudou a distribuição de Z de Fisher a qual não deve ser confundida com a variável Z que é a variável reduzida da curva normal, que historicamente precedeu a de F, sendo que esta foi obtida a partir daquela para se evitar o uso de logaritmos neperianos. Tem-se a seguinte relação entre as duas variáveis.

$$e^{2z} = F \therefore z = \frac{1}{2} \ln F \therefore 2z \cdot \ln F = \log F$$

$$\therefore z = \frac{1}{2} \cdot \frac{\log F}{\log e} = \frac{1}{2} \log_e F \text{ ou } z = \frac{1}{2} \ln F$$

A distribuição F está entre aquelas distribuições de probabilidade mais importantes na estatística com inúmeras aplicações, tendo ainda, um maior destaque na estatística experimental, onde a mais importante delas é o seu emprego nas análises de experimentos. Nesse caso, o investigador científico tem por objetivo comparar os efeitos de dois ou mais tratamentos sob determinadas condições. A hipótese de igualdade de efeitos de tratamentos é testada usando a distribuição de probabilidade F. A definição desta variável é dada por: Sejam U e V duas variáveis aleatórias independentes, cada uma distribuída segundo uma distribuição de qui-quadrado, com v<sub>1</sub> e v<sub>2</sub> graus de liberdade respectivamente. Então a variável aleatória

$$F = \frac{\frac{U}{v_1}}{\frac{V}{v_2}} = \frac{\frac{\chi^2_{v_1}}{v_1}}{\frac{\chi^2_{v_2}}{v_2}}, \text{ possui distribuição de F com } v_1 \text{ e } v_2 \text{ graus de liberdade que são os parâmetros.}$$

A função densidade de probabilidade desta variável aleatória é como segue: Uma variável aleatória F possui distribuição de f com v<sub>1</sub> e v<sub>2</sub> graus de liberdade se sua função densidade de probabilidade for da forma:

$$f(F) = \frac{\frac{(v_1 + v_2 - 2)}{2}}{\left[\frac{v_1 - 2}{2}\right]! \left[\frac{v_2 - 2}{2}\right]!} \cdot v_1^{\left(\frac{v_1}{2}\right)} \cdot v_2^{\left(\frac{v_2}{2}\right)} \cdot \frac{F^{\left(\frac{v_1 - 2}{2}\right)}}{\left(v_1 \cdot F + v_2\right)^{\left(\frac{v_1 + v_2}{2}\right)}}, \quad 0 \leq F \leq \infty$$

Essa função é uma função contínua cuja variável aleatória varia de 0 a +∞, é assimétrica à direita e depende dos parâmetros v<sub>1</sub> e v<sub>2</sub>, que são os graus de liberdade. A notação da variável é dada por: f ∩ F(v<sub>1</sub>, v<sub>2</sub>).

**DISTRIBUIÇÃO DE PROBABILIDADE GAMA- Gama(α, β):** A distribuição Gama de dois e três parâmetros tem sido largamente utilizada em hidrologia com a finalidade de modelar as freqüências de cheias anuais. A função densidade de probabilidade é dada como: Uma variável aleatória contínua x com (0 < x < ∞) se distribui segundo uma distribuição gama de parâmetros α > 0 e β > 0 se sua função de densidade de probabilidade é:

$$f(x) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} & \text{para } 0 < x < +\infty \\ 0 & \text{para caso contrário} \end{cases}$$

Se o parâmetro α for inteiro, esta família de distribuições de probabilidade recebe o nome de distribuições de Erlang. Essa distribuição tem papel de destaque em telefonia e recebeu esse nome em homenagem ao engenheiro dinamarquês Erlang, que empregou essa distribuição pela primeira vez para modelar duração de chamadas telefônicas por volta de 1917.

**DISTRIBUIÇÃO DE PROBABILIDADE DE ERLANG-Erl(α, β):** A distribuição de Erlang é um caso especial da distribuição gama. Se o parâmetro α de uma variável aleatória de Erlang não for um número, porém α > 0, então a variável aleatória terá uma distribuição gama. No entanto, na função densidade de Erlang, o parâmetro α aparece como α fatorial. Por conseguinte, de modo a definir uma variável aleatória gama, requeremos uma generalização da função fatorial. Uma variável aleatória exponencial descreve o comprimento até que a primeira contagem seja obtida em um processo de Poisson. Uma generalização da distribuição exponencial é o comprimento até que α contagens ocorram em um processo de Poisson. A variável aleatória que é igual ao comprimento do intervalo até que α contagens em um processo de Poisson é uma variável aleatória de Erlang. A função densidade de probabilidade [f(X)] é definida como: A variável aleatória X, que é igual ao comprimento do intervalo até que α falhas ocorram em um processo de Poisson, com média β > 0, tem distribuição de Erlang com parâmetros α e β. A função densidade de probabilidade de X é dada por:

$$f(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{(\alpha - 1)!}, \text{ para } x > 0 \text{ e } \alpha = 1, 2, \dots, \text{ Uma variável aleatória de Erlang com } \alpha = 1 \text{ é uma variável}$$

aleatória exponencial. Probabilidades envolvendo as variáveis aleatórias de Erlang são frequentemente determinadas calculando o somatório das variáveis aleatórias de Poisson. Além disso a função densidade de probabilidade de uma variável aleatória de Erlang pode ser usada para determinar probabilidades; entretanto, a integração por partes é frequentemente necessária sendo assim tem-se de ter o cuidado para definir a variável aleatória e o parâmetro em unidades consistentes.

**DISTRIBUIÇÃO DE PROBABILIDADE DE CAUCHY- Cauchy(μ, σ):** A distribuição de Cauchy é uma distribuição simétrica, em forma de sino. Sendo que o interesse nessa distribuição se deve ao fato de não possuir nenhum momento (nem função geradora de momentos), isto é, todos os momentos são divergentes. Representa assim um caso atípico extremo que é usado em estatística para testar hipóteses ou conjecturas. Pode ser também demonstrado que o quociente de duas variáveis aleatórias normais padrões é uma variável aleatória com distribuição de Cauchy. Este fato deve alertar o leitor ou pesquisador para o risco de se obter resultados insperados ao se trabalhar com razões de variáveis aleatórias normais. A função densidade de probabilidade [f(X)] desta variável é dada por:

$$f(X) = \frac{1}{\sigma\pi} \cdot \frac{1}{1 + \left(\frac{x - \mu}{\sigma}\right)^2}, \quad x \in \mathfrak{R}. \text{ Para } \mu = 0 \text{ e } \sigma = 1, \text{ obtém-se a chamada distribuição padronizada, que}$$

é o caso particular da distribuição t de Student com 1 grau de liberdade.

**DISTRIBUIÇÃO DE PROBABILIDADE DE PARETO-PARETO(α):** Esta é um modelo de distribuição de probabilidade frequentemente usada em estudos aplicados de economia principalmente em conexão com problemas de distribuição de rendas de populações. A função densidade de probabilidade [f(x)] desta variável aleatória é dada conforme a seguir. Diz-se que uma variável aleatória contínua tem distribuição de Pareto com parâmetros α > 0 e b > 0, se sua função densidade de probabilidade [f.d.p.] é dada por:



$$f(X) = \begin{cases} \frac{\alpha}{b} \left(\frac{b}{x}\right)^{\alpha+1}, & \text{se } x \geq b \\ 0, & \text{se } x < b \end{cases}$$

Na equação acima, b pode representar algum nível mínimo de renda, e X

é o nível de renda e  $f(X) \cdot \Delta X$  dá a proporção de indivíduos com renda entre X e  $X + \Delta X$ .

Pode-se observar que a equação acima pode de  $f(X)$  pode também ser escrita na seguinte forma:

$$f(X) = \begin{cases} \alpha \cdot b^\alpha \cdot X^{-\alpha-1}, & \text{se } X \geq b \\ 0, & \text{se } X < b \end{cases}$$

**DISTRIBUIÇÃO-F (F-DISTRIBUTION):** A distribuição esperada de uma variável aleatória F, calculada como uma variância sendo que em geral a soma dos quadrados entre grupos ou SQentre grupos dividida por outra, normalmente a soma dos quadrados dentro de grupos, ou SQdentro de grupos.

**DISTRIBUIÇÃO NORMAL PADRÃO (STANDARD NORMAL DISTRIBUTION):** Uma distribuição normal de probabilidades com média = 0 e variância = 1.

**DISTRIBUIÇÃO-T (T-DISTRIBUTION):** É um modelo o qual é uma modificação da distribuição de probabilidades normal padrão. Para tamanhos amostrais pequenos, a distribuição t de Student é leptocúrtica, mas, conforme o tamanho amostral aumenta, a distribuição t se aproxima de uma distribuição normal padrão. As probabilidades de cauda da distribuição-t são usadas para calcular intervalos de confiança.

**DISTRIBUTIVO (DISTRIBUTIVE):** Em matemática, ou na teoria de conjuntos, é a propriedade de que  $c(a + b) = ca + cb$ .

**DEFLATOR PIB:** Medida do nível médio de preços.

**DELINEAMENTO:** Modo de atribuir os tratamentos às parcelas, num experimento. Também conhecido como ensaio.

**DELINEAMENTO:** A parte do ensaio que especifica os tratamentos que serão avaliados, as unidades experimentais, a variável em análise e o modo como tratamentos serão designados às unidades experimentais. Veja também desenho.

**DELINEAMENTO:** Conjunto específico de estímulos conjuntos criado para exibir as propriedades estatísticas específicas de ortogonalidade e balanço.

**DELINEAMENTO AXIAL:** É aquele em que, a partir de um ponto central, variam-se os níveis de cada fator separadamente, um só de cada vez. Também se chama FAO-ANDA.

**DELINEAMENTO AXIAL:** É um tipo de delineamento de uso comum em regiões pioneiras, para barateamento da pesquisa, apresenta como vantagem um número relativamente pequeno de tratamentos, mas tem como defeito o fato de não permitir a avaliação das interações, e é conhecido também como delineamento FAO-ANDA.

**DELINEAMENTO BALANCEADO:** Delineamento de estímulos no qual cada nível dentro de um fator aparece um número igual de vezes.

**DELINEAMENTO DE LIGAÇÃO:** Delineamento de estímulos para muitos fatores ou atributos no qual os atributos são quebrados em grupos menores. Cada grupo de atributo tem alguns atributos contidos em outros grupos, de modo que os resultados de cada um podem ser combinados ou ligados.

**DELINEAMENTO DE PESQUISA EM EPIDEMIOLOGIA:** Abordagem epidemiológica utilizada para a busca dos objetivos de uma investigação, elegendo-se uma metodologia adequada, um instrumento válido de coleta de informações, seguindo-se com a seleção de informações nos trabalhos, de campo, de laboratório ou outros, sobre uma determinada questão de pesquisa.

**DELINEAMENTO EXPERIMENTAL:** É o plano utilizado na experimentação e implica na forma como os tratamentos serão designados às unidades experimentais e em um amplo entendimento das análises a serem feitas quanto todos os dados estiverem disponíveis. Como exemplo, podemos citar delineamento inteiramente casualizado, delineamento em blocos casualizados e outros.

**DELINEAMENTO DE TRATAMENTOS:** É o plano que possibilita o pesquisador combinar nas unidades experimentais os diferentes níveis e combinação de níveis, referente aos diferentes tratamentos a serem ensaiados. Como por exemplo, podem ser citados os experimentos fatoriais e o de parcelas subdivididas.

**DELINEAMENTO DE UMA INVESTIGAÇÃO:** O mesmo que estrutura, arquitetura, configuração, modelo, protocolo, desenho ou design da pesquisa; os traços gerais e específicos da metodologia da investigação, tais como: o tamanho da amostra, a existência ou não de grupo controle, as características dos participantes, as observações necessárias, a frequência da coleta de dados e outros aspectos técnicos, que permitam alcançar os objetivos de estudo.

**DELINEAMENTO EM BLOCOS CASUALIZADOS:** É o tipo de delineamento mais comum e frequentemente o melhor, e é usado praticamente para todos os tipos de experimento. e ainda permite trabalhar com material experimental menos homogêneo. No entanto há redução do número de graus de liberdade do resíduo, por isso só é útil quando é pequena a heterogeneidade das parcelas dentro de cada bloco.

**DELINEAMENTO EM BLOCOS CASUALIZADOS COM ALGUNS TRATAMENTOS COMUNS:** Delineamento de uso generalizado em experimentos de competição de cultivares numerosos. É um delineamento robusto e flexível, de análise fácil, valioso especialmente para competições de cultivares em número elevado. É um delineamento que praticamente não apresenta defeitos, e substitui com vantagem os látices em geral, mas não é apropriado para experimentos de adubação.

**DELINEAMENTO EM BLOCOS CASUALIZADOS COM PARCELAS SUBDIVIDIDAS:** É um tipo de delineamento pouco comuns, sendo recomendados para tipos especiais de experimento, como por exemplo, um ensaio fatorial de cultivares de mangueira versus tipos de podas. Permite maior precisão na comparação de tratamentos atribuídos às subparcelas, mas a análise é mais complicada e ele apresenta dificuldade em certas comparações de médias, e ainda serve para aplicar-se em ensaios fatoriais.

**DELINEAMENTO EXPERIMENTAL:** Plano de pesquisa no qual o pesquisador manipula ou controla diretamente uma ou mais variáveis preditoras ou tratamentos e avalia seus efeitos sobre as variáveis dependentes. Comum nas ciências físicas, está se tomando popular nos negócios e em ciências sociais. Por exemplo, os respondentes são confrontados com anúncios separados que variam sistematicamente em uma característica, tais como diferentes apelos como emocional versus racional ou tipos de apresentação tais como colorido versus preto-e-branco, e são então questionados sobre suas atitudes, avaliações ou sentimentos em relação aos diferentes anúncios.

**DELINEAMENTO EXPERIMENTAL, PASSOS:** i) Obtenha um grande número de sujeitos que são pretestados na variável dependente; ii) Designe cada sujeito ao grupo experimental ou ao grupo controle;

iii) Controle cuidadosamente para detectar diferenças entre os dois grupos; iv) mensure outra vez ambos os grupos na variável dependente, após o tratamento experimental.

**DELINEAMENTO FATORIAL:** Delineamento com mais de um fator ou tratamento. Os delineamentos fatoriais examinam os efeitos de diversos fatores simultaneamente, formando grupos baseados em todas as possíveis combinações de níveis ou valores das diversas variáveis de tratamento.

**DELINEAMENTO FATORIAL:** Método para planejar estímulos para avaliação gerando todas as possíveis combinações de níveis. Por exemplo, uma análise conjunta de três fatores com três níveis por fator (3 x 3 x 3) resultaria em 27 combinações que poderiam atuar como estímulos.

**DELINEAMENTO FATORIAL:** As condições são de tal forma que todos os níveis de cada variável são combinadas com todos os níveis da outra. Exemplo de um experimento com quatro grupos.

	PRETESTE	SEM PRETESTE	
TRATAMENTO	GRUPO 1	GRUPO 3	EFEITO
SEM TRATAMENTO	GRUPO 2	GRUPO 4	PRINCIPAL DE TRATAMENTO
	EFEITO PRINCIPAL DE PRETESTE		

**DELINEAMENTO FATORIAL COMPLETOS:** É um tipo de delineamento bastante comuns, especialmente o fatorial de 3x3x3 e o 4x4. Têm como vantagem a grande eficiência quando as interações são pequenas, mas tem como defeito o fato de que o número de tratamentos cresce rapidamente. Além disso, esse tipo é geralmente usado com confundimento de algumas interações, para reduzir o tamanho dos blocos.

**DELINEAMENTO FATORIAL FRACIONÁRIO:** É um tipo de delineamento de uso restrito, geralmente em experimentos industriais ou ensaios agrícolas com fertilizantes, e tem como vantagem a grande redução do número de tratamentos, mas apresenta alguns defeitos como uma análise mais complicada e o confundimento de certos efeitos de tratamentos, o que pode levar a dubiedade de interpretação.

**DELINEAMENTO FATORIAL FRACIONAL:** Abordagem, como uma alternativa a um delineamento fatorial, que emprega apenas um subconjunto dos possíveis estímulos necessários para estimar os resultados com base na regra de composição assumida. Sua tarefa primária é reduzir o número de avaliações coletadas embora ainda mantenha a ortogonalidade entre os níveis e as estimativas sub-sequentes das utilidades parciais. O delineamento mais simples é um modelo aditivo, no qual apenas efeitos principais são estimados. Se termos de interação selecionados são incluídos, estímulos adicionais são criados. O delineamento pode ser criado consultando fontes publicadas ou usando programas de computador que acompanham a maioria dos pacotes de análise conjunta.

**DELINEAMENTO EM BLOCOS INCOMPLETOS EQUILIBRADOS:** É um tipo de delineamento pouco comum, mas muito úteis em casos especiais, em que os blocos devam ser bem pequenos, sendo assim a análise estatística é fácil e a redução do tamanho dos blocos trazem algumas vantagens. No entanto ele não é um ensaio robusto, pois a perda de tratamentos cria sérios problemas de análise, e ainda pode ser usado em experimentos de adubação.

**DELINEAMENTO EM QUADRADO LATINO:** É o tipo de delineamento pouco usado na experimentação vegetal, mas razoavelmente comum nos ensaios com animais especialmente com vacas leiteiras. Ele

tem dois tipos de blocos que são as linhas e as colunas e permite, pois, elimina melhor a heterogeneidade do material experimental, no entanto há grande redução no número de graus de liberdade do resíduo, e só se usa, em geral, para 4 a 8 tratamentos.

**DELINEAMENTO EM BLOCOS COMPLETOS CASUALIZADOS:** Tipo de delineamento em que sempre que não haja homogeneidade das condições experimentais devemos utilizar o princípio do controle local. Esse delineamento leva em consideração os três princípios básicos da experimentação: repetição, casualização e controle local. As principais características deste delineamento são: i) as parcelas são distribuídas em grupos ou blocos, o mais uniformes possíveis; ii) o número de parcelas por bloco deve ser um múltiplo do número de tratamentos que geralmente corresponde ao próprio número de tratamentos e iii) os tratamentos são designados às parcelas de forma casual. Existe o delineamento em blocos casualizados ou DBC com mais de uma repetição de tratamento por bloco.

**DELINEAMENTO EM BLOCOS CASUALIZADOS BALANCEADOS COM GRUPOS DE TRATAMENTOS DIFERENTES-GROUP BALANCED BLOCK DESIGN (GBBD):** A característica principal deste delineamento é o grupamento de tratamentos dentro de blocos homogêneos baseado em características dos tratamentos. i) Tratamentos pertencentes ao mesmo grupo são sempre testados no mesmo bloco; ii) Tratamentos pertencentes ao mesmo grupo são comparados com um maior grau de decisão do que aqueles pertencentes a grupos diferentes. Este experimento é comumente usado em ensaios de variedades, onde variedades com caracteres morfológicos similares são colocados no mesmo grupo. Dois dos critérios mais comumente usados para agrupar as variedades são: Altura de planta: evitar o efeito de competição, Duração do crescimento: minimizar a competição e facilitar as operações de colheita. Outro tipo de ensaio usando este delineamento é aquele envolvendo controle químico de insetos na qual os tratamentos podem ser subdivididos em operações de pulverização similares para facilitar a aplicação do inseticida no campo.

**DELINEAMENTO EM CLASIFICAÇÃO HIERÁRQUICA OU DELINEAMENTO EM SUB-AMOSTRAS:** Tipo de delineamento que é usado devido a necessidade de algumas vezes ser necessário amostrar subunidades das unidades experimentais para adquirir os dados necessários para um estudo de unidade observacional neste caso é uma sub-amostra tomada de uma unidade experimental maior. As sub-amostras introduz uma outra fonte de variabilidade para as observações em adição aquelas entre tratamentos e entre unidades experimentais.

**DELINEAMENTO EM QUADRADOS LATINOS-DQL:** Tipo de delineamento que apesar de sua alta eficiência, constitui-se no delineamento estatístico menos utilizado na pesquisa agrônoma, por exemplo, por ter menos flexibilidade. Eles levam em consideração três princípios básicos: repetição, casualização e controle local. i) controle local: controla a heterogeneidade do ambiente de suas maneiras diferentes. ii) casualização: os tratamentos são distribuídos nos blocos de tal forma que cada um apareça uma só vez em cada linha e em cada coluna. os quadrados latinos mais usados são de 5 x 5 a 8 x 8. os menores de 5 x 5 devem ser usados com mais de uma repetição e fazer a análise conjunta. os maiores de 8 x 8 torna-se difícil a execução e análise do ensaio. As vantagens são: i) controla a heterogeneidade do ambiente em duas direções; ii) conduz a estimativas menos elevadas do erro experimental e iii) vários delineamentos do mesmo tamanho pode ser combinados. Já as desvantagens são as seguintes: i) o número de linhas, colunas e tratamentos deve ser igual; ii) delineamentos pequenos fornecem poucos graus de liberdade (g.l.) para estimação da variação residual e iii) os quadrados latinos não são apropriados se os efeitos de linha, coluna e tratamentos não são aditivos, ou seja se eles interagem em qualquer combinação. na subamostragem do delineamento em quadrado latino (DQL) a análise de com subamostras é quase idêntica aquele para um

delineamento em blocos completos casualizados (DBC) com subamostras. a única diferença é devido a inclusão de um segundo fator de bloqueamento no DQL.

**DELINEAMENTO EM RETÂNGULO LATINO-DRL:** Tipo de delineamento em que é assumido que um dos fatores de bloqueamento é aumentado, mas para manter as características de equilíbrio necessário, este fator deve ser um múltiplo do número de tratamentos. Suponha que os blocos colunas sejam aumentados para dar um arranjo retangular de UE de tratamento  $t \times mt$ . Por exemplo, um delineamento retângulo latino  $3 \times 6$  com 3 tratamentos é representado pelo seguinte esquema de casualização.

A	C	C	B	B	A
C	B	A	C	A	B
B	A	B	A	C	C

**DELINEAMENTO EM QUADRADO LATINO REPETIDO (DQLR):** Tipo de delineamento em que é assumido que delineamento em quadrado latino (DQL) padrão é repetido  $s$  vezes. Os mesmos  $t$  tratamentos são usados em cada dos  $s$  quadrados. Assim os tratamentos são cruzados com os quadrados de modo que a interação tratamento  $\times$  quadrado possa existir.

**DELINEAMENTO QUADRADO GREGO-LATINO (DQGL):** Tipo de delineamento em que as seguintes condições são requeridas: i) O delineamento Quadrado Grego-Latino permite remover três fontes de variação do erro experimental; ii) Neste delineamento as unidades experimentais são agrupadas em três diferentes maneiras. Estas são agrupadas pelas linhas, colunas e por classificação extra, designado por letra grega; iii) A designação deste agrupamento é restrita, de tal modo que, cada letra grega ocorre uma vez, e apenas uma vez, em cada linha e cada coluna; iv) As letras gregas formam um quadrado latino com relação as linhas e colunas; v) Os tratamentos designados pelas letras latinas, são agora designadas as unidades experimentais de tal maneira que cada tratamento ocorre uma vez, e apenas uma vez, em cada linha, coluna e cada letra grega e vi) Os tratamentos formam assim um quadrado latino diferente nas linhas e colunas, ou que é independente do quadrado formado pelas letras gregas. As vantagens deste tipo de delineamento são as seguintes: i) Ele permite controlar a variabilidade para três fontes de variação. Já as desvantagens são as seguintes: i) O número de unidades experimentais (UE's) aumenta rapidamente quando o número de tratamento aumenta; ii) O número de graus de liberdade (G.L.) do erro é pequeno se o número de tratamento é pequeno; iii) Perda de dados complica extremamente a análise estatística e iv) Cancelamento nos três agrupamentos é raramente possível. As restrições são: i) O DQGL tem sido construído para todo número de tratamento; ii) Teoricamente, estes delineamentos são possíveis para qualquer  $t$  desde que  $t$  seja um número primo ou uma potencia de um número primo. Do ponto de vista prático, os limites são  $5 \leq t \leq 12$ , por que:

Se  $t = 3$  grau de liberdade do erro-GLE (erro) = 0

Se  $t = 4$  GLE = 3, GLE muito pequeno

Se  $t = 12$  GLE = 99 (144 Unidades experimentais-UE's) } número UE's muito grande

Devido suas limitações, o quadrado grego-latino tem sido raramente usado na pesquisa.

Quadrado Grego Latino 4 x 4 (básico)

	1	2	3	4
1	C $\gamma$	A $\beta$	D $\alpha$	B $\delta$
2	D $\delta$	B $\alpha$	C $\beta$	A $\gamma$
3	A $\alpha$	C $\delta$	B $\gamma$	D $\beta$
4	B $\beta$	D $\gamma$	A $\delta$	C $\alpha$

**DELINEAMENTO (EXPERIMENTO) EM FAIXAS:** Tipo de delineamento usado em experimentos onde o ensaio em faixa é especialmente adequado para o experimento com dois fatores em que a precisão desejada para medir o efeito da interação entre os dois fatores é mais alta do que para os efeitos principais de um dos fatores. Isto é realizado com o uso de três tamanhos de parcelas. Parcela com faixa vertical para o primeiro fator. Parcela com faixa horizontal para o segundo fator. Parcela interação para a interação entre os dois fatores. A parcela com faixa vertical e horizontal é sempre perpendicular entre si. Não há nenhum relacionamento entre seus tamanhos. A parcela interação é a menor. Os graus de precisões associados com os efeitos principais de ambos os fatores são sacrificados para aumentar a precisão do efeito da interação. Este experimento é típico em ensaios de irrigação, aplicação de fungicidas, uso de diferentes cortes em campineiras, épocas de colheitas, etc.

**DELINEAMENTO (EXPERIMENTO) EM PARCELAS SUBSUBDIVIDIDAS (SPLIT-SPLIT-PLOT):** O experimento em parcelas subdivididas é uma extensão do experimento em parcelas subdivididas para acomodar um terceiro fator é exclusivamente adequado para experimento com 3 fatores, onde três diferentes níveis de precisão são desejados para os vários efeitos. Este experimento é caracterizado por duas características importantes: i) Há 3 tamanhos de parcelas correspondendo aos três fatores e ii) Há 3 níveis de precisão, com a parcela recebendo o menor grau de precisão e a subparcela recebendo o maior grau de precisão.

**DELINEAMENTO (EXPERIMENTO) EM FAIXAS COM PARCELAS SUBDIVIDIDAS (SPLIT-SPLIT-PLOT):** O tipo de experimento em faixas com parcelas subdivididas é uma extensão do experimento em faixas na qual a parcela intersecção é subdivida para acomodar um terceiro fator. As principais características são: i) Há 4 tamanhos de parcelas- a faixa horizontal, a faixa vertical, a parcela intersecção e a subparcela; ii) ha 4 níveis de precisão com que os efeitos dos vários fatores são medidos com o nível mais alto correspondendo ao fator subparcelas e suas interações com outros fatores. Por exemplo, considere o caso em que A: 3 níveis de N B: 6 variedades C: 2 métodos de plantio. Faça a casualização para o fator vertical e fator horizontal (variedades) e divida cada parcela intersecção em cada uma repetição em subparcelas e faça a casualização seguindo o mesmo esquema do experimento em parcelas subdivididas.

**DELINEAMENTO (EXPERIMENTO) EM BLOCOS BALANCEADOS COM GRUPOS DE TRATAMENTOS DIFERENTES EM PARCELA SUBDIVIDA:** Tipo de delineamento em que a característica principal deste experimento é o agrupamento de tratamentos dentro de blocos homogêneos baseado em características do tratamento. Tratamentos pertencentes ao mesmo grupo são sempre testados no mesmo bloco. Tratamentos pertencentes ao mesmo grupo são comparados com um maior grau de precisão do que aqueles pertencentes a grupos diferentes. Este experimento é comumente usado em ensaio de variedades



onde variedades com caracteres morfológicos similares são colocados mesmo entre. Dois critérios mais comumente usados para agrupar as variedades são: Altura de planta-evitar o efeito de competição; duração do crescimento-minimizar a competição e facilitar as operações de colheita. Outro tipo de ensaio usando este experimento é aquele envolvendo controle químico de insetos na qual os tratamentos podem ser subdivididos em operações de pulverização similares para facilitar a aplicação do inseticida no campo. Por exemplo, as 45 variedades são criadas de acordo com seu período de crescimento:  $S_1$ -menos de 105 dias - $S_2$ -105 a 115 dias- $S_3$ -maior que 115 dias e são testadas em duas quantidades de adubo ( $A_1$  e  $A_2$ ).

**DELINEAMENTOS (EXPERIMENTOS) FATORIAIS:** Tipo de delineamento em que o tipo de experimento multifator mais comumente usado é o experimento fatorial, sendo aqueles nos quais são estudados, ao mesmo tempo, os efeitos de dois ou mais tipos de tratamentos ou fatores. Cada subdivisão de um fator é denominada de nível do fator. Os tratamentos nos experimentos fatoriais consistem de todas as combinações possíveis entre os diversos fatores nos seus diferentes níveis.

**DELINEAMENTOS FATORIAIS COM TRATAMENTOS ADICIONAIS OU FATORIAIS AUMENTADOS:** É qualquer experimento tendo um fatorial completo de tratamento mais um ou mais tratamentos adicionais. Suponha que um experimento fatorial  $3 \times 2 + 2$  foi conduzido na UFERZAM. Três doses de nitrogênio (20, 30 e 40 libras) em combinação com duas doses de fósforos (5 e 10 libras) mais dois tratamentos adicionais:  $C_1$  = tratamento padrão versus fatorial  $3 \times 2$   $C_2$  = Testemunha versus todos os outros, com três repetições foram delineados para os efeitos desses fatores na produção do algodão. Delineamento experimental de um delineamento em blocos completos casualizados (DBCC). Número de tratamentos= $3 \times 2 + 2 = 8$ .

**DELINEAMENTOS (EXPERIMENTO) FATORIAIS COM SUBAMOSTRAS:** São tipos ou modelos de delineamentos usados na pesquisa onde em muitos experimentos vários fatores estão envolvidos, alguns podem ser fatoriais ou cruzados com outros; alguns podem ser subamostrais dentro dos níveis dos outros fatores. Quanto tanto os fatores cruzados e em subamostras aparecem no mesmo experimento ele é chamado experimento fatorial com subamostra.

**DELINEAMENTOS (EXPERIMENTO) COM FATORES EM SUBAMOSTRAS:** Tipo de delineamento quando cada nível da hierarquia representa um fator, neste experimento há uma hierarquia de observação com subamostras das unidades experimentais. Os tratamentos representam os níveis mais altos e as subamostras representam os níveis mais baixos da hierarquia.

**DELINEAMENTOS (EXPERIMENTO) EM SUBAMOSTRAS COM DIFERENTES NÍVEIS DE FATORES:** É um experimento que tem número diferentes de níveis de fatores que são subamostras dentro dos outros fatores. Os níveis dos fator B subamostras dentro do fator A pode variar de um nível do fator A a outra de tal maneira que os graus de liberdades para o QMA e  $\frac{QMB}{A}$  serão mais ou menos iguais.

**DELINEAMENTO (EXPERIMENTO) EM PARCELAS SUBDIVIDIDAS (SPLIT-PLOT):** O experimento em parcelas subdivididas ou Split-Plot é próprio para experimentos com 2 fatores (A e B), em que a resposta de um deles é mais importante. Consiste em colocar nas parcelas um fator que deve ser o menos importante e subdividir esta parcela com o outro fator. Nestes experimentos as parcelas são divididas em pares iguais, denominados de subparcelas, e podem ser distribuídas em qualquer delineamento estatístico, sendo mais utilizados os delineamentos inteiramente casualizados (DICs) e os delineamentos em blocos completos casualizados (DBCCs). De acordo com a estruturação das subparcelas, podemos distinguir dois tipos de

experimentos em parcelas subdivididas: i) Parcelas subdivididas no espaço: quando em cada parcela há uma subdivisão da sua área em subáreas. Constituindo cada uma delas, uma subparcelas. Por exemplo: parcelas, cultivares de milho e sua área subdivididas subáreas, cada uma delas com espaçamentos diferentes. ii) Parcelas subdivididas no tempo: quando as parcelas não se subdividem, mas, periodicamente, são tomados dados em cada uma delas, constituindo estas tomadas as subparcelas. Por exemplo: nas parcelas diferentes cultivares de manga, e a cada ano avaliar a produção de frutos sempre nas mesmas parcelas. Cada constitui uma subparcela. A principal característica dos experimentos em parcelas subdivididas está na forma como é feita a casualização dos dois grupos de tratamentos. A casualização é feita em dois estágios: Primeiro casualiza-se os níveis do fator A nas parcelas. Em seguida casualiza-se os níveis do fator B nas subparcelas. O delineamento em parcelas subdivididas é desejável nas seguintes situações: Ele pode ser usado quando os tratamentos associados aos níveis de um dos fatores exigem maior quantidade de material na unidade experimental do que os tratamentos do outro fator. Por exemplo: Métodos de preparo do solo nas parcelas grandes. O outro fator, variedades nas parcelas menores. O delineamento pode ser usado quando um fator adicional é incorporado num experimento, para ampliar seu objetivo. Por exemplo: Principal objetivo: comparar os efeitos de diversos fungicidas contra infecção por uma doença, neste caso diversas variedades são incluídas para ampliar o objetivo do experimento para ver a resistência da planta. As variedades são distribuídas nas parcelas e os fungicidas nas subparcelas. Largamente usado em ensaios envolvendo: adubação mineral e calagem; irrigação e densidade de plantio; métodos de ensaio e recursos audiovisuais; etc. Devido ao tamanho da parcela e precisão de medidas dos efeitos não serem o mesmo para os fatores, as seguintes diretrizes são sugeridas quando designar um particular fator: i) grau de precisão: para um maior grau de precisão para o fator B do que o fator A, aloque o fator B as subparcelas e o fator A a parcela. ii) Tamanho relativo dos efeitos principais: se o efeito principal de um fator (fator B) é esperado ser maior e mais fácil de detectar do que algum do outro fator (A). O fator B deve ser designado as parcelas e o fator A subparcelas. Práticas de manejo: as práticas cultivares por um fator pode ditar o uso de parcelas grandes. Por conveniência prática, tal fator deve ser designado as parcelas. Experimento de manejo d'água e adubação. As vantagens deste delineamento são: i) melhor utilização dos recursos dando maior eficiência; ii) permitem estudar os efeitos principais dos fatores; iii) é mais prático de distribuir que o fatorial, o que os tornam muitas vezes prescrito pelos pesquisadores. Já as desvantagens são: i) o número de tratamentos cresce rapidamente quando o número de níveis aumenta, dificultando a instalação do experimento e ii) menos eficiente, do ponto de vista estatístico, que os fatoriais pois possuem dois resíduos.

**DELINEAMENTOS EM RETICULADOS QUADRADOS (SQUARE LATTICES OU TWO-DIMENSIONAL LATTICES):** Tipo de delineamento utilizado na experimentação moderna quando o número de tratamentos é elevado. Os reticulados quadrados nos permitem dispor um número  $v=k^2$  de tratamentos em b blocos de k parcelas. O número de tratamentos deve ser um quadrado perfeito, como por exemplo, 25, 36, 49, 64, 81, 100, etc. Nestes delineamentos os tratamentos de um bloco em uma repetição se distribuem por todos os blocos de qualquer das outras repetições. Nos reticulados podem ser realizadas análises intrablocos ou com recuperação da informação interblocos. Além dos reticulados quadrados existem os cúbicos (cubic lattices), os retangulares (rectangular lattices) e outros, todos de menor importância. Há também os quadrados reticulados (lattices squares), em que as parcelas se dispõem em linhas e colunas, como acontece nos delineamentos em quadrados latinos.

**DELINEAMENTO DE REVERSÃO:** Delineamento específico para experimentos de alimentação de vacas leiteiras. Há dois tipos: a) o de reversão simples, com 3 períodos para cada animal, com o mesmo



tratamento aplicado nos períodos 1 e 3; b) o de reversão dupla, com 4 períodos para cada animal, em que se aplica um tratamento aos períodos 1 e 3, e outro aos períodos 2 e 4.

**DELINEAMENTO DE REVERSÃO:** É um tipo usado em pesquisas de alimentação de vacas leiteiras, e tem como vantagem a grande eficiência quando bem conduzido, mas apresenta defeitos tais como efeitos residuais de um tratamento sobre os que se aplicam à mesma vaca, em períodos subseqüentes. É conhecido na língua inglesa como switch back.

**DELINEAMENTO EXPERIMENTAL:** É o mesmo que delineamento do experimento.

**DELINEAMENTO FAO-ANDA:** Ver delineamento axial.

**DELINEAMENTO INTEIRAMENTE CASUALIZADO:** É o tipo bastante comum, e aplica-se a todos os tipos de ensaios, mas de preferência, quando o material experimental é homogêneo e é pequeno o número de tratamentos. Já o número de graus de liberdade do resíduo é o maior possível. No entanto se o material experimental for heterogêneo o desvio padrão residual será alto e o ensaio pouco eficiente.

**DELINEAMENTO ÓTIMO:** Delineamento de estímulos que é ortogonal e balanceado.

**DELINEAMENTO QUADRADOS RETICULADOS:** É um tipo de delineamento raramente usado. Apresentam linhas e colunas, e são péssimos delineamentos, praticamente inúteis.

**DELINEAMENTO RETICULADO (OU LÁTICE) QUADRADO:** É o tipo de delineamento bastante usado para competição de cultivares de milho, mas não serve para ensaios com fertilizantes. Presta-se à competição de cultivares numerosos (100, 121, dentre outros), e o número de tratamentos deve ser um quadrado perfeito. A análise estatística é complicada, e ele não é um ensaio robusto, pois a perda de tratamentos inutiliza o delineamento. Este tipo de delineamento pode ser substituído com grande vantagem, pelos blocos casualizados com alguns tratamentos comuns.

**DELINEAMENTO RETICULADOS (OU LÁTICES) CÚBICOS:** É um tipo de delineamento pouco usados. Aplicam-se apenas, e raramente em ensaios de competição de cultivares muito numerosos. Permitem o uso de blocos bem pequenos o que é uma vantagem, no entanto não é robusto, e pode ser substituído, com vantagem, pelos blocos casualizados com alguns tratamentos comuns.

**DELINEAMENTO RETICULADOS RETANGULARES:** É um tipo de delineamento raramente usado, e se aplica quando há  $K(K+1)$  tratamentos, a serem postos em blocos de  $K$  parcelas, no entanto é um tipo de considerado como um péssimo delineamento, praticamente inútel.

**DELINEAMENTO ROBUSTO:** É um delineamento que pode manter suas propriedades mesmo quando há perdas graves de parcelas, como, por exemplo, perda total de tratamentos ou de blocos.

**DELINEAMENTO ROTATIVO:** Delineamento específico para experimentos de alimentação de vacas leiteiras em que cada vaca recebe vários tratamentos diferentes, em períodos sucessivos.

**DELINEAMENTO ROTATIVO:** Delineamento bastante comum nas pesquisas de alimentação de vacas leiteiras, apresenta como vantagem uma maior precisão, pois se usa cada vaca, em vários períodos, como bloco, mas apresenta defeitos como efeitos residuais de um tratamento sobre os que se aplicam à mesma vaca, em períodos subseqüentes, e são denominados de change over em inglês.

**DELPHI, TÉCNICA:** Tem como objetivo a convergência de opinião de um grupo, buscando consenso estatístico.

**DENDROGRAMA:** Representação gráfica do tipo gráfico em árvore dos resultados de um procedimento hierárquico no qual cada objeto é colocado em um eixo e o outro eixo representa os passos no procedimento hierárquico. Começando com cada objeto representado como um agrupamento separado, o dendrograma mostra graficamente como os agrupamentos são combinados em cada passo do procedimento até que todos estejam contidos em um único agrupamento.

**DENDROGRAMA:** Diagrama em forma de árvore que sumariza um processo de agrupamento de objetos em classes hierárquicas ou análise de cluster. Os casos semelhantes são unidos em grupos cujas posições no dendrograma são determinadas pelo nível de (dis) similaridade entre eles.

**DENDROMETRIA:** Disciplina fundamental da Ciência Florestal que estuda os equipamentos e métodos de medição das dimensões das árvores. É objetivo da dendrometria a determinação do diâmetro, altura, volume, peso, forma do tronco, idade e crescimento das árvores. A maioria das outras disciplinas florestais e dos trabalhos científicos na área utiliza-se de conceitos e métodos da dendrometria. Palavra composta pelos radicais gregos dendron que sigfica árvore e metron que indica medida.

**DENSIDADE APARENTE:** É a densidade da rocha ou mineral de minério, para conversão de volume em tonelagem. Também é conhecido na literatura como fator tonelagem. Veja que este parâmetro é de suma importância, pois qualquer erro na sua determinação reflete-se proporcional e diretamente nas reservas determinadas. Por isso, recomenda-se que a densidade aparente seja rotineiramente determinada, principalmente quando o minério for heterogêneo em sua composição.

**DENSIDADE DA CLASSE:** É a frequência da classe dividida pela sua amplitude.

**DENSIDADE DA CLASSE:** É a frequência da classe dividida pela sua amplitude. Em inglês class density.

**DENSIDADE DA FLORESTA:** Número de indivíduos de todas as espécies arbóreas presentes numa determinada área.

**DENSIDADE DE AMOSTRAGEM:** Refere-se a quantidade e distribuição espacial de amostras num dado corpo de minério, que garanta a estimativa de teor e tonelagem dentro de uma margem aceitável de variação.

**DENSIDADE DE ESPÉCIES:** Número de espécies presente em determinada unidade de área.

**DENSIDADE DE INCIDÊNCIA:** Frequência ou densidade de novos eventos por pessoa-tempo. A densidade de incidência é especialmente útil no estudo de taxas de frequência de doenças ou eventos que ocorram mais de uma vez em um mesmo indivíduo, com otites médias, resfriados ou internações hospitalares.

**DENSIDADE DE INCIDÊNCIA:** Frequência ou densidade de novos eventos por pessoa-tempo como pessoa-meses ou pessoa-anos. A densidade de incidência é especialmente útil quando o evento de interesse como por exemplo, resfriados, otites médias ou infartos do miocárdio pode ocorrer em uma mesma pessoa mais de uma vez, durante o período de estudo.

**DENSIDADE DE VALOR EXTREMO:** Densidades de probabilidades associadas a valores extremos. Em inglês Extreme value density.

**DENSIDADE-DEPENDÊNCIA:** i) Efeitos cujas intensidades mudam com o aumento da densidade da população. ii) Efeitos cujas intensidades aumentam com o aumento da densidade da população, por exemplo, uma mortalidade que se torna muito alta próximo à capacidade de carga. Os efeitos que se tornam menos pronunciados em populações de grandes tamanhos talvez devido à redução da fecundidade causada pela dificuldade de acasalar são algumas vezes chamado inverso densidade-dependente. Efeitos densidade-dependente são frequentemente causa de não-linearidade em um modelo.

**DEPENDÊNCIA ESPACIAL:** Condição que caracteriza a associação existente entre observações devido a causas espaciais. Em inglês Spatial dependence.

**DOBRADIÇA (HINGE):** A posição dos quartis superior e inferior em um gráfico de caule e folha.

**DP:** Ver desvio padrão S.

**DEPRECIAÇÃO:** Perda de valor de bens de capital em consequência de desgaste.

**DERIVA:** Está relacionada ao fato da variável regionalizada apresentar uma certa tendência de variação previsível por uma função determinística. Esta função é geralmente explicada por polinômios de baixo grau (1 ou 2).

**DERIVA GENÉTICA:** Mudanças aleatórias na frequência gênica em uma população, geralmente pequena, devido a erros de amostragem, havendo tendência de fixar-se um determinado alelo.

**DADOS CATEGÓRICOS:** Dados consistindo em contagens ou observações que podem ser classificadas em categorias. As categorias podem ser descritivas.

**DADOS QUALITATIVOS:** Dados derivados de atributos não-numéricos, tais como sexo, origem étnica ou nacionalidade ou outra variável de classificação.

**DADOS QUANTITATIVOS:** Dados na forma de medidas numéricas ou contagens.

**DEFEITO:** Usado em controle estatístico de qualidade, um defeito é um tipo particular de não conformidade em relação a especificações ou requisitos. Algumas vezes, defeitos são classificados em tipos, tais como defeitos de aparência e defeitos funcionais.

**DEMING:** William Edward Deming (1900-1993) foi um líder no uso do controle estatístico de qualidade.

**DESVIO PADRÃO:** A raiz quadrada positiva da variância. O desvio padrão é a medida mais largamente usada para expressar a variabilidade.

**DESVIO PADRÃO DA AMOSTRA:** A raiz quadrada da variância da amostra. O desvio-padrão da amostra é a medida mais largamente usada na variabilidade de dados amostrais.

**DESVIO PADRÃO DA POPULAÇÃO:** Veja desvio padrão.

**DIAGNÓSTICO DE REGRESSÃO:** Técnicas para examinar um modelo ajustado de regressão, de modo a investigar a adequação do ajuste e determinar se qualquer uma das suposições básicas tem sido violada.

**DIAGRAMA CAUSA E EFEITO:** Um gráfico usado para organizar as várias causas potenciais de um problema. Também chamado de diagrama espinha de peixe.

**DIAGRAMA DE CONCENTRAÇÃO DE DEFEITOS:** Uma ferramenta de qualidade que mostra graficamente a localização de defeitos em uma peça ou em um processo.

**DIAGRAMA DE DISPERSÃO:** Um diagrama contendo as observações de duas variáveis, x e y. Cada observação é representada por um ponto mostrado em suas coordenadas x-y. O diagrama de dispersão pode ser muito efetivo em revelar a variabilidade conjunta de x e y ou a natureza da relação entre elas.

**DIAGRAMA DE RAMO E FOLHAS:** Um método de disposição de dados em que o ramo corresponde a uma faixa de valores e a folha representa o próximo dígito. É uma alternativa ao histograma, porém dispõe as observações individuais em vez de colocá-los em retângulos.

**DISPERSÃO:** A quantidade de variabilidade exibida pelos dados.

**DISTÂNCIA DE COOK:** Em regressão, a distância de Cook é uma medida da influência de cada observação individual sobre as estimativas dos parâmetros do modelo de regressão. Ela expressa a distância que o vetor das estimativas dos parâmetros do modelo com a *i*-ésima observação removida está a partir do vetor das estimativas dos parâmetros do modelo baseado em todas as observações. Valores grandes da distância de Cook indicam que a observação é influente.

**DISTRIBUIÇÃO AMOSTRAL:** A distribuição de probabilidades de uma estatística. Por exemplo, a distribuição amostral da média da amostra  $\frac{N-n}{N-1}$  é a distribuição normal.

**DISTRIBUIÇÃO BIMODAL:** Uma distribuição de duas modas.

**DISTRIBUIÇÃO BIVARIADA:** A distribuição de probabilidades conjuntas de duas variáveis aleatórias.

**DISTRIBUIÇÃO BIVARIADA NORMAL:** A distribuição conjunta de duas variáveis aleatórias normais.

**DISTRIBUIÇÃO BIMODAL:** Uma distribuição de duas modas.

**DISTRIBUIÇÃO BIVARIADA:** A distribuição de probabilidades conjuntas de duas variáveis aleatórias.

**DISTRIBUIÇÃO BIVARIADA NORMAL:** A distribuição conjunta de duas variáveis aleatórias normais.

**DISTRIBUIÇÃO CONTÍNUA:** Uma distribuição de probabilidades para uma variável aleatória contínua.

**DISTRIBUIÇÃO DE FREQUÊNCIAS:** Um arranjo das frequências de observações em uma amostra ou população, de acordo com os valores que as observações assumem.

**DISTRIBUIÇÃO DE PROBABILIDADES:** Para um espaço amostral, uma descrição do conjunto de resultados possíveis, juntamente com um método para determinar probabilidades. Para uma variável aleatória, uma distribuição de probabilidades é uma descrição da faixa, juntamente com um método para determinar probabilidades.

**DISTRIBUIÇÃO DE PROBABILIDADES CONDICIONAIS:** A distribuição de uma variável aleatória dada em que o experimento aleatório produz um resultado em um evento. O evento dado deve especificar valores para uma ou mais variáveis aleatórias.

**DISTRIBUIÇÃO DE PROBABILIDADES CONJUNTAS:** A distribuição de probabilidades para duas ou mais variáveis aleatórias em um experimento aleatório. Veja função de probabilidade conjunta e função densidade de probabilidade conjunta.

**DISTRIBUIÇÃO DE PROBABILIDADES MARGINAIS:** A distribuição de probabilidades de uma variável aleatória obtida por meio da distribuição de probabilidade conjunta de duas ou mais variáveis aleatórias.

**DISTRIBUIÇÃO DE REFERÊNCIA:** A distribuição de uma estatística de teste quando a hipótese nula é verdadeira. Algumas vezes, uma distribuição de referência é chamada de distribuição nula da estatística de teste.

**DISTRIBUIÇÃO DISCRETA:** Uma distribuição de probabilidades para uma variável aleatória discreta.

**DISTRIBUIÇÃO F:** A distribuição de uma variável aleatória definida como razão a de duas variáveis aleatórias independentes qui-quadrado, cada uma sendo dividida pelo seu número de graus de liberdade.

**DISTRIBUIÇÃO GAUSSIANA:** Outro nome para a distribuição normal, baseado na forte conexão de Karl Friedrich Gauss para a distribuição normal, frequentemente usada em aplicações de física e de engenharia elétrica.

**DISTRIBUIÇÃO MULTINOMIAL:** A distribuição de probabilidades conjuntas de variáveis aleatórias que conta o número de resultados em cada uma das k classes em um experimento aleatório, com uma série de tentativas independentes com probabilidade constante de cada classe em cada tentativa. Ela generaliza uma distribuição binomial.

**DISTRIBUIÇÃO NULA:** Em um teste de hipóteses, a distribuição da estatística de teste quando a hipótese nula é considerada verdadeira.

**DISTRIBUIÇÃO POSTERIOR:** A distribuição de probabilidades para um parâmetro em uma análise bayesiana calculado a partir de uma distribuição prévia e a distribuição condicional dos dados para um certo parâmetro.

**DISTRIBUIÇÃO PRÉVIA:** A distribuição inicial de probabilidades considerada para um parâmetro em uma análise bayesiana.

**DISTRIBUIÇÃO T:** A distribuição de variável aleatória definida como a razão de duas variáveis aleatórias independentes. O numerador é uma variável aleatória normal padrão e o denominador é a raiz quadrada de uma variável aleatória qui-quadrado dividida por seu número de graus de liberdade.

**DERIVADAS PADRÕES:**

$y = f(x)$	$\frac{dy}{dx}$
$x^n$	$nx^{n-1}$
$\sqrt{x}$	$\frac{1}{2\sqrt{x}}$
$\frac{1}{x}$	$-\frac{1}{x^2}$
$e^{ax}$	$ae^{ax}$

$\log_e x$	$\frac{1}{x}$
$a^x$	$a^x \log_e a$
$y = f(x)$	$\frac{dy}{dx}$
$Sen x$	$Cos x$
$Cos x$	$-Sen x$
$\tan x$	$Sec^2 x$
$Sen^{-1}\left(\frac{x}{a}\right)$	$\frac{1}{\sqrt{a^2 - x^2}}$
$Cos^{-1}\left(\frac{x}{a}\right)$	$-\frac{1}{\sqrt{a^2 - x^2}}$
$\tan^{-1}\left(\frac{x}{a}\right)$	$\frac{a}{(a^2 + x^2)}$

**DESCOBERTA DE CONHECIMENTO EM BASES DE DADOS (KDD):** Extração de nova informação a partir de bases de dados por meio de uma variedade de processos de descoberta de conhecimento.

**DESCONFIRMABILIDADE:** Possibilidade de especificar condições sob as quais uma hipótese é considerada falsa. A menos que uma hipótese seja desconfirmável, não pode ser genuinamente testada.

**DESCONTO:** Redução no valor atual de benefícios que ocorrerão mais tarde ou aumento no custo presente desses benefícios para levar em conta o valor do dinheiro em função do tempo isto é, inflação.

**DESCRIÇÃO:** Registro de eventos na ordem de seus acontecimentos, após um período de observação.

**DESCRIÇÃO DAS VARIÁVEIS:** A variável constitui um primeiro nível de operacionalização de uma construção teórica e, para cada uma, se deve dar, em seguida, uma descrição operacional. Para algumas variáveis a descrição é simples, porém, em outros casos, essa definição é mais complexa. Uma variável contínua, pode ser transformada em discreta e depois em categórica ordinal, por exemplo, idade como diferença entre a data atual e data de nascimento, anos completos, faixas de idade. É recomendável tomar o valor bruto e depois categorizá-lo, isso dá mais flexibilidade ao pesquisador.

**DESENHO:** O mesmo que delineamento. É usado porque tem sonoridade similar a design, o termo da língua inglesa que traduz. No entanto, o termo delineamento é mais adequado.

**DESENHO (DE UMA INVESTIGAÇÃO):** O mesmo que delineamento.

**DESENHOS DE CONSUMIDORES:** Respondentes desenham o que estão sentindo ou como percebem um objeto.



**DESFECHO CLÍNICO DE UM ESTUDO EPIDEMIOLÓGICO:** É o evento, desenlace, do fenômeno saúde-doença que está sendo investigado. Exemplo: morte, doença, deficiência física. Pode ser estático, refletindo apenas um estado de saúde-doença em um determinado momento, ou dinâmico, traduzindo uma mudança de estado saúde-doença.

**DESFECHO DE INCIDÊNCIA:** Desfecho de um delineamento dinâmico de pesquisa epidemiológica, referindo-se a casos novos que ocorrem durante um período de observação. Sinônimo: desfecho de casos incidentes ou de casos novos.

**DESFECHO DE PREVALÊNCIA:** Desfecho de um delineamento estático da pesquisa epidemiológica, referindo-se a captação dos casos existentes em apenas um momento. Por isso, esta medida tende a super-representar aqueles casos de duração mais longa, talvez os de menor risco. Sinônimos: desfecho de casos prevalentes ou de casos existentes.

**DESIGN:** O mesmo que delineamento.

**DESIGUALDADE DE BONFERRONI:** Seja  $\theta = (\theta_1, \dots, \theta_k)$ , com os  $\theta_i$ 's não independentes e seja  $C = C_1 \times \dots \times C_k$ , então:  $\Pr(\theta \in C) \geq 1 - \sum_i \Pr(\theta_i \notin C_i)$ . Em inglês Bonferroni inequality.

**DESIGUALDADE DE BONFERRONI:** Técnica para ajustar o nível alfa selecionado para controle da taxa de erro Tipo 1 geral. O procedimento envolve (1) computar a taxa ajustada como  $\alpha$  dividido pelo número de testes estatísticos a serem executados e então (2) usar a taxa ajustada como o valor crítico em cada teste separado.

**DESIGUALDADE DE CAUCHY- SCHWARZ:** Sejam X e Y duas variáveis aleatórias. Então,  $[E(XY)]^2 \leq E(X^2)E(Y^2)$ .

**DESIGUALDADE DE CHEBYSHEV:** Seja X uma variável aleatória com expectância  $\mu$  e desvio padrão  $\sigma$ . Então  $P(|X - \mu| \geq K\sigma) \leq 1 - \frac{1}{K^2}$ . Em inglês Chebyshev inequality.

**DESIGUALDADE DE CHEBYCHEV:** Se X é uma variável aleatória com variância finita, então  $\Pr\{|X - \mu_x| \geq r\sigma_x\} = \Pr\{(X - \mu_x)^2 \geq (r\sigma_x)^2\} \leq r^{-2}$  para todo  $r > 0$ . Em inglês Chebychev inequality.

**DESIGUALDADE DE CRAMER E RAO:** Uma inequação dando um limite inferior para a variância do estimador de um parâmetro. Se t é um estimador de  $\theta$  para uma distribuição com função de frequência  $f(x, \theta)$  e se o vício  $b(\theta)$  é dado por  $b(\theta) = E(t) - \theta$ , a inequação estabelece que  $\text{Var}(t) \geq E[(1 + \partial b / \partial \theta)^2 / (\partial \log f / \partial \theta)^2]$ . Em inglês Cramer-Rao inequality.

**DESIGUALDADE DE JENSEN:** Se X é uma variável aleatória definida no intervalo I, e g é qualquer função convexa em I, então a desigualdade  $g(E[X]) \leq E[g(X)]$  é conhecida como desigualdade de Jensen. Em inglês Jensen's inequality.

**DESIGUALDADE DE MARKOV:** Seja X uma variável aleatória tal que |X| tenha expectância  $\mu$ . Então, para qualquer constante " $\alpha$ " positiva  $P(|X| \geq \alpha) \leq \frac{\mu}{\alpha}$ . Em inglês Markov inequality.

**DESIGUALDADE DE MINKOWSKI:** Sejam X e Y variáveis aleatórias, então:

$$[E(|X + Y|)^{\alpha}]^{\frac{1}{\alpha}} \leq [E(|X|)^{\alpha}]^{\frac{1}{\alpha}} + [E(|Y|)^{\alpha}]^{\frac{1}{\alpha}}. \text{ Em inglês Minkowski inequality.}$$

**DESIGUALDADE DE TCHEBYCHEV:** Se  $g(x)$  é uma função não-negativa de uma variável x, a desigualdade de Tchebychev (1874) estabelece que para cada  $k > 0$ ,  $\Pr\{g(x) > k\} \leq E[g(x)]/k$ . Em inglês Tchebychev's inequality.

**DESVANTAGENS DA AMOSTRAGEM:** Existem situações nas quais não se justifica todo o esforço de selecionar uma amostra, como por exemplo. i) Quando a população é muito pequena, não se justifica todo o gasto em recursos e tempo necessários para selecionar uma amostra relativamente grande a fim de gerar resultados precisos. ii) Se a característica a ser estudada é de fácil observação pode não compensar a realização de um plano de amostragem mesmo quando a população não seja muito pequena. Pode acontecer também que a informação necessária já esteja disponível como subproduto de alguma outra pesquisa e iii) Quando é preciso estudar características ou conhecer parâmetros que sejam fundamentais para o planejamento de um país ou para servir de referência para a realização de outras pesquisas, justifica-se a realização de um levantamento de informação por meio de um censo.

**DESVIO:** A diferença entre o valor de um conjunto (dado) e a média deste conjunto. Expressa-se como  $e = (X - \bar{X})$ . Conhecido também por erro, afastamento ou resíduo. Em inglês deviation.

**DESVIO:** Diferença entre um escore e a média aritmética. É equivalente à expressão afastamento, a qual, todavia, é menos empregada. Em inglês deviation.

**DESVIO:** Refere-se a diferença  $(x - m)$  entre um valor observado (x) e a média da população respectiva (m).

**DESVIO:** Diferença entre um registro individual e a média do grupo de contemporâneos para dada característica. Estas diferenças somam zero quando a média correta é usada.

**DESVIO ABSOLUTO:** Medida de dispersão igual à soma dos desvios de cada valor em relação à média, dividida pelo número de valores.

**DESVIO ABSOLUTO:** A diferença de um valor com relação a uma constante, sem levar em conta o sinal, que representa uma amostra. Em inglês é denominado de Absolute Deviation.

**DESVIO ABSOLUTO MÉDIO:** Medida de dispersão derivada do desvio médio de observações a partir de uma medida central, sendo os desvios considerados em valor absoluto. Em inglês Mean absolute deviation.

**DESVIO ABSOLUTO MEDIANO:** É uma medida de dispersão dos dados  $x_1, \dots, x_n$ , dada por,  $\text{definida por, } \text{dam} = \text{med}_{1 \leq j \leq n} |x_j - \text{med}_{1 \leq i \leq n}(x_i)|$ . Ou seja, calculamos a mediana dos dados, depois os desvios absolutos dos dados em relação à mediana e, finalmente, a mediana desses desvios absolutos.

**DESVIO EM RELAÇÃO À MÉDIA:** É a diferença entre cada elemento de um conjunto de valores e a média aritmética, ou seja:  $d_i = X_i - \bar{X}$ .

**DESVIO EXPLICADO:** Diferença entre o valor predito y e a média dos valores y, em um par de valores de uma coleção de dados bivariados.

**DESVIO MÉDIO:** Média dos valores absolutos dos desvios de todas as observações em relação à média.

**DESVIO MÉDIO (D):** Medida de variação igual à somas dos desvios de cada valor a contar da média, dividida pelo número de valores.



**DESVIO MÉDIO (D):** É a média aritmética do valor absoluto da diferença entre cada valor e a média, no caso dos dados não agrupados. No caso dos dados agrupados, tem que se levar em conta a frequência absoluta de cada observação.

**DESVIO MÉDIO (PARA DADOS TABULADOS):** Se os valores vierem dispostos em uma tabela de frequências, agrupados ou não em classes, serão usadas as seguintes fórmulas: Cálculo pela média:

$$Dm = \frac{\sum (E|X_i - \bar{X}|f_i)}{Ef_i}; \text{ Cálculo pela mediana: } Dm = \frac{\sum (E|X_i - Md|f_i)}{Ef_i}.$$

**DESVIO MÉDIO ABSOLUTO (PARA DADOS BRUTOS):** É a média aritmética dos valores absolutos dos desvios tomados em relação a uma das seguintes medidas de tendência central: média ou mediana. Símbolo=Dm.

**DESVIO MÉDIO ABSOLUTO:** Média de valores absolutos das distâncias de todos os valores de uma lista à sua média.

**DESVIO MÉDIO ABSOLUTO:** Medida de dispersão com características de que leva em conta apenas a grandeza dos desvios a partir da média e não os sinais desses desvios. Utiliza mais informações, sendo, portanto, mais difícil de calcular. É de difícil manejo. Mede a variação em subperíodos em relação à média do período. É aplicado em vários casos de estudos econômicos, porém com limitações nos resultados.

**DESVIO NÃO-EXPLICADO:** A diferença entre a coordenada y e o valor predito em um par de valores de uma coleção de dados bivariados.

**DESVIO PADRÃO:** É a medida de dispersão associada à medida de tendência central. Dado um conjunto de n observações  $\{X_i, i=1, n\}$  com média  $\bar{X}$ , então o desvio padrão pode ser calculado como:  $s = \sqrt{\frac{1}{N} \sum_{i=1}^n (X_i - \bar{X})^2}$

Freqüentemente esta expressão apresenta-se com o denominador igual a (N-1) ao invés de N, justamente para se ter uma estimativa não enviesada do desvio padrão populacional. Quando N for grande (maior que 40), a variância amostral ( $S^2$ ) aproxima-se da variância populacional ( $\sigma^2$ ) o que equivale a deixar o denominador igual a N.

**DESVIO PADRÃO:** Medida de dispersão absoluta muito usada em análise estatística. É igual à raiz quadrada positiva da variância. Em inglês Standard deviation.

**DESVIO-PADRÃO (DP):** Raiz quadrada da variância.

**DESVIO PADRÃO:** Uma medida de dispersão de dados que mostra o quanto eles estão perto da média da distribuição.

**DESVIO PADRÃO:** Uma medida de variabilidade que é a raiz quadrada positiva da variância.

**DESVIO PADRÃO:** O desvio padrão é uma medida de dispersão usada com a média. Mede a variabilidade dos valores à volta da média. O valor mínimo do desvio padrão é 0 indicando que não há variabilidade, ou seja, que todos os valores são iguais à média. A fórmula de cálculo do desvio padrão para os valores  $x_1, x_2, x_3, \dots, x_n$  de uma amostra é a seguinte:

$$\sqrt{\frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}$$

onde  $\bar{X}$  é a média da amostra. Por convenção, usa-se a letra grega  $\sigma$  (sigma) para o desvio padrão da população e s para o desvio padrão da amostra.

**DESVIO PADRÃO:** Medida de dispersão com características de que utiliza maior quantidade de informação e é mais fácil de manejar do que o intervalo ou do que o desvio médio absoluto. É uma medida de dispersão absoluta com limitadas possibilidades de uso em comparações como em amostras que possuem diferentes unidades de medidas. Tem como pressuposto a distribuição normal da série. É a medida de dispersão mais frequentemente utilizada, com propriedades muito úteis.

**DESVIO PADRÃO:** Afastamento padrão em relação à média; informa a dispersão dos valores em torno da média.

**DESVIO PADRÃO:** Raiz quadrada da variância.

**DESVIO PADRÃO:** Raiz quadrada da variância.

**DESVIO PADRÃO:** Parâmetro estatístico que expressa o desvio médio de um conjunto de dados em torno da média. Corresponde à raiz quadrada da variância.

**DESVIO PADRÃO:** Medida de dispersão absoluta muito usada para comparar o grau de variabilidade entre duas ou mais amostras ou populações. É igual à raiz quadrada positiva da variância.

**DESVIO PADRÃO:** Medida de dispersão ou variabilidade de uma distribuição de frequência. É igual à raiz quadrada da soma dos desvios ao quadrado, dividida pelo número de casos. Em outras palavras, é a raiz quadrada da média dos desvios elevados ao quadrado. Em inglês standard deviation.

**DESVIO PADRÃO:** É a medida de dispersão mais geralmente empregada, pois leva em consideração a totalidade dos valores da variável em estudo. É um indicador de variabilidade bastante estável. O desvio padrão baseia-se nos desvios em torno da média aritmética e a sua fórmula básica pode ser traduzida como: a raiz quadrada da média aritmética dos quadrados dos desvios e é representada por S.

**DESVIO PADRÃO:** É a raiz quadrada positiva da variância.

**DESVIO PADRÃO:** É a raiz quadrada positiva do quadrado médio do resíduo. Também se chama desvio padrão a raiz quadrada ( $\sigma$ ) da variância ( $\sigma^2$ ).

**DESVIO PADRÃO:** É a raiz quadrada da soma dos quadrados dos desvios dividida pelo número de elementos, ou dito de outra forma, é raiz quadrada da média aritmética dos quadrados dos desvios. Em inglês **standard deviation**.

**DESVIO PADRÃO:** Medida de dispersão calculada ao subtrair a média de uma série de cada valor em uma série, elevando-se ao quadrado cada resultado, somando-os, dividindo a soma pelo número de itens menos um e obtendo e a raiz quadrada desse valor, ou a raiz quadrada da soma dos desvios ao quadrado da média, dividido pelo número de observações menos um.

**DESVIO PADRÃO:** A medida de dispersão mais usada, que pode ser considerada como uma variabilidade dos dados com uma medida de variabilidade de dados de uma distribuição de frequências. Isto é, o desvio padrão mede a dispersão dos valores individuais em torno da média. Para seu cálculo, deve-se obter a média da distribuição e, a seguir, determinar os desvios para mais e para menos a partir da mesma.

Assim, o desvio padrão é a média aritmética de uma distribuição de frequências, ou seja, é a raiz quadrada da média aritmética dos quadrados dos desvios, esses tomados a partir da média aritmética.

**DESVIO PADRÃO AMOSTRAL:** É o desvio padrão calculado para uma amostra particular. Em inglês Sample standard deviation.

**DESVIO PADRÃO AMOSTRAL:** Desvio-padrão calculado com base nos valores de uma mostra; usado para estimar o valor do desvio padrão da população.

**DESVIO PADRÃO DA POPULAÇÃO:** O desvio padrão de uma variável para toda a população.

**DESVIO PADRÃO POPULACIONAL:** Desvio padrão calculado quando se conhecem todos os elementos da população. Ver desvio padrão amostral.

**DESVIO QUARTIL (Dq):** Também chamado de amplitude semi-interquartil e é baseada nos quartis.

**DESVIO QUARTIL REDUZIDO:**  $D_{qr} = \frac{(Q_3 - Q_1)}{2Md}$  ou  $\frac{[(Q_3 - Q_1)]}{2Md} \times 100$  para resultado em %.

**DESVIO SEMI-INTERQUARTIL:** O mesmo que intervalo semi-interquartil. Em inglês semi-interquartile deviation.

**DESVIO TOTAL:** Soma do desvio explicado e do desvio não explicado para determinado par de valores em uma coleção de dados bivariados.

**DETERMINANTE:** Número que se calcula com os elementos de uma matriz quadrada, segundo certas regras.

**DETERMINANTE:** O determinante de uma matriz é a combinação das entradas da matriz com a propriedade de que ele é zero se e só se a matriz não é inversível. Em inglês determinant.

**DE WIJS:** Modelo de variograma teórico, descrito pela seguinte equação:  $\gamma(h) = C_0 + a \ln(h)$ .

**DF:** Algoritmo não derivativo.

**DF-REML:** REML obtido via algoritmo DF.

**DFBETA:** Medida da mudança em um coeficiente de regressão quando uma observação é omitida da análise. O valor de DFBETA está relacionado ao próprio coeficiente; uma forma padronizada (SDFBETA) também está disponível. Nenhum limitante de referência pode ser estabelecido para DFBETA, embora o pesquisador possa procurar valores substancialmente diferentes das observações remanescentes para avaliar a influência potencial. Os valores de SDFBETA são medidos por seus erros padrão, apoiando assim a ideia de cortes de 1 ou 2, correspondentes aos níveis de confiança de 0,10 ou 0,05, respectivamente.

**DFFIT:** Medida do impacto de uma observação no ajuste geral do modelo, que também tem uma versão padronizada (SDFFIT). A melhor norma prática é classificar como influentes quaisquer valores padronizados (SDFFIT) que excedam a  $\frac{2}{\sqrt{\frac{p}{n}}}$ , onde p é o número de variáveis independentes + 1 e n é o tamanho

da amostra. Não existe qualquer valor de referência para a medida DFFIT.

**DIAGNÓSTICO:** É um trabalho de triagem e síntese, em estilo direto, visando não resolver controvérsias e dúvidas, mas coloca-las às claras, em nítido contraste com o que é sabido e pacífico.

**DIAGRAMA CARTESIANO:** Representação gráfica por meio de dois eixos perpendiculares que são as coordenadas cartesianas: a abscissa corresponde ao eixo horizontal e ordenada ao eixo vertical. Em inglês cartesian chart.

**DIAGRAMA DE ÁRVORE:** Uma útil representação gráfica para definir os pontos amostrais de um experimento que envolve múltiplas etapas.

**DIAGRAMA DE ÁRVORE:** Em inglês tree diagram.

**DIAGRAMA DE CAIXA:** Representação gráfica da distribuição de um conjunto de dados. Em inglês Boxplot.

**DIAGRAMA DE CAULE-E-FOLHAS:** O mesmo que separador de frequências.

**DIAGRAMA DE CONTROLE:** É um dispositivo gráfico destinado ao acompanhamento, no tempo, semana a semana, mês a mês, da evolução dos coeficientes de incidência, com o objetivo de se estabelecer e implementar medidas profiláticas que possam manter a doença sob controle.

**DIAGRAMA DE DISPERSÃO:** É uma representação gráfica para os dados bivariados, em que cada par de dados  $(x_i, y_i)$  é representado por um ponto de coordenadas  $(x_i, y_i)$ , num sistema de eixos coordenados.

**DIAGRAMA DE DISPERSÃO:** Um conjunto de pontos determinados pela tabulação de conjunto de dados bivariados. Cada indivíduo ou elemento tem dois escores, um para cada variável.

**DIAGRAMA DE DISPERSÃO:** É a representação num gráfico cartesiano de duas variáveis aleatórias, para as quais se deseja estudar uma provável correlação. Caso exista alguma correlação e se for definida uma relação funcional entre as variáveis, então a variável da ordenada é a dependente e aquela da abscissa é a independente.

**DIAGRAMA DE DISPERSÃO:** Gráfico que, por meio das coordenadas cartesianas, é constituído com o fim de mostrar a relação entre os escores de dois testes obtidos por indivíduos de certo grupo. Os escores de um teste são representados ao longo da abscissa e os escores do outro ao longo da ordenada. Se os escores nas duas variáveis são altamente correlacionados, os pontos do diagrama tendem a situar-se ao longo de uma reta. Em inglês scattergram ou scatter diagram.

**DIAGRAMA DE DISPERSÃO:** Diagrama que mostra a relação entre duas grandezas; uma delas é medida ao longo do eixo vertical, e a outra, ao longo do eixo horizontal; cada observação é representada por um ponto.

**DIAGRAMA DE DISPERSÃO:** É a representação gráfica da nuvem de pontos plotados num sistema de eixos ortogonais de um conjunto de pares ordenados de valores  $(x, y)$ , onde cada par ordenado corresponde a uma observação.

**DIAGRAMA DE DISPERSÃO:** Representação da relação entre duas variáveis métricas que descreve que os valores conjuntos de cada observação em um gráfico bidimensional.

**DIAGRAMA DE EXTREMOS E QUARTIS:** É um tipo de representação gráfica, em que se realçam algumas características da amostra. O conjunto dos valores da amostra compreendidos entre o 1º e o 3º quartis,  $Q_{.25}$  e  $Q_{.75}$  é representado por um retângulo que é a caixa com a mediana indicada por uma barra. Consideram-se seguidamente duas linhas que unem os lados dos retângulos com os valores máximo e mínimo, respectivamente.

**DIAGRAMA DE EXTREMOS E QUARTIS:** É um diagrama que representa os valores extremos e os quartis de uma variável estatística.

**DIAGRAMA DE FREQUÊNCIA:** Diagrama em barras que ilustra a distribuição das observações por diversas categorias.

**DIAGRAMA DE RAMOS E FOLHAS:** É um tipo de representação que se pode considerar entre a tabela e o gráfico, uma vez que são apresentados os verdadeiros valores da amostra, mas numa apresentação sugestiva, que faz lembrar um histograma. Consiste em escrever do lado esquerdo de uma linha vertical, o dígito ou dígitos da classe de maior grandeza, seguidos dos restantes.

**DIAGRAMA DE RAMO-E-FOLHAS:** Uma variante do histograma que fornece uma descrição visual da distribuição de variável, bem como a enumeração dos valores reais dos dados.

**DIAGRAMA EM ÁRVORE:** Gráfico pictórico dos diferentes resultados possíveis em um evento composto.

**DIAGRAMA IC:** Diagrama indicado quando em uma amostra há relação entre duas variáveis, sendo uma categórica nominal ou ordinal e outra numérica discreta ou contínua. O diagrama descreve a estimação da média e o respectivo intervalo de confiança com 95% ou 99% de probabilidade de confiança da variável de cada categoria.

**DIAGRAMA PONTUAL (DOTPLOT):** Diagrama que descreve graficamente duas ou mais amostras sendo até sete no máximo, permitindo a visualização pontual dos escores, em sentido vertical, e as respectivas comparações.

**DIAGRAMA DE CAMINHOS:** Representação gráfica do conjunto completo de relações entre os construtos do modelo. Relações causais são representadas por setas retilíneas, apontando da variável preditora para a variável ou construto dependente. Setas curvas correspondem a correlações entre construtos ou indicadores, mas nenhuma causalidade é implicada.

**DIAGRAMA DE PERFIL:** Representação gráfica dos dados que ajuda a projetar observações atípicas ou a interpretar a solução final de agrupamento. Normalmente, as variáveis da variável estatística de agrupamento ou aquelas usadas para a validação são listadas ao longo do eixo horizontal e a escala de valores é usada para o eixo vertical. Linhas separadas representam os escores originais ou padronizados para objetos individuais ou centróides de agrupamentos em um plano gráfico.

**DIAGRAMA DE VENN:** Em inglês Venn diagram.

**DIAGRAMA VERTICAL:** Representação gráfica de agrupamentos. Os objetos separados são mostrados horizontalmente ao longo do topo do diagrama, e o processo de agrupamento hierárquico é representado verticalmente em combinações de agrupamentos. Esse diagrama é semelhante a um dendrograma invertido e auxilia na determinação do número apropriado de agrupamentos na solução.

**DIAGRAMAS:** São gráficos geométricos dispostos em duas dimensões. São os mais usados na representação de séries estatísticas.

**DIAGRAMAS DE DISPERSÃO:** Os diagramas de dispersão são representações gráficas de uma possível relação entre duas variáveis. Se num sistema de eixos representarmos um conjunto de pontos  $(x_i, y_i)$ ,  $i=1, \dots, n$  podemos visualizar e ter uma idéia grosseira da relação e da correlação entre as 2 variáveis.

**DIÂMETRO:** Numa circunferência, o diâmetro é o comprimento de qualquer corda que lhe passe pelo centro.

**DIÂMETRO MÉDIO:** Atributo de um povoamento florestal é o valor do diâmetro de um indivíduo representativo do povoamento. É obtido pela média aritmética dos diâmetros medidos.

**DIÂMETRO MÉDIO QUADRÁTICO:** Atributo de um povoamento florestal, corresponde ao diâmetro da árvore de área transversal média. O diâmetro médio quadrático de um povoamento é sempre maior que o diâmetro médio, e essa diferença depende do desvio padrão dos diâmetros.

**DIÁRIO DA TEORIA DE CONSTRUÇÃO:** Agenda que documenta em detalhe os pensamentos, premissas, hipóteses e revisões de pensamento de um pesquisador humanista.

**DICOTOMIA:** Uma divisão dos membros de uma população, ou amostra, em dois grupos. A definição dos grupos pode ser em termos de uma variável mensurável, porém é usualmente baseada em características quantitativas ou atributos. Em inglês Dichotomy.

**DIFERENÇA DE COEFICIENTE:** Semelhante a diferença de riscos, mas aplicada a coeficientes.

**DIFERENÇA DE RISCO:** Risco no grupo exposto menos o risco no grupo não-exposto.

**DIFERENÇA MÍNIMA SIGNIFICATIVA (D.M.S.):** Valor mínimo que deve ter a diferença entre as estimativas de duas médias de tratamentos para que se considere não ser devida ao acaso.

**DIFERENÇAS ESTATISTICAMENTE SIGNIFICATIVAS:** Diferenças não aleatórias, que não são imputadas ao acaso, algum fator que, não o acaso, explica as diferenças, por exemplo, entre as médias de duas amostras.

**DIFERENÇAS INDIVIDUAIS:** A maioria das pesquisas sociais envolve grupos e conclusões baseadas em médias de grupos. O risco dessa abordagem é a perda de informações importantes sobre leis de comportamento e diferenças individuais.

**DIFERENÇAS NÃO-ESTATISTICAMENTE SIGNIFICATIVAS:** As diferenças são triviais e imputadas ao acaso; os dois grupos comparados são semelhantes, do ponto de vista estatístico.

**DIFERENTE (DIFERENÇA):** Símbolo matemático dado por  $\neq$ .

**DIFERENCIAL DE SELEÇÃO:** Diferença entre a média, para uma determinada característica, do grupo de indivíduos selecionados e a média da população do qual ele proveio. A resposta esperada de seleção para um caractere é igual ao diferencial de seleção multiplicado pela herdabilidade da característica.

**DIMENSÃO:** Aspecto de uma variável; por exemplo, podemos falar da dimensão da crença da variável religiosidade ou da dimensão de política externa da variável orientação política.



**DIMENSÃO:** Atributo de um elemento de dados, como idade ou sexo de um cliente ou data de um registro histórico.

**DIMENSÃO OBJETIVA:** Características físicas ou tangíveis de um objeto que têm uma base objetiva de comparação. Por exemplo, um produto tem tamanho, forma, cor, peso e assim por diante.

**DIMENSÃO PERCEBIDA:** Uma atribuição subjetiva, por parte do respondente, de aspectos a um objeto, a qual representa suas características intangíveis. Exemplos incluem qualidade, caro e boa aparência. Essas dimensões percebidas são únicas do respondente individual e podem exibir pouca correspondência com dimensões objetivas reais.

**DIMENSÃO SUBJETIVA:** Ver dimensão percebida.

**DIMENSIONALIDADE INICIAL:** Um ponto de partida para selecionar a melhor configuração espacial para dados antes de iniciar um procedimento MDS (escalonamento multidimensional), o pesquisador deve especificar quantas dimensões ou características estão representadas nos dados.

**DIMENSÕES:** Características de um objeto. Pode-se imaginar que objeto específico possui dimensões percebidas ou subjetivas, por exemplo, caro, frágil e objetivas, por exemplo, cor, preço, características.

**DINÂMICA DE GRUPO:** A interação entre as pessoas de um grupo.

**DIREÇÃO DE ANISOTROPIA:** É dada pela direção de menor variabilidade do depósito. Geralmente, é convencionalizado para ser medido no sentido anti-horário, a partir do eixo das abscissas, como no ciclo trigonométrico.

**DISCAGEM ALEATÓRIA:** Método de gerar listas de números telefônicos de forma aleatória.

**DISCAGEM RANDÔMICA DE DÍGITOS:** Uso da tabela de números randômicos ou aleatórios para gerar números de telefone com a finalidade de contatar entrevistados em potencial.

**DISCRETIZAÇÃO:** É o processo utilizado para determinar o teor médio de um painel ou bloco de uma jazida, a partir da sua discretização, ou seja, a subdivisão em pequenos sub-painéis ou sub-blocos, os quais são avaliados individualmente e então compostos para o domínio original. Esta técnica foi originalmente desenvolvida pela krigagem de bloco, porém recentemente foi sendo adaptada a outros métodos computacionais de avaliação de reservas, como, por exemplo, o método da ponderação pelo inverso da potência da distância. Para discretização do domínio a ser estimado.

dimensão do domínio	número de pontos
uma	10
duas	6 x 6
três	4 x 4 x 4

**DISPARIDADES:** Diferenças entre as distâncias geradas no computador que representam similaridades e as distâncias fornecidas pelo respondente.

**DISPERSÃO:** Grau de espalhamento de uma distribuição.

**DISTÂNCIA:** Medida de distância entre dois objetos ou intervalo. Em análise multivariada (AMV), não se trata necessariamente de medidas usando réguas ou algo equivalente. Medidas de distância aqui estão relacionadas a dissimilaridade entre objetos.

**DISTÂNCIA:** É dada pela separação entre dois pontos quaisquer no espaço. Pode ser calculada a partir das coordenadas dos pontos. Se  $(x_1, y_1)$  e  $(x_2, y_2)$  são dois pontos no plano, então a distância é dada por:

$d = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2}$  e se  $(x_1, y_1, z_1)$  e  $(x_2, y_2, z_2)$  forem as coordenadas de dois pontos no espaço, a distância é dada pela seguinte equação:

$$d = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2 + (Z_1 - Z_2)^2}.$$

**DISTÂNCIA DE COOK (D):** Medida resumo da influência de um único caso ou observação baseada nas mudanças totais em todos os outros resíduos quando o caso é eliminado do processo de estimação. Valores grandes, geralmente maiores que 1 indicam influência substancial pelo caso afetando os coeficientes de regressão estimados.

**DISTÂNCIA DE MAHALANOBIS (D):** Medida de unicidade de uma observação baseada em diferenças entre os valores da observação e os valores médios para todos os outros casos ao longo de todas as variáveis independentes. A fonte de influência em resultados de regressão ocorre quando o caso é muito diferente em uma ou mais variáveis preditoras, causando assim um desvio da equação da regressão como um todo.

**DISTÂNCIA DE MAHALANOBIS (D<sup>2</sup>):** Forma padronizada de distância euclideana. Resposta de escalonamento em termos de desvios padrão que padroniza os dados, com ajustes feitos para intercorrelações entre as variáveis.

**DISTÂNCIA EUCLIDEANA:** Medida mais comumente usada da similaridade entre dois objetos. Essencialmente, é uma medida do comprimento de um segmento de reta desenhado entre dois objetos.

**DISTORÇÃO DE NÃO-RESPOSTA:** Erro que resulta de uma diferença sistemática entre aqueles que respondem e que não respondem a um instrumento de mensuração.

**DISTORÇÃO DE RESPOSTA:** Erro que resulta da tendência das pessoas em responder a uma pergunta de forma incorreta, por meio de má interpretação deliberada ou falsificação inconsciente.

**DISTORÇÃO DE SELEÇÃO:** Diferenças sistemáticas entre o grupo de teste e o grupo de controle devido a um processo de seleção distorcido.

**DISTORÇÃO DO INSTRUMENTO DE MENSURAÇÃO:** Erro que resulta do projeto do questionário ou do instrumento de mensuração.

**DISCAGEM DIGITAL ALEATÓRIA:** Um aspecto do ETAC System (Computer-Assisted Telephone Interview), que é um sistema de entrevista assistida por telefone com o auxílio do computador. Computadores podem ser usados para gerar aleatoriamente conjuntos de números com sete algarismos, constituindo a amostra do survey, em vez de selecionar nomes e números de assinantes telefônicos.

**DISCREPANTE (OUTLIER):** Valor, leitura ou medida fora dos limites esperados e, por isto, colocado em dúvida ou considerado erro.

**DISPERSÃO:** Discrepância entre valores observados.



**DISPERSÃO:** Distribuição de valores em torno de algum valor central, como uma média. A amplitude é um exemplo simples de medida de dispersão. Poderíamos afirmar que a média de idade num grupo é de 37,9 anos e que amplitude varia de 12 a 89 anos.

**DISPERSÃO:** O grau de dispersão demonstrado por observações. Usualmente é medida como um promédio em torno de um valor central ou de uma estatística de ordem, podendo ser, também, a média de desvios dos valores entre eles. Em inglês Dispersion.

**DISPERSÃO: Termo que indica o** quanto um conjunto de dados está espalhado. A dispersão é normalmente avaliada em torno da média, por meio de da variância, do desvio padrão e do desvio médio. Mas também pode ser definida pela amplitude que é a diferença entre o máximo e o mínimo do conjunto ou ainda pela amplitude inter-quartilica, isto é, a diferença entre o terceiro e o primeiro quartil. Em inglês dispersion.

**DISPERSÃO OU VARIABILIDADE:** É a maior ou menor diversificação dos valores de uma variável em torno de um valor de tendência central, como a média ou a mediana tomado como ponto de comparação. A média ainda que considerada como um número que tem a faculdade de representar uma série de valores não pode por si mesma, destacar o grau de homogeneidade ou heterogeneidade que existe entre os valores que compõem o conjunto.

**DISTÂNCIA DE MAHALANOBIS:** Uma medida que fornece o quanto um ponto ou um dado se afasta da média da amostra ou centróide no espaço das variáveis independentes utilizadas no ajuste de um modelo de regressão múltipla. Ela fornece uma forma de descobrir pontos que estão mais afastados do que os demais no espaço multidimensional. Em inglês Mahalanobis distance.

**DISTÂNCIA INTERQUARTÍLICA:** Distância entre o primeiro e o terceiro quartís. Esta amplitude contém a metade da frequência total e fornece uma medida simples da dispersão que é útil em estatística descritiva. Em inglês Interquartile range.

**DISTRIBUÍDA COMO:** Termo que se refere ao fato de uma variável aleatória contínua possuir determinada distribuição de probabilidade. É indicado pelos símbolos  $\sqcup$  ou  $\cap$ .

**DISTRIBUIÇÃO:** O total de observações ou um conjunto de dados numa variável, quando se tabula as observações de acordo com as frequências de cada observação, temos uma distribuição de frequência.

**DISTRIBUIÇÃO AGREGADA:** É o tipo de distribuição onde verifica-se que, uma das características mais comuns em infecções de hospedeiros invertebrados e vertebrados por parasitas é a tendência destes últimos apresentarem-se distribuídos de maneira agregada nos hospedeiros, sendo que a maioria dos hospedeiros geralmente alberga poucos parasitas, enquanto que poucos hospedeiros albergam a maior proporção da população de parasitas. São nesses poucos hospedeiros que processos dependentes da densidade exercem sua influência, via supressão da fecundidade ou sobrevivência parasita, ou por meio da influência do parasita sobre a sobrevivência ou fecundidade do hospedeiro. O padrão de dispersão dos parasitas tem sido considerado. Uma das características mais comuns em infecções de hospedeiros invertebrados e vertebrados por parasitas é a tendência destes últimos apresentarem-se distribuídos de maneira agregada nos hospedeiros. A maioria dos hospedeiros geralmente alberga poucos parasitas, enquanto que poucos hospedeiros albergam a maior proporção da população de parasitas. A causa primária da distribuição agregada de parasitas dentro da população de hospedeiros está associada principalmente a fatores estocásticos ambientais, antes de fatores de natureza demográfica.

**DISTRIBUIÇÃO A POSTERIORI MARGINAL:** Em um vetor de parâmetros  $\theta=(\theta_1,\dots,\theta_k)$ , a distribuição a posteriori marginal de um parâmetro  $\theta_i$  é a distribuição a posteriori livre do efeito dos outros parâmetro. Em inglês Marginal posterior distribution.

**DISTRIBUIÇÃO A POSTERIORI:** Depois de ter observado um conjunto de dados,  $X$ , a quantidade de informação conhecida sobre um parâmetro  $\theta$ , relacionado aos dados por meio de  $p(x|\theta)$ , está contida na distribuição a posteriori,  $p(\theta|X)$ , que pode ser calculada por meio do teorema de Bayes. Em inglês Posterior distribution.

**DISTRIBUIÇÃO A PRIORI:** Representação probabilística da informação inicial existente sobre um parâmetro de interesse antes de realizar algum experimento relacionado com tal parâmetro. Em inglês Prior distribution.

**DISTRIBUIÇÃO AMOSTRAL:** Distribuição de probabilidade de uma estatística, tal como a média amostral  $\bar{X}$ .

**DISTRIBUIÇÃO AMOSTRAL DA MÉDIA  $\bar{X}$ :** Se  $X_1, X_2, \dots, X_n$  é uma amostra de  $n$  observações independentes de uma população normal, com média  $\mu_X$  e variância  $\sigma_X^2$ , e se  $\bar{X}$  é a média desta amostra, então eesa média  $\bar{X}$  tem uma distribuição amostral ou de probabilidade normal com média  $\mu_X$  e variância

$$\frac{\sigma_X^2}{n}, \text{ isto é } \bar{X} \cap N\left(\mu_X, \frac{\sigma_X^2}{n}\right).$$

**DISTRIBUIÇÃO AMOSTRAL DE MÉDIAS AMOSTRAIS:** Distribuição das médias amostrais obtida quando extraímos repetidas amostras de mesmo tamanho e da mesma população.

**DISTRIBUIÇÃO AMOSTRAL:** A distribuição de uma estatística ou série de estatísticas em todas as amostras possíveis que podem ser encontradas de acordo com um esquema de amostragem específico. A expressão quase sempre está relacionada com um esquema de amostragem envolvendo seleção aleatória e muitas vezes se refere à distribuição da função de um número fixo  $n$  de variáveis independentes. Em inglês Sampling distribution.

**DISTRIBUIÇÃO AMOSTRAL:** Distribuição de estimadores tais como a da média amostral, com o objetivo de que as características de uma amostra se aproximarão mas não de forma perfeita ou exata das características da população da qual ela foi selecionada, isto se refere às aproximações que se produziriam se fosse selecionado um grande número de amostras, cada uma dando uma estimativa desta característica para toda a população.

**DISTRIBUIÇÃO ASSIMÉTRICA:** Histograma que apresenta uma distribuição de frequências de forma acentuadamente assimétrica, apresentando valores substancialmente menores num dos lados, relativamente ao outro.

**DISTRIBUIÇÃO ASSIMÉTRICA:** Uma distribuição de frequência fora do normal com alguns valores extremos, altos ou baixos, que faz com que as três medidas de tendência central como a média, a mediana e a moda desviem uma da outra.

**DISTRIBUIÇÃO ASSINTÓTICA:** A forma limite de uma distribuição de frequência ou probabilidade dependente de um parâmetro, tal como tamanho de amostra ou tempo, quando o parâmetro tende ao limite, geralmente infinito. Do inglês Asymptotic distribution.

**DISTRIBUIÇÃO BIDIMENSIONAL:** É a representação de uma variável bidimensional  $(x_i, x_j)$ , sendo que  $x_i$  e  $x_j$  duas variáveis unidimensionais.

**DISTRIBUIÇÃO BIMODAL:** Distribuição de frequência dotada de duas modas, ou seja, de dois valores que se distinguem dos demais por sua alta frequência. Graficamente, a distribuição bimodal possui duas saliências ou corcovas, que podem não apresentar a mesma altura. Tal fenômeno ocorre quando a população é formada de dois grupos diferentes, cujas médias são por isso mesmo diversas. Em inglês bimodal distribution.

**DISTRIBUIÇÃO BIMODAL:** Distribuição de frequências que possui duas modas. Em inglês Bimodal distribution.

**DISTRIBUIÇÃO BETA:** Em inglês Beta distribution.

**DISTRIBUIÇÃO BETA:** O termo distribuição beta é geralmente aplicado quando uma variável aleatória tem densidade  $f(x)=x^{\alpha-1}(x-1)^{\beta-1}/B(\alpha,\beta)$ ,  $0 \leq x \leq 1$ ;  $\alpha, \beta > 0$ . Em inglês Beta distribution.

**DISTRIBUIÇÃO BETA-BINOMIAL:** Quando uma variável aleatória  $X$  tem função de massa de probabilidade dada por  $f(x)=n!B(\alpha+x,\beta+n-x)/[x!(n-x)!B(\alpha,\beta)]$ , para  $x=0,1,\dots,n$ ,  $n \geq 1$ , diz-se que  $X$  tem distribuição beta-binomial. Em inglês Beta-binomial distribution.

**DISTRIBUIÇÃO BIDIMENSIONAL:** É a representação de uma variável bidimensional  $(x_i, x_j)$ , sendo  $x_i$  e  $x_j$  duas variáveis unidimensionais.

**DISTRIBUIÇÃO BINOMIAL:** Se um evento tem probabilidade  $P$  de acontecer em qualquer tentativa, a probabilidade de  $r$  ocorrências em  $n$  tentativas independentes é  $\frac{n!}{x!(n-x)!} P^x q^{n-x}$  onde  $q=1-p$ . Esta

última expressão fornece as probabilidades para  $r=0, 1, \dots, n$ , e é conhecida como distribuição binomial. Em inglês Binomial distribution.

**DISTRIBUIÇÃO BINOMIAL:** A variável aleatória  $X$ , inteira e não negativa, tem uma distribuição binomial  $B(n, p)$  quando é igual ao número de sucessos de um dado acontecimento em  $n$  tentativas independentes, em que esse acontecimento tem probabilidade de ocorrência  $p$  em cada uma delas, sendo a respectiva função de probabilidade dada por:

$$P_x = \binom{n}{x} p^x (1-p)^{n-x} \text{ onde a variável } x \text{ toma valores em } (X = 0, 1, 2, 3, \dots, n).$$

**DISTRIBUIÇÃO BINOMIAL:** Distribuição em que se observam as frequências de dois acontecimentos mutuamente exclusivos, com probabilidades fixas  $p$  e  $q$ , sendo  $p + q = 1$ .

**DISTRIBUIÇÃO BINOMIAL:** Modelo de distribuição de probabilidades de variáveis aleatórias discretas, consistindo de  $n$  provas idênticas e independentes, cada uma apresentando apenas dois resultados: sucesso ( $p$ ) e insucesso ( $q=1-p$ ).

**DISTRIBUIÇÃO BINOMIAL:** É o modelo de distribuição teórica ou especial de probabilidade para descrever o comportamento da distribuição de variáveis aleatórias discretas ou descontínuas de Bernoulli, onde representa apenas duas possibilidades de ocorrência sucesso ou fracasso em  $n$  provas ou repetições de um ensaio de Bernoulli, ou seja, a probabilidade de se verificarem ao longo das  $n$  experiências,  $x$  realizações de um evento Sucesso  $A$  e  $n-x$  realizações de um evento fracasso  $B$ , numa ordem dada, é  $p^x q^{n-x}$

, mas, na realidade, é possível verificar  $x$  realizações de  $A$  ao longo de  $n$  experiências em  $C_x^n$  ordens diferentes, pois as diferentes ordens possíveis excluem-se mutuamente. Daí resulta que a probabilidade de se verificarem ou ocorrerem  $x$  realizações de  $A$  e  $n - x$  realizações de  $B$ , numa ordem qualquer, é, para  $x = 0, 1, 2, 3, 4, \dots, n$ ; igual a  $P(X = x) = P_x = C_x^n P^x q^{n-x}$ . Essa lei chama-se lei binomial de probabilidade e a variável  $x$  é uma variável aleatória discreta que possui uma distribuição binomial, ou, mais simplesmente, uma variável binomial. A expressão  $C_x^n P^x q^{n-x}$ , é, com efeito, o termo geral do desenvolvimento do binômio  $(p + q)^n$ , também conhecido como binômio de Newton ou binômio de Tartália. Uma vez que  $p+q=1$ , deduz-se imediatamente que a relação  $\sum_{x=0}^n P_x = \sum_{x=0}^n C_x^n P^x q^{n-x} = 1$  se verifica facilmente. A distribuição teórica ou

especial de probabilidade binomial, é usada para descrever o comportamento da distribuição de variáveis aleatórias discretas ou descontínuas de Bernoulli, ou seja que possuem apenas duas possibilidade de ocorrência sucesso que é o um (1) e Insucesso ou fracasso que é o zero (0), cuja notação é dada como segue:  $X \sim \text{Bin}(n; p)$  ou  $X \sim B(n; p)$ , a variável aleatória discreta  $X$  possui distribuição teórica ou especial de probabilidade binomial com parâmetros  $n$  e  $p$ .

$$P(X = xi) = C_x^n \cdot p^x \cdot q^{n-x} = \frac{n!}{x!(n-x)!} \cdot p^x q^{n-x} \text{ , onde } x = \{0, 1, 2, \dots, n\} \text{ e } C_x^n = \frac{n!}{x!(n-x)!} \text{ ,}$$

sendo  $p + q = 1$ ,  $q = 1-p$ ,

$$(p + q)^n = \sum_{x=0}^n \frac{n!}{x!(n-x)!} P^x q^{n-x} = 1, 0$$

$$P(X= 0) + P(X= 1) + P(X= 2) + P(X= 3) + \dots + P(X= n) = \sum_{i=1}^n P(x = x_i) = 1$$

$$P(x \geq 1) = 1 - [P(x = 0)], P(x \leq 2) = P(x = 0) + P(x = 1) + P(x = 2)$$

A forma generalizada da distribuição binomial é dada a seguir.

X = Número de Sucessos	P(X=x <sub>i</sub> )
0	$\frac{n!}{0!(n)!} P^0 q^n$
1	$\frac{n!}{1!(n-1)!} P^1 q^{n-1}$
2	$\frac{n!}{2!(n-2)!} P^2 q^{n-2}$
3	$\frac{n!}{3!(n-x)!} P^3 q^{n-3}$
⋮	⋮
x	$\frac{n!}{x!(n-x)!} P^x q^{n-x}$
⋮	⋮

$n$	$\frac{n!}{n!(0)!} p^n q^0$
$\text{Total} = \sum_{x=0}^n \frac{n!}{x!(n-x)!} p^x q^{n-x} = 1$	

**DISTRIBUIÇÃO BINOMIAL DE PROBABILIDADE:** A distribuição de probabilidade que mostra a probabilidade de x sucessos em n ensaios de um experimento binomial.

**DISTRIBUIÇÃO BINOMIAL NEGATIVA:** Distribuição em que as frequências relativas ou probabilidades são definidas por um binômio com índice negativo. Por exemplo, se os valores da variável aleatória são 0,1,2, dentre outras, a frequência para x=j é o coeficiente de  $t^j$  na expansão de  $(1-pt)^{-n}(1-p)^n$  em potências de t. Esta distribuição é também chamada de Pascal. Em inglês Negative binomial distribution.

**DISTRIBUIÇÃO BINOMIAL NEGATIVA:** Uma distribuição que é parametrizada por uma média m e um parâmetro de agregação k que é grande quando a agregação é pequena; de fato, quando k torna-se grande, a distribuição binomial negativa aproxima-se da distribuição de Poisson.

**DISTRIBUIÇÃO BINOMIAL NEGATIVA:** Dado um acontecimento que ocorre com probabilidade de sucesso p, o número de tentativas falhadas X necessárias para obter r sucessos verifica a chamada distribuição de Pascal ou binomial negativa:  $P_x(X) = \binom{n+r-1}{x} p^r (1-p)^x$  ( $X = 0, 1, 2, 3, \dots, \infty$ ).

**DISTRIBUIÇÃO BIVARIADA:** A distribuição de um par de variáveis  $x_1$  e  $x_2$ , escrita para o caso contínuo como  $dF = f(x_1, x_2) dx_1 dx_2$ . Em inglês Bivariate distribution.

**DISTRIBUIÇÃO CONTÍNUA:** Uma distribuição de probabilidade é dita contínua se o seu conjunto de valores da variável aleatória X(S) é um conjunto contínuo. Em inglês continuous distribution.

**DISTRIBUIÇÃO CONTÍNUA:** Acontece quando um conjunto de fenótipos de uma determinada característica não forma classes distintas.

**DISTRIBUIÇÃO CASUAL DOS TRATAMENTOS:** Processo de designar os tratamentos aos pacientes ao acaso usando, por exemplo, uma tabela de números aleatórios. Este procedimento só é adotado em ensaios clínicos casualizados. Veja distribuição randômica dos tratamentos.

**DISTRIBUIÇÃO CHI-QUADRADA:** Distribuição de variável aleatória contínua da forma  $dF = e^{-x/2} x^{v/2-1} dx / [2^{v/2} \Gamma(v/2)]$ ,  $0 \leq x < \infty$ . Em inglês Chi-square distribution.

**DISTRIBUIÇÃO CONDICIONAL DE  $\theta$  DADO Y:**  $\theta | Y$ .

**DISTRIBUIÇÃO CONJUNTA:** É a distribuição de probabilidade de duas ou mais variáveis aleatórias. O termo é equivalente a distribuição multivariada. Em inglês Joint distribution.

**DISTRIBUIÇÃO DISCRETA:** Uma distribuição de probabilidade é dita discreta ou descontínua se o conjunto de valores que a variável aleatória assume é um conjunto finito ou infinito porém discreto ou enumerável. Em inglês discrete distribution.

**DISTRIBUIÇÃO DA AMOSTRAGEM DA MÉDIA DA AMOSTRA:** Distribuição de frequência das médias de muitas amostras extraídas de uma população que tem uma distribuição normal.

**DISTRIBUIÇÃO DA POPULAÇÃO:** Distribuição de frequência de todos os elementos de uma população.

**DISTRIBUIÇÃO DE AMOSTRA:** Uma distribuição de amostra consiste dos valores de uma estatística computados de todas possíveis amostras de um dado tamanho. Veja o Teorema do Limite Central.

**DISTRIBUIÇÃO DE BERNOULLI:** Caso especial da distribuição binomial na qual é realizado um único experimento de Bernoulli. Em inglês Bernoulli distribution.

**DISTRIBUIÇÃO DE CAU CHY:** Nome geralmente usado para denotar a distribuição contínua  $dF = dx / \pi(1+x^2)$ ,  $-\infty \leq x \leq \infty$ . Em inglês Cauchy distribution.

**DISTRIBUIÇÃO DE CONTÁGIO:** Ver distribuição agregada.

**DISTRIBUIÇÃO DE CORRELAÇÃO CONSTANTE:** Uma distribuição multivariada é de correlação constante se a correlação entre os pares de variáveis é igual a uma constante  $\rho$ . Em inglês Equally correlated distribution.

**DISTRIBUIÇÃO DE DIÂMETRO:** A distribuição de diâmetro (DAP) das árvores de um povoamento florestal representa a estrutura de tamanho das árvores, sendo normalmente apresentada na forma de tabela de frequência de classes de diâmetro, na forma de histograma ou na forma de um modelo de distribuição estatística ajustada aos dados. As distribuições estatísticas geralmente utilizadas são: distribuição Weibull, distribuição exponencial e distribuição (SB) ou distribuição Spetial Beta de Johnson.

**DISTRIBUIÇÃO DE DIRICHLET:** O K-análogo variado da Distribuição Beta. Em inglês Dirichlet distribution.

**DISTRIBUIÇÃO DE ERLANG:** Distribuição do tipo Gama de forma  $f(t) = \lambda(\lambda t)^{k-1} e^{-\lambda t} / \Gamma(k)$ , proposta por Erlang para o tempo entre chegadas e tempo de atendimento em problemas de filas. Em inglês Erlangian distribution.

**DISTRIBUIÇÃO DE ERROS:** Segue geralmente uma distribuição normal, se os erros forem aleatórios.

**DISTRIBUIÇÃO DE FREQUÊNCIA:** Uma tabela em que se lista todos os escores em uma coluna e o número de indivíduos que recebeu cada escore ou frequência na outra. Exemplo:

ESCORE	FREQUÊNCIA
15	1
14	0
13	3
12	2
11	6
10	5
9	2
8	4
7	1



Um arranjo tabular em que se apresenta a frequência com que cada escore aparece num conjunto de dados. É uma apresentação de dados sobre uma certa variável, de forma resumida.

**DISTRIBUIÇÃO DE FREQUÊNCIA:** Tabela que indica as frequências com que ocorrem os casos correspondentes a cada intervalo. Distribuição de escores e distribuição de frequências são expressões praticamente sinônimas. Em inglês frequency distribution

**DISTRIBUIÇÃO DE FREQUÊNCIA:** Uma apresentação sumária da frequência de resposta para cada categoria da variável.

**DISTRIBUIÇÃO DE FREQUÊNCIA:** Especificação da forma como as frequências dos membros de uma população são distribuídas de acordo com os valores que assumem as variáveis. Em inglês Frequency distribution.

**DISTRIBUIÇÃO DE FREQUÊNCIA:** Descrição do número de vezes que os vários atributos de uma variável são observados numa amostra. Relatar que 53% de uma amostra são homens e 47% mulheres é um exemplo de distribuição de frequência. Outro exemplo é informar que 15 das cidades estudadas tinham populações abaixo de 10.000, 23, populações entre 10.000 e 25.000, e assim por diante.

**DISTRIBUIÇÃO DE FREQUÊNCIAS:** Termo que indica a resposta da ordenação, classificação e agrupamento em classes de um conjunto de medidas, a partir da qual é possível ter-se uma idéia da distribuição dos dados, em termos do valor ou valores mais freqüentes, bem como da forma desta distribuição. A representação gráfica da distribuição de frequências é feita comumente sob a forma de histogramas e curvas acumulativas. O modelo teórico da distribuição de freqüências é interpretado a partir do histograma, sobre o qual podem ser ajustados ou a distribuição normal ou a lognormal, que são as distribuições mais comuns na natureza.

**DISTRIBUIÇÃO DE FREQUÊNCIAS:** O mesmo que tabela de frequências.

**DISTRIBUIÇÃO DE FREQUÊNCIAS:** Diagrama em que se coloca o ponto referente ao valor de cada dado em um dos eixos e a freqüência com que ocorre aquele valor no outro eixo.

**DISTRIBUIÇÃO DE FREQUÊNCIA COM INTERVALOS DE CLASSE:** Quando o tamanho da amostra é elevado é mais racional efetuar o agrupamento dos valores em vários intervalos de classe.

**DISTRIBUIÇÃO DE FREQUÊNCIA CUMULATIVA:** Um sumário tabular de dados quantitativos que mostra o número de observações com valores menores ou iguais ao limite de classe superior de cada uma das classes.

**DISTRIBUIÇÃO DE FREQUÊNCIA PERCENTUAL CUMULATIVA:** Um sumário tabular de dados quantitativos que mostra a porcentagem de observações com valores menores ou iguais ao limite de classe superior de cada uma das classes.

**DISTRIBUIÇÃO DE FREQUÊNCIA PERCENTUAL:** Um sumário tabular dos dados que mostra a porcentagem das observações em cada uma das classes não sobrepostas.

**DISTRIBUIÇÃO DE FREQUÊNCIA RELATIVA CUMULATIVA:** Um sumário tabular de dados quantitativos que mostra a fração ou proporção de observações com valores menores ou iguais ao limite de classe superior de cada uma das classes.

**DISTRIBUIÇÃO DE FREQUÊNCIA SEM INTERVALOS DE CLASSE:** É a simples condensação dos dados conforme as repetições de seus valores. Para um rol de tamanho razoável esta distribuição de frequência é inconveniente, já que exige muito espaço.

**DISTRIBUIÇÃO DE LAPLACE:** Distribuição de frequências expressada como a seguinte equação:  $dF=(2\sigma)^{-1}\exp\{-|x-m|/\sigma\}dx$ ,  $-\infty\leq x\leq\infty$  e  $\sigma>0$ . Em inglês Laplace distribution.

**DISTRIBUIÇÃO DE PARASITAS:** Padrão segundo o qual números de parasitas são particionados entre hospedeiros avaliáveis. Se a fração da população que tem  $i$  parasitas é  $p(i)$  então a distribuição é o conjunto de números  $p(0)$ ,  $p(1)$ ,  $p(2)$ , ... . Estatística descritiva desta distribuição inclui prevalência  $(1-p(0))$  e média de intensidade  $(p(1)+2p(2)+3p(3)+\dots)$ . As distribuições úteis aqui incluem a de Poisson e a binomial negativa.

**DISTRIBUIÇÃO DE PARETO:** Uma relação empírica que descreve o número de pessoas  $y$  cuja renda é  $x$ , primeiramente definida por Pareto (1897) com a forma  $y=Ax^{-(1+\alpha)}$ ,  $0\leq x\leq\infty$ . A expressão é usada para denotar qualquer distribuição de frequência desta forma, se relacionadas à renda ou não. A variável  $x$  pode ser medida de algum valor arbitrário, não necessariamente zero. Em inglês Pareto distribution.

**DISTRIBUIÇÃO DE POISSON:** Uma distribuição discreta com frequências relativas em valores variáveis  $0, 1, 2, \dots, x, \dots$  dada por  $\frac{e^{-\lambda}\lambda^x}{x!}$ . Em inglês Poisson distribution.

**DISTRIBUIÇÃO DE POISSON:** Modelo de distribuição de probabilidades de variáveis aleatórias discretas, cujos eventos são raros e referentes ao tempo e ao espaço. A média é igual à variância.

**DISTRIBUIÇÃO DE POISSON:** Distribuição discreta de probabilidade que se aplica a ocorrência de um evento ao longo de um intervalo específico de tempo, distância, área, volume ou outra unidade semelhante.

**DISTRIBUIÇÃO DE POISSON:** Distribuição típica do espalhamento de parasitas entre numa população hospedeira. Do número de formigueiro por hectare, do número de atendimentos a pequenos animais em uma semana em um hospital veterinário, dentre outros.

**DISTRIBUIÇÃO DE POISSON:** Distribuição discreta de probabilidade que dá a freqüência de ocorrência de certos tipos de evento aleatório; pode servir como aproximação da distribuição binomial.

**DISTRIBUIÇÃO DE POISSON:** Distribuição teórica de probabilidade para variáveis aleatórias discretas oriundas de eventos raros. Em inglês Poisson distribution.

**DISTRIBUIÇÃO DE POISSON:** A distribuição de Poisson  $P(\lambda)$  resulta da passagem ao limite quando  $n \rightarrow \infty$  de uma distribuição binomial  $B(n, p)$  em que  $\lambda = np$  se mantém constante. A equação é dada por:

$$P_x(X) = \frac{\lambda^x e^{-\lambda}}{x!} \quad (X = 0, 1, 2, 3, \dots, \infty)$$

**DISTRIBUIÇÃO DE PÓLYA-EGGENBERGER:** Distribuição especial onde se considera um esquema probabilístico do seguinte tipo: de uma urna que contém,  $m$  bolas, apenas diferentes na sua cor, das quais,  $M_p$ , são brancas e,  $M_q$ , são pretas,  $0 < p < 1$ ,  $p + q = 1$ , tiram-se,  $N$ , bolas sucessivamente e após cada extração a bola é repostada na urna assim como,  $\Delta$ , bolas da cor da bola retirada; qual a probabilidade de obter,  $X$ , bolas brancas?

Tem-se que:



$$P(X = x) = \binom{N}{x} \frac{M_p (M_p + \Delta) \cdots (M_p + (x-1)\Delta) M_q (M_q + \Delta) \cdots (M_q + (N-x-1)\Delta)}{M(M+\Delta) \cdots (M+(N-1)\Delta)}$$

$$0 \leq x \leq N, \quad M + (N-1)\Delta > 0$$

Variável aleatória com função de probabilidade dada pela expressão anterior diz-se que tem distribuição de Pólya-Eggenberger. Esta distribuição tem como casos particulares i) a distribuição Binomial com parâmetros, N, p; quando,  $\Delta = 0$ ; ii) a distribuição Hipergeométrica com parâmetros, M, N, p; quando,  $\Delta = -1$ ; iii) a distribuição Uniforme discreta em (N+1) pontos; quando, iv) a distribuição Binomial Negativa: quando,  $N \rightarrow \infty, p \rightarrow 0, e \frac{\Delta}{M} \rightarrow 0$ , de modo que,  $Np, \frac{N\Delta}{M}$ , tendam para valores finitos não iguais a zero.

**DISTRIBUIÇÃO DE PROBABILIDADE:** Distribuição que fornece a probabilidade de um valor x como função do próprio valor; ou de forma mais geral, a probabilidade de ocorrências conjuntas do conjunto de variáveis  $x_1, \dots, x_p$  como função das suas quantidades. Em inglês Probability distribution.

**DISTRIBUIÇÃO DE PROBABILIDADES:** É uma fórmula matemática abstrata que relaciona os valores da característica com a sua probabilidade de ocorrência na população.

**DISTRIBUIÇÃO DE PROBABILIDADE:** A coleção dos pares [x, f(x)] é denominado de distribuição de probabilidade de uma variável aleatória X, isto é, uma distribuição de probabilidade é na realidade o seu gráfico. Em inglês probability distribution.

**DISTRIBUIÇÃO DE PROPABILIDADE:** Conjunto de valores de uma variável aleatória juntamente com suas probabilidades correspondentes.

**DISTRIBUIÇÃO DE QUI QUADRADO**  $\left[\chi_n^2\right]$ : É uma distribuição de probabilidade gama com  $\alpha = \frac{n}{2}$  e  $\lambda = \frac{1}{2}$ . A soma de n variáveis aleatórias normais unitárias independentes tem esta distribuição.

**DISTRIBUIÇÃO DE WEIBULL:** Distribuição da forma  $f(x) = k(x - \varepsilon)^{k-1} (v - \varepsilon)^{-k} \exp\left\{-\left[\frac{(x - \varepsilon)}{(v - \varepsilon)}\right]^k\right\}$  com  $x, v > \varepsilon$  e  $k > 1$ . Em inglês Weibull distribution.

**DISTRIBUIÇÃO DE WISHART:** A distribuição conjunta de variâncias e covariâncias em amostras de uma população Normal p-variada, dada pelo estatístico Escocês John Wishart (1898-1956) em Wishart (1928). Ver teorema de Wishart. Em inglês Wishart distribution.

**DISTRIBUIÇÃO DE VALORES EXTREMOS GENERALIZADOS (GEV):** De acordo com o teorema de Fisher-Tippett (1928) temos inicialmente o seguinte: considere uma sequência de variáveis aleatórias independentes e identicamente distribuídas. Se existirem seqüências de constantes normalizadoras  $c_n > 0$  e  $d_n \in \mathfrak{R}$  e uma distribuição não degenerada H tal que  $\frac{M_n - d_n}{c_n} \xrightarrow{d} H$ , onde  $\xrightarrow{d}$  representa

convergência em distribuição, então h é do tipo de uma das três distribuições ou funções densidade de probabilidade. Gumbel dada por:  $\lambda_1, \lambda_2, \dots, \lambda_k$ ; Fréchet dada por:  $\lambda_1 X_1, \lambda_2 X_2, \dots, \lambda_k X_k = 0$  e Weibull, dada por:

$$H_{III,\alpha}(x) = \begin{cases} \exp\{-(-x)^{-\alpha}\} & x \leq 0 \\ 1, & x > 0, \alpha < 0 \end{cases} . \text{Esses três tipos de distribuições podem ser vistos como membros}$$

de uma única família de distribuições: a distribuição de valores extremos generalizada (Generalized Extreme Value, GEV) padrão, a qual é a representação das três distribuições de valores extremos em uma única família de função densidade de probabilidade de um único parâmetro  $\xi$ . É denotada por  $H_\xi$

e é dada por

$$H_\xi(y) = \begin{cases} \exp\left\{-\left(1 + \xi y\right)^{\frac{-1}{\xi}}\right\} & \text{se } \xi \neq 0 \\ \exp\{-\exp\{-y\}\} & \text{se } \xi = 0, \end{cases}$$

Onde  $1 + \xi y > 0$ , a família de locação e escala correspondente  $H_{\xi,\mu,\sigma}$  é obtida substituindo-se y por  $\frac{(y-\mu)}{\sigma}$ , para  $\mu \in \mathfrak{R}$  e  $\sigma > 0$ . Se o índice de cauda  $\xi = 0$  (caso limite quando  $\xi \rightarrow 0$ ),  $H_\xi$  corresponde

à distribuição Gumbel. Para se provar que  $H_\xi(x) \rightarrow H_\lambda(x)$  basta usar o resultado  $(1 + \gamma x)^{\frac{1}{\gamma}} \rightarrow \exp(x)$  quando  $\lambda \rightarrow 0$ . Os casos onde  $\xi < 0$  ou  $\xi > 0$  correspondem respectivamente às distribuições Weibull ou a Fréchet. A reparametrização  $\xi = \frac{1}{\alpha}$  faz com que as 3 distribuições se tornem um modelo contínuo,

unificado. A distribuição GEV foi introduzida por Jenkinson (1955) e é também conhecida como a distribuição de Von Mises- Jenkinson. Sua densidade ( $\mu = 0$ ) é dada por:

$$H_{\xi,\sigma}(y) = \begin{cases} \exp\left\{-\left(1 + \xi \frac{y}{\sigma}\right)^{\frac{-1}{\xi}}\right\} \frac{1}{\sigma} \left(1 + \xi \frac{y}{\sigma}\right)^{\frac{-1}{\xi}-1}, & \text{se } \xi < 0 \text{ e } -\infty < y < \left(\frac{-\sigma}{\xi}\right); \\ \text{ou } \xi > 0 \text{ e } y \geq \left(\frac{-\sigma}{\xi}\right) \\ \exp\left\{-\exp\left\{-\frac{y}{\sigma}\right\}\right\} \frac{1}{\sigma} \exp\{-y\} & \text{se } \xi = 0 \text{ e } y \in \mathfrak{R}. \end{cases}$$

O parâmetro de forma  $\xi$  pode ser usado para modelar um grande número de comportamento de caudas, como por exemplo, grandes cheias em barragens, chuvas torrenciais, dentre outros.

**DISTRIBUIÇÃO DO MESMO TIPO:** Duas distribuições de probabilidade  $F$  e  $F^*$  e são do mesmo tipo se existirem constantes a e b tais que  $F^*(ax + b) = F(x) \quad \forall x$ . Por exemplo, suponha  $Z = F = N(0,1)$ , uma variável aleatória normal padrão, e  $Z^* = aZ + b$ . Então  $Z^* = F^* = N\left(b, a^2\right)$  e  $F^*(ax + b) = F(x)$ .

**DISTRIBUIÇÃO DESCONTÍNUA:** Tipo de distribuição que acontece quando um conjunto de fenótipos de uma determinada característica pode ser agrupado em classes distintas.

**DISTRIBUIÇÃO EM EPIDEMIOLOGIA:** É o estudo da variabilidade da freqüência das doenças ao nível coletivo, em função de variáveis ligadas ao tempo, aos espaços ambiental e populacional e à pessoa.

**DISTRIBUIÇÃO EXPONENCIAL:** Modelo de distribuição de probabilidades de variáveis aleatórias contínuas, referente ao intervalo de tempo decorrido entre eventos raros e discretos.

**DISTRIBUIÇÃO EXPONENCIAL:** Uma distribuição cuja função de densidade é dada por  $f(x) = \sigma^{-1} \exp\left\{-\frac{(x-m)}{\sigma}\right\}$ ,  $m \leq x \leq \infty$ . O parâmetro  $\sigma$  é o desvio padrão da distribuição. Em inglês

Exponential distribution.

**DISTRIBUIÇÃO EXPONENCIAL DUPLA:** A distribuição com função de densidade  $f(x) = ae^{b|x-c|}$ ,  $-\infty \leq x \leq \infty$ ,  $a, b$  e  $c$  constantes,  $b < 0$ . A distribuição pode ser considerada como uma exponencial ordinária deslocada até o ponto  $c$ . Um exemplo de distribuição exponencial dupla é a distribuição da semi-amplitude de uma população retangular. Em inglês Double exponential distribution.

**DISTRIBUIÇÃO EXPONENCIAL DUPLA:** A definição desse tipo de modelo probabilístico será feita por meio de um exemplo. Consideremos  $X$  e  $Y$  duas variáveis aleatórias independentes com distribuição exponencial de parâmetro  $\lambda$ . Como podemos determinar a distribuição de probabilidade da variável  $W = X - Y$ ? A distribuição exponencial é um modelo que descreve tempo de espera, por exemplo, além de poder ser usada em outras circunstâncias. Podemos olhar para este problema como a determinação da distribuição de probabilidade da diferença entre dois tempos de espera, por exemplo, da passagem de dois automóveis com velocidades diferentes no mesmo local. Assim, sejam

$$f_x(x) = \lambda e^{-\lambda x}, \quad x > 0; \quad e \quad f_y(y) = \lambda e^{-\lambda y}, \quad y > 0.$$

Pelo teorema que acabamos de ver, a densidade de probabilidade da diferença  $W = X - Y$  é dada por  $f_w(W) = \int_{-\infty}^{+\infty} f_y(X - W) f_x(x) dx$ . A função integrada só é diferente de 0 quando se tem simultaneamente

os valores  $x > 0$  e  $x - w > 0$ . Sendo assim, se  $w \geq 0$  tem-se

$$f_w(W) = \int_w^{+\infty} \lambda e^{-\lambda(x-w)} \lambda e^{-\lambda x} dx = \frac{\lambda}{2} e^{\lambda w} \int_w^{+\infty} 2\lambda e^{-2\lambda x} dx = \frac{\lambda}{2} e^{-\lambda w}.$$

Por outro lado, quando  $w < 0$ , vem  $f_w(W) = \int_0^{+\infty} \lambda e^{-\lambda(x-w)} \lambda e^{-\lambda x} dx = \frac{\lambda}{2} e^{\lambda w} \int_0^{+\infty} 2\lambda e^{-2\lambda x} dx = \frac{\lambda}{2} e^{\lambda w}$ . De uma forma mais resumida, temos então que a densidade da diferença de duas exponenciais independentes com o mesmo parâmetro  $\lambda$  é dada por:  $f_w(W) = \frac{\lambda}{2} e^{-\lambda|w|}$ ,  $-\infty < w < +\infty$ . A esta distribuição dá-se nome de distribuição

de Laplace, segundo o matemático francês Pierre Simon de Laplace, e o seu gráfico existe para vários valores de  $\lambda$ . A função de distribuição associada a esta densidade pode ser facilmente obtida integrando a densidade, isto é,  $F_w(W) = \int_{-\infty}^w \frac{\lambda}{2} e^{-\lambda|x|} dx$ . O cálculo desta integral simplifica-se se tivermos o cuidado de separar

os casos em que o argumento é negativo ou positivo. Assim, quando  $w < 0$ ,  $F_w(W) = \int_{-\infty}^w \frac{\lambda}{2} e^{\lambda x} dx = \frac{1}{2} e^{\lambda w}$ . Se  $w > 0$ , então  $F_w(W) = \frac{1}{2} + \int_0^w \frac{\lambda}{2} e^{-\lambda x} dx = 1 - \frac{e^{-\lambda w}}{2}$ . A esta distribuição também se chama exponencial dupla

porque o seu gráfico dobrado em torno do eixo dos  $yy$  é proporcional á densidade exponencial. Mais rigorosamente, isto corresponde ao fato de que  $A > C$  tem distribuição exponencial, já que, com  $w > 0$

, temos que,  $F_{|w|}(W) = P\{|W| \leq w\} = P\{-w \leq W \leq w\} = F_w(W) - F_w(-W)$

$$F_{|w|}(W) = 1 - \frac{e^{-\lambda w}}{2} - \frac{e^{-\lambda w}}{2}$$

$$F_{|w|}(W) = 1 - e^{-\lambda w}.$$

**DISTRIBUIÇÃO EM J:** Distribuição de frequências acentuadamente assimétrica quer positiva quer negativamente. Sua representação gráfica assemelha-se a um jota de pé ou invertido. Em inglês J-shaped distribution.

**DISTRIBUIÇÃO F:** É o quociente  $F_{nm}$  das variáveis aleatórias independentes  $X_n \sim \chi_n^2$  e  $X_m \sim \chi_m^2$  divididas pelos respectivos números de graus de liberdade ( $n$  corresponde ao numerador), ou seja:

$$F_{nm} = \frac{\frac{X_n}{n}}{\frac{X_m}{m}}$$

**DISTRIBUIÇÃO F:** É a distribuição teórica de probabilidades da razão de duas quantidades ou variáveis aleatórias independentes, cada uma distribuída como a variância de uma amostra normal.

**DISTRIBUIÇÃO F:** A distribuição da razão de duas quantidades independentes, cada uma distribuída como a variância de uma amostra Normal. Em inglês F-distribution.

**DISTRIBUIÇÃO F:** Distribuição de variável aleatória contínua relacionada com a distribuição qui-quadrado; é usada em processos de análise da variância e em análise da regressão.

**DISTRIBUIÇÃO F DE SNEDECOR:** Um termo popular para a distribuição da razão de variâncias. Em inglês Snedecor F distribution.

**DISTRIBUIÇÃO FIDUCIAL:** A distribuição de um parâmetro necessário para inferência fiducial sobre esse parâmetro. Em inglês Fiducial distribution.

**DISTRIBUIÇÃO GAMA:** Uma distribuição cuja densidade é da forma  $f(x) = \frac{e^{-x} X^{\lambda-1}}{\Gamma(\lambda)}$ ,  $0 \leq x < \infty$ . É

também conhecida como Tipo III de Pearson ou simplesmente como distribuição do tipo III. Em inglês Gamma distribution.

**DISTRIBUIÇÃO GAMA  $[G(\alpha, \lambda)]$ :** É a distribuição de uma variável aleatória positiva, cuja função geradora de momentos é dada por:

$$G(t) = \left[1 - \frac{t}{\lambda}\right]^{-\alpha}.$$

**DISTRIBUIÇÃO GAUSSIANA INVERSA:** Seja  $X$  uma variável com distribuição Normal, então,  $X^{-1}$  terá distribuição Normal inversa, também chamada de Gaussiana inversa. Em inglês Inverse Gaussian distribution.

**DISTRIBUIÇÃO GAUSSIANA:** Nome alternativo para a distribuição Normal. Em inglês Gaussian distribution.

**DISTRIBUIÇÃO GENERALIZADA DE VALORES ESTREMOS (GEV):** Atualmente uma das grandes preocupações da humanidade é em relação ao aquecimento global e suas conseqüências para a vida no planeta. Um dos efeitos mais visíveis é o aumento da ocorrência e da amplitude de eventos extremos como, por exemplo, chuvas torrenciais seguidas de grandes períodos de estiagem, aumentos e quedas bruscas da temperatura. Esses fatores devem ser levados em consideração no planejamento de construções e de atividades relacionadas à gestão e saúde, por exemplo. Neste contexto a previsão do comportamento destes fenômenos extremos torna-se de suma importância para governos e populações em geral, tornando o

estudo da teoria de valores extremos cada dia mais imprescindível em estudos de variáveis ambientais. Os fundamentos da teoria de valores extremos foram inicialmente expostos por Fisher e Tippet (1928), que definiram os três tipos possíveis (I, II e III) de distribuições assintóticas de valores extremos, conhecidas como de Gumbel, de Fréchet e de Weibull, respectivamente. No entanto, o primeiro a estudar e formalizar a aplicação estatística destas distribuições foi Gumbel (1958), cuja metodologia tem sido freqüentemente aplicada à máxima anual de séries de dados referentes a precipitações e vazões de rios. A distribuição de Gumbel inicialmente foi apresentada pelo próprio autor como distribuição de dupla exponencial em 1958. A função de probabilidade é definida por meio dos valores de máximos e mínimos. Desta forma, as funções de probabilidade, respectivamente aos valores de máximos e mínimos, são :

$$f(x) = \alpha \cdot \exp(-y - e^{-y}) \text{ e } f(x) = \alpha \cdot \exp(y - e^y); y = \alpha(x - u)$$

Em ambas expressões, o parâmetro u é denominado como parâmetro de posição, cujo domínio é dado por  $-\infty < u < \infty$ . O parâmetro de escala é representado por  $\alpha$ , sendo este restrito a  $\alpha > 0$  e x indica a variável em estudo. Jenkinson (1955) propôs que os três tipos de distribuições de valores extremos (Gumbel, de Fréchet e de Weibull) poderiam ser representados numa forma paramétrica única, designada por distribuição generalizada de valores extremos (GEV), cuja função de distribuição de probabilidade acumulada é dada pela seguinte expressão,

$$F(x) = \exp \left\{ - \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]^{\left( \frac{1}{\xi} \right)} \right\}$$

a qual é definida em,  $-\infty < x < \mu - \sigma/\xi$ , para  $\xi < 0$  e  $\mu - \sigma/\xi < x < \infty$ , para  $\xi > 0$ , sendo  $\mu, \sigma$  e  $\xi$  os parâmetros de posição, escala e de forma respectivamente, com  $\sigma > 0$ . A distribuição GEV tem sido utilizada com grande freqüência em estudos de fenômenos ambientais principalmente para solucionar problemas relacionados à áreas de Engenharia, entre os quais, velocidades máximas de ventos, temperaturas máximas e mínimas e em estudos de precipitação pluvial máxima. Desde 1975, a Natural Environmental Research Council (NERC) recomenda a distribuição GEV para a análise de freqüências de enchentes no Reino. Para fazer inferências sobre esses parâmetros, inúmeras sugestões foram propostas, entre elas, técnicas gráficas, estimadores baseados no método dos momentos; métodos de estatísticas de ordem; método dos momentos de probabilidade ponderada; método máxima verossimilhança; método máxima verossimilhança penalizada e métodos Bayesianos.

**DISTRIBUIÇÃO GEOMÉTRICA:** Seja uma variável aleatória X que representa o número de tentativas de um experimento até obter um resultado de interesse, X tem distribuição geométrica se sua função de massa de probabilidade é dada por  $f(x) = (1-p)^{x-1}p$ , sendo p a probabilidade de obter o resultado de interesse. Em inglês Geometric distribution.

**DISTRIBUIÇÃO HIPERGEOMÉTRICA:** Uma distribuição de uma variável discreta geralmente associada com a amostragem sem reposição de uma população finita. A freqüência de r sucessos e n-r fracassos em uma amostra de n obtida de uma população de N em que existem Np sucessos e Nq fracassos (p+q=1) é  $\frac{n!(Np)^r(Nq)^{n-r}}{(r!(n-r)!N^n)}$  onde  $N^{[r]} = N(N-1)...(N-r+1)$ . Em inglês Hypergeometric distribution.

**DISTRIBUIÇÃO HIPERGEOMÉTRICA:** Modelo de distribuição de probabilidades de variáveis aleatórias discretas, com as mesmas características da distribuição binomial. O tamanho da amostra, contudo,

é relativamente grande em relação ao tamanho da população, alterando-se em grau acentuado a probabilidade dos elementos do universo pela retirada de cada unidade sem reposição.

**DISTRIBUIÇÃO HIPERGEOMÉTRICA:** Este modelo de distribuição em inglês denomina-se hypergeometric distribution.

**DISTRIBUIÇÃO HIPERGEOMÉTRICA:** Distribuição discreta de probabilidades, aplicável quando se faz amostragem sem reposição.

**DISTRIBUIÇÃO HIPERGEOMÉTRICA:** Uma variável aleatória X inteira não negativa tem uma distribuição hipergeométrica H(N, n, p) quando é igual ao número de sucessos em n tentativas sendo o número de sucessos máximo possível Np, o número máximo de insucessos N(1-p) e cada tentativa faz reduzir de uma unidade o número máximo possível de sucessos ou de insucessos na tentativa seguinte, conforme o resultado. Por exemplo, quando se tem uma urna inicialmente com N bolas, das quais  $N_1 = Np$  são brancas e  $N_2 = N(1-p)$  são pretas, e se extraem sem reposição n bolas, o número x de bolas brancas que se podem extrair segue essa distribuição, onde p tem de ser um número racional, porque Np tem de ser inteiro. A distribuição teórica ou especial de probabilidade é dada pela seguinte equação:

$$P_x(X) = \frac{\binom{N_1}{x} \binom{N_2}{n-x}}{\binom{N}{n}}$$

$$x = 0, 1, 2, 3, \dots, \text{Min}(N_1, n)$$

**DISTRIBUIÇÃO HIPERGEOMÉTRICA K-DIMENSIONAL:** É aquele modelo usado para representar probabilidades de ocorrências ou eventos aleatórios quando considera-se uma população com N elementos contendo  $M_1$  elementos com a característica  $A_1$ ,  $M_2$  elementos com a característica  $A_2$ , ...,  $M_k$  elementos com a característica  $A_k$ , e retira-se uma amostra de n elementos sem reposição. O vetor aleatório  $X = \{X_1, X_2, \dots, X_{k-1}\}$ , onde  $X_i$  é o número de elementos com a característica  $A_i$  na amostra n, sendo  $i = \{1, 2, 3, \dots, K - 1\}$ , segue uma distribuição hipergeométrica k-dimensional de parâmetros  $N, n, p_1 = \frac{M_1}{N}, \dots, p_{k-1} = \frac{M_{k-1}}{N}$ , onde  $p_i$  representa a proporção inicial de elementos com a característica

$A_i$  na população, e escreve-se  $X \cap H(N, n, p_1, p_2, \dots, p_{k-1})$ . O modelo matemático da sua função de massa de probabilidade é dada por:

$$f(x_1, x_2, \dots, x_{k-1}) = P(x_1, x_2, \dots, x_{k-1}) = \frac{C_{x_1}^{Np_1} \dots C_{x_{k-1}}^{Np_{k-1}} C_{x_k}^{Np_k}}{C_n^N}, \text{ onde } x_k = n - \sum_{i=1}^{k-1} x_i$$

$$\text{e } p_k = 1 - \sum_{i=1}^{k-1} p_i, \text{ para } \max\{0, n - N(1 - p_i)\} \leq x_i \leq \min\{n, Np_i\}, i = 1, \dots, k, \text{ e } 0 < p_1, \dots, p_k < 1$$

**DISTRIBUIÇÃO HOMOCEDÁSTICA:** Condição pela qual, para todos os valores de uma variável, as distribuições dos valores da outra variável têm aproximadamente a mesma variabilidade.

**DISTRIBUIÇÃO INVERSA:** Esta expressão é usada para denotar a distribuição da recíproca de uma variável. Em inglês Inverse distribution.



**DISTRIBUIÇÃO LOGNORMAL:** Se o logaritmo de uma variável está distribuído Normalmente, então a variável é chamada de variável com distribuição lognormal. Em inglês Log-Normal distribution.

**DISTRIBUIÇÃO LOGNORMAL:** É aquela encontrada geralmente em muitos problemas de avaliação de reservas. A distribuição de frequências apresenta-se com uma assimetria positiva, onde ocorre uma grande quantidade de valores baixos e uns poucos valores altos, que praticamente definem a jazida, principalmente em casos de metais raros. Formalmente, a distribuição lognormal é definida como uma distribuição contínua caracterizada pela propriedade que os logaritmos das observações seguem uma distribuição normal. A função densidade de probabilidade da distribuição lognormal é dada por:

$$f(X) = \frac{1}{X\beta\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left[\frac{\ln(X)-\alpha}{\beta}\right]^2}$$

onde:  $\alpha$  é a média dos logaritmos de  $X$ ;  $\beta$  é o desvio padrão dos logaritmos de  $X$ . Os dois parâmetros:  $\alpha$  e  $\beta^2$ , definem a forma da curva de distribuição de probabilidades. A distribuição lognormal é sempre assimétrica para a direita, sendo que o grau de assimetria depende somente do valor de  $\beta^2$ , a variância dos logaritmos das observações. O objetivo básico da transformação não linear (logarítmica) embutida na função densidade de probabilidade, é, a mudança da forma da distribuição para uma distribuição normal ou aproximadamente normal. Entretanto, algumas distribuições de frequências apresentam-se após a transformação logarítmica, com certa assimetria negativa (assimétrica para a esquerda), que pode ser corrigida pela adição de uma constante o terceiro parâmetro às observações originais, antes da transformação logarítmica.

A função densidade de probabilidade da distribuição lognormal a três parâmetros é descrita por:

$$f(X) = \frac{1}{(X+C)\beta\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{\ln(X+C)-\alpha}{\beta}\right]^2}$$

A constante  $C$ , o terceiro parâmetro da distribuição lognormal, pode ser estimada segundo Landim (1986), como:

$$C = \frac{m^2 - P_1 P_2}{P_1 + P_2 - 2m}$$

onde:  $m$  é a mediana, ou seja, o valor de teor correspondente a 50 % da distribuição;  $p_1$  é o valor de teor correspondente a um percentil entre 5 e 20 %;  $p_2$  é o valor de teor correspondente a um percentil entre 100- $p_1$ .

**DISTRIBUIÇÃO MARGINAL:** Para um vetor de variáveis aleatórias, a função de distribuição de uma variável independente das outras. Em inglês Marginal distribution.

**DISTRIBUIÇÃO MULTINOMIAL:** A distribuição discreta associada com eventos que podem ter mais de dois resultados: é a generalização da distribuição binomial. Se há  $k$  resultados possíveis incompatíveis e exaustivos de algum evento para o qual as probabilidades separadas são  $p_i$  ( $i=1, 2, \dots, k$ ) então, em  $n$  experimentos, a distribuição de  $x_1$  eventos de primeiro tipo,  $x_2$  de segundo tipo...  $x_k$  de  $k$ -ésimo tipo é

$$f(x_1, x_2, \dots, x_k) = n! \prod_i \frac{\binom{x_i}{p_i}}{x_i!}, 0 \leq x_i \leq n. \text{ Em inglês Multinomial distribution.}$$

**DISTRIBUIÇÃO MULTINOMIAL:** É a generalização da distribuição binomial de duas ou mais variáveis: se os acontecimentos  $A_1, A_2, \dots$  forem mutuamente exclusivos e tiverem probabilidades de ocorrência  $p_1, p_2, \dots, p_k$  (em que  $p_1 + p_2 + \dots + p_k = 1$ ), o vetor aleatório  $X_1, X_2, \dots, X_k$ , que é igual ao número de ocorrências de  $A_1, A_2, \dots, A_k$  em  $n$  tentativas independentes, tem uma distribuição multinomial  $M(n, p_1, p_2, \dots, p_k)$ .

**DISTRIBUIÇÃO MULTINOMIAL:** É aquele modelo matemático onde considera-se a realização de  $n$  execuções independentes, e sob as mesmas condições, de uma experiência aleatória, denominadas de  $n$  provas de Bernoulli, em cada uma das quais se verifica a ocorrência ou não dos acontecimentos mutuamente exclusivos  $A_1, A_2, \dots, A_k$ , denominados sucessos, sendo  $p_i = P(A_i)$  constante ao longo das  $n$  execuções ou repetições da experiência aleatória. O vetor aleatório  $X = (X_1, X_2, \dots, X_{k-1})$ , onde  $X_i$  é o número de ocorrências do acontecimento ou evento  $A_i$  nas  $n$  execuções da experiência,  $i = 1, 2, \dots, k-1$ ; segue a distribuição multinomial de parâmetros  $n, p_1, p_2, \dots, p_{k-1}$ , e escreve-se a notação na forma  $X \sim M(n, p_1, p_2, \dots, p_{k-1})$ . A sua função de probabilidade é dada por:

$$f(x_1, x_2, \dots, x_{k-1}) = \frac{n!}{x_1! \dots x_{k-1}! x_k!} p_1^{x_1} \dots p_{k-1}^{x_{k-1}} p_k^{x_k}, \text{ onde } x_k = n - \sum_{i=1}^{k-1} x_i \text{ e } p_k = 1 - \sum_{i=1}^{k-1} p_i, \text{ para}$$

$$0 \leq x_1, x_2, \dots, x_k \leq n, \text{ e } 0 < p_1, p_2, \dots, p_k < 1.$$

**DISTRIBUIÇÃO MULTIVARIADA:** A distribuição simultânea de  $p$  variáveis ( $p > 1$ ): ou de forma equivalente, a distribuição de probabilidades de  $p$  variáveis. Em inglês Multivariate distribution.

**DISTRIBUIÇÃO NÃO SINGULAR:** Distribuição multivariada de  $p$  variáveis que não pode, por nenhuma transformação linear das suas variáveis, ser convertida em uma distribuição com menos de  $p$  variáveis. Em inglês Non-singular distribution.

**DISTRIBUIÇÃO NORMAL:** Também conhecida como distribuição Gaussiana, tem a forma de um sino, e se caracteriza pela média e desvio padrão da distribuição. Esta curva surge quando os dados são afetados por muitas variáveis independentes em que nenhuma delas seja mais importante que as outras, por exemplo, que é função de muitos processos, tais como genética e ambiente. Muitos testes estatísticos partem da suposição de que os dados estão normalmente distribuídos no caso da estatística paramétrica.

**DISTRIBUIÇÃO NORMAL:** É sob esta forma que ocorre a maioria das distribuições de frequências de variáveis aleatórias encontradas na natureza e, por isso, é de longe a distribuição teórica mais utilizada., muita observações são normalmente, ou no mínimo aproximadamente normalmente distribuídas. Ainda segundo estes autores, mesmo para observações que estão longe de serem normais, a distribuição normal pode ser aplicada se um problema pode ser resolvido considerando o comportamento de uma variável formada pelo cálculo de estatísticas de um grupo de observações do que considerando o comportamento das observações originais. Além disso, segundo aqueles autores, pode-se utilizar uma função matemática para transformar cada observação numa nova observação que segue a distribuição normal. A função densidade de probabilidade, que descreve matematicamente a distribuição normal é dada por:

$$f(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{(X-\mu)}{\sigma}\right]^2}, \text{ onde: } f(X) \text{ é a função densidade de probabilidade; } X \text{ é a variável aleatória; } m$$

e  $s$ , respectivamente a média e o desvio padrão, definem a forma da curva. Ver curva de Gauss, distribuição Lognormal e teorema do limite central.

**DISTRIBUIÇÃO NORMAL:** Dados distribuídos na curva simétrica em forma de sino, onde moda, mediana e média aritmética têm o mesmo valor.



**DISTRIBUIÇÃO NORMAL:** Uma família de distribuições simétricas e em formato de sino, cuja curva é descrita matematicamente por uma equação geral, também chamada função probabilística normal de Laplace-Gaussian.

**DISTRIBUIÇÃO NORMAL:** Curva normal ou de Gauss, em forma de sino; distribuição simétrica de freqüências, determinada pelo conhecimento da média e do desvio-padrão. Propriedades: a média  $\pm 1$  desvio-padrão abrange 68% das observações; a média  $\pm 2$  desvios-padrão inclui 95% das observações; a médias  $\pm 3$  desvios-padrão encerra, praticamente, todas as observações (99%).

**DISTRIBUIÇÃO NORMAL:** É a propriedade de um conjunto de dados de se distribuir simetricamente em torno da média. Isto significa que o dado mais freqüente tem o valor da média, ao passo que os demais valores maiores e menores decrescem em freqüência à medida que se afastam da média.

**DISTRIBUIÇÃO NORMAL:** Modelo de distribuição de probabilidades de variáveis aleatórias contínuas de largo emprego em estatística, caracterizando-se por ser simétrica, em forma de sino, assintótica, cuja área sob a curva é igual à unidade.

**DISTRIBUIÇÃO NORMAL:** A mais importante distribuição de variável contínua, sua função de densidade tem a forma de um sino, muitas populações seguem a distribuição normal.

**DISTRIBUIÇÃO NORMAL:** Distribuição contínua de probabilidade puramente teórica na qual o eixo horizontal representa todos os valores possíveis de uma variável e o eixo vertical representa a probabilidade desses valores ocorrerem. Os valores probabilísticos sobre as variáveis estão agrupados em torno da média em um padrão simétrico, unimodal, conhecido como curva normal, ou forma de sino.

**DISTRIBUIÇÃO NORMAL:** A distribuição de freqüências de variáveis contínuas com amplitude infinita cuja função de densidade é:

$$f(x) = \frac{\exp\left(\frac{-(x-m)^2}{2\sigma^2}\right)}{\sqrt{2\pi\sigma^2}}, \quad -\infty \leq x \leq \infty, \text{ onde } m \text{ é a média e } \sigma \text{ o desvio padrão. Em inglês Normal distribution.}$$

**DISTRIBUIÇÃO NORMAL:** Uma distribuição contínua, simétrica e na forma de sino sobre a média, sendo a média, a mediana e a moda iguais. Sessenta e oito por cento das observações enquadram-se em mais ou menos 1 desvio-padrão da média, aproximadamente 95% enquadram-se mais ou menos em dois desvios-padrão e, aproximadamente, 99,5% enquadram-se em mais ou menos 3 desvios-padrão.

**DISTRIBUIÇÃO NORMAL:** O teorema do limite central afirma que a distribuição da soma de  $n$  variáveis aleatórias independentes de médias e variâncias finitas se aproxima, quando  $\bar{X}$ , da chamada distribuição normal ou gaussiana  $N(\mu, \sigma^2)$ , de média e desvio padrão  $\mu$  e  $\sigma$ , cuja densidade é dada por:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

**DITRIBUIÇÃO NORMAL:** Também chamada distribuição de Gauss, possui as seguintes características: é unimodal, simétrica, campanular. Muitas medidas antropométricas distribuem-se normalmente. Em inglês normal distribution.

**DISTRIBUIÇÃO NORMAL:** Em muitos dos histogramas oriundos de problemas resolvidos com dados numéricos, deve ser verificado que algumas variáveis têm distribuições que podem ser descritas como simétricas em torno da média e com forma de sino. Seria útil se estas distribuições fossem descritas de modo empírico como uma equação numérica de uma curva que aproximasse bem a curva simétrica com a forma de sino tantas vezes vista. Uma boa aproximação para muitas dessas variáveis é dada pela distribuição normal. Esta é descrita pela equação:

$$f(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2},$$

onde  $\mu$  e  $\sigma$  são respectivamente a média e o desvio padrão da distribuição. A distribuição normal é simétrica em torno da média e tem a forma de sino. Para distribuições com um pequeno desvio padrão o sino é alto e apertado, e para distribuições com um grande desvio padrão o sino é pequeno e alargado. Se uma variável é normalmente distribuída então 95% dos seus valores dentro de 2 desvios ou mais, precisamente 1,96 desvios padrões de cada lado da média.

**DISTRIBUIÇÃO NORMAL:** Curva normal ou de Gauss, em forma de sino; distribuição simétrica de freqüências, determinada pelo conhecimento da média e do desvio-padrão. Propriedades: a média  $\pm 1$  desvio-padrão abrange 68% das observações; a média  $\pm 2$  desvios-padrão inclui 95% das observações; a média  $\pm 3$  desvios-padrão encerra, praticamente, todas as observações, cerca de 99%.

**DISTRIBUIÇÃO NORMAL BIVARIADA:** Distribuição de dados emparelhados em que, para qualquer valor fixo de uma variável, os valores da outra variável se distribuem segundo a normal. Cujas expressões da função densidade de probabilidade normal bivariada é apresentada a seguir conforme a equação.

$$f(x_1, x_2) = \frac{1}{2\pi\sqrt{\sigma_{11}\sigma_{22}(1-\rho_{12}^2)}} \exp\left\{\frac{-1}{2(1-\rho_{12}^2)}\left[\left(\frac{X_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2 + \left(\frac{X_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2 - 2\rho_{12}\left(\frac{X_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{X_2-\mu_2}{\sqrt{\sigma_{22}}}\right)\right]\right\}$$

Se  $X_1$  e  $X_2$  não são correlacionadas, então  $\rho_{12} = 0$ , e a densidade conjunta pode ser escrita como o produto das densidades normais univariadas, ou seja,  $f(x_1, x_2) = f(x_1)f(x_2)$ , além do que  $X_1$  e  $X_2$  são ditas independentes.

**DISTRIBUIÇÃO NORMAL DE PROBABILIDADE:** Uma distribuição contínua de probabilidade. Sua função de densidade da probabilidade tem a forma de sino e é determinada pela média  $\mu$  e pelo desvio-padrão  $\sigma$ .

**DISTRIBUIÇÃO NORMAL MULTIVARIADA:** Diz-se que um conjunto de  $n$  variáveis aleatórias reais  $X_i$  ( $i = 1, \dots, n$ ) apresenta uma distribuição normal multivariada se e somente se a combinação linear dessas variáveis tiver uma distribuição normal, quaisquer que sejam os coeficientes  $a_i$  dessa combinação:  $\sum_{i=1}^n a_i x_i \cap N(\mu, \sigma)$

**DISTRIBUIÇÃO NORMAL MULTIVARIADA:** Generalização da distribuição normal univariada para o caso de  $p$  variáveis. Uma distribuição normal multivariada de grupos de amostras é uma suposição básica exigida para a validade dos testes de significância em MANOVA.

**DISTRIBUIÇÃO NORMAL PADRONIZADA:** Distribuição normal com média 0 e desvio-padrão 1.

**DISTRIBUIÇÃO NORMAL PADRONIZADA:** A distribuição normal com média 0 e variância 1.

**DISTRIBUIÇÃO NORMAL PADRÃO:** A distribuição normal com média de zero e desvio padrão de 1.

**DISTRIBUIÇÃO NORMAL PADRÃO:** A distribuição normal padrão ou reduzida que tem média zero e variância 1.

**DISTRIBUIÇÃO RANDÔMICA DOS TRATAMENTOS:** O mesmo que distribuição casual dos tratamentos.

**DISTRIBUIÇÃO SIMÉTRICA:** Histograma que apresenta uma distribuição de frequências de forma aproximadamente simétrica, relativamente a uma classe média.

**DISTRIBUIÇÃO t:** Uma distribuição de variável aleatória contínua relacionada com as distribuições normal e qui-quadrado; usada para cálculo de intervalos de confiança e teste de hipóteses sobre pequenas amostras.

**DISTRIBUIÇÃO t DE STUDENT:** Distribuição da razão da média amostral proveniente de uma distribuição Normal e uma variável  $\chi^2$ . Em inglês Student t distribution.

**DISTRIBUIÇÃO t DE STUDENT:** Um ajuste à distribuição normal para levar em conta pequenas amostras ( $n < 30$ ).

**DISTRIBUIÇÃO t DE STUDENT:** Distribuição em forma de sino, em geral associada a experimentos com pequenas amostras; chamada também distribuição t.

**DISTRIBUIÇÃO t DE STUDENT:** É a função de distribuição de frequências de amostragem da estatística:  $t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$ , onde:  $\mu$  é a média populacional; S é o desvio padrão amostral.

Ao contrário da distribuição normal que é uma curva, a distribuição t é uma família de curvas baseada em um único parâmetro que são os graus de liberdade. A função densidade de probabilidade da distribuição t de Student (pseudônimo de W.S. Gosset) é:

$$f(X) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{2\pi}\Gamma\left(\frac{\nu}{2}\right)} \cdot \left(1 + \frac{X^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

**onde:** X é a variável aleatória;  $\nu$  são os graus de liberdade;  $\Gamma\left(\frac{\nu}{2}\right) = \left(\frac{\nu}{2} + 1\right)!$  se  $\nu$  é par;

$$\Gamma\left(\frac{\nu}{2}\right) = \left(\frac{\nu}{2} - 1\right) \cdot \left(\frac{\nu}{2} - 2\right) \cdots \left(\frac{1}{2}\right) \Gamma\left(\frac{1}{2}\right), \text{ se } \nu \text{ é ímpar; e } \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}.$$

**DISTRIBUIÇÃO T DE STUDENT:** Distribuição cuja variável aleatória é o quociente  $T_n$  das variáveis aleatórias independentes  $Z \sim N(0, 1)$  e da raiz quadrada de  $X \sim \chi_n^2$  dividida pelo respectivo número de graus de liberdade, ou seja:

$$T_n = \frac{Z}{\sqrt{\frac{X}{n}}}$$

**DISTRIBUIÇÃO TRINOMIAL:** Suponha-se que um determinado experimento aleatório tem três resultados possíveis: resultados dos tipos 1, 2 ou 3 (por exemplo, vitória, derrota ou empate) com probabilidade p, q e r,  $p + q + r = 1$ . Suponha-se que essa experiência seja repetida independentemente n vezes, onde n é algum número inteiro fixado. Representa-se por X e Y, respectivamente, os números de resultados do tipo 1 e do tipo 2 nessa seqüência de experiências. Conseqüentemente, n-X-Y representa o número de resultados do tipo 3. As variáveis aleatórias x, Y são tidas como possuindo uma distribuição trinomial. O contradomínio conjunto consiste de todos os pares (i, j) de números inteiros não negativos que satisfazem  $0 \leq i + j \leq n$ . Para calcular a probabilidade  $p(i, j)$  de que  $X = i$  e  $Y = j$ , deve-se determinar primeiro, o número de maneiras em que i das experiências poderia ser selecionado para ser do tipo 1,  $\binom{n}{i}$  maneiras, deve-se multiplicar esse valor pelo número de maneiras em que j das experiências restantes poderia ser selecionado para ser do tipo 2,  $\binom{n-i}{j}$  maneiras, e deve-se multiplicar o resultado pela

probabilidade de um dado conjunto de i resultados do tipo 1, j do tipo 2, e  $n - i - j$  do tipo 3, que é  $p^i q^j r^{n-i-j}$ . Assim, tem-se que

$$p(i, j) = \binom{n}{i} \binom{n-i}{j} p^i q^j r^{n-i-j} = \frac{n!}{i! j! (n-i-j)!} p^i q^j r^{n-i-j} \text{ para } 0 \leq i + j \leq n. \text{ Essa distribuição é}$$

uma extensão da distribuição binomial. Se a fórmula  $p_i = \sum_j p(i, j)$ ,  $q_j = \sum_i p(i, j)$  for usada juntamente com o teorema binomial, verifica-se que individualmente, X e Y têm distribuições binomiais.

$p_i = \binom{n}{i} p^i (1-p)^{n-i}$ ,  $q_j = \binom{n}{j} q^j (1-q)^{n-j}$ . Note-se que, neste caso,  $p(i, j) \neq p_i q_j$ , e portanto, que X e Y não são independentes.

**DISTRIBUIÇÃO UNIFORME:** Distribuição de probabilidades que fornece a mesma probabilidade de ocorrência a cada um dos possíveis valores da variável. Em inglês Uniform distribution.

**DISTRIBUIÇÕES A POSTERIORI CONDICIONAIS:** Sejam dois parâmetros  $\theta$  e  $\phi$  com distribuição a posteriori  $p(\theta, \phi|x)$ , a distribuição a posteriori condicional de  $\theta$  é  $p(\theta|\phi, x) = p(\theta, \phi|x) / p(\phi|x)$ , sendo  $p(\phi|x)$  a distribuição marginal de  $\phi$ . Em inglês Conditional posterior distributions.

**DISTRIBUIÇÕES ESTÁVEIS:** Uma variável aleatória (ou função de probabilidade ou distribuição de probabilidade) é estável se, para variáveis aleatórias independentes e identicamente distribuídas  $X, X_1, X_2$ , satisfaz  $c_1 X_1 + c_2 X_2 \stackrel{d}{=} b(c_1, c_2) X + a(c_1, c_2)$  para todos números não negativos  $c_1, c_2$  e números reais apropriados  $b(c_1, c_2) \in \mathfrak{R}$ . A expressão  $c_1 X_1 + c_2 X_2 \stackrel{d}{=} b(c_1, c_2) X + a(c_1, c_2)$  representa de distribuições que são fechadas a menos de variações na locação e escala) sob convolução e multiplicação com números reais. Por exemplo, a convolução de duas distribuições Poisson é Poisson. Contudo, a distribuição de Poisson não satisfaz  $c_1 X_1 + c_2 X_2 \stackrel{d}{=} b(c_1, c_2) X + a(c_1, c_2)$ .

Isso mostra que  $c_1 X_1 + c_2 X_2 \stackrel{d}{=} b(c_1, c_2) X + a(c_1, c_2)$  é mais restritivo do que simplesmente ser uma classe fechada em relação à convolução. Considere  $S_n$  uma soma de variáveis aleatórias independentes e identicamente distribuídas (v.a.s. i.i.d.). Por  $c_1 X_1 + c_2 X_2 \stackrel{d}{=} b(c_1, c_2) X + a(c_1, c_2)$  temos constantes reais  $a_n$  e  $b_n > 0$  e  $X = X_1$ ,  $S_n = X_1 + \cdots + X_n \stackrel{d}{=} b_n X + a_n$ ,  $n = 1$ , que podemos reescrever como  $b_n^{-1} (S_n - a_n) \stackrel{d}{=} X$ . Isto prova que, se uma distribuição é estável, então ela é distribuição limite para soma

de variáveis aleatórias independentes e identicamente distribuídas (v.a.s.i.i.d.). O teorema a seguir mostra que não existem outras possibilidades para a distribuição limite. Teorema: (propriedade limite das distribuições estáveis). A classe das distribuições estáveis (não degeneradas) coincide com a classe de todas as possíveis distribuições limite (não degeneradas) para as somas de variáveis aleatórias independentes e identicamente distribuídas (v.a.s.i.i.d.) (apropriadamente centradas e normalizadas).

**DISTRITALIZAÇÃO:** Processo político-organizacional de reorientação do sistema de saúde, em nível local, capaz de facilitar a implantação e o desenvolvimento de modelos assistenciais alternativos, como base para a orientação do sistema único de saúde (SUS).

**DISTRITOS SANITÁRIOS:** Representam uma resposta, para o município, que visa organizar serviços e estabelecimentos numa verdadeira rede, com mecanismos de comunicação e integração, destacando-se os procedimentos de referência e contra-referência e a instauração de modelos assistenciais alternativos de base epidemiológica. O distrito sanitário tem sido reconhecido como unidade operacional e administrativa mínima do sistema de saúde, definida com critérios geográficos, populacionais, epidemiológicos, administrativos e políticos, onde se localizam recursos de saúde, públicos e privados, organizados por meio de um conjunto de mecanismos políticos institucionais com a participação da sociedade organizada para desenvolver ações integrais de saúde capazes de resolver a maior quantidade possível de problemas de saúde.

**DIVERSIDADE DE ESPÉCIES:** Medida da variabilidade de espécies presente em uma comunidade, que considera tanto a riqueza de espécies quanto a distribuição de abundância dessas espécies. Foram desenvolvidos diversos índices para quantificar a diversidade de espécies, como o de Shannon, Simpson, Berger-Parker dentre outros, que dão diferentes pesos a esses atributos. Cada índice de diversidade possui sua unidade particular dependendo de como o índice é formulado. Cada índice de diversidade possui uma notação própria.

**DISTRIBUIÇÃO DE PROBABILIDADE DISCRETA:** Uma distribuição de probabilidade discreta refere-se ao conjunto de todos os possíveis valores de uma variável aleatória discreta e suas probabilidades associadas. O conjunto dos 6 resultados no lançamento de um dado e suas probabilidades associadas é um exemplo de uma distribuição de probabilidade discreta. A soma das probabilidades associadas com todos os valores que uma variável discreta pode assumir é sempre igual a 1.

**DISTRIBUIÇÃO DE PROBABILIDADE:** Uma distribuição de probabilidade refere-se às probabilidades clássicas ou a priori associadas com todos os valores que uma variável aleatória pode assumir. Como essas probabilidades são aceitas a priori e sem qualquer experimentação, a distribuição de probabilidade é normalmente apresentada como uma distribuição de frequência teórica relativa. Esta difere de uma distribuição de frequência empírica relativa, a qual refere-se à razão entre o número de vezes que cada resultado realmente ocorre e o número total de experimentos ou observações reais. Por exemplo, lançando realmente um dado um certo número de vezes, não é provável que obtenhamos cada resultado exatamente  $\frac{1}{6}$  das vezes. Todavia, quando o número de lançamentos aumenta, a distribuição de frequência empírica relativa se estabiliza em uma distribuição de frequência relativa teórica ou de probabilidade (uniforme) de  $\frac{1}{6}$ .

**DMS:** Termo usado nos testes estatísticos para diferença entre médias, ou seja, é a diferença mínima significativa a uma taxa experimental de erro alfa, geralmente 5% de probabilidade.

**DOMÍNIO DE UMA FUNÇÃO:** Numa função  $f$  de  $A$  em  $B$ , chamamos de domínio de  $f$  e indicamos por  $D(f)$  ou  $\text{Dom}(f)$  o conjunto dos elementos de  $A$  que estão representados em  $B$ , isto é, o conjunto dos elementos  $a \in A$  para os quais existe  $b \in B$  tal que  $f(a) = b$ . Note que o domínio de uma função  $g$  de  $A$  em  $B$  coincide com o conjunto de partida  $A$  se, e somente se cada elemento do conjunto de partida  $A$  admite um representante em  $B$ , isto é, se, e só se  $g$  é função total. Seja a função  $f : A \rightarrow B; Y = f(X)$ . Nestas condições, temos  $X \in A$  e  $Y \in B$ . Os valores de  $X$  constituem o domínio da função  $f$  e os valores de  $Y$  constituem o conjunto imagem da função  $f$ . O conjunto  $B$  é chamado de contradomínio. O domínio de uma função também é chamado de campo de definição ou campo de existência da função, e é representado pela letra  $D$ .

**DOMÍNIO DO VARIOGRAMA:** O variograma é para ser válido dentro do campo geométrico definido. O campo geométrico, considerado geologicamente homogêneo, deveria ser até certo ponto intrínseco ou independente da posição nas características representando a variabilidade da variável regionalizada. Quando esta hipótese é verificada e reconhecida uma lei de dispersão única no campo mineralizado, esta lei é denominada lei intrínseca, segundo estes autores. Entretanto, esta característica intrínseca não se mantém quando se move o campo para a zona de borda, pois nesta zona a mineralização não existe e, por isso, o valor da variável regionalizada pode ser zero e o conceito de valor médio nesse campo de um dado tamanho tornaria-se insignificante.

**DRAGRAGEM DE DADOS:** Análise de conjuntos de dados por meio de técnicas modernas de computação, permitindo a avaliação de centenas de associações possíveis entre as variáveis em estudo. A menos que o valor de alfa seja ajustado, a testagem de múltiplas hipóteses aumenta o risco de erros falso-positivos.

**DU(i, j):** Distribuição discreta uniforme.

**DUPLO CEGO:** Processo usado em um experimento, segundo o qual a pessoa não sabe se está recebendo um tratamento ou um placebo; a pessoa que administra também não sabe.

**DUPLO-CEGO:** Avaliação na qual o observado desconhecem certos detalhes do que é avaliado por exemplo, a que grupo o avaliado pertence, se exposto/não-exposto ou doente/não-doente. Ver avaliação cega.

**DIAGRAMA DE ZEUNER:** Sistema convencional de representação gráfica, usado na teoria formal da população, composto de três eixos formando um triedro trirretângulo, dos quais,  $OX$  é o eixo das Idades,  $OY$  é o eixo das datas de nascimento e  $OZ$  é o eixo das quantidades de pessoas vivas.

$\delta_{\bar{X}}$  : Desvio absoluto médio dos valores de um conjunto em relação a média  $\bar{X}$ .

$\delta_{X_0}$  : Desvio absoluto médio dos valores de um conjunto em relação a  $X_0$ .

$\vec{D}$  : Convergência em distribuição.

$d_k$  : Valor do decil amostral de ordem  $K$ .





# E

**E:** Símbolo matemático dado por  $\wedge$ .

**EB:** Estimador bayesiano.

**EBLUP:** Melhor preditor linear não viciado empírico.

**ECONOMETRIA:** O termo Econometria é constituído por duas palavras gregas: Economia e o verbo Metro que significa medir, ou seja, medir as relações econômicas. A econometria usa: elementos da teoria econômica e métodos matemáticos, além de métodos estatísticos e dados estatísticos além de meios informáticos para analisar, quantificar e confirmar empiricamente as relações econômicas.

**ECONOMETRIA:** Ramo da economia que utiliza a análise estatística para estudar e fazer previsões em economia.

**ECONOMETRIA:** Econometria é a principal ferramenta quantitativa das Ciências Econômicas e Sociais, utilizando-se de métodos estatísticos e funções matemáticas aplicadas à economia, consegue traduzir em números os objetos de estudo econômicos. Exemplo: A Renda (R) de uma família influencia no consumo (C) da mesma? Para responder a esta questão de nível teórico, primeiramente, deve-se utilizar do modelo matemático para esta função como hipótese nula.

**EDIÇÃO DE DADOS:** Processo de revisar dados com a finalidade de detectar deficiências ou erros no modo como eles foram registrados ou colecionados.

**EDITORAÇÃO:** O processo de checagem à procura do entrevistador.

**EFEITO:** Do tratamento  $y_i$  sobre a variável  $\xi$ , é o valor da diferença  $\hat{A} = \alpha_1 - \alpha$  entre a massa aleatória m.a.  $\alpha$  da população dos valores  $\xi$  e a massa aleatória m.a.  $\alpha_1$  da subpopulação dos valores  $\xi$  que receberam aquele tratamento.

**EFEITO ALEATÓRIO:** Efeito de algum fator não sistemático com probabilidade de ocorrência determinada por alguma distribuição de probabilidades. Em inglês Random effect.

**EFEITO ALEATÓRIO:** São aqueles representativos de tratamentos ou fatores de uma amostra oriunda de uma determinada população. Por exemplo, se de um rebanho for retirada uma amostra de animais, estes animais entraram no ensaio representando todo o rebanho.

**EFEITOS FIXOS:** Também chamados de parâmetros, são aqueles cujas considerações são limitadas, isto é, são restritas aos tratamentos ou níveis ensaiados. Por exemplo níveis de nutrientes, espaçamento para determinada cultura, etc.

**EFEITO AQUIESCÊNCIA:** Descreve o fenômeno de certos respondentes mais frequentemente endossarem uma afirmativa positiva que discordarem com o seu oposto. Assim as respostas se tornam mais ou menos predizíveis.

**EFEITO ATOR:** Quando um indivíduo é convidado a ser respondente num estudo de pesquisa, ele tende a assumir um papel do tipo de pessoa que ele pensa que deve ser na situação. Considerando que esse papel difere do seu comportamento natural em situações similares na vida real, diminui-se a validade de pesquisa.

**EFEITO COBAIA:** Quando um sujeito sente que precisa colaborar com a pesquisa, fazendo o máximo de que é capaz, ou quando o método de coleta de dados o faz agir fora do seu natural, tornando-o defensivo, antagonista ou muito cooperativo.

**EFEITO COORTE:** Devido a uma geração; as gerações têm, em geral, risco diferenciado de adquirir uma doença ou um outro resultado, diferença esta que é denominada efeito coorte.

**EFEITO DE ESTILO RESPOSTA:** Série de respostas sistemáticas de um respondente que refletem um viés ou padrão consistente. Exemplos incluem as respostas de que um objeto sempre em desempenho excelente ou ruim, ao longo de todos os atributos, com pouca ou nenhuma variação.

**EFEITO DE EXPECTÂNCIA:** O sujeito experimental, por meio de alguns sinais sutis e não intencionais do pesquisador, desempenha de maneira que ele acha que o pesquisador espera que ele desempenhe.

**EFEITO DE HALO:** Tendência que tem uma característica irrelevante de uma unidade de estudo a influenciar a característica relevante numa direção favorável ou desfavorável.

**EFEITO DE INTERAÇÃO:** Em delineamentos fatoriais, os efeitos conjuntos de duas variáveis de tratamento em adição com os efeitos principais individuais. Isso significa que a diferença entre grupos sobre uma variável de tratamento varia de acordo com o nível da segunda variável de tratamento. Por exemplo, considere que os respondentes foram classificados por renda (três níveis) e sexo (homens versus mulheres). Uma interação significativa seria encontrada quando as diferenças entre homens e mulheres sobre a(s) variável (eis) independente(s) variassem substancialmente ao longo dos três níveis de renda.

**EFEITO DE REGRESSÃO:** Tendência dos escores estimados de situar-se mais perto da média do que os escores estimadores, quando se tratar de correlação imperfeita. Por exemplo, pode-se prever o peso a partir da altura, por que ambas as variáveis se correlacionam, se bem que de modo imperfeito. O indivíduo mais alto do grupo estará entre os mais pesados. Da mesma forma, o filho dos pais inteligentes será inteligente, mas menos afastado da média do que os pais. O efeito de regressão é o que se chama regressão rumo à média. Quando a correlação é perfeita não há regressão. Em inglês regression effect.

**EFEITO DE REGRESSÃO:** A seleção de sujeitos, por serem eles particularmente altos ou baixos em alguma variável, para serem medidos em uma forma equivalente de variável, poderá levá-los a regressarem à média, devido à correlação imperfeita entre as duas medidas da variável. Grupos extremos, num segundo teste, comportam como se os sujeitos mais brilhantes ficassem menos brilhantes e os mais obscuros ficassem menos obscuros. Quanto mais baixa a correlação, tanto maior a regressão à média. Grupos selecionados com base em escores extremos em uma variável, mas avaliados numa variável dependente não relacionada são bastante livres do efeito de regressão na segunda variável.

**EFEITO DE REGRESSÃO ESTATÍSTICA:** O mesmo que Regressão em direção à média.

**EFEITO DE SER OBSERVADO (OU EFEITO HAWTHORNE):** O próprio ato de observar um indivíduo influencia a conduta da pessoa que se sabe observada; ela muda de atitude e distorce os resultados.

**EFEITO DE VIZINHANÇA:** Em estatística espacial, especificamente quando são utilizados os efeitos CAR, o efeito de vizinhança assume que, para um particular local só os vizinhos o influenciam. Isto é uma extensão da propriedade markoviana para o caso espacial. Em inglês Neighboring effect.

**EFEITO DO TRATAMENTO:** Em ensaios clínicos, a diferença entre os resultados observados no grupo experimental e no grupo controle submetido a placebo.

**EFEITO DO TESTE:** Efeito que representa um subproduto do processo de pesquisa e não a variável experimental.

**EFEITO EXPERIMENTAL:** O efeito do tratamento variável sobre a variável dependente.

**EFEITO FURO:** Modelo de variograma teórico, descrito pela seguinte equação:

$$\gamma(h) = C_0 + C \left[ \frac{1 - \text{Sen}(r)}{r} \right], \text{ com } r \text{ em rad. Este modelo desenha um variograma cíclico, dado pela alter-}$$

nância de camadas amostradas ao longo de um furo de sondagem.

**EFEITO HALO (OU AURÉOLA):** A tendência da parte do observador de ser influenciada, indevidamente, por uma característica, favorável ou desfavorável, da pessoa observada, de modo a alterar seu julgamento que é o do observador sobre as demais características do observado.

**EFEITO HAWTHORNE:** Num estudo de eficiência industrial, executado na Fábrica de Hawthorne em Chicago nos Estados Unidos da América, observou-se que ao se escolher um grupo de operários para participar de um projeto de pesquisa, eles agiram diferentemente dos demais, o que resultou em consistente aumento de produtividade, a despeito das mudanças nas condições de trabalho. Sabe-se que o conhecimento que um sujeito tem de que ele ou ela está participando de um experimento pode alterar as respostas a intervenções. Se as respostas dos sujeitos são muito alteradas, os efeitos de intervenção não se generalizarão a sujeitos que não possuem tal conhecimento.

**EFEITO MODELADOR:** Efeito no qual uma terceira variável independente (a variável moderadora) faz com que a relação entre um par de variáveis dependente/ independente mude, dependendo valor da variável moderadora. Também é conhecido como um efeito interativo e semelhante ao efeito de interação visto em métodos de análise de variância.

**EFEITO MULTIPLICATIVO:** Alguns fatores são representados por meio de efeitos que se multiplicam nas equações que formam um modelo, estes efeitos são chamados multiplicativos e expressam a influência do fator como uma proporção de aumento, ou diminuição do valor médio. Em inglês Multiplicative effect.

**EFEITO PENEIRA:** É o nome que se dá a ocorrência de pesos negativos resultantes da krigagem ordinária, tendo em vista a particular localização das amostras em relação ao ponto a ser estimado. Como o sistema de equações de krigagem reconhece a disposição espacial das amostras, tem-se que as amostras mais próximas passantes na malha da peneira receberão pesos positivos e numericamente superiores em detrimento daquelas distantes retidas na malha da peneira que geralmente resultam em pesos negativos. Tais valores, a princípio, não deveriam causar distorções na estimativa de teores e variâncias, desde que os valores da variável de interesse não apresentassem extremos, comuns em distribuições lognormais. Nestes casos, os teores podem resultar em valores negativos ou então em valores superiores ou inferiores aos valores originais. Pontos ou blocos com teores negativos são inadmissíveis, pois em depósitos minerais

caso estejam sobre minério o teor deverá ser positivo, acima de um teor de corte predefinido, e se estiverem sobre o estéril, ou rocha encaixante, o teor deverá ser positivo, abaixo do teor de corte, podendo inclusive ser igual a zero, mas nunca negativo.

**EFEITO PEPITA:** É o valor de variância identificada na origem do variograma, geralmente atribuído a problemas analíticos, pois duas amostras tomadas no mesmo ponto deveriam, em princípio, ter os mesmos resultados. Contudo, ultimamente este efeito tem sido atribuído a variabilidade natural do depósito, sendo o efeito pepita proporcional a esta variabilidade, podendo migrar para o efeito pepita puro com o aumento do coeficiente de variação.

**EFEITO PLACEBO:** Um placebo é um estímulo inerte ou neutro dado a sujeitos como se fosse a própria variável de tratamento. Em pesquisa médica, o placebo é às vezes uma injeção de água esterilizada ou uma pílula quimicamente inócua, aplicada ao sujeito de forma que ele não possa distinguir entre o ingrediente ativo e o neutro. Se as condições não forem as mesmas o placebo pode produzir efeitos.

**EFEITO PLACEBO:** Efeito produzido por placebo.

**EFEITO PLACEBO:** Efeito que ocorre quando um indivíduo não tratado acredita, incorretamente, que está recebendo um tratamento real e reporta uma melhora de sintomas.

**EFEITO PREÂMBULO:** As instruções ou prefácio motivando respostas a um teste ou questionário induzem certas atitudes ao respondente, que podem persistir até depois da situação de pesquisa.

**EFEITO PRINCIPAL:** Em delineamentos fatoriais, o efeito individual de cada variável de tratamento sobre a variável dependente.

**EFEITO PROXY:** Ver proxy.

**EFEITO RESIDUAL:** De um tratamento sobre certa unidade experimental, quando esse tratamento é aplicado repetidas vezes sobre essa mesma unidade, é a parte do efeito que, permanecendo, de uma para outra aplicação, vai influir sobre o resultado da aplicação subsequente.

**EFEITOS DE INTERAÇÃO:** Efeitos de uma combinação de características relacionadas, também conhecidas como termos de interação. Ao avaliar o valor, uma pessoa pode designar um único valor a combinações específicas de características que vai contra a regra de composição aditiva. Por exemplo, considere que uma pessoa está avaliando produtos bucais descritos pelos fatores ou atributos de cor e marca. Considere ainda que essa pessoa tem uma preferência média por vermelho e pela marca X. Quando essa combinação específica de níveis (vermelho e marca X) é avaliada com a mesma regra de composição aditiva, como todas as outras combinações, o produto vermelho de marca X tem uma avaliação de preferência geral esperada em algum lugar no meio de todos os possíveis estímulos. Se, porém, a pessoa realmente prefere o produto bucal vermelho de marca X mais do que qualquer outro estímulo, mesmo algum que esteja acima de outras combinações de atributos (cor e marca) que tenham melhores avaliações das características individuais, então uma interação é percebida. Essa avaliação única de uma combinação que é maior (ou poderia ser menor) do que o esperado com base nos julgamentos separados indica uma interação em dois sentidos. Interações de ordem mais alta (três sentidos ou mais) podem ocorrer entre mais combinações de níveis.

**EFEITOS DE REGRESSÃO:** É o problema que ocorre na extensão de teores para uma grande área de influência, sem levar em consideração a correlação espacial com teores de amostras vizinhas. Os efeitos

de regressão traduzem-se na subestimativa sistemática de reservas quando os teores são baixos e na superestimativa sistemática quando altos.

**EFEITOS PRINCIPAIS:** Efeito direto de cada fator que é a variável independente sobre a variável dependente. Pode ser complementado por efeitos de interação em situações específicas.

**EFETIVIDADE:** Resultados ou conseqüências de determinado procedimento ou tecnologia médica quando aplicados na prática. A efetividade de um programa de saúde distingue-se da sua eficácia pelo fato de fazer referência a situações reais, enquanto o segundo termo se aplica apenas aos resultados obtidos em condições ideais.

**EFETIVIDADE (EFFECTIVENESS):** A extensão de quanto uma determinada intervenção, procedimento, programa, regime ou serviço produz um resultado benéfico, quando empregado no mundo real, em uma população definida; é o resultado verdadeiramente observado nas condições habituais de uso, ou seja, a sua utilidade em condições usuais.

**EFICÁCIA:** Resultados ou conseqüências de um tratamento, medicamento, tecnologia ou programa e saúde desde um ponto de vista estritamente técnico ou numa situação de utilização ideal, por exemplo, quando todos os indivíduos aderem à terapêutica.

**EQUAÇÕES ESTRUTURAIS (OU COMPORTAMENTAIS):** São aquelas que descrevem a estrutura de uma economia ou o comportamento de alguns agentes econômicos, como consumidores ou produtores. Existe uma equação estrutural para cada variável endógena do sistema. Os coeficientes das equações estruturais são chamados parâmetros estruturais e expressam o efeito direto de cada variável explicativa sobre a variável dependente.

**EQUAÇÕES SIMULTÂNEAS VIESADAS:** São aquelas que se referem à superestimação dos parâmetros estruturais obtidos pela aplicação do método de mínimos quadrados simples (MMQ) às equações estruturais do modelo de equações simultâneas. Este viés é resultante do fato que aquelas variáveis endógenas do sistema, as quais são também variáveis explicativas, são correlacionadas com os termos erros, violando assim a quinta hipótese do método de mínimos quadrados simples (MMQ).

**EQUAÇÕES NAS FORMAS REDUZIDAS:** São aquelas obtidas pela resolução do sistema de equações estruturais e expressam cada variável endógena do sistema apenas como função de variáveis exógenas ou pré-determinadas. Como as variáveis exógenas do sistema são não correlacionadas com o termo erro, o método de mínimos quadrados simples (MMQ) fornece estimativas consistentes dos parâmetros na forma reduzida. Isto permite medir os efeitos diretos e indiretos totais de uma mudança nas variáveis exógenas sobre as variáveis endógenas e pode ser usado para obter parâmetros estruturais consistentes.

**ERROS NAS VARIÁVEIS:** Os erros nas variáveis referem-se ao caso em que as variáveis do modelo de regressão incluem erros de medida. Estes provavelmente são muito comuns, em vista da maneira que a maioria dos dados são coletados e elaborados. Erros de medida na variável dependente são incorporados no termo residual, deixando as estimativas por parâmetros de mínimos quadrados não viesadas e consistentes embora ineficientes ou maior do que a variância mínima. Contudo, com erros de medida nas variáveis explicativas, a quarta hipótese do MMQ de independência das variáveis explicativas com o termo erro é violada, fazendo com que as estimativas dos parâmetros sejam viesadas e inconsistentes. Em uma regressão simples, é viesado para baixo, enquanto é viesado para cima. Não há nenhum teste formal para detectar a presença de erros nas variáveis. Somente a teoria econômica e o conhecimento de como os

dados foram coletados pode algumas vezes dar alguma indicação da importância do problema. Um método de obter estimativas consistentes dos parâmetros pelo método dos mínimos quadrados (MMQ), porém ainda viesadas e ineficientes é substituir a variável explicativa sujeita a erros de medida por outra variável que é altamente correlacionada com a variável explicativa em questão, porém que seja independente do termo erro. No mundo real, poderia ser difícil encontrar desta maneira uma variável instrumental, e nunca poderíamos ter certeza que ela seria independente do termo erro. A variável instrumental mais popular é o valor defasado da variável explicativa em questão. Erros de medida na variável explicativa somente podem ser corrigidos pela inversão dos mínimos quadrados. Isto implica em fazer regressão de X contra Y. Então,  $\hat{b}_0 = -\frac{\hat{b}_0}{\hat{b}_1}$  e  $\hat{b}_1 = \frac{1}{\hat{b}_1}$ , onde e são estimativas consistentes dos parâmetros intercepto e declividade da regressão.

**ERROS NAS VARIÁVEIS:** Erros nas variáveis referem-se ao caso em que as variáveis no modelo de regressão incluem erros de medida. Erros de medida na variável dependente são incorporados no termo residual e não criam nenhum problema especial. Entretanto, erros nas variáveis explicativas conduzem a estimativas viesadas e inconsistentes dos parâmetros. Um método de se obter estimativas através do método dos mínimos quadrados (MMQ) consistentes dos parâmetros é substituir a variável explicativa, sujeita a erros de medida, por outra variável chamada variável instrumental, que é altamente correlacionada com a variável explicativa original, mas é independente do termo erro. Isto é relativamente difícil de fazer e, até certo ponto, arbitrário. A variável instrumental mais comum é a geralmente a variável explicativa defasada na questão. Outro método usado quando somente X está sujeito a erros de medida, envolve a regressão de X contra Y.

**ESTIMAÇÃO DE MÍNIMOS QUADRADOS INDIRETOS:** Mínimos quadrados indiretos (MMQI) é um método de cálculo dos valores dos parâmetros estruturais para equações exatamente identificadas. O MMQI envolve o uso de mínimos quadrados simples (MMQ) para estimar as equações na forma reduzida do sistema, após o que utilizam-se os coeficientes estimados para calcular os parâmetros estruturais. Contudo, não é fácil calcular os erros padrão dos parâmetros estruturais, nem pode-se usar MMQI nos casos de superidentificação.

**ESTIMAÇÃO DE MÍNIMOS QUADRADOS DE DOIS ESTÁGIOS:** Mínimos quadrados de dois estágios (MMQ2E) é um método de estimação de parâmetros estruturais consistentes, de equações superidentificadas para equações exatamente identificadas, o MMQ2E mostra os mesmos resultados que o MMQI, porém ele também mostra os erros padrão dos parâmetros estruturais estimados. O MQ2E envolve a regressão de cada variável endógena sobre todas variáveis exógenas do sistema; após o que, tomam-se os valores previstos das variáveis endógenas para estimar as equações estruturais do modelo.

**EVENTOS MUTUAMENTE EXCLUSIVOS:** Dois ou mais eventos são mutuamente exclusivos, ou separados, se a ocorrência de um deles exclui ou impede a ocorrência do(s) outro(s). Quando um evento ocorre, o(s) outro(s) não ocorre(m). Por exemplo, em um único lançamento de uma moeda, pode-se obter cara ou coroa, mas não ambos. Caras e coroas são, portanto eventos mutuamente exclusivos. Em um único lançamento de um dado, pode-se obter um e somente um de seus possíveis resultados: 1, 2, 3, 4, 5 ou 6. Os resultados são, portanto mutuamente exclusivos. Uma carta de um baralho extraída ao acaso deve ser de único naipe: ouro, copas, paus, ou espada. Uma criança que nasce, ou é do sexo masculino ou do sexo feminino. Um artigo produzido em uma linha de produção ou é bom ou é defeituoso.

**EVENTOS NÃO MUTUAMENTE EXCLUSIVOS:** Dois ou mais eventos são não mutuamente exclusivos se eles podem ocorrer ao mesmo tempo. A ocorrência de um dos eventos não exclui a ocorrência do(s) outro(s). Por exemplo: uma carta extraída ao acaso de um baralho pode ser tanto um ás como ser de paus. Portanto, ases e paus não são eventos mutuamente exclusivos, porque pode se extrair um ás de paus. Uma vez que se pode ter inflação e recessão ao mesmo tempo, inflação e recessão não são eventos mutuamente exclusivos.

**EVENTOS INDEPENDENTES:** Dois ou mais eventos são independentes se a ocorrência de um deles não afeta de maneira alguma a ocorrência do(s) outro(s). Por exemplo, em dois sucessivos lançamentos de uma moeda equilibrada, o resultado do segundo lançamento não depende do resultado do primeiro lançamento. O mesmo é verdade para dois lançamentos sucessivos de um par de dados, ou extração de duas cartas de um baralho com reposição.

**EVENTOS DEPENDENTES:** Dois ou mais eventos são dependentes se a ocorrência de um deles afeta a probabilidade de ocorrência do(s) outro(s). Por exemplo, se extrairmos uma carta de um baralho e não a recolocarmos, a probabilidade de extrair a mesma carta numa segunda extração é 0. Todas as outras probabilidades também são afetadas, desde que há agora somente 51 cartas no baralho. Semelhantemente, se a proporção de artigos defeituosos é maior para o turno noturno do que para o turno diurno, a probabilidade de um artigo extraído ao acaso da produção noturna ser defeituosa é maior do que a da produção diurna.

**EFICÁCIA (EFFICACY):** A extensão de quanto uma determinada intervenção, procedimento, programa, regime ou serviço produz um resultado benéfico, em condições ideais de observação; trata-se do resultado no laboratório, ou seja, da sua utilidade em condições ideais.

**EFICIÊNCIA:** O conceito de eficiência em estimação estatística é atribuído ao matemático inglês Sir Ronald Aylmer Fisher (1921) como uma tentativa de medir o mérito relativo de vários estimadores possíveis. Um estimador é mais eficiente que outro se tem menor variância e, se existe um estimador de variância mínima,  $V$ , a eficiência de outro estimador com variância  $V_1$  é definida como a razão  $\frac{V}{V_1}$ . Em inglês Efficiency.

**EFICIÊNCIA:** Relação entre os recursos utilizados e os resultados obtidos em determinada atividade. A produção eficiente é aquela que maximiza os resultados obtidos com um dado nível de recursos ou minimiza os recursos necessários para obter determinado resultado.

**EFICIÊNCIA:** Medida da sensibilidade de um teste não-paramétrico, em comparação com um teste paramétrico correspondente.

**EFICIÊNCIA (EFFICIENCY):** Refere-se aos efeitos alcançados em relação ao esforço despendido em termos de recursos e tempo; é o resultado tendo em conta os insumos utilizados; trata-se do rendimento dos recursos; os estudos sobre eficiência, por exemplo, de custo benefício ou custo-efetividade são empregados para fazer comparações entre duas ou mais intervenções ou procedimentos, programas, regimes, serviços dentre outros, entre si, de modo a escolher a que é mais apropriada, ou seja, que tenha melhor relação custo benefício ou custo efetividade, para ser recomendada ou implementada

**EFICIÊNCIA DE AJUSTE:** Teste estatístico realizado para determinar se a distribuição de probabilidade hipotética para uma população deve ser rejeitada.



**EFICIÊNCIA DE DELINEAMENTO:** Grau em que um delineamento condiz com um delineamento ortogonal. Essa medida é usada principalmente para avaliar e comparar delineamentos quase ortogonais. Os valores da eficiência de delineamento variam de 0 a 100, o qual denota um delineamento ótimo.

**EFICIÊNCIA ASSINTÓTICA:** De um estimador; o mesmo que eficiência. A expressão é empregada pelos autores que consideram um tipo de eficiência existente no finito, à qual reservam o nome de eficiência propriamente dita.

**EIGENVALOR:** Eigenvalor é o conjunto de valores ao longo de uma fileira ou coluna de uma matriz quadrada simétrica. Há tantos eigenvalores quantas são as fileiras ou colunas em uma matriz. Uma descrição realística de um eigenvalor demanda um bom conhecimento de álgebra linear. Entretanto, eles podem ser conceitualmente considerados como uma medida de força que diz respeito ao comprimento relativo de um eixo derivado da matriz quadrada simétrica. Eigenvalores são também conhecidos como variáveis latentes.

**EIGENVECTOR (OU FACTOR LOADING):** Cada eigenvalor tem um eigenvetor associado. Se um eigenvalor é o comprimento de um eixo, o eigenvetor é o que determina a orientação espacial deste eixo. Os valores em um eigenvetor não são únicos porque qualquer coordenada que descreva a mesma orientação seria aceitável. Normalmente elas são padronizadas de algum modo, isto é seus valores quadrados somam 1. Os eigenvetores são normalmente usados para auxiliar na interpretação de uma análise multivariada usando a técnica de análise dos componentes principais.

**EIXO DE ANISOTROPIA:** É dado pela direção de menor variabilidade no corpo ou direção do eixo de anisotropia.

**ELEMENTO:** Em amostragem, os elementos são as unidades que compõem a população, como indivíduos ou eleitores.

**EFEITO DO TRATAMENTO:** A mudança da média para uma resposta devido à presença do tratamento.

**EFEITO PRINCIPAL:** Uma estimativa do efeito de um fator ou variável que expressa, independentemente, a mudança na resposta devido a uma mudança naquele fator, sem considera os outros fatores que podem estar presentes no sistema.

**EFICIÊNCIA:** Um conceito em estimação de parâmetros que usa as variâncias de estimadores diferentes; essencialmente, um estimador é mais eficiente que outro se ele tiver menor variância. Quando estimadores são tendenciosos, o conceito requer modificação.

**EFICIÊNCIA RELATIVA ASSINTÓTICA (ERA):** Usada para comparar testes de hipóteses. A ERA de um teste relativo a outro é a razão limitante dos tamanhos de amostra necessários para obter probabilidades idênticas de erros nos dois procedimentos.

**ELIMINAÇÃO REGRESSIVA:** Um método de seleção de variáveis em regressão que começa com todas as variáveis regressoras candidatas no modelo, eliminando um regressor insignificante de cada vez, até que só restem regressores significativos.

**EQUAÇÃO:** Formulação na qual duas expressões matemáticas são colocadas em situação de equivalência. Por exemplo:  $[y = 5x + 3t - 12w^2]$ .

**EQUAÇÃO ESTRUTURAL:** Tipo de equação ou modelo que representa a força e a natureza das relações hipotéticas entre conjuntos de variáveis numa pesquisa científica.

**EQUAÇÕES NORMAIS:** O conjunto de equações lineares simultâneas geradas na estimação de parâmetros pelo método dos mínimos quadrados.

**EDA EFEITO DO ARCO (HORSESHOE EFFECT):** Uma propriedade indesejável de muitos métodos de ordenação que acabam exagerando a média e comprimindo os extremos de um gradiente ambiental. Ele leva a um gráfico de ordenação que é curvo, em forma arco, ferradura, ou círculos.

**EFEITO PRINCIPAL (MAIN EFFECT):** O efeito individual de uma única variável. Na ausência de qualquer interação os efeitos principais são aditivos: a resposta de um experimento pode ser acuradamente predita conhecendo-se os efeitos principais de cada fator individual. Comparar com interação.

**EFEITOS ALEATÓRIOS (RANDOM EFFECTS):** Em um delineamento de análise de variância, o conjunto de níveis de tratamento que representam uma variável, um subconjunto aleatório de todos os níveis de tratamento possíveis. Também conhecido como fator aleatório. Comparar com efeitos fixos.

**EFEITOS FIXOS (FIXED EFFECTS):** Em um delineamento de análise de variância, o conjunto de níveis dos tratamentos que representa todos os níveis de tratamentos possíveis. Também conhecido como fator fixo. Comparar com efeitos aleatórios.

**EIXO MAIOR (MAJOR AXIS):** Em uma análise de componentes principais, é o primeiro eixo principal. Além disso, é o eixo mais longo de uma elipse.

**EIXO PRINCIPAL (PRINCIPAL AXIS):** Uma nova variável criada por uma ordenação em que as variáveis originais são ordenadas.

**EIXO-X (X-AXIS):** A linha horizontal em um gráfico, também chamado de abscissa. O eixo-x em geral representa a variável causal, independente ou preditora em um modelo estatístico.

**EIXO-Y (Y-AXIS):** A linha vertical em um gráfico, também chamado de ordenada. O eixo-y geralmente representa o resultado ou variável dependente ou resposta em um modelo estatístico.

**ELEMENTO (ELEMENT):** Um valor em uma matriz. Na maioria das vezes é indicado com  $a_{ij}$ , onde  $i$  e  $j$  são os números da linha e da coluna, respectivamente, da matriz.

**EFEITO PLACEBO:** Termo que é usado para indicar um aumento na condição de saúde de um indivíduo que é submetido a medicação ou tratamento inócua, ou testemunha que é conhecido também como placebo. Muito comum em pesquisas médicas, este efeito pode ser observado nos indivíduos do grupo controle, que imaginam estar sendo tratado de forma efetiva, embora não o estejam.

**EMPÍRICO:** Que advém da experiência e da observação da realidade ou da prática. Contrasta-se com teórico.

**EXPERIMENTO:** É um processo de investigação científica de alguns procedimentos, feito para responder a determinadas perguntas.

**ENSAIO:** Termo usado para designar artigo teórico, experimento clínico ou agrônomo ou biológico.

**ENSAIO CLÍNICO CONTROLADO:** Tipo de estudo ou experimento típico da área médica, no qual existem dois grupos de sujeitos: um grupo experimental, que receberá o tratamento ou medicamento que está sendo avaliado pela pesquisa, e um grupo controle, que receberá tratamento convencional ou uma substância placebo.

**ENSAIO CLÍNICO CRUZADO:** Tipo de experimento ou estudo experimental da área médica, no qual dois grupos de sujeitos são submetidos aos mesmos procedimentos em tempos diferentes. Por exemplo: numa primeira etapa, um grupo recebe uma droga experimental, enquanto o outro grupo recebe um placebo. Após um período de repouso, inverte-se o procedimento, o grupo que havia inicialmente recebido a droga agora passa a receber um placebo, enquanto o grupo placebo passa a receber a droga experimental.

**ENTREVISTA ESTRUTURADA:** Tipo de entrevista cuja estrutura prevê um roteiro de perguntas preestabelecidas e que serão aplicadas a todos os indivíduos respondentes da entrevista.

**ENTREVISTA EXECUTIVA:** Tipo de entrevista utilizada nas pesquisas de mercado, realizada no ambiente de trabalho e referindo-se geralmente a produtos e serviços.

**ENTREVISTA NÃO ESTRUTURADA:** Tipo de entrevista no qual não há um roteiro preestabelecido de perguntas a serem feitas aos indivíduos.

**ENTREVISTA PAINEL:** Técnica muito usada em pesquisa de mercado, onde se aplica um tipo de entrevista não estruturada realizada simultaneamente com vários sujeitos, que são levados a opinar sobre determinado tema, num contexto informal.

**ENTREVISTA SEMI-ESTRUTURADA:** Tipo de entrevista usada em pesquisa de mercado que possui componentes estruturados e não estruturados, ou seja, há um roteiro de perguntas preestabelecidas a serem feitas ao respondente, mas há também um espaço para a discussão livre e informal de determinado tema do interesse do pesquisador.

**ENUNCIADO DE HIPÓTESES:** Etapa do método de pesquisa que ocorre em seguida à formulação do problema de pesquisa.

**EPIDEMIOLOGIA:** Disciplina científica que tem por objetivo investigar as causas das doenças nas populações. Conjunto de métodos e técnicas para descrever e analisar a ocorrência, distribuição, extensão e progressão da saúde e das doenças em geral nas populações humanas.

**ERRO EXPERIMENTAL:** É a medida da variação existente quando duas parcelas que receberam o mesmo tratamento, não apresentam necessariamente a mesma resposta.

**ERRO  $\alpha$  ( OU RISCO  $\alpha$ ):** No teste de hipóteses, um erro incorrido por rejeitar uma hipótese nula quando ela é realmente verdadeira, também chamado de erro tipo I ou de primeira espécie.

**ERRO ALEATÓRIO:** Um erro, geralmente um termo em um modelo estatístico, que se comporta como se fosse retirado ao acaso de uma distribuição particular de probabilidades.

**ERRO  $\beta$  (OU RISCO  $\beta$ ):** No teste de hipóteses, um erro incorrido por falhar em rejeitar uma hipótese nula quando ela é realmente falsa, também chamado de erro tipo II ou de segunda espécie.

**ERRO DE ESTIMAÇÃO:** A diferença entre o valor estimado e o valor verdadeiro.

**ERRO-PADRÃO:** O desvio padrão do estimador de um parâmetro. O erro padrão é também o desvio-padrão da distribuição amostral do estimador de um parâmetro.

**ERRO QUADRÁTICO MÉDIO:** O valor esperado do desvio ao quadrado de um estimador em relação ao valor verdadeiro do parâmetro por ele estimado. O erro quadrático médio pode ser decomposto na variância do estimador mais o quadrado da tendenciosidade; isto é,

$$MSE(\Theta) = E(\Theta - \theta)^2 = V(\Theta) + [E(\Theta) - \theta]^2.$$

**ELIMINAÇÃO REGRESSIVA (BACKWARD ELIMINATION):** Um método de regressão gradativa em que você começa com um modelo contendo todos os parâmetros possíveis e os elimina um por vez. Ver também regressão gradativa (stepwise regression); comparar com seleção progressiva (forward selection).

**ERRO (ERROR):** Um valor que não representa a medida ou observação original. Ele pode ser causado por desatenção no campo, instrumentos com mau funcionamento ou erros de digitação cometidos ao transferir os dados do caderno de campo ou folhas de dados para as planilhas.

**ERRO PADRÃO DA MÉDIA (STANDARD ERROR OF THE MEAN):** Uma estimativa assintótica do desvio padrão da população. Com frequência apresentada em publicações em vez do desvio padrão amostral. O erro padrão da média é menor que o amostral e pode dar a falsa impressão de menos variabilidade nos dados.

**ERRO PADRÃO DA REGRESSÃO (STANDARD ERROR OF REGRESSION):** Para um modelo de regressão, a raiz quadrada da soma dos quadrados dos resíduos dividida pelos graus de liberdade do modelo.

**ERRO TIPO I (TYPE I ERROR):** Rejeição errônea de uma hipótese nula estatística verdadeira. Em geral, é indicado pela letra grega alfa ( $\alpha$ ).

**ERRO TIPO II (TYPE II ERROR):** Aceitar incorretamente uma hipótese nula estatística que é falsa. Em geral, indicado pela letra grega beta ( $\beta$ ). O poder estatístico é igual a  $1 - \beta$ .

**ERROS COMUNS QUE SE COMETEM NA FORMULAÇÃO DE UMA PESQUISA:** Tipo de erro decorrente do uso dos seguintes fatores (RICHARDSON et al., 2009): i) escolha do tema sem fazer uma análise crítica de sua importância, originalidade e validade; ii) seleção de um problema muito vago e abrangente; iii) hipóteses mal formuladas e difíceis de testar e iv) ausência de um plano de pesquisa que oriente a coleta e análise dos dados.

**ERROS COMUNS NA REVISÃO DE LITERATURA QUE SE COMETEM NA ETAPA DE UMA PESQUISA:** Tipo de erro decorrente do uso dos seguintes fatores (RICHARDSON et al., 2009): i) revisão de literatura muito rápida deixando fora trabalhos que podem melhorar o projeto de pesquisa; ii) uso exagerado de fontes secundárias, que são referências que determinado autor faz de outros autores; iii) leitura concentrada na análise dos dados, sem considerar a valiosa informação que pode aparecer na descrição da metodologia e técnicas de coleta de dados; iv) revisão concentrada em uma área, sem procurar áreas conexas; v) revisão muita ampla ou muito restrita de literatura disponível; vi) referências bibliográficas incorretas, dificultando a sua localização posterior e vii) fichas muito extensas, não permitindo distinguir a informação importante daquela menos importante.

**ERROS COMUNS QUE SE COMETEM NA COLETA DE DADOS EM UMA ETAPA DE UMA PESQUISA:** Tipo de erro decorrente do uso dos seguintes fatores (RICHARDSON et al., 2009): i) falhas no

relacionamento com os sujeitos da pesquisa; ii) mudanças no plano de pesquisa devido a conveniências administrativas; iii) falta de avaliação das medidas disponíveis, antes de decidir aquelas que serão utilizadas na pesquisa; iv) escolha de medidas pouco adequadas, produzindo erros de medição e v) escolha de medidas sem os conhecimentos suficientes que garantem a sua correta aplicação.

**ERROS COMUNS QUE SE COMETEM NA APLICAÇÃO DE TESTES PADRONIZADOS EM UMA ETAPA DE UMA PESQUISA:** Tipo de erro decorrente do uso dos seguintes fatores (RICHARDSON et al., 2009): i) não se determina a validade dos instrumentos, na situação em que serão utilizados; ii) uso de inventários de personalidade ou outros instrumentos de autoavaliação, em situações que permitem respostas falsas; iii) uso de testes sem o devido conhecimento e treinamento; iv) falha no cálculo do tempo de duração de testes aplicando, desnecessariamente, instrumentos muito longos e v) inexistência de pré-teste para avaliar o instrumento a ser utilizado.

**ERROS COMUNS QUE SE COMETEM NO USO DA ESTATÍSTICA EM UMA ETAPA DE UMA PESQUISA:** Tipo de erro decorrente do uso dos seguintes fatores (RICHARDSON et al., 2009): i) escolha de testes estatísticos inadequados para análise; ii) procura de técnicas estatísticas, após coletar a informação; iii) uso de apenas uma técnica estatística, quando os dados permitem a aplicação de diversos coeficientes; iv) uso de estatística quando não se tem nem os dados suficientes, nem a informação adequada; v) considerar diferenças que não são significativas e esquecer aquelas significativas; vi) uso incorreto das técnicas de correlação e vii) transformação de variáveis intervalares em nominais, para simplificar a análise, perdendo informação que pode ser valiosa.

**ERROS COMUNS QUE SE COMETEM NO PLANO DE PESQUISA E NA METODOLOGIA EM UMA ETAPA DE UMA PESQUISA:** Tipo de erro decorrente do uso dos seguintes fatores (RICHARDSON et al., 2009): i) inexistência de um plano de pesquisa adequado ao problema em estudo; ii) não se define a população da pesquisa; iii) escolha de amostras muito pequenas que não permitem trabalhar subgrupos de interesse; iv) mudanças no planejamento para facilitar a coleta de dados, mas que debilitam a pesquisa; v) instrumentos muito cansativos que prejudicam a colaboração dos entrevistados; vi) intentos de fazer em seis meses o trabalho de dois anos; vii) falta de um plano de coleta detalhado que evite trabalhar em excesso e perder tempo e viii) coleta de dados sem realizar um pré-teste ou uma avaliação dos instrumentos e procedimentos.

**ERROS QUE SE COMETEM NA PESQUISA HISTÓRICA:** Tipo de erro decorrente do uso dos seguintes fatores (RICHARDSON et al., 2009): i) escolha de um tema que não dispõe de evidências suficientes; ii) excesso de fontes secundárias, particularmente em estudos referentes a acontecimentos passados; iii) problema de pesquisa mal formulado; iv) inadequação na avaliação dos dados históricos; v) viés pessoal nos procedimentos de pesquisa e vi) relatório que apenas registra fatos sem integrá-los a uma teoria.

**ERROS QUE SE COMETEM NA PESQUISA DESCRITIVA:** Tipo de erro decorrente do uso dos seguintes fatores (RICHARDSON et al., 2009): i) objetivos específicos poucos claros; ii) coleta de dados pouco adequada aos objetivos, não obtendo a informação necessária para analisar o problema; iii) amostragem por ocorrência; iv) planos de pesquisa elaborados após a coleta de dados e v) instrumentos mal elaborados.

**ERROS QUE SE COMETEM NA PESQUISA TIPO ENQUETES:** Tipo de erro decorrente do uso dos seguintes fatores (RICHARDSON et al., 2009): i) amostras pouco adequadas ao problema de pesquisa; ii) uso de questionários para problemas que precisam de outras técnicas, tais como entrevistas, observações, etc.; iii) questionários mal elaborados, sem pré-testes; iv) questionários com muitas perguntas inúteis; v)

apresentação pouco adequada do instrumento de coleta e vi) falta de análise da amostra de sujeitos que não responderam ao questionário, para determinar possíveis vieses.

**ERROS QUE SE COMETEM NA PESQUISA TIPO ENTREVISTAS:** Tipo de erro decorrente do uso dos seguintes fatores (RICHARDSON et al., 2009): i) plano de entrevista pouco adequado; ii) falta de treinamento; iii) falta de controle de possíveis vieses por parte do entrevistador; iv) não se faz análise de confiabilidade dos dados; v) linguagem da entrevista pouco compreensível para o entrevistado e vi) informações solicitadas, pouco conhecidas pelo entrevistado.

**ERROS QUE SE COMETEM NA PESQUISA DOS ESTUDOS OBSERVACIONAIS:** Tipo de erro decorrente do uso dos seguintes fatores (RICHARDSON et al., 2009): i) observadores pouco treinados; ii) pauta de observação muito complexa, exigindo muito do observador; iii) falta de controle do observador e iv) intentos de avaliar comportamentos pouco comuns, prejudicando a confiabilidade dos dados.

**ERROS QUE SE COMETEM EM UMA PESQUISA NA ANÁLISE DE CONTEÚDO:** Tipo de erro decorrente do uso dos seguintes fatores (RICHARDSON et al., 2009): i) escolha de conteúdos fáceis de analisar, mas inadequados aos objetivos de pesquisa; ii) não se estabelece a confiabilidade das técnicas utilizadas e iii) categorias poucos claras.

**ERROS QUE SE COMETEM EM UMA PESQUISA NA ANÁLISE DE RELAÇÕES ENTRE VARIÁVEIS:** Tipo de erro decorrente do uso dos seguintes fatores (RICHARDSON et al., 2009): i) amostras inadequadas para a comparação de variáveis; ii) intentos de análise correlacional após coletar os dados, em vez de coletar os dados necessários para determinada pesquisa; iii) insistência na análise de relações comprovadamente pouco úteis; iv) seleção de variáveis sem revisar as teorias existentes e v) uso de técnicas de correlação simples quando o problema exige correlação parcial ou múltipla.

**ERROS QUE SE COMETEM NA PESQUISA EXPERIMENTAL:** Tipo de erro decorrente do uso dos seguintes fatores (RICHARDSON et al., 2009): i) existência de diferenças entre o tratamento do grupo experimental e do grupo de controle, produzindo resultados errados; ii) utilização de poucos casos produzindo erros amostrais e iii) emparelhamento (matching) dos registros em base e variáveis não suficientemente correlacionadas com a variável dependente e iv) intentos de emparelhamento em três ou quatro variáveis, perdendo muitos sujeitos.

**ERROS QUE SE COMETEM NO PROCESSAMENTO DE DADOS EM UMA PESQUISA:** Tipo de erro decorrente do uso dos seguintes fatores (RICHARDSON et al., 2009): i) inexistência de uma pauta para codificar e registrar os dados; ii) não se registram detalhes e variações nos procedimentos de codificação, surgindo dificuldades quando o pesquisador tenta descrever a metodologia da pesquisa; iii) não se revisa a codificação para detectar erros e iv) mudanças nos procedimentos de qualificação dos dados.

**ERROS QUE SE COMETEM NA PREPARAÇÃO DO RELATÓRIO EM UMA PESQUISA:** Tipo de erro decorrente do uso dos seguintes fatores (RICHARDSON et al., 2009): i) esperar até que a pesquisa termine para preparar o relatório; ii) organização da revisão da literatura cronologicamente, em vez de prepará-la por temas; iii) não se incluem os resultados da revisão da literatura; iv) uso de muitas citações que não correspondem ao tema tratado; v) descrição inadequada da amostra e procedimentos metodológicos e vi) análise de resultados pouco significativos, esquecendo aspectos importantes da pesquisa.

**ETAPAS NO PLANEJAMENTO E REALIZAÇÃO DE UMA PESQUISA CIENTÍFICA:** Passos ou procedimentos que o pesquisador ou investigador científico tem que implementar ou seguir ao conduzir



um trabalho de pesquisa usando o método científico (RICHARDSON et al., 2009): i) identificação do problema de pesquisa; ii) revisão da literatura relacionada com o tema; iii) formulação do problema em termos específicos; iv) formulação das hipóteses; v) definição dos conceitos e variáveis; vi) especificação dos pressupostos na interpretação dos resultados; vii) elaboração do plano de pesquisa, e nesta etapa tem-se o uso das seguintes técnicas: amostragem, controle de variáveis, critérios de avaliação dos resultados e elaboração dos instrumentos; viii) procedimentos de coleta dos dados; ix) metodologia da análise dos dados; x) execução da pesquisa e xi) avaliação dos resultados e conclusões.

**ESCALAR (SCALAR):** Um único número, não uma matriz.

**ESCALONAMENTO MULTIDIMENSIONAL NÃO MÉTRICO (NMDS-NON-METRIC MULTIDIMENSIONAL SCALING):** Um método de ordenação que preserva a ordem da classificação das distâncias, ou dissimilaridades, originais entre as observações. Este método geral e robusto de ordenação pode usar qualquer tipo de medida de distância ou dissimilaridade.

**ERRO TIPO I:** Em teste de hipóteses, um erro cometido ao rejeitar uma hipótese nula quando, na verdade ela é verdadeira, também chamado de erro  $\alpha$ .

**ERRO TIPO II:** Em teste de hipóteses, um erro cometido ao falhar em rejeitar uma hipótese nula quando, na verdade, ela é falsa.

**ESPAÇO AMOSTRAL:** O conjunto de todos os resultados possíveis de um experimento aleatório.

**ESPAÇO DA AMOSTRA:** Chama-se espaço da amostra ao conjunto de todas as matrizes  $Y$  possíveis de dimensão  $(n \times 1)$  no caso univariado ou de dimensão  $(n \times p)$  no caso multivariado. Designando por  $A$  o espaço da amostra e por  $B$  o espaço dos parâmetros, um estimador é uma aplicação de  $A$  em  $B$ .

**ESPAÇO DOS PARÂMETROS:** Designa-se por espaço dos parâmetros o conjunto de todos os vectores  $\theta$  que satisfazem as restrições do modelo. Se  $Y$  é uma variável aleatória unidimensional, o espaço da amostra é, geralmente,  $\mathfrak{R}^n$  ou um subconjunto de  $\mathfrak{R}^n$ . Se  $Y$  é uma variável aleatória  $p$  dimensional, o espaço da amostra é  $\mathfrak{R}^{np}$  ou um seu subconjunto. O espaço dos parâmetros é  $\mathfrak{R}^q$ , se as restrições sobre  $\theta$  se limitarem à dimensão,  $q$ . Designando por  $A$  o espaço da amostra e por  $B$  o espaço dos parâmetros, um estimador é uma aplicação de  $A$  em  $B$ .

**ELIMINAÇÃO REGRESSIVA (BACKWARD ELIMINATION):** Um método de regressão gradativa em que você começa com um modelo contendo todos os parâmetros possíveis e os elimina um por vez. Ver também regressão gradativa (stepwise regression); comparar com seleção progressiva (forward selection).

**ERRO (ERROR):** Um valor que não representa a medida ou observação original. Ele pode ser causado por desatenção no campo, instrumentos com mau funcionamento ou erros de digitação cometidos ao transferir os dados do caderno de campo ou folhas de dados para as planilhas.

**ERRO PADRÃO DA MÉDIA (STANDARD ERROR OF THE MEAN):** Uma estimativa assintótica do desvio padrão da população. Com frequência apresentada em publicações em vez do desvio padrão amostral. O erro padrão da média é menor que o amostral e pode dar a falsa impressão de menos variabilidade nos dados.

**ERRO PADRÃO DA REGRESSÃO (STANDARD ERROR OF REGRESSION):** Para um modelo de regressão, a raiz quadrada da soma dos quadrados dos resíduos dividida pelos graus de liberdade do modelo.

**ERRO TIPO I (TYPE I ERROR):** Rejeição errônea de uma hipótese nula estatística verdadeira. Em geral, é indicado pela letra grega alfa ( $\alpha$ ).

**ERRO TIPO II (TYPE II ERROR):** Aceitar incorretamente uma hipótese nula estatística que é falsa. Em geral, indicado pela letra grega beta ( $\beta$ ). O poder estatístico é igual a  $1 - \beta$ .

**ESCALAR (SCALAR):** Um único número, não uma matriz.

**EXPERIMENTAÇÃO:** Termo usado para indicar o método de se verificar a adaptação de conhecimentos ou tecnologias a situações diversas daquelas nas quais foram criadas ou desenvolvidas. Por exemplo, a realização de um ensaio de competição de linhagens e/ou cultivares em ambiente diverso àquele no qual foram criadas.

**ENSAIO:** É o conjunto de todos os tratamentos, aplicados de forma repetida. Quando mais de um fator estiver sendo estudado, o ensaio é chamado de ensaio ou experimento fatorial. Também conhecido como experimento.

**ESCALA NOMINAL:** É aquela em que o número assinalado à observação serve somente como um nome à categoria à qual a observação pertence. São exemplos os números de placas de automóveis e código de endereçamento postal, estado civil, classificação ocupacional e cor.

**ESCALA DE POSIÇÃO:** É aquela onde o valor numérico da medida é usado somente como um meio de arranjar os elementos medidos em ordem, do menor para o maior. Podem ser citados nesta categoria os graus escolares, a ordem de produção e a intensidade da cor.

**ESCALA DE INTERVALO:** É aquela onde o ponto zero é definido por convenção, e os intervalos podem ter significado empírico, como datas, quociente de inteligência, temperatura em graus Celsius ou Fahrenheit. O número assinalado à observação considera como informação pertinente não somente a ordem relativa das medidas, mas também o tamanho do intervalo entre as medidas.

**ESCALA DE PROPORÇÃO:** É aquela onde o ponto zero é verdadeiro. Alguns exemplos são os tratamentos medidos em graus Kelvin e as quantidades físicas medidas em unidades tais como metro, quilo, segundo. Nesta escala, a única diferença em relação à anterior é aquela relativa ao ponto zero de medida. Podem também ser incluídas nesta escala as observações derivadas da ocorrência de algum fenômeno quantificado por meio de contagem.

**ESCALA ABSOLUTA:** Tipo de escala referente a contagem ou enumeração de ocorrências. O ponto zero absoluto nesta escala pressupõe que todas as operações matemáticas, tal como na escala de razão podem ser aplicadas. A escala absoluta poderia ser tomada como um mero caso especial da escala de razão, onde a única transformação matemática permitida para a escala absoluta é a transformação identidade, isto é, transformá-la nela mesma. Porém, ela é considerada uma categoria de escala a parte, para que se possa mantê-la destacada. Num certo sentido, a escala absoluta é a escala mais informativa possível, já que ela é única. Seria uma espécie de escala fundamental. A quantidade de pessoas alocadas em uma equipe de trabalho tipifica uma variável na escala absoluta de mensuração, assim como a quantidade de ovos que uma ave põe. Por exemplo, o número de insetos numa planta é enumerável. Observe que o número total de insetos não pode ser transformado em nenhum outro número e permanecer sendo discreto, com o significado de número de insetos. É por isso que a escala absoluta é diferente da escala de razão.



**ESTATÍSTICA:** A ciência de coletar, analisar, interpretar e retirar conclusões a partir de dados.

**ESTATÍSTICA DE TESTE:** Uma função de amostra de observações que fornece a base para testar uma hipótese estatística.

**ESTATÍSTICA RESUMIDA:** Um valor resumido calculado a partir de uma amostra de observações. Geralmente, uma estatística é um estimador de algum parâmetro de uma população.

**ESTATÍSTICA SQRP:** Em uma análise de regressão, a soma dos quadrados residual prevista (SQRP). Despreze cada ponto e estime os parâmetros do modelo a partir dos dados restantes. Estime o ponto desprezado desse modelo. Cada ponto é estimado uma vez e a soma dos quadrados desses erros é calculada.

**ESTATÍSTICA SUFICIENTE:** Um estimador é dito ser uma estatística suficiente para um parâmetro desconhecido se a distribuição da amostra, dada a estatística, não depende do parâmetro desconhecido. Isso significa que a distribuição do estimador contém toda a informação útil acerca do parâmetro desconhecido.

**ESTIMADOR BIPONDERADO:** O estimador bponderado ou biquadrado, é o estimador de localização  $T_n$  que é a solução de  $\sum_{i=1}^n \psi(u_i) = \mathbf{0}$ , onde  $\psi(u) = \begin{cases} u(1-u)^2 & |u| \leq 1 \\ \mathbf{0} & |u| > 1 \end{cases}$ ,  $u_i = \frac{x_i - T_n}{s_n}$ .

**ESTIMADOR RESISTENTE:** Um estimador é resistente se é afetado apenas até um certo limite quer por um pequeno número de erros grosseiros, quer por qualquer número de pequenos arredondamentos e erros de agrupamento.

**ESTIMADOR ROBUSTO E EFICIENTE:** Um estimador tem robustez de eficiência relativamente a um conjunto de distribuições se a sua variância, ou se o estimador for enviesado, o seu quadrático médio não se afastar muito do estimador de variância mínima para cada distribuição.

**ESTATÍSTICAS ORDINAIS:** As estatísticas ordinais de uma amostra são as observações arranjadas por ordem crescente de amplitude. Seja  $X_1, X_2, \dots, X_n$  uma amostra aleatória de dimensão n. As observações  $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ , arranjadas por ordem crescente, dá-se o nome de estatística ordinais da amostra, e a  $X_{(i)}$  dá-se o nome de i-ésima estatística ordinal.

**ESTIMADOR-L:** Sejam  $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$  as estatísticas ordinais de uma amostra de dimensão n.

Sejam  $a_1, a_2, \dots, a_n$  números reais,  $\mathbf{0} \leq a_i \leq \mathbf{1}$   $i = 1, 2, 3, \dots, n$ , tais que  $\sum_{i=1}^n a_i = \mathbf{1}$ . diz-se que T é um

estimador-L com ponderações  $a_1, a_2, \dots, a_n$  se  $T = \sum_{i=1}^n a_i X_{(i)} = \mathbf{1}$ .

**ESTIMADOR-M:** O estimador-M  $T_n(x_1, x_2, \dots, x_n)$  para a função  $\rho$  e para a amostra  $x_1, x_2, \dots, x_n$  é o valor de t que minimiza a função objetivo  $T = \sum_{i=1}^n \rho(x_i, t)$ . Frequentemente se escreve apenas  $T_n$  quando

a referência à amostra não for essencial. Quando a derivada de  $\rho$  em ordem a t, função que a menos de uma constante multiplicativa, denota-se por  $\psi$ , for conhecida, pode ser mais conveniente calcular  $T_n$  encontrando o valor de t que satisfaz  $\sum_{i=1}^n \psi(x_i, t) = \mathbf{0}$ . Assume-se implicitamente que tem derivada em

ordem a t para todos os valores de t. No entanto, para alguma  $\rho$  a derivada pode não existir num conjunto finito de pontos. Em geral se espera que seja contínua. O estimador-M mais conhecido é a média amostral, estimador de localização de mínimos quadrados. Para a estimação pelo mínimos quadrados,  $\rho$  é o quadrado do resíduo  $\rho(x, t) = (x - t)^2$ . Logo minimiza-se  $\sum_{i=1}^n (x_i - t)^2$ . Derivando a função objetivo em ordem a t, a menos da constante 2, obtém-se a equação  $\sum_{i=1}^n (x_i - t) = \mathbf{0}$ . Como  $t = \sum_{i=1}^n \frac{x_i}{n}$  resolve esta

equação,  $T_n$  é de fato a média amostral.

**ESTIMADOR DE ESCALA:** Um estimador de escala w é uma função não negativa da amostra  $w(x_1, x_2, \dots, x_n)$ , tal que  $w(a + bx_1, \dots, a + bx_n) = |b|w(x_1, x_2, \dots, x_n)$ . Um estimador de escala ignora

assim translações da amostra, e comporta-se como seria de esperar relativamente a alterações uniformes no que respeita a dispersão da amostra. Ele é invariante no diz respeito a localização, e equivariante no diz respeito a escala.

**ESTIMADOR EQUIVARIANTE:** Quando se faz uma translação em toda a mostra, de amplitude  $\alpha$ , qualquer estimador razoável de localização deve sofrer uma translação de igual amplitude. Especificamente, o que se quer é,  $T_n(x_1 + \alpha, \dots, x_n + \alpha) = T_n(x_1, x_2, \dots, x_n) + \alpha$  e assim tal estimador se diz equivariante para a localização.

**ESTIMADOR INVARIANTE:** Se for incluída alguma medida de escala na amostra, denominada de  $S_n$  na maioria dos estimadores-M, é razoável exigir que uma translação de toda a amostra mantenha o valor de  $S_n$  inalterado, neste caso  $S_n$  é invariante para a localização, ou seja  $S_n(x_1 + \alpha, \dots, x_n + \alpha) = S_n(x_1, x_2, \dots, x_n)$

e que  $S_n$  siga as alterações ou mudanças de escala  $S_n(bx_1, \dots, bx_n) = |b|S_n(x_1, x_2, \dots, x_n)$ , isto é,  $S_n$  é

equivariante para a escala. O valor absoluto de b parece no lado direito da equação  $S_n(bx_1, \dots, bx_n) = |b|S_n(x_1, x_2, \dots, x_n)$  porque as medidas de escala são negativas.

**ESTIMADOR NÃO PARAMÉTRICO:** Esta forma simples de se estimar as funções de confiabilidade e de taxas de falha não é necessário de se especificar as distribuições para a variável que identifica o tempo de falha. Neste procedimento não é possível o uso de censura nem a presença de variáveis que representa tratamentos que podem ser grupos, classes, dentre outras.

**ESTIMAÇÃO DE MÁXIMA VEROSSIMILHANÇA:** Um método de estimação de parâmetros que maximiza a função verossimilhança de uma amostra.

**ESTIMAÇÃO DE PARÂMETROS:** O processo de estimação de parâmetros de uma população ou distribuição. Estimação de parâmetros, juntamente com testes de hipóteses, é uma das maiores técnicas de inferência estatística.

**ESTIMAÇÃO INTERVALAR:** A estimação de um parâmetro por uma faixa de valores entre os limites inferior e superior, em contraste com a estimação pontual, em que o parâmetro é estimado por um simples valor numérico. Um intervalo de confiança é um procedimento típico de estimação intervalar.

**ESTIMADOR (OU ESTIMADOR PONTUAL):** Um procedimento para produzir uma estimativa de um parâmetro de interesse. Um estimador é geralmente uma função somente de valores de dados amostrais e, quando esses valores dos dados estão disponíveis, ele resulta em uma estimativa do parâmetro de interesse.

**ESTIMADOR BAYESEANO:** Um estimador para um parâmetro obtido a partir do método bayesiano, que usa uma distribuição prévia para o parâmetro juntamente com a distribuição condicional dos dados uma vez fornecido o parâmetro para obter a distribuição posterior do parâmetro. O estimador é obtido a partir da distribuição posterior.

**ESTIMADOR CONSISTENTE:** Um estimador que converge em probabilidade para o valor verdadeiro do parâmetro estimado quando o tamanho da amostra aumenta.

**ESTIMADOR DE MÍNIMOS QUADRADOS:** Qualquer estimador obtido pelo método dos mínimos quadrados.

**ESTIMADOR DE MOMENTO:** Um método de estimar parâmetros igualando momentos da amostra a momentos da população. Uma vez que os momentos da população serão funções de parâmetros desconhecidos, isso resulta em equações que podem ser resolvidas para estimativas dos parâmetros.

**ESTIMADOR DE PONTO:** Veja estimador.

**ESTIMADOR NÃO-TENDENCIOSO:** Um estimador que tem seu valor esperado igual ao parâmetro que está sendo estimado é dito não-tendencioso.

**ESTIMADOR TENDENCIOSO:** Veja estimador não-tendencioso.

**ESTIMATIVA (OU ESTIMATIVA DE PONTO):** O Valor numérico de um estimador pontual.

**ESCORE OU VALOR DO COMPONENTE PRINCIPAL (PRINCIPAL COMPONENT SCORE):** O valor de cada observação, que é uma combinação linear das variáveis originais medidas para aquela observação. É determinado por uma análise de componentes principais.

**ESCORES AJUSTADOS DOS LOCAIS (FITTED SITE SCORES):** Um resultado da análise de redundância. Os escores dos locais ajustados são a matriz  $Z$ , que resulta de uma ordenação dos dados originais em relação às características ambientais.

**ESCORES DAS ESPÉCIES (SPECIES SCORES):** Um resultado da análise de redundância. Os escores das espécies são a matriz de autovetores  $A$ , que resulta da análise de componentes principais da matriz padronizada de variância-covariância dos valores preditos da variável resposta  $(\hat{Y})$ . É usada para gerar tanto os escores dos locais quanto os escores ajustados dos locais.

**ESCORES DOS LOCAIS (SITE SCORES):** Um resultado da análise de redundância. Os escores dos locais são a matriz  $F$  resultante de uma ordenação dos dados originais em relação a características ambientais  $F = YA$ , onde  $Y$  é a resposta multivariada e  $A$  é a matriz de autovetores que resulta da análise de componentes principais dos valores esperados (ajustados) da variável resposta. Comparar com escores ajustados dos locais e escores das espécies.  $Z = \hat{Y}A$ , onde  $\hat{Y}$  são os valores preditos da variável resposta multivariada e  $A$  é a matriz de autovetores, que resulta de uma análise de componentes principais dos valores esperados (ajustados) da variável resposta. Comparar com escores dos locais e escores das espécies.

**ESCORE-Z (Z-SCORE):** O resultado de transformar uma variável pela subtração de sua média amostral e dividindo a diferença pelo desvio padrão amostral.

**ESCORE DO COMPONENTE PRINCIPAL (PRINCIPAL COMPONENT SCORE):** O valor de cada observação, que é uma combinação linear das variáveis originais medidas para aquela observação. É determinado por uma análise de componentes principais.

**ESCORES AJUSTADOS DOS LOCAIS (FITTED SITE SCORES):** Um resultado da análise de redundância. Os escores dos locais ajustados são a matriz  $Z$ , que resulta de uma ordenação dos dados originais em relação às características ambientais.

**ESCORES DAS ESPÉCIES (SPECIES SCORES):** Um resultado da análise de redundância. Os escores das espécies são a matriz de autovetores  $A$ , que resulta da análise de componentes principais da matriz padronizada de variância-covariância dos valores preditos da variável resposta  $(\hat{Y})$ . É usada para gerar tanto os escores dos locais quanto os escores ajustados dos locais.

**ESCORES DOS LOCAIS (SITE SCORES):** Um resultado da análise de redundância. Os escores dos locais são a matriz  $F$  resultante de uma ordenação dos dados originais em relação a características ambientais  $F = YA$ , onde  $Y$  é a resposta multivariada e  $A$  é a matriz de autovetores que resulta da análise de componentes principais dos valores esperados (ajustados) da variável resposta. Comparar com escores ajustados dos locais e escores das espécies.  $Z = \hat{Y}A$ , onde  $\hat{Y}$  são os valores preditos da variável resposta multivariada e  $A$  é a matriz de autovetores, que resulta de uma análise de componentes principais dos valores esperados (ajustados) da variável res- posta. Comparar com escores dos locais e escores das espécies.

**ESCORE-Z (Z-SCORE):** O resultado de transformar uma variável pela subtração de sua média amostral e dividindo a diferença pelo desvio padrão amostral.

**ESFERICIDADE (SPHERICITY):** A premissa dos métodos de teste de hipóteses multivariado de que as covariâncias de cada grupo são iguais umas às outras.

**ESPAÇO AMOSTRAL (SAMPLE SPACE):** O conjunto de todos os resultados possíveis ou resultados de um evento ou experimento.

**ESFERICIDADE (SPHERICITY):** A premissa dos métodos de teste de hipóteses multivariado de que as covariâncias de cada grupo são iguais umas às outras.

**ESPAÇO AMOSTRAL (SAMPLE SPACE):** O conjunto de todos os resultados possíveis ou resultados de um evento ou experimento.

**ESTUDO ANALÍTICO:** Um estudo em que uma amostra de uma população é usada para fazer inferência para uma população futura. É necessário considerar a estabilidade. Veja estudo enumerativo.

**ESTUDO DE CAPACIDADE DE PROCESSO:** Um estudo que coleta dados para estimar a capacidade de processo. Veja capacidade de processo, razão de capacidade de processo, PCR e  $PCR_k$ .

**ESTUDO DE OBSERVAÇÃO:** Um sistema é observado e dados devem ser coletados, porém mudanças não são feitas no sistema. Veja experimento.

**ESTUDO ENUMERATIVO:** Um estudo no qual uma amostra proveniente de uma população é usada para fazer inferência para a população. Veja estudo analítico.

**EVENTO:** Um subconjunto de um espaço amostral.

**EVENTOS MUTUAMENTE EXCLUDENTES:** Uma coleção de eventos cujas interseções são vazias.

**EVENTOS NÃO MUTUAMENTE EXCLUDENTES:** Dois eventos A e B são ditos não mutuamente excludentes quando têm alguns elementos em comum, isto é  $A \cap B \neq \emptyset$  Exemplo: Lançamento de um dado, e sejam os eventos:

**A = Ocorre face par = {2,4,6}; e**

**B = ocorre face menor que três = {1,2}. Sendo assim  $A \cap B = \{2\}$ .**

**EXAUSTIVA:** Uma propriedade de uma coleção de eventos que indica que sua união é igual ao espaço amostral.

**EXPERIMENTO FATORIAL:** Um tipo de planejamento de experimentos em que cada nível de um fator é testado em combinação com cada nível de outro fator. Em geral, em um experimento fatorial, todas as combinações dos níveis dos fatores são testadas.

**EXPERIMENTO:** Uma série de testes em que mudanças são feitas no sistema sob estudo.

**EXPERIMENTO ALEATÓRIO:** Um experimento que pode resultar em diferentes resultados, muito embora seja repetido da mesma maneira toda vez.

**EXPERIMENTO COMPARATIVO:** Um experimento em que os tratamentos que são as condições experimentais que devem ser estudados são incluídos no experimento. Os dados do experimento são usados para avaliar os tratamentos.

**EXPERIMENTOS CEGOS:** São aqueles onde o técnico deve tomar uma unidade experimental e entregar a unidade ao pesquisador, que fará as medições sem saber a que grupo pertence a unidade, evitando desta forma a tendenciosidade.

**EXPERIMENTOS DUPLAMENTE CEGOS:** São aqueles feitos com pessoas onde recomendam-se certos cuidados como: i) não se deve informar à a pessoa o grupo para o qual foi designada; ii) devem ser mantidos alheios ao resultado do sorteio, todos os profissionais envolvidos no trato dessas pessoas, para que seus comentários não afetem o moral delas; iii) o pesquisador que faz as observações ou medições devem fazê-lo sem saber a que grupo pertencem a pessoa que é examinada.

**EXPERIMENTO DE OTIMIZAÇÃO:** Um experimento conduzido para melhorar ou otimizar um sistema ou processo. Assume-se que os fatores importantes sejam conhecidos.

**EXPERIMENTO EXPLORATÓRIO:** Um experimento planejado e conduzido para a finalidade de explorar ou isolar um conjunto promissor de fatores para experimentos futuros. Muitos experimentos exploratórios são fatoriais fracionários, tais como os planejamentos fatoriais fracionários de dois níveis.

**EXPERIMENTO FATORIAL  $2^k$ :** Um experimento fatorial completo com k fatores, sendo cada fator testado em dois níveis.

**EXPERIMENTO FATORIAL  $2^{k-p}$ :** Um experimento fatorial completo ou fracionário com k fatores testados em uma fração  $2^{-p}$ , sendo cada fator testado em somente dois níveis.

**EXPERIMENTO FATORIAL COM DOIS NÍVEIS:** Um experimento fatorial completo ou fracionário com todos os fatores testados com somente dois níveis para cada fator. Veja Experimento Fatorial  $2^k$ .

**EXPERIMENTO FATORIAL FRACIONÁRIO:** Um tipo de experimento fatorial em que nem todas as combinações possíveis de tratamentos são corridas. Isso é geralmente feito para reduzir o tamanho de um experimento com vários fatores.

**EXPERIMENTO PLANEJADO:** Um experimento em que os testes são planejados antes e os planos geralmente incorporam modelos estatísticos. Veja experimento.

**EXPERIMENTOS ROTACIONAIS:** É aquele em que se utiliza o princípio dos blocos. Nestes casos porém cada participante, como uma pessoa, um paciente é um bloco, como por exemplo em um ensaio na área de odontologia, onde se pretenda comparar dois métodos de escovação, A e B, neste caso se cada participante do experimento for tomado como bloco, é preciso sortear a ordem dos métodos (AB ou BA) para cada participante. Em inglês, cross-over experiments.

**EXTRAPOLAÇÃO DISFARÇADA:** Uma extrapolação é uma previsão em análise de regressão que é feita no ponto  $(x_1, x_2, \dots, x_k)$ , que está fora da região dos dados usados para gerar o modelo. Extrapolação disfarçada ocorre quando não é óbvio que o ponto esteja fora da região. Isso pode ocorrer quando a multicolinearidade está presente nos dados usados para construir o modelo.

**ELEMENTO DE DADOS:** Nível de dados guardados em um banco de dados. Elementos de dados separados podem representar diferentes níveis de agregação de resumo, variando de dados primitivos a dados agregados.

**ELIMINAÇÃO BACKWARD:** Método de seleção de variáveis numa regressão linear que consiste em ajustar o modelo com todas as covariáveis e eliminar, uma a uma, as menos significativas. Em inglês Backward elimination.

**ELIMINAÇÃO BACKWARD:** Método de seleção de variáveis para inclusão no modelo de regressão que começa incluindo todas as variáveis independentes no modelo e então elimina as que não oferecem uma contribuição significativa para previsão. Poderíamos traduzir como eliminação retroativa, mas a expressão em inglês é de uso corrente em nosso idioma.

**ELIMINAÇÃO DE PARÂMETROS:** Em alguns estudos são considerados modelos que possuem parâmetros que não são úteis à análise realizada. Estes parâmetros podem ser eliminados por meio do uso de distribuições marginais ou condicionais. Em inglês Elimination of parameters.

**ELIPSE:** É uma das cônicas. É uma curva fechada de segunda ordem. Pode ser definida como o conjunto de pontos, para os quais a soma das distâncias de dois pontos fixos, focos, é constante. Outra forma é defini-la como o conjunto de pontos do plano que satisfaz a equação:  $ax^2 + by^2 = 1$ . Em inglês ellipse.

**EM:** Algoritmo expectation-maximization.

**EMM:** Equações de modelo misto.

**EMMSE:** Estimador de mínimo erro quadrático médio.

**EM-REML:** REML obtido via algoritmo EM.



**ESPAÇOS AMOSTRAIS EQUIPROVÁVEIS:** É o espaço amostral discreto e finito  $S = \Omega = \{a_1, a_2, a_3, a_4, \dots, a_n\}$  onde todos os resultados do espaço amostral têm a mesma probabilidade de ocorrer. Em outras palavras, tem-se que todos os eventos elementares têm a mesma probabilidade, ou seja,  $P(B) = \frac{|B|}{|S|}$ , em que  $|B|$  é o número de

elementos do conjunto B e  $|S|$  é o número de elementos do conjunto S. Por exemplo, se  $B = \{a_1, a_3\}$ , então  $P(B) = \frac{2}{n}$ . Assim, se o experimento aleatório tem um espaço amostral equiprovável, o cálculo de probabili-

dades de eventos se reduz a contar o número de elementos desses conjuntos. Naturalmente, isso só se aplica a espaços amostrais discretos e finitos. Note que se o espaço amostral for discreto e infinito, ou contínuo, o valor obtido de  $|S|$  será  $|S| = \infty$ .

**ESPAÇOS AMOSTRAIS EQUIPROVÁVEIS:** É o espaço amostral discreto e finito  $S = \Omega = \{a_1, a_2, a_3, a_4, \dots, a_n\}$  onde todos os resultados do espaço amostral têm a mesma probabilidade de ocorrer. Em outras palavras, tem-se que todos os eventos elementares têm a mesma probabilidade, ou seja,  $P(B) = \frac{|B|}{|S|}$ , em que  $|B|$  é o número de elementos do conjunto B e  $|S|$  é o número de elementos do

conjunto S. Por exemplo, se  $B = \{a_1, a_3\}$ , então  $P(B) = \frac{2}{n}$ . Assim, se o experimento aleatório tem um

espaço amostral equiprovável, o cálculo de probabilidades de eventos se reduz a contar o número de elementos desses conjuntos. Naturalmente, isso só se aplica a espaços amostrais discretos e finitos. Note que se o espaço amostral for discreto e infinito, ou contínuo, o valor obtido de  $|S|$  será  $|S| = \infty$ .

**ESQUEMA DE PARES CASADOS (MATCHED-PAIRS LAYOUT):** Uma forma de delineamento em blocos aleatorizados na qual as réplicas são selecionadas para serem similares em tamanho corporal, ou outra característica, e então são atribuídas de maneira aleatória aos tratamentos.

**ESTATÍSTICA ASSINTÓTICA (ASYMPTOTIC STATISTICS):** Tipos de estimadores estatísticos que são calculados com base na premissa de que, se o experimento for repetido infinitas vezes, o estimador poderia convergir no valor calculado. Ver também inferência frequentista.

**ESTATÍSTICA DESCRITIVA (SUMMARY STATISTICS):** Números que descrevem a posição, a dispersão, a assimetria e a curtose ou achatamento de dados.

**ESTATÍSTICA-F (F-STATISTIC):** O resultado, em uma análise de variância (ANOVA) ou em regressão, da divisão da soma dos quadrados entre grupos, (SQentre grupos) pela soma dos quadrados dentro de grupos (SQdentro de grupos). Ela tem dois graus de liberdade diferentes – os do numerador, associados com a SQentre grupos e os graus de liberdade do denominador são associados com a SQdentro de grupos. Esses dois valores também definem a forma da distribuição-F. Para determinar sua significância estatística, ou valor de P, a estatística-F é comparada ao valor crítico da distribuição-F, dado o tamanho amostral, a probabilidade de erro Tipo I e os graus de liberdade do numerador e do denominador.

**ESTATÍSTICA NÃO PARAMÉTRICA (NON-PARAMETRIC STATISTICS):** Um ramo da análise estatística que não depende dos dados terem sido tirados de uma variável aleatória definida e com distribuição de probabilidade conhecida. É frequentemente substituída por análises de Monte Carlo.

**ESTATÍSTICA DO TESTE (TEST STATISTIC):** O resultado numérico da manipulação de dados que são usados para examinar hipóteses estatísticas. As estatísticas do teste são comparadas a valores que poderiam ser obtidas de cálculos idênticos se a hipótese nula estatística de fato fosse verdadeira.

**ESQUEMA DE PARES CASADOS (MATCHED-PAIRS LAYOUT):** Uma forma de delineamento em blocos aleatorizados na qual as réplicas são selecionadas para serem similares em tamanho corporal, ou outra característica, e então são atribuídas de maneira aleatória aos tratamentos.

**ESTATÍSTICA ASSINTÓTICA (ASYMPTOTIC STATISTICS):** Tipos de estimadores estatísticos que são calculados com base na premissa de que, se o experimento for repetido infinitas vezes, o estimador poderia convergir no valor calculado. Ver também inferência frequentista.

**ESTATÍSTICA DESCRITIVA (SUMMARY STATISTICS):** Números que descrevem a posição, a dispersão, a assimetria e a curtose ou achatamento de dados.

**ESTATÍSTICA-F (F-STATISTIC):** O resultado, em uma análise de variância (ANOVA) ou em regressão, da divisão da soma dos quadrados entre grupos, (SQentre grupos) pela soma dos quadrados dentro de grupos (SQdentro de grupos). Ela tem dois graus de liberdade diferentes – os do numerador, associados com a SQentre grupos e os graus de liberdade do denominador são associados com a SQdentro de grupos. Esses dois valores também definem a forma da distribuição-F. Para determinar sua significância estatística, ou valor de P, a estatística-F é comparada ao valor crítico da distribuição-F, dado o tamanho amostral, a probabilidade de erro Tipo I e os graus de liberdade do numerador e do denominador.

**ESTATÍSTICA NÃO PARAMÉTRICA (NON-PARAMETRIC STATISTICS):** Um ramo da análise estatística que não depende dos dados terem sido tirados de uma variável aleatória definida e com distribuição de probabilidade conhecida. É frequentemente substituída por análises de Monte Carlo.

**ESTATÍSTICA DO TESTE (TEST STATISTIC):** O resultado numérico da manipulação de dados que são usados para examinar hipóteses estatísticas. As estatísticas do teste são comparadas a valores que poderiam ser obtidas de cálculos idênticos se a hipótese nula estatística de fato fosse verdadeira.

**EMPARELHAMENTO (OU PAREAMENTO):** Formação de pares ou duplas de indivíduos com características semelhantes, por exemplo, de sexo e idade.

**EMPRESAS DE ADMINISTRAÇÃO DE CAMPO:** Empresas que fornecem serviços de suporte, tais como formatação de questionários, selecionadores por escrito e coleta de dados para as companhias de pesquisa de serviços totais.

**EMPRESA DE PESQUISA DE MARKETING AD HOC OU PERSONALIZADAS:** Companhias de pesquisa que executam pesquisa de marketing personalizada para atender a projetos específicos de clientes empresariais.

**EMPRESAS DE PESQUISA DE SERVIÇOS PERSONALIZADOS:** Empresas que coletam, montam e vendem os mesmos dados gerais de pesquisa de mercado para várias outras.

**ÊNFASE INADEQUADA:** O uso de palavras ou frases em negrito, itálico, maiúsculos ou sublinhadas no contexto de uma pergunta, que pode servir para influenciar o entrevistado.

**ENSAIO:** Ver Experimento.



**ENSAIO CLÍNICO CONTROLADO E CASUALIZADO:** Ensaio clínico que envolve pelo menos um tratamento em teste e um tratamento controle, com recrutamentos e seguimentos simultâneos de todos os grupos, e onde os tratamentos são designados aos pacientes por processo aleatório, de tal maneira que nem os pacientes nem os responsáveis pela seleção e tratamentos desses pacientes possam influenciar a alocação de tratamentos e onde as alocações permanecem desconhecidas dos pacientes e do pessoal clínico até o final. A alocação é conhecida dos pacientes e dos clínicos apenas por códigos, de preferência numéricos. Em inglês randomized controlled clinical trial (rcct).

**ENSAIO CLÍNICO:** O mesmo que investigação ou estudo clínico. Em metodologia científica, a denominação tem a conotação de pesquisa clínica em condições controladas, o que tende a ser sinônimo de ensaio clínico randomizado.

**ENSAIO CLÍNICO:** Atividade de pesquisa que envolva a administração de um tratamento em teste (por exemplo, uma droga, um procedimento cirúrgico ou um dispositivo médico) para uma unidade experimental, com a finalidade de avaliar o tratamento. Na maioria dos casos, a unidade experimental é o homem, mas pode ser um animal experimental. Veja unidade experimental.

**ENSAIOS CLÍNICOS:** Designação genérica dos desenhos experimentais que podem validar, com maior precisão e controle, os enunciados que avançam relações de causa e efeito, ou seja, as hipóteses que surgem etiologias para doenças. Na prática, esses estudos são muito utilizados na avaliação de ensaios terapêuticos nos testes de novas drogas.

**ENSAIOS INDEPENDENTES:** Os ensaios sucessivos de um evento são considerados independentes se a probabilidade do resultado de qualquer ensaio for independente do resultado dos outros. A expressão é geralmente limitada aos casos onde a probabilidade é a mesma para todos os ensaios. A amostragem de características, vista como uma série de ensaios é freqüentemente chamada de Ensaios de Bernoulli. Em inglês Independent trials.

**ENSAIO:** Qualquer ação experimental feita com a finalidade de obter dados para julgamentos ou conclusão. O mesmo que experimento.

**ENSAIOS CLÍNICOS CONTROLADOS RANDOMIZADOS (ECCR):** Estudos delineados para testar uma medida terapêutica. Os sujeitos são designados randomicamente para um dos seguintes grupos: i) grupo de intervenção, que receberá o tratamento experimental, ou ii) grupo-controle, que receberá o tratamento não-experimental, que pode consistir tanto em um placebo ou substância inerte, como em um método-padrão de tratamento.

**ENSAIOS DE CAMPO CONTROLADOS E RANDOMIZADOS (ECaCR):** Estudos que são semelhantes aos ECCRs, anteriormente descritos, mas que não delineados para testar uma medida preventiva, como uma vacina. Pessoas suscetíveis são randomizadas em 2 grupos e recebem então a vacina ou um placebo, geralmente no começo da estação de alto risco. Testar a eficiência das vacinas por ECaCR é dispendioso, mas pode ser necessário quando uma nova vacina está sendo introduzida.

**ESTIMADOR DESTENDENCIADO (UNBIASED ESTIMATOR):** Um parâmetro estatístico calculado, que não é consistentemente maior nem menor que o valor real.

**ESTIMADORES-M (M-ESTIMATORS):** Um método de regressão para minimizar o efeito de dados destoantes. Eles calculam inclinações de regressões após dar menos peso às maiores somas dos quadrados.

**ESTRESSE (STRESS):** Uma medida de quão bem os resultados de um escalonamento multidimensional não métrico prediz os valores observados. Além disso, o resultado de gastar muito tempo em análises estatísticas e não tempo suficiente no campo.

**EVENTO (EVENT):** Uma simples observação ou processo com começo e fim claramente definidos.

**EVENTO COMPLEXO (COMPLEX EVENT):** Na teoria de conjuntos, eventos complexos são coleções de eventos simples. Se A e B são dois eventos independentes, então  $C = A$  ou  $B$ , expresso como  $C = A \cup B$  e lê-se A união com B, será um evento complexo. Se A e B forem eventos independentes, com probabilidades associadas  $P_A$  e  $P_B$ , então a probabilidade de C é  $P_C = P_A + P_B$ . Comparar com eventos compartilhados.

**EVENTOS COMPARTILHADOS (COMPLEX EVENT):** Na teoria de conjuntos, os eventos compartilhados são coleções de eventos simples. Se A e B são dois eventos independentes, então  $C = A$  e  $B$ , escrito como  $C = A \cap B$  e lido como A intercessão com B, é um evento compartilhado. Se A e B são dois eventos independentes, cada um associado a uma probabilidade  $P_A$  e  $P_B$ , então a probabilidade de C ( $P_C$ ) é igual a  $P_A \times P_B$ . Comparar com evento complexo.

**ENTRADA DE DADOS:** Processo de teclar os dados para armazenamento eletrônico.

**ENTRADA DE DADOS:** O processo de entrar com os dados brutos dos questionários concluídos em um computador.

**ENTRADA DE DADOS INTELIGENTES:** A checagem lógica de informações que estão sendo inseridas em um dispositivo de banco de entrada de dados por determinada máquina ou por outra conectada a ela.

**ESTIMADOR DESTENDENCIADO (UNBIASED ESTIMATOR):** Um parâmetro estatístico calculado, que não é consistentemente maior nem menor que o valor real.

**ESTIMADORES-M (M-ESTIMATORS):** Um método de regressão para minimizar o efeito de dados destoantes. Eles calculam inclinações de regressões após dar menos peso às maiores somas dos quadrados.

**EXCLUSIVO (EXCLUSIVE):** Um conjunto de resultados é dito exclusivo se não houver nenhuma sobreposição em seus valores. Comparar com completo.

**EXPECTATIVA (EXPECTATION):** A média aritmética, média ou valor esperado de uma distribuição de probabilidades.

**EXPERIMENTO (EXPERIMENT):** Um conjunto de observações replicadas, em geral realizadas sob condições controladas.

**EXPERIMENTO DE PRESSÃO (PRESS EXPERIMENT):** Um tipo de ensaio ou experimento manipulativo no qual um tratamento é aplicado e reaplicado ao longo da duração do experimento com objetivo de manter a força do tratamento. Comparar com experimento de pulso.

**EXPERIMENTO DE PULSO (PULSE EXPERIMENT):** Um experimento manipulativo em que o tratamento é aplicado apenas uma vez, e então a réplica tratada é deixada para se recuperar da manipulação. Comparar com experimento de pressão.

**EXPERIMENTO DE TRAJETÓRIA (TRAJECTORY EXPERIMENT):** Um tipo de experimento natural em que todas as réplicas são amostradas repetidamente em muitos pontos no tempo. A replicação é

baseada na variabilidade temporal. Os experimentos de trajetória frequentemente são analisados usando séries temporais ou modelos autorregressivos. Comparar com experimento fotográfico.

**EXPERIMENTO FOTOGRÁFICO (SNAPSHOT EXPERIMENT):** Um tipo de experimento natural em que todas as réplicas são amostradas em um ponto no tempo. A replicação é fundamentada na variabilidade espacial. Comparar com experimento de trajetória.

**EXPERIMENTO MANIPULATIVO (MANIPULATIVE EXPERIMENT):** Aquele no qual o investigador deliberadamente aplica um ou mais tratamentos a uma população amostral, ou populações, e então observa o resultado do tratamento. Comparar com experimento natural.

**ESTRESSE (STRESS):** Uma medida de quão bem os resultados de um escalonamento multidimensional não métrico prediz os valores observados. Além disso, o resultado de gastar muito tempo em análises estatísticas e não tempo suficiente no campo.

**EVENTO (EVENT):** Uma simples observação ou processo com começo e fim claramente definidos.

**EVENTO COMPLEXO (COMPLEX EVENT):** Na teoria de conjuntos, eventos complexos são coleções de eventos simples. Se A e B são dois eventos independentes, então  $C = A$  ou B, expresso como  $C = A \cup B$  e lê-se A união com B, será um evento complexo. Se A e B forem eventos estatisticamente independentes, com probabilidades associadas  $P_A$  e  $P_B$ , então a probabilidade de C é  $P_C = P_A + P_B$ . Comparar com eventos compartilhados.

**EVENTOS COMPARTILHADOS (COMPLEX EVENT):** Na teoria de conjuntos, os eventos compartilhados são coleções de eventos simples. Se A e B são dois eventos independentes, então  $C = A$  e B, escrito como  $C = A \cap B$  e lido como A intercessão com B, é um evento compartilhado. Se A e B são dois eventos independentes, cada um associado a uma probabilidade  $P_A$  e  $P_B$ , então a probabilidade de C ( $P_C$ ) é igual a  $P_A \times P_B$ . Comparar com evento complexo.

**ENTREVISTA:** Encontro para coleta de dados, no qual uma pessoa (um entrevistador) faz perguntas a outra que é um respondente. Entrevistas podem ser feitas pessoalmente ou por telefone.

**ENTREVISTA:** É um instrumento de pesquisa utilizado na fase de coleta de dados.

**ENTREVISTA:** Embora possa ser usada como uma técnica de pesquisa, a entrevista é mais frequentemente usada em conexões com atividades de não pesquisa. Como uma técnica de pesquisa, ela é uma conversação com a finalidade definida de obter certas informações. Para que ela tenha algum poder de generalização, é preciso, entretanto, que se obedeça a certos critérios comuns da estratégia científica. Muitas vezes a entrevista é usada no lugar de questionários, porém seu melhor uso parece ser o de instrumento complementar usado, antecipadamente, com objetivos exploratórios.

**ENTREVISTA (VALIDADE):** A validade da entrevista como uma técnica da pesquisa se centra primariamente na competência do entrevistador. Daí ser uma das suas principais fraquezas quaisquer tendências que o entrevistador possa ter.

**ENTREVISTA POR TELEFONE ASSISTIDAS POR COMPUTADOR (ETAC):** Adaptação de computadores para estruturar e facilitar entrevistas por telefone, incluindo seleção de amostras, apresentação de questionários, registros e análise de dados.

**ENTREVISTADO:** A pessoa que responde às perguntas no instrumento de pesquisa.

**ENTREVISTAS DE ABORDAGEM EM SHOPPING:** Entrevistas conduzidas em áreas públicas de shopping centers, nas quais os compradores são interceptados e entrevistados face a face.

**ENTREVISTAS EM PROFUNDIDADE:** Entrevistas individuais que investigam e extraem respostas detalhadas a perguntas, muitas vezes com o uso de técnicas não direcionadas para descobrir motivações ocultas.

**ENTREVISTAS EXECUTIVAS:** O equivalente industrial das entrevistas porta a porta.

**ENTREVISTAS PESSOAIS:** Entrevistas nas quais as informações são solicitadas diretamente dos entrevistados numa situação face a face.

**ENTREVISTAS PORTA A PORTA:** Entrevistas conduzidas pessoalmente com os consumidores em suas residências.

**ENTREVISTAS TELEFÔNICAS ASSISTIDAS POR COMPUTADOR (CATI):** Entrevistas telefônicas em local central, nas quais os entrevistadores inserem respostas diretamente no computador.

**ENTREVISTAS TELEFÔNICAS EM LOCAL CENTRAL:** Entrevistas conduzidas por telefonemas a respondentes de uma central localizada de pesquisa de marketing.

**ENUMERÁVEL:** Um conjunto é dito enumerável ou contável se puder ser posto em correspondência um a um com os números naturais, isto é, se existe uma função bijetora entre o conjunto e os números naturais. Em inglês enumerable.

**ENVELOPE:** Uma curva é o envelope ou curva limite ou ainda envoltória de uma família de curvas se ela toca que a é tangente a cada curva da família. Em inglês envelope.

**EXCLUSIVO (EXCLUSIVE):** Um conjunto de resultados é dito exclusivo se não houver nenhuma sobreposição em seus valores. Comparar com completo.

**EXPECTATIVA (EXPECTATION):** A média aritmética, média ou valor esperado de uma distribuição de probabilidades.

**EXPERIMENTO NATURAL (NATURAL EXPERIMENT):** A comparação entre dois ou mais grupos que não foram manipulados de nenhuma forma pelo investigador; pelo contrário, o delineamento recai sobre a variação natural entre grupos. Os experimentos naturais idealmente comparam grupos que são iguais em todas as formas, exceto para um único fator causal de interesse. Embora os experimentos naturais possam ser analisados com muitas das mesmas estatísticas que os manipulativos, as inferências resultantes podem ser mais fracas por causa de variáveis confundidas que não foram controladas.

**EXTENSÃO (EXTENT):** A área espacial total ou amplitude temporal que é coberta por todas as unidades amostrais em um estudo experimental. Comparar com grão.

**EXTRAPOLAÇÃO (EXTRAPOLATION):** A estimativa de valores não observados fora da gama dos valores observados. Em geral, é menos confiável que a interpolação.

**EXPERIMENTO (EXPERIMENT):** Um conjunto de observações replicadas, em geral realizadas sob condições controladas.

**EXPERIMENTO DE PRESSÃO (PRESS EXPERIMENT):** Um tipo ensaio ou de experimento manipulativo no qual um tratamento é aplicado e reaplicado ao longo da duração do experimento com objetivo de manter a força do tratamento. Comparar com experimento de pulso.

**EXPERIMENTO DE PULSO (PULSE EXPERIMENT):** Um experimento manipulativo em que o tratamento é aplicado apenas uma vez, e então a réplica tratada é deixada para se recuperar da manipulação. Comparar com experimento de pressão.

**EXPERIMENTO DE TRAJETÓRIA (TRAJECTORY EXPERIMENT):** Um tipo de experimento natural em que todas as réplicas são amostradas repetidamente em muitos pontos no tempo. A replicação é baseada na variabilidade temporal. Os experimentos de trajetória frequentemente são analisados usando séries temporais ou modelos autorregressivos. Comparar com experimento fotográfico.

**EXPERIMENTO FOTOGRÁFICO (SNAPSHOT EXPERIMENT):** Um tipo de experimento natural em que todas as réplicas são amostradas em um ponto no tempo. A replicação é fundamentada na variabilidade espacial. Comparar com experimento de trajetória.

**EXPERIMENTO MANIPULATIVO (MANIPULATIVE EXPERIMENT):** Aquele no qual o investigador deliberadamente aplica um ou mais tratamentos a uma população amostral, ou populações, e então observa o resultado do tratamento. Comparar com experimento natural.

**EQUAÇÃO DE REGRESSÃO:** Equação que fornece a estimativa dos valores de uma variável em função dos valores de outra variável, desde que exista correlação entre ambas. Quando as estimativas partem de diversas variáveis, os valores são preditos pela equação de regressão múltipla.

**ERRO DE AMOSTRAGEM:** Diferença entre os valores obtidos numa amostra que são as estatísticas e os valores da população que são os parâmetros. Se um escore for usado como estimador dos escores de toda a população, a diferença entre esse escore e a média de todos os valores registrados será um erro de amostragem.

**ERRO PADRÃO DE MEDIDA:** Erro padrão de qualquer medida, devido às flutuações da amostragem ou aos erros de observação.

**ESCALA:** É uma série de números definida segundo um critério capaz de indicar os vários graus de certo traço ou característica.

**ESCALA DE PERCENTIS:** Escala que indica os valores correspondentes aos vários percentis. Tais escalas são usadas na apresentação das normas dos testes.

**ESTATÍSTICA:** Qualquer número usado para descrever um aspecto da amostra. Assim, a média, o desvio-padrão, o coeficiente de correlação entre dois conjuntos de medidas dos membros de uma amostra são exemplos de estatística. A cada estatística na amostra correspondente um parâmetro na população. As estatísticas são representadas por letras do nosso alfabeto ao que os parâmetros são representados pelas letras gregas correspondentes.

**ESTATÍSTICA DESCRITIVA:** Trata da apuração, apresentação, análise e interpretação dos dados observados. Em outras palavras, descreve as amostras ou a população.

**ESTATÍSTICA INDUTIVA:** Método que parte do particular para o geral, ou seja, o processo pelo qual são feitas generalizações para a população, a partir das amostras. É também conhecida como inferência estatística.

**ESPAÇO AMOSTRAL:** Conjunto de todos os possíveis resultados de uma experiência aleatória, resultados esses que podem ser de natureza quantitativa ou qualitativa.

**EVENTO:** Qualquer subconjunto do espaço amostral, isto é, qualquer resultado ou conjunto de resultados do espaço amostral.

**EVENTO SIMPLES:** Formado por um único elemento do espaço amostral.

**EVENTO COMPOSTO:** Todo aquele formado por dois ou mais elementos do espaço amostral.

**EVENTO CERTO:** Diz-se daquele que ocorre sempre, ou seja, em todas as realizações da experiência.

**EVENTO IMPOSSÍVEL:** Aquele que nunca ocorre em nenhuma realização de experiência.

**EVENTO SOMA:** É o evento que consiste na realização de pelo menos um dos eventos  $E_1$  e  $E_2$ . É também conhecido como evento união.

**EVENTO PRODUTO:** Também chamado evento intersecção, consiste na realização de ambos os eventos  $E_1$  e  $E_2$ , isto é, eles devem ocorrer simultaneamente.

**EVENTO CONDICIONADO:** Consiste na realização do evento  $E_1$ , sob a condição de ter-se realizado o evento  $E_2$ , isto é, com a informação que o evento  $E_2$  já ocorreu.

**EVENTOS MUTUAMENTE EXCLUSIVOS:** Dois eventos, associados a uma experiência aleatória, são ditos mutuamente exclusivos se a ocorrência de um deles exclui a possibilidade de ocorrência do outro, ou seja, não ocorrem simultaneamente.

**EVENTO COMPLEMENTAR:** Define-se evento complementar de um evento  $E$ , associado a uma experiência aleatória, aquele que só ocorre se  $E$  deixar de ocorrer, isto é, é o evento formado por todos os elementos do espaço amostral  $\Omega$  que não pertencem a  $E$ .

**EVENTOS INDEPENDENTES:** Dois eventos, associados a uma experiência aleatória, são ditos independentes, quando a ocorrência de um não depende (ou não é condicionada, ou não se vincula) à ocorrência do outro, isto é, a informação adicional de que um dos eventos já ocorreu em nada altera a possibilidade de ocorrência do outro.

**EXPERIÊNCIA ALEATÓRIA:** O conceito de experiência aleatória é considerado como primitivo, muito embora subentendido que duas condições devem ser satisfeitas: Deve sempre ser possível repetir a experiência indefinidamente, fixadas certas condições; Mesmo mantendo as condições iniciais, deve ser impossível influenciar no resultado de uma particular repetição da experiência.

**EXATIDÃO:** Corresponde a boa precisão e razoável ausência de vício.

**EXPERIMENTO NATURAL (NATURAL EXPERIMENT):** A comparação entre dois ou mais grupos que não foram manipulados de nenhuma forma pelo investigador; pelo contrário, o delineamento recai sobre a variação natural entre grupos. Os experimentos naturais idealmente comparam grupos que são iguais em todas as formas, exceto para um único fator causal de interesse. Embora os experimentos naturais



possam ser analisados com muitas das mesmas estatísticas que os manipulativos, as inferências resultantes podem ser mais fracas por causa de variáveis confundidas que não foram controladas.

**EXTENSÃO (EXTENT):** A área espacial total ou amplitude temporal que é coberta por todas as unidades amostrais em um estudo experimental. Comparar com grão.

**EXTRAPOLAÇÃO (EXTRAPOLATION):** A estimativa de valores não observados fora da gama dos valores observados. Em geral, é menos confiável que a interpolação.

**EPIDEMIOLOGIA:** O estudo dos fatores que determina a frequência e a distribuição das doenças nas coletividades humanas conforme o Guia de Métodos de Ensino da Associação Internacional de Epidemiologia.

**EPIDEMIOLOGIA:** Ciência que estuda o processo saúde-doença em coletividades humanas, analisando a distribuição e os fatores determinantes das enfermidades, danos à saúde e eventos associados à saúde coletiva, propondo medidas específicas de prevenção, controle, ou erradicação de doenças, e fornecendo indicadores que sirvam de suporte ao planejamento, administração e avaliação das ações de saúde.

**EPIDEMIOLOGIA:** Estudos dos fatores que influenciam a ocorrência e a distribuição das doenças nas populações humanas.

**EPIDEMIOLOGIA CLÍNICA:** Epidemiologia que tem como contexto as urgências sócio-sanitárias de povos superexplorados. Apresenta nas seguintes características: i) enfrenta postulados teóricos-metodológicos e práticos da saúde oficial e da medicina hegemônica; ii) não se reduz ao uso progressista de conceitos, técnicos e linhas de ação convencionistas, tampouco à adaptação terceiro-mundista de modalidades simplificadas do saber dos centros hegemônicos; iii) surge em torno do pensamento científico emancipador como uma expressão particular de uma autárquica que tem como correspondente a necessidade popular; iv) crescimento e aprofundamento especializados de revolução filosófica que esteve na periferia dos campos técnicos.

**EPIDEMIOLOGIA MOLECULAR:** Estudo da disseminação de fragmentos de material genético que tenha relevância para a compreensão do comportamento de determinados agravos na comunidade.

**EPIDEMIOLOGIA PSIQUIÁTRICA:** Método epidemiológico aplicado à psiquiatria visando: i) ao diagnóstico comunitário; ii) ao estudo de funcionamento dos serviços de saúde mental, formais e informais; iii) à estrutura dos quadros clínicos e das síndromes mais comuns na área; iv) à estimativa das populações de maior risco: adolescentes, idosos, minorias, trabalhadores em função estigmatizada, trabalhadores em função muito insalubre, trabalhadores sem produto; (v) aos processos de determinação em saúde mental.

**EPIDEMIOLOGIA SOCIAL:** deve ser um conjunto de conceitos, métodos e formas de ação prática que se aplicam ao conhecimento e transformação do processo saúde-doença na dimensão coletiva ou social.

**EPSILON ( $\epsilon$ ):** Se for desejado corrigir o teste F univariado utiliza-se o epsilon de Huynh-Feldt ou de Greenhouse-Geisser. Lembrando que F é a razão entre os quadrados médios entre grupos e dentro dos grupos e que os graus de liberdade entre os grupos é k-1, enquanto que o dentro dos grupos é n-k-1. Para corrigir o valor de F, uma vez que se tenha percebido a falta de esfericidade, deve-se multiplicar o grau de liberdade entre os grupos pelo valor de epsilon. Para violações mais severas da esfericidade ( $\epsilon < 0,75$ ) o epsilon de Greenhouse-Geisser, mais conservador, deve ser utilizado. Em inglês epsilon.

**EQM:** Erro quadrático médio.

**EQS:** Programa de computador alternativo ao LISREL usado para estimar modelos de equações estruturais. O EQS coloca condições menos restritivas quanto à normalidade multivariada dos dados.

**EQUABILIDADE:** Proporção entre a diversidade observada e a máxima diversidade, configurando-se uma medida da uniformidade da distribuição de abundância entre as espécies de uma comunidade. Unidade: grandeza adimensional.

**EQUAÇÃO DA MÉDIA PONDERADA:** Utilizada no cálculo do valor da variável de interesse como média ponderada dos valores de amostras vizinhas:

$$V = \frac{\sum_{i=1}^n V_i P_i}{\sum_{i=1}^n P_i}, \text{ onde: } V_i \text{ é o valor da } i\text{-ésima amostra; } P_i \text{ é o ponderador associado a } i\text{-ésima amostra.}$$

**EQUAÇÃO DAS DIFERENÇAS, MÉTODO DA:** Método de construção de curvas de índice de sítio, que necessita de dados de parcelas com remedições. Parte da curva da altura média das árvores dominantes em função da idade do povoamento e, utilizando duas ocasiões de medição, faz-se a diferença (modelos lineares) ou a razão (modelos não-lineares) das equações nas duas ocasiões. A expressão obtida é então reorganizada de modo a expressar a altura na segunda ocasião de medição em função da altura na primeira ocasião e das idades na primeira e segunda ocasião de medição. Os parâmetros do modelo são estimados via regressão linear ou não-linear utilizando a expressão da diferença ou razão obtida.

**EQUAÇÃO DE AMOSTRAGEM DE GY:** É a relação que determina a massa da amostra em função das suas características mineralógicas, granulométricas, de liberação e de forma das partículas, do diâmetro das partículas e da variância de amostragem. A equação de amostragem de Gy, é:

$$V = \frac{\sum_{i=1}^n V_i \cdot P_i}{\sum_{i=1}^n P_i}, \text{ onde: } M \text{ é a massa requerida em gramas; } C \text{ é a constante de amostragem do material a ser}$$

analisado, expressa em g/cm<sup>3</sup>, calculado a partir de quatro parâmetros, os quais são característicos do material:  $C = f \cdot g \cdot l \cdot m$   $f$  é o fator de forma igual a 0,5 em todos os casos práticos, exceto para ouro, onde é igual a 0,2;  $g$  é o fator de distribuição granulométrica, geralmente igual a 0,25;  $l$  é o fator de liberação, com valores entre 0 e 1, sendo 0 para materiais completamente homogêneos e 1 para materiais completamente heterogêneos;  $m$  é o fator de composição mineralógica, expresso em g/cm, que pode ser calculado como segue:

$$V = \frac{\sum_{i=1}^n V_i P_i}{\sum_{i=1}^n P_i}, \text{ ou } m = \frac{r}{a} \text{ para teores baixos; onde } r \text{ e } t \text{ são as densidades médias do mineral de minério}$$

e do mineral de ganga, respectivamente;  $a$  é a concentração média do mineral de minério no material amostrado;  $d$  é a dimensão em cm da maior partícula do material amostrado;  $s$  é a medida do erro fundamental cometido na amostragem, ou o erro que pode ser tolerado no resultado analítico da amostra. O fator de liberação tem sido determinado a partir de vários estudos práticos conforme a tabela abaixo.

Valores do fator de liberação  $l$ , em função da razão entre o diâmetro da maior partícula na amostra ( $d$ ) e o diâmetro de liberação do minério ( $L$ ).



1	0,8	0,4	0,2	0,1	0,05	0,02
d/L	1	4	10	40	100	

**EQUAÇÃO DE BIOMASSA:** Modelo empírico utilizado para predição da biomassa de árvores individuais, geralmente em função do DAP e da altura.

**EQUAÇÃO DE DIFERENÇAS:** É uma equação que envolve as diferenças entre os sucessivos valores de uma função de variável inteira. Uma equação deste tipo pode ser entendida como a versão discreta de uma equação diferencial. Por exemplo, a equação  $f(n+1)-f(n)=g(n)$  é uma versão discreta da equação diferencial  $f'(x)=g(x)$ . Em inglês difference equation.

**EQUAÇÕES DIFERENCIAIS (ORDINÁRIAS):** Quando é dada a derivada  $y' = f'(t)$  de uma função desconhecida  $y = f(t)$ , normalmente se tem que encontrar a antiderivada. Às vezes, a derivada  $y'$  não é dada como uma função de t, mas está envolvida em uma equação que contém também a função desconhecida  $y = f(t)$ . Como exemplo pode-se ter a seguinte equação  $y' = ay + b + c$  com coeficientes conhecidos a, b e c. Tal equação é chamada equação diferencial, já que ela contém não somente a função desconhecida, mas também sua derivada. O problema consiste em se encontrar uma função apropriada, que satisfaça à equação diferencial. As equações diferenciais ocorrem frequentemente na pesquisa biológica em análise de sistemas fisiológicos e de sistemas ecológicos. Na análise de sistemas quando uma quantidade varia em uma parte de um sistema, sua taxa de variação normalmente depende de quantidades em outras partes do sistema. Além disso, qualquer variação de uma quantidade pode influenciar indiretamente a própria quantidade, sendo este fenômeno chamado feedback ou retroalimentação. O estudo dos sistemas de feedback originou-se na engenharia, mas sua aplicação nas pesquisas biológicas e outras tornou-se mais produtiva. Para uma melhor compreensão das equações diferenciais e de suas soluções veja um mecanismo conceitual. Interpreta-se  $y' = \frac{dy}{dt}$  como um coeficiente angular em um sistema t,y de coordenadas

retangulares, então, uma equação diferencial dada associa um coeficiente angular  $y'$  a cada ponto (t,y). Como um exemplo considere a equação  $y' = y - t^2$ . Aqui  $y'$  está univocamente associado com um ponto (t,y). Portanto, pode-se traçar uma reta por cada ponto (t,y), com coeficiente angular determinado por  $y'$ . Então tem-se o plano repleto de coeficientes angulares ou direções. Este plano é chamado um campo de inclinações ou de direções. A tarefa de solucionar uma equação diferencial é frequentemente chamada integração de uma equação diferencial. Uma solução é também chamada uma integral e o gráfico de tal integral é dito uma curva integral. Consequentemente, as linhas de correntes e curvas integrais significam a mesma coisa. Uma das equações diferenciais mais simples é dada por  $\frac{dy}{dt} = a'y$  onde a é

uma constante dada. A integração normalmente é feita por um procedimento um tanto simbólico. Sabe-se que  $\frac{dy}{dt}$  é o limite do quociente de diferenças  $\frac{\Delta y}{\Delta t}$ , quando  $\Delta t \rightarrow \infty$ . Entretanto, não se define  $\frac{dy}{dt}$

como um quociente de duas quantidades  $dy$  e  $dt$ . Trata-se  $dy$  e  $dt$  meramente como símbolos. Agora, negligenciando este fato, multiplica-se a equação  $\frac{dy}{dt} = a'y$  por  $dt$  e tem-se simbolicamente  $dy = ay \cdot dt$ .

Em seguida, dividindo-se por y, a equação se torna  $\frac{dy}{y} = a dt$  ( $y \neq 0$ ). aqui a variável y ocorre somente no

primeiro membro e t somente no segundo membro. Diz-se que separou-se as variáveis. A integração da  $\int \frac{dy}{y} = \int a dt$  ou, de acordo com a fórmula  $\int \frac{1}{x} dx = \ln|x| + C$  ( $x \neq 0$ ).

Ou seja,  $\ln|y| = at + C$  ( $y \neq 0$ ) onde C é uma constante arbitrária. Pode-se remover o logaritmo natural, aplicando a função inversa. Consequentemente, a solução explícita da equação  $\frac{dy}{dt} = a'y$  é  $|y| = e^{at+c}$ .

Entretanto, a solução é raramente escrita nesta forma. Já que  $e^{u+v} = e^u \cdot e^v$ , e já que y pode tomar valores positivos e negativos, reescreve-se a solução na forma  $y = c \cdot e^{at}$  ou  $y = c \cdot \exp(at)$  onde c substitui  $\pm e^c$ . A solução é uma função exponencial com coeficiente a dado e constante arbitrária c. Já que c é indeterminado, chama-se  $y = c \cdot e^{at}$  ou  $y = c \cdot \exp(at)$  de solução geral da equação diferencial  $y' = ay$ . Como foi obtido a solução por um método simbólico, o resultado deve ser verificado. De  $y = c \cdot e^{at}$  ou  $y = c \cdot \exp(at)$  tem-se por derivação, que  $\frac{dy}{y} = cae^{at}$ .

Que pode ser reescrita na forma  $\frac{dy}{dt} = ay$ . Consequentemente, a equação  $y = c \cdot e^{at}$  ou  $y = c \cdot \exp(at)$

satisfaz realmente a equação diferencial  $\frac{dy}{dt} = a'y$ . Essas equações podem ser aplicadas em experimentos

envolvendo crescimento de uma célula, processo de nascimento, processo de nascimento e morte, desintegração radioativa, tecido vivo exposto à radiação ionizante, traçador radioativo, diluição de uma substância, cinética química, etc.

**EQUAÇÕES DIFERENCIAIS DE PRIMEIRA ORDEM:** É aquela onde ocorre somente a primeira derivada.

**EQUAÇÕES DIFERENCIAIS DE SEGUNDA ORDEM:** É aquela onde ocorre a segunda derivada de uma função desconhecida.

**EQUAÇÃO DIFERENCIAL EXPLÍCITA:** É aquela onde a derivada de ordem mais elevada for dada como uma função de todas as outras variáveis.

**EQUAÇÃO DIFERENCIAL IMPLÍCITA:** É aquela onde a derivada de ordem mais elevada não for dada como uma função de todas as outras variáveis.

**EQUAÇÃO DIFERENCIAL LINEAR:** É aquela que é escrita como um agregado linear igualando a zero. Como exemplo considere o seguinte:  $\frac{dy}{dt} + py + q = 0$ . Aqui p e q são constantes ou são funções de t, isto

é,  $p = p(t)$  e  $q = q(t)$ . A equação de segunda ordem  $y'' = ay' + by + c$  é também linear se a, b, c são constantes ou são funções conhecidas da variável independente.

**EQUAÇÃO DIFERENCIAL NÃO LINEAR:** É aquela que não é escrita como um agregado linear igualando a zero.

**EQUAÇÃO DIFERENCIAL NÃO LINEAR HOMOGÊNEA:** É aquela que não é escrita como um agregado linear igualando a zero e se o termo particular que não contém a função desconhecida estiver ausente. Então  $y' = ay + c$  ou  $y'' = ay' + by + c$  são homogêneas se, e somente se, C=0.

**EQUAÇÃO DIFERENCIAL NÃO LINEAR HETEROGÊNEA:** É aquela que não é escrita como um agregado linear igualando a zero e se o termo particular que não contém a função desconhecida estiver presente. Então  $y' = ay + c$  ou  $y'' = ay' + by + c$  são homogêneas se, e somente se, C=0.

**EQUAÇÃO DIFERENCIAL PARCIAL:** É aquela que contém as chamadas derivadas parciais.

**EQUAÇÃO DIFERENCIAL ORDINÁRIA:** É aquela que não contém as chamadas derivadas parciais.

**EQUAÇÃO DE ESTIMAÇÃO NÃO-VIESADA:** Diz-se a que exprime o estimador de um parâmetro em função de estimadores não-viesados dos parâmetros da respectiva população, o que não impede que a sua solução seja um estimador viesado.

**EQUAÇÃO DE FORMA:** Modelo empírico que expressa o diâmetro do tronco da árvore em função da altura ao longo do tronco que é a posição no tronco a partir do solo, DAP e altura total da árvore. Geralmente a equação de forma é ajustada como a razão da área transversal a diferentes alturas pela área transversal à altura do peito.

**EQUAÇÃO DE LAPLACE:** É uma equação diferencial parcial. Em duas dimensões ela é dada por:  $\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = 0$  e em três dimensões por:  $\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} = 0$ . Em inglês Laplace's equation.

**EQUAÇÃO DE PRODUÇÃO:** Modelo empírico que realiza a predição da produção por unidade de área) do povoamento florestal diretamente em função de outros atributos do povoamento, como por exemplo: área basal, árvores por hectare e índice de sítio.

**EQUAÇÃO DE REGRESSÃO:** Equação real ou aproximada que liga uma ou mais variáveis quantitativas como, por exemplo, dose de nutriente, espaçamento, dentre outras. Aquela que se quer estimar, como por exemplo, produção de milho, altura das plantas, dentre outras.

**EQUAÇÃO DE REGRESSÃO:** Equação gerada para explicar e prever os valores de uma variável independente com base nos valores de uma ou mais variáveis dependentes. Se Y for sempre o dobro de X, a equação de regressão seria  $y = 2x$ . A maioria das equações de regressão é um pouco mais complexa.

**EQUAÇÃO DE REGRESSÃO:** Equação que fornece a estimativa dos escores de uma variável, desde que exista correlação entre ambas. Quando as estimativas partem de diversas variáveis, os escores são preditos pela equação de regressão múltipla. Em inglês regression equation.

**EQUAÇÃO DE REGRESSÃO LINEAR MÚLTIPLA ESTIMADA:** A estimativa da equação de regressão múltipla baseada em dados amostrais e no método dos mínimos quadrados.

**EQUAÇÃO DE REGRESSÃO LINEAR SIMPLES ESTIMADA:** A estimativa da equação de regressão desenvolvida a partir de dados da amostra usando o método dos mínimos quadrados; .

**EQUAÇÃO DE REGRESSÃO LINEAR SIMPLES:** A equação que descreve como o valor médio ou esperado da variável dependente está relacionado com a variável independente; .

**EQUAÇÃO DE REGRESSÃO MÚLTIPLA:** Equação que traduz uma relação linear entre uma variável dependente y e duas ou mais variáveis independentes ( $x_1, x_2, \dots, x_k$ ); .

**EQUAÇÃO DE REGRESSÃO:** Equação algébrica que descreve o relacionamento entre variáveis.

**EQUAÇÃO DE VOLUME:** Modelo empírico para a predição do volume de árvores individuais a partir de atributos de fácil medição na árvore, geralmente diâmetro à altura do peito (DAP) e altura, mas também pode incluir um quociente de forma ou uma segunda medida do diâmetro do tronco acima da altura do

peito. A equação de volume local ou de simples entrada é aquela cujo volume é função apenas do DAP. Na equação de volume de dupla entrada, o volume é função do DAP e da altura. Já na equação de volume por classe de forma, volume é função do DAP, da altura e de uma medida de forma, geralmente um quociente de forma ou uma segunda medida de diâmetro do tronco.

**EQUAÇÃO DIFERENCIAL:** Formulação matemática correspondente a um modelo temporal contínuo. Equação envolvendo derivadas.

**EQUAÇÃO DIFERENCIAL:** Formulação matemática correspondente a um modelo temporal discreto.

**EQUAÇÃO DIFERENCIAL:** É uma equação que envolve as derivadas de primeira ordem ou acima de uma função. Se a equação envolve apenas as derivadas de primeira ordem então ela é denominada linear. Se a equação envolver apenas uma variável então ela é denominada de equação diferencial ordinária. Equações envolvendo mais de uma variável são rotuladas de equações diferenciais parciais. Em inglês Differential equation.

**EQUAÇÃO DIFERENCIAL PARCIAL:** É uma equação que envolve derivadas com respeito a mais do que uma variável. Muitas das equações utilizadas na física são deste tipo. Em inglês Partial differential equation.

**EQUAÇÕES BÁSICAS DE RESERVAS:** Para determinação da reserva em peso, podem ser:

$$V = \frac{\sum_{i=1}^n V_i P_i}{\sum_{i=1}^n P_i}, \text{ onde: } R \text{ é a reserva em peso; } V \text{ é o volume; } D \text{ é a densidade aparente; } T \text{ é o teor em porcen-}$$

tagem; A é a área superficial; E é a espessura, para depósitos em que a densidade aparente pode ser determinada, e:

$$V = \frac{\sum_{i=1}^n V_i P_i}{\sum_{i=1}^n P_i}, \text{ onde: } T \text{ é o teor expresso em peso por unidade de volume, para depósitos em que a densi-}$$

dade aparente não pode ser determinada.

**EQUAÇÃO DIOFANTINA:** É uma equação que deve apresentar apenas soluções inteiras. Em inglês Diophantine equation.

**EQUAÇÕES ÍNDICE DE VOLUME COMERCIAL, SISTEMA DE:** Sistema de equações para realizar predição do sortimento, composto de três equações: i) equação de volume total, ii) equação da razão do volume comercial em função da razão dos diâmetros, e iii) equação da razão do volume comercial em função da razão das alturas. A razão do volume comercial é sempre a razão do volume comercial até um dado diâmetro ou dada altura pelo volume total da árvore. A razão dos diâmetros é a razão do diâmetro em diferentes alturas do tronco pelo diâmetro à altura do peito (DAP). A razão das alturas é da diferença da dada altura ao longo do tronco pela altura total, dividida pela altura total.

**EQUAÇÕES NORMAIS:** Conjunto de equações simultâneas construídas para a estimação de parâmetros via método de mínimos quadrados. Em inglês Normal equations.

**EQUIDADE:** Distribuição justa de determinado atributo junto com a eficiência, a liberdade da escolha pelo consumidor e a maximização da saúde. A equidade é um dos objetivos mais importantes seguidos

pelos sistemas de saúde modernos. O conceito de equidade não necessariamente equivalente à igualdade, embora os termos sejam por vezes usados como sinônimos. Quando se define equidade na presença da saúde em termos de igualdade, o termo envolve duas dimensões importantes: a equidade horizontal, que é o tratamento igual de indivíduos que se encontram numa situação de saúde igual, e a equidade vertical, que se refere ao tratamento apropriadamente desigual de indivíduos em situações de saúde distintas.

**EQUILÍBRIO:** Em teoria de filas, sistemas em equilíbrio estatístico são aqueles em que o número de clientes ou itens em espera oscila de forma que a média e a distribuição permanecem constantes durante um longo período. Em inglês Equilibrium.

**EQUILÍBRIO DE HARDY-WEINBERG:** Uma população encontra-se em equilíbrio quando suas frequências gênicas e genotípicas não se alteram ao longo das sucessivas gerações.

**EQUITABILIDADE:** Ver equabilidade.

**ERGODICIDADE:** Geralmente, esta palavra denota uma propriedade de certos sistemas que se desenvolvem ao longo do tempo de acordo com leis probabilísticas. Sob certas circunstâncias um sistema tenderá em probabilidade a uma forma limite que é independente da posição inicial. Esta é a propriedade de ergodicidade. Em inglês Ergodicity.

**ERGÓDICO:** Termo usado em vários sentidos, para descrever um processo em que uma única realização normalmente propicia informação suficiente com referência ao processo. Um processo pode ser ergódico em muitos aspectos, tais como ergódico na média ou ergódico na distribuição.

**ERRO:** É definido como a diferença entre o valor observado (O) e o valor aceito (A).

**ERRO:** É o mesmo que desvio ou afastamento. Por diversas vezes este termo é considerado sinônimo de resíduo.

**ERRO:** Mesmo que desvio. Em inúmeras vezes este termo é considerado como sinônimo de resíduo.

**ERRO:** Em geral, um erro no sentido coloquial. Em um sentido mais limitado, erro é usado em estatística para denotar a diferença entre um valor observado e seu verdadeiro valor. Em inglês Error.

**ERRO:** Termo aleatório incluído na regressão e na análise da variância, que representa os efeitos de todos os fatores outros que não os especificamente incluídos na análise.

**ERRO:** Um evento randômico como, por exemplo, registro mal feito, soma incorreta, dentre outros, que tende a aumentar a variabilidade de escores de forma assistemática. Uma solução típica ao problema do erro consiste em aumentar o tamanho da amostra, uma vez que isto teria o efeito de assegurar o cancelamento de erros entre grupos.

**ERRO ABSOLUTO:** O erro absoluto de uma observação  $x$  é o desvio absoluto de  $x$  do seu valor real. Em inglês absolute error.

**ERRO ALEATÓRIO:** Produz achados que são demasiadamente altos ou demasiadamente baixos em quantidades aproximadamente iguais devido a fatores casuais.

**EFEITO:** Uma medida da comparação de interesse. Consultar também efeito de tratamento.

**EFEITO DE MODIFICAÇÃO:** Consultar interação.

**EFEITO DE TRATAMENTO:** Uma especificação do parâmetro da comparação do tratamento de interesse; por exemplo, a diferença nas médias entre o grupo tratado e o grupo de controle.

**EFEITO FIXO:** Os níveis ou categorias do fator de interesse compreendem toda a população como por exemplo, tratamentos em um teste clínico.

**EFEITO PLACEBO:** A resposta induzida por sugestões em parte dos tratadores dos animais ou dos pesquisadores, quando o animal recebe uma imitação (dummy) do tratamento.

**EFEITOS ALEATÓRIOS:** Os níveis do fator de interesse representam membros individuais de uma amostra da população de interesse como por exemplo, suínos individuais amostrados de uma população de suínos Gloucester Old-Spot.

**ELEMENTO:** Um único objeto ou indivíduo ou unidade de investigação.

**EQUAÇÃO DE REGRESSÃO LINEAR:** Descreve a relação linear entre duas variáveis quando uma variável é dependente da outra.

**EQUAÇÃO DE REGRESSÃO LINEAR MÚLTIPLA:** Uma forma de análise multivariada, normalmente conhecida como regressão múltipla, a qual é uma expressão matemática que descreve a relação linear entre duas ou mais variáveis explicativas e uma variável dependente.

**EQUAÇÕES DE ESTIMATIVA GENERALIZADA (EEG):** Utilizadas para estimar os parâmetros e seus erros padrão em um modelo de regressão que representa uma estrutura hierárquica de dois níveis; não assumem uma distribuição de probabilidade específica para efeitos aleatórios.

**ERRO AMOSTRAL:** A parte do erro da estimativa do parâmetro populacional que está presente, pois tomamos apenas uma amostra de observações da população, não toda a população.

**ERRO HUMANO:** Variabilidade nas mensurações em razão de erro humano.

**ERRO PADRÃO DA ESTIMATIVA:** Uma medida da precisão da estatística da amostra como uma estimativa do parâmetro da população. É igual ao desvio-padrão da distribuição amostral da estatística.

**ERRO PADRÃO DA MÉDIA (EPM):** O desvio-padrão da distribuição amostral da média; é uma medida de dispersão das médias amostrais e da precisão da média amostral como uma estimativa da média da população.

**ERRO PADRÃO DA MENSURAÇÃO:** Uma estimativa da variabilidade das medidas individuais em um estudo de repetição; é igual à raiz quadrada de metade da variância das diferenças entre os pares de medidas.

**ERRO PROPORCIONAL OU DE ESCALA:** Um erro técnico em que a magnitude da imprecisão da mensuração aumenta ou diminui de acordo com a magnitude dos valores.

**ERRO SISTEMÁTICO:** Quando o valor registrado encontra-se sistematicamente acima ou abaixo de seu valor verdadeiro.

**ERRO TIPO I:** Rejeitamos a hipótese nula quando não deve ser rejeitada, ou seja, quando é verdadeira. A probabilidade máxima de cometer um erro Tipo I é o nível de significância do teste.



**ERRO TIPO II:** Quando não rejeitamos a hipótese nula quando esta deveria ter sido rejeitada, ou seja, quando é falsa.

**ERRO ZERO:** Um erro técnico em que o instrumento falha ao registrar uma leitura de zero verdadeira.

**ERROS ALEATÓRIOS:** Os valores registrados estão igualmente distribuídos acima e abaixo do valor verdadeiro; isso se deve a fontes inexplicáveis.

**ERROS DE ARREDONDAMENTO:** Imprecisões que ocorrem em razão do arredondamento do número para um número menor de casas decimais ou algarismos relevantes.

**ERROS INSTRUMENTAIS:** Imprecisões introduzidas por mecanismos mecânicos ou eletrônicos utilizados nas mensurações.

**ERROS OU VARIAÇÕES TÉCNICAS:** Variabilidade nas medidas em razão da variedade de causas instrumentais e em razão de erros humanos.

**ERROS PADRÃO ROBUSTOS:** Estimções do erro padrão baseadas mais na variabilidade do conjunto de dados do que assumidas pelo modelo de regressão; portanto, são robustas perante a violação das hipóteses do modelo de regressão.

**ESCALA CONTÍNUA:** Teoricamente todos os valores são possíveis às vezes delimitados por um limite superior e/ou inferior, como por exemplo, altura, peso.

**ESCALA DISCRETA (DESCONTÍNUA):** Os dados podem ser apenas valores inteiros, normalmente contagens; por exemplo, número de filhotes, número de ovos no ninho, número de partições.

**ESCALA NOMINAL:** As categorias distintas que definem a variável são desordenadas e recebem uma classificação, por exemplo, cor do pelame.

**ESCALA ORDINAL:** As categorias que constituem a variável têm alguma ordem intrínseca, mas os intervalos entre as várias categorias não podem ser interpretados de maneira consistente.

**ESCOLHA ALEATÓRIA:** É o mesmo que uma distribuição aleatória. Além disso, um conjunto de objetos é considerado aleatório quando disponibilizados de maneira aleatória.

**ESCOLHA ALEATÓRIA BLOQUEADA:** Um método de distribuição aleatória dos indivíduos nos tratamentos, com objetivo de obter números aproximadamente iguais em cada grupo de tratamento. Também denominada escolha aleatória restrita.

**ESCOLHA ALEATÓRIA ESTRATIFICADA:** Fracionamos a população em diferentes estratos, de acordo com a categorização das variáveis-chave potencialmente de “confusão”; então, dentro de cada camada, distribuimos aleatoriamente os animais em cada um dos grupos de tratamento.

**ESCOLHA ALEATÓRIA POR AGRUPAMENTO:** Consultar escolha aleatória por grupo.

**ESCOLHA ALEATÓRIA POR GRUPO:** A escolha aleatória é aplicada a todo o grupo de unidades, não apenas às unidades individuais dentro de um grupo. Também conhecida como escolha aleatória por agrupamento.

**ESCOLHA ALEATÓRIA RESTRITA:** Consultar escolha aleatória bloqueada.

**ESCOLHA ALEATÓRIA SIMPLES:** Distribuimos os animais aleatoriamente nos diferentes grupos de tratamento, sem utilizar quaisquer refinamentos ou restrições.

**ESPECIFICIDADE:** A efetividade de um teste diagnóstico ou de triagem para identificar animais que não estão doentes. É a proporção de negativos verdadeiros identificados pelo teste como negativos.

**ESTABILIDADE:** Está relacionada à repetição de medidas a longo prazo.

**ESTATÍSTICA:** i) É uma estimativa amostral de um parâmetro da população. Às vezes denominada estatística amostral; uma vez que uma estatística sempre está relacionada com a amostral, a palavra amostra é redundante; ii) Definida de maneira simples, é a habilidade em manipular dados e análises; geralmente em ciências biológicas implica em um procedimento numérico mais amplo à ciência (consultar biometria). Em inglês *Statistic*

**ESTATÍSTICA DA RAZÃO DE VEROSSIMILHANÇA:** Utiliza a razão de duas verossimilhanças para comparar dois modelos de regressão.

**ESTATÍSTICA INFERENCIAL:** Conjunto de métodos e procedimentos utilizados para se realizar inferências ou extrapolação acerca de uma população ou universo estatístico, com base em estudos e avaliações obtidos numa amostra representativa dessa mesma população.

**ESTATÍSTICAS DESCRITIVAS:** A parte da estatística que descreve e caracteriza a distribuição de dados e resume e expõe todos os achados.

**ESTATÍSTICAS INFERENCIAIS:** O ramo da estatística que envolve a obtenção de conclusões acerca de uma população utilizando as informações da amostra.

**ESTIMAÇÃO:** O procedimento que fornece uma aproximação de um parâmetro da população utilizando dados da amostra.

**ESTIMAÇÃO DA VEROSSIMILHANÇA MÁXIMA (EVM):** Um procedimento que propicia estimativas dos parâmetros em um modelo, pela maximização da verossimilhança.

**ESTIMATIVA NÃO-TENDENCIOSA DO PARÂMETRO DA POPULAÇÃO:** A média da distribuição amostral da estatística da amostra coincide com o parâmetro da população da qual a estatística está sendo estimada.

**ESTUDO COORTE:** Uma forma de estudo observacional. Iniciamos definindo os grupos (coortes) de animais pela exposição dos animais dos grupos aos fatores de interesse; normalmente acompanhamos esses animais durante certo tempo e observamos o resultado, por exemplo, doença.

**ESTUDO CRUZADO:** Estudo em que se utilizam todas as medidas dos indivíduos incluídos no estudo, simultaneamente.

**ESTUDO DE CASO CONTROLE:** Uma forma de estudo observacional. No início da investigação, identificamos os animais como sendo doentes que são os casos ou saudáveis que são os controles. Então, avaliamos se os animais dos dois grupos têm diferenças depois da exposição aos vários fatores de risco.

**ESTUDO DE EQUIVALÊNCIA:** Utilizado para mostrar que a eficácia clínica de um tratamento é similar àquele do tratamento existente.



**ESTUDO DE INFERIORIDADE:** Utilizado para mostrar que a eficácia clínica de um tratamento não é pior do que aquela do tratamento já existente.

**ESTUDO DE NÃO-INFERIORIDADE:** Utilizado para mostrar que a eficácia clínica de um tratamento não é substancialmente pior do que aquela do tratamento existente.

**ESTUDO EPIDEMIOLÓGICO:** Envolve a investigação da etiologia de uma doença, verificando se vários fatores conhecidos como fatores de risco estão associados com a ocorrência e distribuição da doença na população.

**ESTUDO EXPERIMENTAL:** Intervimos no estudo; em seguida, observamos o efeito de nossa intervenção na resposta de interesse, normalmente com objetivo de estabelecer se uma mudança na resposta pode ser diretamente atribuível a nossa ação.

**ESTUDO LONGITUDINAL:** Permite avaliar as alterações nos mesmos indivíduos com o passar do tempo.

**ESTUDO LONGITUDINAL PROSPECTIVO:** O estudo é conduzido ao longo do tempo, a partir de um ponto de partida definido.

**ESTUDO LONGITUDINAL RETROSPECTIVO:** O estudo é conduzido observando-se o passado a partir de um ponto de partida definido.

**ESTUDO OBSERVACIONAL:** Apenas observamos os animais no estudo e registramos as medidas relevantes desses animais; não tentamos interferir, por exemplo, administrando medicamentos ou excluindo fatores que achamos que possam influenciar o curso da enfermidade.

**ESTUDO SEQUENCIAL:** O tamanho da amostra não é definido previamente, mas depende dos resultados, à medida que se tornam disponíveis. Deve-se planejar uma abordagem formal à análise; leva em consideração o nível de significância e a precisão e depende das regras de interrupção, as quais permitem a finalização do teste em favor de um determinado tratamento, caso algumas condições sejam atendidas.

**ESTUDO-PILOTO:** Investigação preliminar de pequena escala.

**ESTUDOS DE BIOEQUIVALÊNCIA:** Utilizados para mostrar que duas formulações de um medicamento têm biodisponibilidade semelhante quando cada formulação propicia a mesma quantidade de medicamento no organismo.

**EXATIDÃO:** Refere-se ao quão bem os valores observados de determinada quantidade estão de acordo com o valor verdadeiro.

**EXPERIMENTOS LABORATORIAIS:** Uma forma particular de estudo experimental realizado em laboratórios, no qual a intervenção experimental é altamente regulada e controlada.

**ERRO DE ESTIMAÇÃO:** É a diferença entre o verdadeiro valor do parâmetro e o valor calculado a partir do dados de uma amostra. Este depende diretamente do grau de dispersão ou variabilidade da variável em estudo e inversamente ao tamanho da amostra.

**ERRO ALEATÓRIO:** Erro que afeta a mensuração de maneira transitória e inconsistente.

**ERRO ALEATÓRIO:** Erro de pesquisa indutiva, associado à variabilidade amostral, uma vez que se estuda apenas uma amostra de um todo, e os resultados podem diferir de uma amostra para outra. Sinônimo: erro amostral.

**ERRO ALEATÓRIO:** Veja erro analítico.

**ERRO ALEATÓRIO OU AMOSTRAGEM ALEATÓRIA:** Erro que resulta de uma variação ao acaso.

**ERRO ALFA:** Ou erro do tipo I, ou de primeira espécie, ou ainda nível de significância de um teste de hipótese ou de significância.

**ERRO ALFA:** O mesmo que erro tipo I, o qual consiste em rejeitar a hipótese  $H_0$  quando na realidade ela for verdadeira.

**ERRO AMOSTRAL:** Viés observado nas amostras aleatórias, decorrente da natural variabilidade dos elementos constituintes das populações, assinalando-se o fato de que nem todas as unidades do universo participam da amostra. O erro amostral é reduzido pelo aumento do tamanho amostral e medido pelo erro padrão.

**ERRO AMOSTRAL:** Resto da subtração que tem por minuendo o valor de uma característica populacional e, para subtraendo, uma sua estimativa. Também se diz erro de amostragem.

**ERRO AMOSTRAL:** Diferença entre um resultado amostral e o verdadeiro resultado populacional; resultados de flutuações amostrais aleatórias.

**ERRO AMOSTRAL:** A diferença entre o valor da amostra e o valor real da média da população.

**ERRO BETA:** Erro do tipo II ou probabilidade de aceitar uma hipótese de nulidade quando na realidade ela for falsa..

**ERRO BETA:** O mesmo que erro do tipo II.

**ERRO DE AMOSTRA:** Um erro de amostra é uma diferença entre um valor de população, ou parâmetro, e um dado valor de amostra. Se  $\mu$  for o valor da média da população e  $\bar{X}_i$ , uma estimativa baseada em uma amostra de tamanho  $n$ , então, a diferença será  $\mu - \bar{X}_i = e_i$ , em que  $e_i$  é o erro de amostra.

**ERRO DE AMOSTRAGEM:** Erro na estimativa do parâmetro populacional. É determinado pelo cálculo do intervalo de confiança.

**ERRO DE AMOSTRAGEM:** Associado a uma amostra de rocha representa a diferença composicional entre a amostra de rocha e a parte do corpo de rocha que a amostra é esperada para representá-la.

**ERRO DE AMOSTRAGEM:** A probabilidade de qualquer amostra extraída cientificamente conter certas diferenças inevitáveis em relação à população da qual ela faz parte.

**ERRO DE AMOSTRAGEM:** O mesmo que erro amostral.

**ERRO DE AMOSTRAGEM:** Diferenças entre os escores obtidos numa amostra que são as estatísticas e os valores da população que são os parâmetros. Se um escore for usado como estimador dos escores de toda a população, a diferença entre esse escore e a média de todos os escores registrados seria um erro de amostragem. Em inglês *sampling error*.

**ERRO DE AMOSTRAGEM:** O valor absoluto da diferença entre um estimador por ponto sem viés e o parâmetro da população correspondente. Para uma média da amostra, o desvio-padrão da amostra e a proporção da amostra, os erros de amostragem são  $e$  e  $p$ , respectivamente.

**ERRO DE AMOSTRAGEM PERMITIDO:** A quantidade de erro de amostragem que um pesquisador está disposto a aceitar.

**ERRO ANALÍTICO:** É o erro de determinação da quantidade real do elemento analisado existente na amostra. As causas deste tipo de erro, podem ser associadas a: i) imperfeição intrínseca do objeto a ser medido, neste caso, entenda-se pela variabilidade natural do material amostrado e inadequadamente preparado; ii) métodos e procedimentos impróprios; iii) condições do operador: acuidade visual, atenção, treinamento, honestidade, dentre outros; iv) condições do ambiente: pressão, campos elétricos, temperatura, umidade, dentre outros; v) variações causadas pela passagem do tempo. Os erros analíticos podem ser classificados em dois tipos: i) erros sistemáticos; ii) erros aleatórios. Os erros sistemáticos são introduzidos principalmente por imperfeições nos aparelhos de medida, métodos de medição e objeto a ser medido. Como causas de efeitos menores, aparecem condições ambientais e erros do operador. Esses erros apresentam sempre grandeza e sinal definidos, característica que permite introduzir correção para erros detectáveis e mensuráveis. Há outros erros sistemáticos com causas determináveis que podem ser estimados, mas apenas dentro de certos limites. A contribuição proveniente dessa origem deve ser introduzida no resultado do processo de medição. E há ainda erros sistemáticos de natureza indeterminada que podem ser avaliados somente por métodos indiretos ou então por novos processos de medições. O problema da variabilidade do operador, introduzida pelas pessoas que realizam as medidas experimentais, deveria ser sempre avaliada. A variabilidade é introduzida porque diferentes operadores naturalmente farão diferentes observações e ainda o mesmo operador tenderá a alguma inconsistência dia após dia ou ainda durante um mesmo dia. Todos os equipamentos de medida, tais como balanças e balões volumétricos, devem ser calibrados rotineiramente, bem como a pureza dos reagentes controlada, que seguramente irão ao menos atenuar os erros sistemáticos. Os erros aleatórios, são desvios introduzidos por qualquer das causas já mencionadas, apenas sabendo-se que os mecanismos de ação não são detectáveis e influenciáveis. O exemplo clássico é o do mesmo operador usando os mesmos instrumentos, nas mesmas condições, medindo uma mesma propriedade, seguidamente. Os resultados não serão os mesmos, havendo incerteza quanto ao valor real da propriedade. Erros aleatórios variam em grandeza e sinal, e não são determináveis individualmente. O uso da técnica estatística permite assegurar confiabilidade ao resultado final calculado, especialmente quando se aumenta o número de observações. Essa confiabilidade pode ser inclusive válida nos casos em que há variações com o tempo.

**ERRO DE ESPECIFICAÇÃO:** Omissão de uma variável chave da análise, que causa um impacto sobre os efeitos estimados de variáveis incluídas.

**ERRO DE ESPECIFICAÇÃO:** Erro na previsão da variável dependente causado pela exclusão de uma ou mais variáveis independentes relevantes. Essa omissão pode distorcer os coeficientes estimados das variáveis incluídas, bem como diminuir o poder preditivo geral do modelo de regressão.

**ERRO DE ESPECIFICAÇÃO:** Falta de qualidade de ajuste do modelo resultante da omissão de uma variável relevante do modelo proposto. Testes para erro de especificação são bastante complicados e envolvem numerosas tentativas entre modelos alternativos. O pesquisador pode evitar erro de especificação em um

eleva do grau pelo uso de apenas bases teóricas para a construção do modelo proposto. Desse modo, o pesquisador está menos sujeito a desconsiderar um construto relevante para o modelo.

**ERRO DE ESPECIFICAÇÃO DA POPULAÇÃO:** Erro que resulta em definir incorretamente a população ou universo da qual uma amostra é escolhida.

**ERRO DE ESTIMAÇÃO:** É a diferença entre o verdadeiro valor do parâmetro e o valor calculado a partir dos dados de uma amostra. Este depende diretamente do grau de dispersão (variabilidade) da variável em estudo e inversamente ao tamanho da amostra.

**ERRO DE ESTIMATIVA DA MÉDIA:** É a diferença da média amostral pela média populacional:

**ERRO DE ESTRUTURA:** Erro resultante de uma estrutura de amostra imprecisa ou incompleta.

**ERRO DE INFORMAÇÃO SUBSTITUTA:** Erro que resulta de uma discrepância entre a informação necessária para resolver para resolver um problema e aquela procurada pelo pesquisador.

**ERRO DE MEDIDA:** Imprecisão na mensuração dos valores verdadeiros das variáveis devido à falibilidade do instrumento de medida, ou seja, escalas de resposta inapropriadas, erros na entrada de dados ou enganos dos respondentes.

**ERRO DE MEDIDA:** Grau em que os valores dos dados não medem verdadeiramente a característica representada pela variável. Por exemplo, quando se questiona sobre a renda familiar total, há muitas fontes de erro de medida, por exemplo, relutância em responder a quantia total e erro na estimativa da renda total que tornam os valores imprecisos.

**ERRO DE MENSURAÇÃO:** Grau em que as variáveis que podemos medir que são as variáveis manifestas não descrevem perfeitamente o(s) construto(s) latente(s) de interesse. Fontes de erro de mensuração podem variar de simples erros de entrada de dados à definição de construtos, por exemplo, conceitos abstratos tais como patriotismo ou lealdade que significam muitas coisas para diferentes pessoas que não são perfeitamente caracterizados por qualquer conjunto de variáveis manifestas. Para todos os fins práticos, todos os construtos têm algum erro de mensuração, mesmo com as melhores variáveis indicadoras. No entanto, o objetivo do pesquisador é minimizar a quantia de erro de mensuração, sem poder levar em conta erro de mensuração a fim de fornecer estimativas mais precisas das relações causais.

**ERRO DE MENSURAÇÃO:** Erro que resulta de uma variação entre a informação procurada e a que é realmente obtida pelo processo de mensuração.

**ERRO DE PADRÃO DA ESTIMATIVA (SEE):** Medida da variação nos valores previstos que pode ser usada para desenvolver intervalos de confiança em torno de qualquer valor previsto. É análogo ao desvio padrão de uma variável em torno de sua média. SEE é a sigla em inglês para standard error of the estimate.

**ERRO DE PREPARAÇÃO:** É aquele envolvido na redução/homogeneização da amostra para fins de análise, devido a perdas sistemáticas, contaminação, mistura, concentração, quarteamento e alteração. As perdas sistemáticas ocorrem quando as amostras são pulverizadas, perdendo-se a cada operação uma determinada quantidade de material. Tais perdas poderiam ser controladas, tomando-se o peso das amostras a cada etapa da preparação, lembrando que a amostra terá um peso maior por meio da contaminação, ou um peso menor devido à perda. A contaminação pode ocorrer durante a preparação, com a introdução de material estranho à amostra, diminuindo o teor do elemento de interesse ou também

aumentando o teor quando a contaminação for de material mineralizado. A contaminação proposital com o objetivo de aumentar os teores do elemento de interesse nas amostras é chamada salgamento. Uma outra forma de contaminação pode ocorrer durante os processos de britagem/moagem da amostra com a introdução de traços de metais utilizados nas ligas desses equipamentos. A mistura tem por objetivo a homogeneização do material para retirada de quantidade adequada para análise. Entretanto, devido a grande variabilidade dos materiais geológicos é extremamente difícil conseguir uma mistura ideal de todos os constituintes. Para garantir a melhor amostragem nesse caso, a homogeneização deve ser feita com o material reduzido a uma granulação mínima que garanta a distribuição homogênea de todos os constituintes no volume da amostra. O quarteamento é a melhor técnica para redução da quantidade de material. O maior problema desta técnica é garantir que o material dividido apresente a mesma quantidade do elemento de interesse em todas as suas porções, ou seja, igual probabilidade de apresentar o elemento de interesse nas porções quarteadas. A alteração é problema específico a certos minerais que podem reagir durante a preparação que envolve a secagem em estufas, oxidação em contato com o ar, dentre outros, deteriorando o resultado analítico.

**ERRO DE PREVISÃO:** Diferença entre os valores reais e os previstos da variável dependente, para cada observação na amostra. Ver resíduo.

**ERRO DE PROCESSAMENTO:** Erro que resulta da transferência incorreta de informações de um documento ao computador.

**ERRO DE SELEÇÃO:** Erro que resulta de seguir procedimentos incompletos ou inadequados de amostragem ou de não seguir os procedimentos corretos.

**ERRO DE SOMA DE QUADRADOS (SSE):** A variação não explicada pela regressão.

**ERRO DE TRATAMENTO:** Uma relação aparente entre duas variáveis que na verdade é o resultado da interação de uma terceira variável.

**ERRO DIFERENCIAL:** Erro consistente, não-aleatório ou sistemático no qual os valores tendem a não ser acurados em uma determinada direção. O mesmo que viés ou tendenciosidade.

**ERRO DO OBSERVADOR:** O mesmo que variação do observador.

**ERRO FALSO NEGATIVO:** O mesmo que erro tipo II.

**ERRO FALSO POSITIVO:** O mesmo que erro tipo I.

**ERRO GEOMÉTRICO:** É o erro associado a estimativa de reservas devido a interpretação errônea do corpo de minério, a partir dos dados de amostragem. Deve-se salientar que é o erro de maior magnitude, pois afeta diretamente o volume determinado e convertido em tonelagem de minério. As causas desse erro estão sempre associadas à amostragem. Assim, caso a densidade de amostragem não seja satisfatória, o quadro resultante de reservas deverá ser classificado entre inferida ou indicada, deixando a reserva medida para situações em que a densidade de amostragem justificar esta atitude.

**ERRO MÁXIMO DE ESTIMATIVA:** Veja margem de erro.

**ERRO NÃO AMOSTRAL:** Erros conseqüentes de fatores externos não relacionados com a amostragem.

**ERRO NÃO AMOSTRAL:** Imperfeições de qualidade dos dados, resultantes de outros fatores que não erros de amostragem. Exemplos: má compreensão das questões pelos respondentes, registros errôneos por entrevistadores e codificadores, erros de digitação dentre outros.

**ERRO NÃO DIFERENCIAL:** É aquele que produz achados que são demasiadamente altos ou demasiadamente baixos em quantidades aproximadamente iguais, devido a fatores aleatórios.

**ERRO OU DISTORÇÃO DO ENTREVISTADOR:** Erro que resulta de o entrevistador consciente ou inconscientemente influenciar um respondente.

**ERRO PADRÃO:** Desvio padrão de uma distribuição de amostragem.

**ERRO-PADRÃO:** Desvio padrão estimado para uma estatística amostral.

**ERRO PADRÃO:** Desvio padrão da distribuição dos diversos valores de uma estatística, em conseqüência da flutuação das amostras. Se, de uma população, forem tiradas muitas amostras e calculadas as respectivas médias, haverá uma diferença entre cada média e a média da população. Se for computada a média de tais médias, seu desvio padrão será o erro da amostragem que é a margem de erro a ser esperada toda vez que se empregar tal processo de amostragem. Em inglês standard error.

**ERRO PADRÃO:** É o desvio padrão de uma população de médias amostrais, o qual corresponde ao quociente do desvio padrão do universo de onde as amostras foram retiradas pela raiz quadrada do tamanho amostral. Quando se desconhece o desvio padrão paramétrico, utiliza-se no cálculo o desvio padrão da amostra.

**ERRO PADRÃO:** Distribuição esperada de um coeficiente de regressão estimado, por exemplo. O erro padrão é semelhante ao desvio padrão dos valores originais dos dados. Denota a amplitude esperada do coeficiente em múltiplas amostras dos dados. É útil em testes estatísticos de significância que procuram ver se o coeficiente é significativamente diferente de zero, ou seja, se a amplitude esperada do coeficiente contém o valor de zero em um dado nível de confiança, no entanto a frase correta seria ou seja, se a amplitude esperada do coeficiente não contém o valor de zero. O valor t de um coeficiente de regressão é o coeficiente dividido por seu erro padrão.

**ERRO PADRÃO:** Medida da dispersão das médias ou das diferenças das médias esperada devido à variação amostral. O erro padrão é usado no cálculo da estatística t.

**ERRO PADRÃO:** O erro padrão é uma medida da precisão da média amostral calculada. O erro padrão obtém-se dividindo o desvio padrão pela raiz quadrada do tamanho da amostra. Ou seja,  $\frac{\sigma}{\sqrt{n}}$ . Quando

não se conhece o desvio padrão da população, usa-se o desvio padrão da amostra (s) ficando a formula:  $\frac{S}{\sqrt{n}}$ . Se de uma população, com média  $\mu_e$  desvio padrão  $\sigma$ , se retirarem muitas amostras todas do

mesmo tamanho n, e para cada amostra se calcular a respectiva média, a distribuição de todas essas médias é normal com média  $\mu_e$  desvio padrão  $\frac{\sigma}{\sqrt{n}}$ . Assim, o erro padrão não é mais do que o desvio padrão

da distribuição das médias das amostras de uma população.



**ERRO PADRÃO DA MÉDIA:** É o valor que expressa a exatidão da média e é calculado com base no Teorema do Limite Central, em que é por si uma variável aleatória com média  $m$  e desvio padrão  $s$ , quando  $\frac{S^2}{n}$  é pequeno a distribuição é normal e, portanto, significa que a diferença, o erro de estimativa da

média, estará em algum lugar entre mais ou menos.

**ERRO PADRÃO DA MÉDIA:** O desvio-padrão de uma distribuição de médias de amostras.

**ERRO PADRÃO DA MEDIDA:** Erro padrão de qualquer medida, devido às flutuações da amostragem ou aos erros de observação. Quando se trata de um teste, o erro padrão de medida é dado pela fórmula:

$$E = s\sqrt{1-r}, \text{ em que}$$

$E$  = erro padrão da medida

$r$  = coeficiente de precisão ou fidedignidade do teste

O erro padrão da medida varia de zero, quando o coeficiente de precisão é igual a 1, até o valor do desvio padrão dos escores quando o coeficiente de precisão é igual a zero. Em inglês standard error of measurement.

**ERRO PADRÃO DE MEDIDA:** Quando se refere a uma média  $\hat{m}$  de  $n$  parcelas, é o desvio padrão ( $s$ ) dividido por  $\sqrt{n}$ , isto é  $S_{(\bar{x})} = \frac{S}{\sqrt{n}}$ .

**ERRO PADRÃO DE MEDIDA:** É o erro de uma dada coleção de resultados da aplicação de um instrumento ou processo de medir, especialmente de um teste, ou seja, constitui uma estimativa da média quadrática das discrepâncias entre as medidas obtidas e os valores verdadeiros as correspondentes magnitudes.

**ERRO PADRÃO PARA O COEFICIENTE:** Na análise de regressão, é o desvio padrão estimado para o coeficiente estimado; um pequeno valor do erro-padrão significa que o coeficiente estimado é uma estimativa mais precisa do verdadeiro coeficiente.

**ERRO PORCENTUAL:** É o erro definido como uma porcentagem do valor aceito.

**ERRO POST HOC:** depois disto, portanto, devido a isto. Muitos eventos seguem um ao outro sem terem relacionamentos de causa e efeito. As razões para isso variam de simples coincidência a complexos relacionamentos com outros fatores. Por exemplo, um aumento da temperatura no oceano está diretamente relacionado ao número de afogamentos.

**ERRO PROVÁVEL:** É o intervalo dentro do qual metade de uma série de leituras da mesma quantidade provavelmente estarão. Para uma distribuição de Gauss, o erro provável é 0,674 vezes o desvio padrão.

**ERRO QUADRADO MÉDIO:** Em forma geral, o quadrado médio de um conjunto de dados é a média aritmética dos quadrados de suas diferenças com relação a um valor central. Em inglês Mean squared error.

**ERRO QUADRÁTICO MÉDIO (EQM):** Grandeza utilizada na regressão para estimar o valor desconhecido da variância do termo erro.

**ERRO SISTEMÁTICO:** Erro que resulta de uma distorção constante nas mensurações.

**ERRO SISTEMÁTICO:** Veja erro analítico.

**ERRO SISTEMÁTICO OU DISTORÇÃO:** Erro que resulta do projeto de pesquisa ou de execução.

**ERRO TIPO I (ALFA= $\alpha$ ):** Um termo alternativo para erro  $\alpha$  ou Erro de primeiro tipo. Consiste em rejeitar uma hipótese verdadeira. Em inglês Type I error.

**ERRO TIPO I (ALFA= $\alpha$ ):** Probabilidade de rejeitar incorretamente a hipótese nula na maioria dos casos, isso significa dizer que existe uma diferença ou correlação. Quando na verdade não é o caso. Também chamado de alfa. Níveis comuns são de 5% ou 1%, chamados de nível 0,05 ou 0,01, respectivamente.

**ERRO TIPO I (ALFA= $\alpha$ ):** Rejeição da hipótese de nulidade quando ela é verdadeira.

**ERRO TIPO I (ALFA= $\alpha$ ):** Erro que consiste em rejeitar a hipótese nula quando ela é verdadeira.

**ERRO TIPO I:** É o erro que o pesquisador comete em rejeitar a hipótese nula ou de nulidade  $H_0$  quando esta é verdadeira. A probabilidade de cometer este erro é denotada por  $\alpha$  e recebe o nome de erro de primeira espécie ou nível de significância do teste.

**ERRO TIPO I:** Termo usado na estatística inferencial que indica a fração ou nível de significância que representa o risco, que corre o pesquisador, de rejeitar a hipótese de irrelevância, quando ela é de fato verdadeira.

**ERRO TIPO I:** Erro que ocorre quando os dados levam à conclusão de que algo é verdadeiro quando na realidade não é verdadeiro. O erro tipo I é também chamado de erro alfa ou de erro falso-positivo.

**ERRO TIPO I:** Erro que ocorre no teste de hipótese e que consiste em rejeitar a hipótese nula quando ela é verdadeira.

**ERRO DO TIPO I:** O erro associado a uma tomada de decisão com base nos dados da amostra.

**ERRO TIPO I:** Probabilidade de rejeitar a hipótese nula quando ela deveria ser aceita, ou seja, concluir que duas médias são significativamente diferentes quando de fato são a mesma. Valores pequenos de alfa, por exemplo, 0,05 ou 0,01, também denotados como  $\alpha$ , levam à rejeição da hipótese nula e aceitação da hipótese alternativa de que as médias das populações não são iguais.

**ERRO TIPO I:** É a decisão de rejeitar a hipótese nula quando esta é verdadeira. A probabilidade de cometer este erro é denotada por  $\alpha$  e recebe o nome de nível de significância do teste.

**ERRO DO TIPO I:** No teste de hipóteses consiste em rejeitar a hipótese nula quando ela é verdadeira. Em inglês Alpha error.

**ERRO DO TIPO I:** Erro falso-positivo; encontram-se diferenças estatisticamente significativas, nos resultados de um estudo, quando, na verdade, trata-se de diferenças ao acaso; leva a afirmar-se que há associação entre exposição e doença ao acaso; leva a afirmar que há associação entre exposição e doença ou há diferença entre dois tratamentos quando ela, de fato, não existe. A de cometer este erro é designada por alfa ou nível de significância estatística.

**ERRO DO TIPO I:** Quando se aplica um teste de hipótese pode-se cometer dois tipos de erro: rejeitar a hipótese nula quando ela é verdadeira que é o erro do tipo I ou aceitar a hipótese nula quando ela é falsa que é o erro do tipo II. O valor máximo que se estabelece para o erro do tipo I é que aceita-se ou rejeita-se a hipótese nula ( $H_0$ ). ou seja, se o  $p$  for menor do que o valor escolhido rejeita-se a  $H_0$ , caso contrário aceita-se a  $H_0$ . esse valor máximo é escolhido antes de se realizar o teste. o valor mais frequentemente usado é o 0,05.



**ERRO DO TIPO I (ERRO ALFA):** Consiste em rejeitar a hipótese de nulidade quando ela é verdadeira.

**ERRO DO TIPO II:** Erro falso negativo; não se encontram diferenças estatisticamente significativas só variações ao acaso, nos resultados de um estudo, quando, na verdade, tais diferenças são reais; leva a afirmar-se que não há associação entre exposição e doença ou não há diferença entre dois tratamentos quando ela, de fato, existe. A probabilidade de cometer este erro é designada por beta. O complemento de beta, ou seja, 1-beta é denominado poder estatístico do teste.

**ERRO DO TIPO II:** Quando realizamos um teste de hipótese podemos cometer dois tipos de erros: rejeitar a hipótese nula quando ela é verdadeira, denominado erro do tipo I ou aceitar a hipótese nula quando ela é falsa, conhecido como erro do tipo II. O valor máximo que estabelecemos para o erro do tipo I é que estabelece se aceitamos ou rejeitamos a hipótese nula ( $H_0$ ). ou seja, se o p for menor do que o valor escolhido rejeitamos a  $H_0$ , caso contrário aceitamos a  $H_0$ . esse valor máximo é escolhido antes de se realizar o teste. o valor mais frequentemente usado é o 0.05.

**ERRO DO TIPO II:** No teste de hipóteses consiste em aceitar a hipótese nula quando ela é falsa. Em inglês Beta error.

**ERRO DO TIPO II:** Erro que ocorre no teste de hipóteses e que consiste em aceitar a hipótese nula quando ela é falsa.

**ERRO DO TIPO II:** O erro associado ao fato de se ser excessivamente conservador e não agir sobre os dados da amostra.

**ERRO TIPO II:** Erro que ocorre quando os dados levam à conclusão de que algo é falso quando na realidade é verdadeiro.

**ERRO TIPO II:** Termo utilizado na estatística inferencial que consiste na aceitação da hipótese nula, quando ela está errada. Quando se diminui o risco do erro tipo I, aumenta-se o risco de se perder os efeitos reais de variáveis, a menos que se aperfeiçoe a precisão do experimento ou se aumente o número de replicações.

**ERRO NÃO-SISTEMÁTICO:** Erro devido ao acaso.

**ERRO TIPO-II:** É a decisão de aceitar a hipótese nula quando esta é falsa. A probabilidade de cometer este erro é denotada por.

**ERRO TIPO II (BETA=  $\beta$ ):** Um termo alternativo para erro  $\beta$  ou Erro de segundo tipo. . No teste de hipóteses consiste em aceitar uma hipótese de nulidade falsa. Em inglês Type II error.

**ERRO TIPO II (BETA=  $\beta$ ):** Probabilidade de falhar incorretamente na rejeição da hipótese nula – em termos simples, a probabilidade de não encontrar uma correlação ou diferença na média quando ela existe. Também chamado de beta, está inversamente relacionado ao erro tipo I. O valor 1 menos o erro tipo II é definido como poder.

**ERRO TIPO II (BETA=  $\beta$ ):** Aceitação da hipótese de nulidade quando ela é falsa.

**ERRO TIPO II:** Probabilidade de falhar na rejeição da hipótese nula quando ela deveria ser rejeitada, ou seja, concluir que duas médias não são significantemente diferentes quando de fato são. Também conhecido como o erro beta ( $\beta$ ).

**ERRO TIPO II:** É o erro que o pesquisador comete em aceitar a hipótese nula ou de nulidade  $H_0$  quando esta é falsa. A probabilidade de cometer este erro é denotada por  $\beta$  e recebe o nome de erro de segunda espécie.

**ERRO DO TIPO II (ERRO BETA):** Consiste em aceitar a hipótese de nulidade quando ela é falsa.

**ERRO PADRÃO:** O desvio padrão de uma distribuição de médias de amostra, ao contrário da distribuição de dados brutos de uma amostra única.

**ERRO-PADRÃO:** Medida da variação que se espera ocorrer, apenas pelo efeito do acaso, quando somente seleciona-se uma amostra aleatória para estimar um parâmetro populacional. Dois erros-padrão para um lado e para outro de um valor estatístico, por exemplo, da prevalência, é o procedimento usado, comumente, para construir o intervalo de confiança.

**ERRO-PADRÃO (EP):** Desvio padrão (dp) de uma população de médias amostrais e não da média das observações individuais. o ep é calculado sendo o dp observado dividido pela raiz quadrada n.

**ERRO-PADRÃO DA DIFERENÇA ENTRE AS MÉDIAS (EDP):** Raiz quadrada da soma das variâncias das populações respectivas, cada uma delas dividida pelo tamanho da sua própria amostra.

**ERRO-PADRÃO DA ESTIMATIVA:** Medida da dispersão dos pontos amostrais em torno da reta de regressão.

**ERRO-PADRÃO DA MÉDIA:** Desvio-padrão de todas as médias amostrais  $\bar{X}$  possíveis.

**ERRO PADRÃO DA MÉDIA:** Quando se refere a uma média m de n parcelas, é o desvio padrão dividido por  $(n)^{1/2}$ , ou ainda  $S_{(\bar{x})} = \frac{S}{\sqrt{n}}$ .

**ERRO QUADRADO MÉDIO:** É a soma dos quadrados dos desvios entre os valores do estimador que é a variável e o parâmetro que ele se propõe a estimar. Em inglês mean square error.

**ERRO SISTEMÁTICO:** Erro da pesquisa indutiva representa a distorção dos achados, resultante de uma diferença sistemática entre a população em estudo, definida no início da investigação. Sinônimos: vício, viés, tendenciosidade.

**ERRO SISTEMÁTICO:** Não é devido ao acaso; o mesmo que tendenciosidade ou viés.

**ERROS PESSOAIS:** Naquelas tarefas ou observações em que entra julgamento pessoal, com freqüência diferentes indivíduos obterão sistematicamente diferentes resultados. Vale distinguir este tipo de erro de enganos, que consistem em erros de natureza computacional e de descuidos em geral.

**ERROS SISTEMÁTICOS:** Aqueles que permanecem os mesmos para cada observação com dado instrumento e dado método, ou são certas funções definidas do valor da quantidade sendo observada.

**ESCALA:** Tipo de medida composta compreendendo diversos itens tendo uma estrutura lógica ou empírica entre si. Exemplos de escalas: distância social de Borgardus, Guttman, Likert e Thurstone. Diferente de índice.

**ESCALA:** Conjunto de símbolos ou números construídos de tal forma que podem ser designados por uma regra para indivíduos, seus comportamentos e suas atitudes.

**ESCALA:** É uma série de números definida segundo um critério capaz de indicar os vários graus de certo traço ou característica. Em inglês scale.

**ESCALA, TIPO (USO):** Uma das principais considerações do pesquisador deve ser o tipo de escala que vai usar para medir a variável dependente. Ele usará técnicas não paramétricas, se a mensuração for em escala ordinal ou nominal.

**ESCALA BALANCEADA:** Escala com o mesmo número de categorias positivas e negativas.

**ESCALA COMPARATIVA:** Escala que compara outro objeto, conceito ou pessoa; as escalas de classificação por ordem são comparativas.

**ESCALA DE CLASSIFICAÇÃO:** Quando se quer registrar a intensidade com que o comportamento aparece, usa-se a escala de classificação. Constitui-se ela de uma série de comportamentos seguida de uma escala numérica ou conceitual.

**ESCALA DE CLASSIFICAÇÃO GRÁFICA:** Escala que mostra o contínuo gráfico, normalmente ancorado por duas extremidades.

**ESCALA DE CLASSIFICAÇÃO POR ITENS:** Uma escala que mostra um número limitado de categorias ordenadas.

**ESCALA DE CLASSIFICAÇÃO POR ORDEM:** Escala que classifica um objeto, conceito ou pessoa em certa ordem.

**ESCALA DE COMPARAÇÃO EMPARELHADA:** Uma escala que mostra uma série de apenas dois objetos como opção.

**ESCALA DE DISTÂNCIA SOCIAL DE BORGARDUS:** Técnica de medição para determinar a disposição das pessoas em participarem de relações sócias, em vários graus de intimidade, com outros tipos de pessoas. É uma técnica particularmente eficiente na medida em que diversas respostas discretas podem ser resumidas sem perda de nenhum detalhe original dos dados.

**ESCALA DE GUTTMAN:** Tipo de medida composta usada para resumir varias observações discretas e representar uma variável mais geral.

**ESCALA DE INTENÇÃO DE COMPRA:** A escala que classifica a intenção de compra dos consumidores.

**ESCALA DE INTERVALO:** Uma escala ordinal com intervalos iguais entre os pontos para mostrar quantidades relativas, que poderão incluir um zero arbitrário.

**ESCALA DE LIKERT:** Uma série contínua de respostas medida do extremo positivo até o extremo negativo ou vice versa em cinco, sete ou nove categorias.

**ESCALA DE OPINIÃO:** Quando se usa escalas desse tipo, pede-se ao indivíduo para dar uma reação pessoal a algum estímulo. A abrangência de estímulos sujeita a uma resposta avaliativa pessoal é quase infinita. Não existe obviamente uma resposta correta.

**ESCALA DE PONTOS:** Tipo de escala frequentemente usada em métodos de análise de avaliação sensorial de alimentos, para determinação do nível de qualidade de vários alimentos, porque é flexível, simples e facilita a análise estatística. Este sistema tem sido intensivamente usado pelas indústrias de laticínios,

ovos, carnes e de panificação. O sistema de escala de pontos deve ser usado, quando se tem um painel, cujos membros foram bem treinados para análise do produto particular. A pressuposição fundamental do método é que os números ou pontos sejam proporcionais ao grau de qualidade do produto e à importância da característica avaliada. Isto, porque neste tipo de escala, são incluídas outras características de qualidade além do sabor e odor, as quais podem ser medidas objetivamente. A escala de pontos deve ser de tal forma, que uma diferença no escore reflita uma diferença semelhante na característica de qualidade medida. Os painelistas, nesse caso, devem ser treinados, de modo a conhecer as características dos diferentes graus de qualidade do padrão, para que, testando as amostras, confirmem-lhes escores apropriados. A tabela a seguir exemplifica o tipo de escala de pontos para o produto pão. O valor para cada característica, apresentado na tabela, pode variar de região para região, isto é, pode ser aumentado o peso de uma característica e diminuído o de outra. Por essa escala, os produtos são classificados de acordo com o número de pontos recebidos pelas amostras nos painéis, em classes de qualidade, como excelente, ótimo, bom, regular, deficiente, inaceitável, dentre outros. Normalmente, há necessidade de repetição dos testes, sendo utilizados os resultados médios. O melhor número de repetições deve ser estudado para cada caso. Como dado prático, pode-se utilizar de 2 a 10 repetições ou amostras.

**Tabela de avaliação de pães. Valor em pontos do pão típico:**

<b>Volume</b>	<b>15</b>
<b>Cor e natureza da crosta</b>	<b>5</b>
<b>Simetria</b>	<b>5</b>
<b>Uniformidade do assado</b>	<b>5</b>
<b>Textura</b>	<b>15</b>
<b>Cor do miolo</b>	<b>10</b>
<b>Grãos</b>	<b>10</b>
<b>Aroma</b>	<b>15</b>
<b>Sabor</b>	<b>20</b>
<b>Total</b>	<b>100</b>

**Volume:** expresso em centímetros cúbicos, é usualmente determinado pelo método de deslocamento de sementes. O volume pode ser ou muito grande, muito pequeno ou satisfatório.

**Cor da crosta:** Deveria, normalmente, apresentar um marrom dourado profundo, no topo, e um marrom dourado claro nos lados do pão.

**Simetria:** Além da sua própria atração ao olhar, ela indica a condição de descanso de massa e seu levantamento ao forno.

**Uniformidade do assado:** O pão ideal possui uma crosta de grossura uniforme e com sua cor dourada, variando somente num grau limitado.

**Textura:** É inteiramente determinada pelo senso do tato. Ao se pressionar levemente a superfície cortada do pão e esfrega-lo por meio do grão, a sensação produzida pode ser descrita como aveludada, seda, macia, elástica ou áspera, dura, esfarelada, empelotada e grudenta.

**Cor de miolo:** depende principalmente da cor natural do endosperma do trigo e das partículas da película do grão na farinha de trigo. É também afetado pelo grão do miolo.

**Grão:** Estrutura celular das fatias de pão. Open grain são células relativamente grandes, enquanto close grain, são células pequenas. Estas células podem ser arredondadas ou alongadas, em que a última forma é aceita como sendo a mais desejável. As paredes grossas da célula, normalmente, são acompanhadas de grãos mais grossos, enquanto que paredes de células mais finas são características de pão de grão mais fino.

**Aroma:** Pode apresentar-se com o cheiro de trigo, de nozes, maltado, doce, fumaça, fungo, rançoso, dentre outros.

**Sabor:** O pão usualmente tem um leve sabor agridoce, embora possa adquirir também sabores distintamente ácidos e rançosos. Parte da sensação do sabor inclui também a qualidade de mastigação do pão.

**ESCALA DE PERCENTIS:** Escala que indica os escores correspondentes aos vários percentis. Tais escalas são usadas na apresentação das normas dos testes. Em inglês percentile scale.

**ESCALA DE PROPORÇÃO:** Escala interna de intervalo com ponto zero significativo, de modo que as magnitudes podem ser comparadas aritmeticamente.

**ESCALA DE SIMILARIDADE:** Escala arbitrária, por exemplo, de -5 a +5, que permite a representação de uma relação ordenada entre objetos que vai do mais semelhante ou mais próximo ao menos similar ou mais distante. Esse tipo de escala é adequado apenas para representar uma única dimensão.

**ESCALA DE SOMA CONSTANTE:** Uma escala na qual um conjunto de pontos é distribuído entre dois ou mais atributos.

**ESCALA DE THURSTONE:** Tipo de medida composta construída de acordo com pesos atribuídos por juízes a diversos indicadores de algumas variáveis.

**ESCALA DE RAZÕES:** Caracteriza-se por apresentar unidade constante e comum de mensuração, atribuindo-se um número real a cada escore, havendo uma razão conhecida entre dois intervalos quaisquer, um quociente conhecido entre dois valores quaisquer e um verdadeiro ponto zero como origem. Exemplo: peso em gramas.

**ESCALA DIFERENCIAL SEMÂNTICA:** Escala que classifica pares opostos de palavras ou frases em um contínuo, que é em seguida plotado como perfil ou imagem.

**ESCALA FATUAL:** Instrumentos em que se empregam itens de teste para os quais existe uma resposta correta, de ampla concordância.

**ESCALA HEDÔNICA:** Este é um método de graduação da preferência em níveis de qualidade para alimentos, podendo ser usado como teste de qualidade para outros produtos não alimentícios, em que há necessidade de avaliação subjetiva ou sensorial. O método consiste basicamente em apresentar as amostras dos produtos, de maneira inteiramente ao acaso, aos provadores e perguntar-lhes sobre a preferência entre elas, segundo uma escala previamente estabelecida, baseada nos atributos gosta e desgosta. Os pontos da escala são distinguíveis verbalmente, de modo que possam ser associados a valores numéricos, possibilitando análise estatística dos resultados. Sua grande vantagem é que pode ser usada para provadores não treinados, amostras de consumidores e também para provadores treinados. É usada com mais sucesso para pessoal não treinado. As amostras são apresentadas, em ordem inteiramente aleatória aos provadores, que são perguntados sobre a de que mais gosta ou desgosta, de acordo com a escala. Os provadores devem ser instruídos sobre as principais características de qualidade do produto em estudo e

sobre o que devem fazer para fornecer as respostas, nas folhas próprias, e isso no caso de consumidores ou de provadores não treinados. A escolha das palavras ou frases que vão identificar os intervalos na escala é de grande importância, uma vez que essa associação verbal não somente deverá dar uma idéia de ordem sucessiva dos intervalos na escala, como também facilitar a decisão do provador em suas respostas. Sempre se evita o uso de expressões ambíguas, que podem causar confusão e dificultar a decisão do provador. Um exemplo de escala hedônica bastante usada é a que tem nove categorias; gosta extremamente, gosta muito, gosta moderadamente, gosta pouco, indiferente (não gosta nem desgosta), desgosta pouco, desgosta moderadamente, desgosta muito e desgosta extremamente. Essas frases são colocadas, em linha, na escala vertical ou horizontalmente. São usadas diferentes formas de escala hedônica, sem maiores efeitos sobre os valores dos resultados. Algumas alternativas na forma de escala são: i) exclusão da categoria neutra; ii) uso de mais categorias gosta que desgosta; iii) redução do número de categorias; (o uso de menos que cinco categorias onde classes não é recomendado). As variações, na forma de escala, causam alterações na distribuição das respostas, e, conseqüentemente, nos parâmetros média, variância, dentre outros, todavia as médias relativas tendem a permanecer constantes.

**ESCALA INTERVALAR:** Semelhante à escala de razões, havendo apenas uma razão conhecida entre dois intervalos quaisquer, sendo a unidade de medição e o ponto zero arbitrários. Exemplo: temperatura em graus centígrados.

**ESCALA INTERVALAR:** Tipo de escala usada em operações empíricas de determinação de igualdade ou diferenças de intervalos que se estabelecem sobre a propriedade medida. Esta unidade física é um critério que permite sua aplicação sucessivamente com os mesmos resultados. Podem-se aplicar todas as medidas estatísticas usuais, além das medidas indicadas a seguir: média aritmética, variância, desvio padrão, correlação de ordem, correlação de posto e técnicas paramétricas e não paramétricas.

**ESCALA INTERVALAR:** Uma escala de mensuração que, além de ordenar os dados, estabelece uma distância unitária na escala, de tal forma que a magnitude entre os escores possa ser conhecida. É também chamada uma escala de igual unidade.

**ESCALA LIKERT:** Tipo de medida composta desenvolvida por Rensis Likert para melhorar os níveis de medição em pesquisa social pelo uso de categorias padronizadas de resposta em questionários de survey. Itens Likert são os que usam categorias de resposta como concordo fortemente, concordo, discordo e discordo fortemente. Tais itens podem ser usados na montagem de genuínas escalas Likert ou na construção de outros tipos de medidas compostas.

**ESCALA LIKERT:** Escala que mostra uma série de atitudes para com o indivíduo e que recebem valores numéricos que vão de favorável a desfavorável.

**ESCALA LIKERT, EXEMPLOS:**

Muito satisfatório	Muito bom
Satisfatório	Bom
Indeciso	Sem opinião
Insatisfatório	Ruim
Muito insatisfatório	Muito ruim
Altamente apropriado	Altamente favorável
Apropriado	Favorável



<b>Neutro</b>	<b>Sem opinião</b>
<b>Desapropriado</b>	<b>Desfavorável</b>
<b>Altamente inapropriado</b>	<b>Altamente desfavorável</b>

**ESCALA MULTIDIMENSIONAL:** Uma escala que mede mais de um atributo.

**ESCALA NÃO BALANCEADA:** Escala com um número irregular de categorias positivas e negativas.

**ESCALA NÃO-COMPARATIVA:** Escala que não compra objetos, conceitos ou pessoas, tais como escalas de classificação gráfica e de classificação por itens.

**ESCALA NOMINAL:** Tipo de escala usada em operações empíricas de determinação de similitude. As estatísticas aplicáveis são a contagem do número de casos no interior de cada categoria, contendo vários elementos e anotando sua frequência. As comparações entre as categorias poderão ser feitas pelas medidas de proporções, porcentagens e razões. Quando o número de casos é suficientemente grande podem-se testar escalas nominais e, no caso de dicotomias, tratar as escalas nominais como escalas de intervalo. Quando o número de casos é pequeno, pode-se empregar a distribuição binomial. As relações entre duas ou mais escalas nominais podem ser tratadas pelo teste de qui quadrado [ $\chi^2$ ] e de Phi de Yule. O tipo apropriado de técnicas estatísticas é o não paramétrico. Exemplos de técnicas estatísticas são a moda e coeficiente de contingência.

**ESCALA NOMINAL:** Escala de valores categóricos que podem ser considerados permutáveis e totalmente carentes de estrutura hierárquica. Em inglês Nominal scale.

**ESCALA NOMINAL:** Escala na qual as diversas categorias ou modalidades de uma variável são contadas. É a mais simples das escalas, havendo relação de equivalência entre e dentre as categorias. Exemplo: estado civil: solteiro, casado, divorciado e viúvo.

**ESCALA NOMINAL:** Uma escala de mensuração que simplesmente classifica os elementos em duas ou mais categorias, indicando quais os elementos são diferentes, mas não quanto à ordem ou magnitude deles.

**ESCALA NOMINAL:** Escala que divide os dados em categorias mutuamente exclusivas e coletivamente exaustivas.

**ESCALA ORDINAL:** Escala de valores categóricos ordenados por alguma hierarquia. Em inglês Ordinal scale.

**ESCALA ORDINAL:** Escala nominal que pode ordenar dados.

**ESCALA ORDINAL:** Tipo de escala usada em operações empíricas de ordenação por posto que resulte da distinção dos elementos de acordo com o maior ou menor grau com que possuem determinada característica. As estatísticas aplicáveis são: a mediana para localizar a posição do caso médio quando os dados seguem um determinado ordenamento. Decis, quartis e percentis indicando as posições dos dados ao longo do contínuo estabelecido. Teste de hipóteses como o teste de qui quadrado [ $\chi^2$ ] e as correlações de posto. A análise de variância pode ser empregada sempre que se esteja verificando uma relação entre uma escala ordinal e uma de intervalo. As estatísticas que incluem média e desvio padrão não deverão ser utilizadas porque estas supõem a igualdade de extensão dos sucessivos intervalos de escala, o que não se aplica na variável ordinal. O tipo apropriado de técnicas estatísticas é o não paramétrico. Exemplos: correlação de Spearman, R de Kendall e W de Kendall.

**ESCALA ORDINAL OU POR POSTOS:** Escala na qual as modalidades de uma variável são ordenadas em graus ou em magnitudes convencionadas, havendo uma relação matemática 'maior do que' ou 'menor do que' dos elementos entre as diversas categorias e de equivalência das unidades dentre cada modalidade. Exemplo: conceitos escolares: excelente, bom, regular e insuficiente.

**ESCALA RACIONAL (RATIO SCALE):** Uma escala de mensuração que, além de possuir unidades iguais, também estabelece um zero absoluto na escala.

**ESCALA RAZÃO:** Tipo de escala usada em operações empíricas de variáveis que reúnem todas as propriedades dos números naturais. Supõem um zero absoluto. As estatísticas aplicáveis para esta variável pode ser todo tipo de medidas estatísticas, como média, variância e desvio padrão.

**ESCALA STAPEL:** Escala que proporciona uma descrição única no centro, que é geralmente medido por mais ou menos cinco pontos.

**ESCALA TIPO LIKERT:** Um procedimento, comumente usado para mensurar atitudes, que requer uma resposta gradualizada a cada item ou afirmativa. Designam-se valores numéricos às respostas alternadas de cada item. O escore do indivíduo é a soma dos valores numéricos.

**ESCALA UNIDIMENSIONAL:** Escala que mede somente um atributo.

**ESCALA VISUAL ANALÓGICA:** É o tipo de escala em os pacientes avaliam o grau de algo real, por exemplo, dor, ansiedade, fome, dentre outras. O pesquisador pode mostrar ao participante da pesquisa o desenho de um segmento de reta com estados extremos no início e no final da linha (nenhuma dor à esquerda e dor insuportável à direita) e pedir para que o paciente marque com um X no segmento de reta.

**ESCALAR:** É uma quantidade que apresenta apenas tamanho. Por exemplo, os números reais são escalares. Em inglês scalar.

**ESCALAS DE MENSURAÇÃO (TIPOS):** Nominal (sexo e raça), ordinal (classe social e grau de satisfação), intervalar (na temperatura a posição do zero é estabelecida por convenção) e razão (em peso e altura o zero significa ausência).

**ESCALAS MÚLTIPLAS:** Método de combinação de diversas variáveis que medem o mesmo conceito em uma única variável como tentativa de aumentar a confiabilidade da medida por meio de medida multivariada. Na maioria dos exemplos, as variáveis separadas são somadas e em seguida seu escore total ou médio é usado na análise.

**ESCALONAMENTO:** Procedimentos para tentar determinar mensurações quantitativas para conceitos subjetivos ou abstratos.

**ESCALONAMENTO MULTIDIMENSIONAL:** O escalonamento multidimensional (MDS) se refere a uma série de técnicas que ajudam o pesquisador a identificar dimensões, chave inerentes a avaliações feitas por respondentes quanto a objetos. É também conhecido como mapeamento perceptual, é um procedimento que permite a um pesquisador determinar a imagem relativa percebida de um conjunto de objetos como por exemplo, empresas, produtos, ideias ou outros itens associados a percepções comumente consideradas. O objetivo do MDS é transformar julgamentos de consumidores quanto a similaridade ou preferência, por exemplo, preferência sobre lojas ou marcas em distâncias representadas em espaço multidimensional. Se os respondentes julgam os objetos A e B os mais semelhantes, comparados com todos



os outros possíveis pares de objetos, técnicas de escalonamento multidimensional colocarão os objetos A e B de tal forma que a distância entre eles no espaço multidimensional seja menor do que a distância entre quaisquer outros pares de objetos. Os mapas perceptuais resultantes exibem a posição relativa de todos os objetos, mas análises adicionais são necessárias para descrever ou avaliar quais atributos ditam a posição de cada objeto.

**ESCANEAMENTO ÓPTICO:** Dispositivo de processamento de dados que pode ler as respostas dos questionários.

**ESCAVAÇÃO:** Acesso a dados mais detalhados usados para criar dados agregados. Por exemplo, após rever cálculos aritméticos de vendas semanais para um conjunto de lojas, o pesquisador escavaria para ver relatos diários para vendas ou relatos para lojas individuais. Pode-se escavar ao mais elementar nível de detalhes que são os dados primitivos para elementos de dados disponíveis.

**ESCOLHA DA POTÊNCIA:** É o parâmetro mais importante para interpolação usando o inverso da potência da distância, pois deve-se ter em mente que potências baixas tendem a suavizar os extremos, enquanto as potências elevadas tendem a realçá-los.

**SCORE BRUTO:** O número inicialmente obtido na correção de um teste, antes de qualquer transformação em score padrão. Em se tratando de testes objetivos, o score bruto é representado pelo número de respostas certas menos certa fração das erradas, pelo tempo gasto, pelo número de erros, dentre outras. Em inglês raw score.

**SCORE DERIVADO:** Score em que se transforma o score bruto por meio da aplicação de processos estatísticos que têm por fim gerar uma escala de valores suscetíveis de interpretação mais fácil. O mesmo que score transformado. Em inglês derived score.

**SCORE DE CORTE:** Critério ou score contra o qual cada score Z discriminante individual é comparado para determinar a pertinência prevista em um grupo. Quando a análise envolve dois grupos, a previsão de grupo é de terminada computando-se um único score de corte. Elementos com scores Z abaixo dessa marca são designados a um grupo, enquanto aqueles com scores acima são classificados no outro. Para três ou mais grupos, funções discriminantes múltiplas são usadas, com um score de corte diferente para cada função.

**SCORE DE CORTE ÓTIMO:** Valor de score Z discriminante que melhor separa os grupos em cada função discriminante para fins de classificação.

**SCORE FATORIAL:** Medida composta criada para cada observação sobre cada fator extraído na análise fatorial. Os pesos fatoriais são usados em conjunção com os valores da variável original para calcular o score de cada observação. O score fatorial pode então ser usado para representar o(s) fator(es) em análises subseqüentes. Os scores fatoriais são padronizados para que tenham uma média de 0 e um desvio padrão de 1.

**SCORE NORMALIZADO:** Score padronizado, cuja escala foi transformada numa distribuição normal, independentemente da forma de distribuição dos scores originários. Em inglês normalized score.

**SCORE PADRÃO:** Qualquer score derivado de scores brutos de modo a formar uma escala padrão facilmente interpretável. Um exemplo de score-padrão é o score z, que traduz cada score bruto como

um desvio positivo ou negativo da média de todos os scores numa escala cuja unidade é um desvio-padrão. A fórmula do score z é:

$$z = \frac{X - \bar{X}}{s}, \text{ em que}$$

$z$  = score padrão

$X$  = score bruto

$\bar{X}$  = média dos scores brutos

$s$  = desvio-padrão

**Outras espécies de score-padrão são:** score T, estanino, score normalizado. Em inglês standard score.

**SCORE PADRÃO:** Os scores mais poderosos, estatisticamente, são os scores padrões, derivados das propriedades da curva normal probabilística. Eles mantêm as diferenças absolutas entre scores, uma característica que percentis e estaninos não possuem. Além disso, podem ser usados para calcular médias e correlações, produzindo os mesmos resultados dos scores originais. O score padrão mais comum é o Z que permite expressar o score bruto em termos de sua distância em unidades de desvio padrão. O score z é dado por:

$$Z = \frac{\text{Score bruto} - \text{Média}}{\text{Desvio} - \text{padrão}} = \frac{X - \mu}{\sigma}$$

**SCORE PADRONIZADO:** Número de desvio-padrão em que determinado valor está acima ou abaixo da média; também chamado score z.

**SCORE T:** Porque os scores Z têm média zero e envolvem decimais, assim como sinais positivos e negativos, são difíceis de interpretar. O score T contorna esse problema, estabelecendo uma nova média igual a 50 e um desvio-padrão igual a 10. Com as propriedades do score Z, o score T reúne as vantagens de ambos. O score T é dado por:  $T = Z(10) + 50$ .

**SCORE TRANSFORMADO:** Expressão equivalente a score derivado. Em inglês transformed score.

**SCORE REVERSO:** Processo de reversão dos scores de uma variável, embora mantenha as características de distribuição, para mudar as relações ou correlações entre duas variáveis. Usado na construção de escala múltipla para evitar um cancelamento entre variáveis com cargas fatoriais positivas e negativas no mesmo fator.

**SCORE VERDADEIRO:** Valor teórico do score resultante de um número infinito de medidas independentes do mesmo traço, por meio de formas equivalentes de um teste. É o score isento de qualquer erro. Um score efetivamente obtido na aplicação de um teste é igual ao score verdadeiro mais o valor positivo ou negativo do erro de medida associado ao teste. Em inglês true score.

**SCORE Z:** Número de desvios-padrão em que determinado valor está acima ou abaixo da média.

**SCORE Z DISCRIMINANTE:** Score definido pela função discriminante para cada objeto na análise e geralmente dado em termos padronizados. Também conhecido como score Z, é calculado para cada objeto em cada função discriminante e usado em conjunção com o score de corte para determinar pertinência prevista ao grupo. É diferente da terminologia score z usada para variáveis padronizadas.

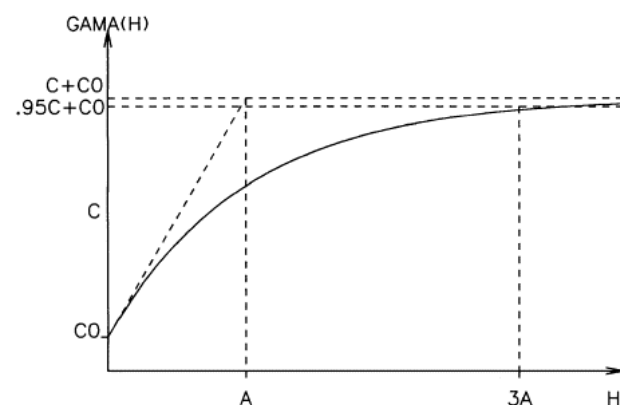
**ESCORES BRUTOS:** Simplesmente, o número total de pontos que um indivíduo ganha num teste, antes de serem convertidos a um sistema de escores formais ou padronizados. Um escore bruto, em si mesmo, não é interpretável, uma vez que não é possível compara-lo com qualquer outra coisa.

**ESCORES PADRONIZADOS:** Transformações de escores brutos em escores  $z$ , onde a média é igual a zero e o desvio padrão igual à unidade (1), simbolizados por  $N(0,1)$ .

**ESFERICIDADE:** Propriedade de que a matriz das variâncias/covariâncias de um conjunto de dados multivariados é um múltiplo escalar da matriz identidade.

**ESFERICIDADE:** O teste de esfericidade checa se a matriz de correlação é igual a matriz identidade, ou seja, na diagonal formados por uns e zero fora da diagonal.

**ESFÉRICO:** Modelo de variograma teórico, descrito pelo seguinte gráfico e equação:



É o modelo comum para a maioria dos depósitos minerais. Para ajuste deste modelo, desenha-se primeiro o patamar e em seguida a reta que une os dois primeiros pontos do variograma. A intersecção entre esta reta e o patamar dá-se a  $\frac{2}{3}$  da amplitude.

**ESGOTAMENTO DE PAINEL:** Quando, em survey de painel, alguns respondentes não participam de todas as etapas do estudo.

**ESPAÇO AMOSTRAL:** A série de pontos amostrais correspondentes a todas as amostras possíveis. O domínio permissível de variação de um ponto amostral. Às vezes, conhecido como descrição do espaço ou espaço de eventos. Em inglês Sample space.

**ESPAÇO AMOSTRAL ( $\Omega$ ):** É o conjunto de todos os resultados ou eventos possíveis de um determinado evento aleatório. Por exemplo: i) Lançar uma moeda para cima e observar a face que irá ficar virada para cima após a queda. O espaço amostral é Cara ou Coroa. ii) De uma urna com 10 bolas vermelhas (v) e 5 bolas brancas (b) retirarmos 2 bolas. O espaço amostral é vv ou vb ou bv ou bb.

**ESPAÇO AMOSTRAL:** Conjunto de todos os resultados ou eventos de um experimento que não pode mais ser subdivididos.

**ESPAÇO AMOSTRAL(L):** É o conjunto de todos os resultados possíveis de um experimento aleatório. Geralmente representado por S. Algumas vezes por E ou ainda por  $\Omega$ . Em inglês sample space.

**ESPAÇO AMOSTRAL:** É o conjunto de todos os resultados possíveis de um experimento ou ensaio aleatório, casual ou estocástico, e é simbolizado por  $\Omega$  ou S.

**ESPAÇO AMOSTRAL:** Das amostras de tamanho  $n$ ,  $\xi: \xi_1, \xi_2, \dots, \xi_n$ , é o conjunto dos pontos que têm por coordenadas o conjunto dos valores, a priori considerados possíveis, daquela ênupla.

**ESPAÇO AMOSTRAL:** Ver espaço de probabilidade.

**ESPAÇO DE BUSCA:** Domínio de potenciais soluções definidas para vários modelos de aprendizado como algoritmos genéticos ou redes neurais. As soluções podem ser especificadas em termos de dimensões consideradas, elementos de dados incluídos ou critérios para soluções prováveis ou não.

**ESPAÇO DE DECISÃO:** Em análise seqüencial e na teoria de funções de decisão, o espaço de decisão é o conjunto de todas as possíveis decisões. Em inglês Decision space.

**ESPAÇO DE PROBABILIDADE:** Conjunto de todos os resultados possíveis de um experimento aleatório. O mesmo que espaço amostral.

**ESPAÇO VETORIAL:** Conjunto definido sobre um corpo de escalares para o qual valem as leis da álgebra vetorial. Alguns exemplos de espaços vetoriais sobre  $R$  são o  $R^2$  e o  $R^3$ . Em inglês vector space.

**ESPECIFICAÇÃO (MODELO DE ELABORAÇÃO):** Em geral, processo por meio do qual conceitos se tornam mais específicos. Termo técnico usado em conexão com o modelo de elaboração, representando o resultado da elaboração no qual uma relação inicialmente observada entre duas variáveis é replicada entre alguns subgrupos criados pela variável de controle e não entre outros. Nesta situação, terão sido específicas as condições sob as quais a relação original existe, por exemplo, entre homens mas não entre mulheres.

**ESPECIFICAÇÕES (CONSTRUÇÃO DO QUESTIONÁRIO):** Conjunto de instruções detalhadas sobre o tratamento adequado de situações ambíguas ou complexas surveys. Sempre que o entrevistador não sabe como lidar com determinada situação ou resposta, a solução deve ser encontrada nas especificações.

**ESFERICIDADE:** Propriedade de que a matriz das variâncias/covariâncias de um conjunto de dados multivariados é um múltiplo escalar da matriz identidade. Em inglês sphericity.

**ESPECIFICIDADE:** Em ensaios ou em experimentos clínicos, a proporção de negativos corretamente identificados pelo teste clínico. Em inglês Specificity.

**ESPECIFICIDADE:** Percentagem de indivíduos sem o evento, como por exemplo sem a afecção investigada, cujo teste-tuberculínico, pode ser do tipo negativo (-).

**ESPECIFICIDADE:** Proporção de testes negativos entre os sadios. Diferenciar de sensibilidade.

**ESQUEMA FATORIAL 2 X 2 (2 X 2 FACTORIAL DESIGN):** Investiga dois fatores, em dois níveis, em uma única investigação; por exemplo, duas drogas e duas dosagens.

**ESS:** Sistema especialista estatístico. Em inglês Expert Statistical System.

**ESPERANÇA:** Valor médio que ocorreria se uma variável aleatória fosse observada muitas vezes; também chamada valor esperado, ou média, e simbolizada por  $\mu$ .

**ESPERANÇA MATEMÁTICA:** Também conhecida por valor esperado ou expectância, ou média de uma variável aleatória, a qual é obtida no caso de variável aleatória discreta ou descontínua, somando o produto de cada valor assumido pela variável pela sua probabilidade de ocorrência, ou a integral no domínio da função densidade de probabilidade do produto da variável pela valor da função  $dx$ , no caso de variável aleatória contínua. É um conceito estatístico que caracteriza a soma das probabilidades de cada possibilidade de saída de um experimento aleatório multiplicada pelo seu valor. Isto é, representa o valor médio esperado de uma experiência se ela for repetida muitas vezes. Esperança ou valor esperado é a média ponderada dos valores que a variável aleatória ou função assume, usando-se, como pesos para ponderação, as probabilidades correspondentes a cada valor. Esperança matemática ou valor esperado ou média de uma variável aleatória  $\{E(x) = \mu_{(x)} = \mu_x = \mu = m\}$ , é determinada como: Para a variável aleatória discreta (vad)  $E(x) = X_1 P_1 + X_2 P_2 + \dots + X_n P_n = \sum_{i=1}^n X_i P_i$

É importante destacar que sob o ponto de vista científico a esperança matemática corresponde ao que se espera que aconteça, em média. Já para a variável aleatória contínua (vac) é dada por:

$$E(x) = \int_{-\infty}^{\infty} x f(x) dx \text{ ou } \int_{-\infty}^{\infty} x f(x) dx$$

Pode acontecer que esta integral (imprópria) não convirja. Consequentemente diremos que  $E(X)$  existirá se e somente se  $\int_{-\infty}^{\infty} |x| f(x) dx$  for finita.

**ERRO TIPO II:** É a probabilidade ou risco do pesquisador, aceitar uma hipótese de nulidade ( $H_0$ ) quando ela for na verdade falsa, o qual é dado por  $\beta$ .

**EXPERIMENTO (ENSAIO):** É um trabalho previamente planejado, que segue determinados princípios básicos que são a repetição, a causalidade e o controle local, e no qual se faz a comparação dos efeitos dos tratamentos, por meio de variáveis respostas.

**ENSAIO DE UNIFORMIDADE OU EM BRANCO:** É aquele onde não há tratamentos (sem tratamentos). Nestes ensaios o pesquisador cultiva o local experimental com uma determinada variedade, tratando toda a área durante todo o ciclo da cultura da maneira mais uniforme possível. Na colheita uma bordadura substancial é removida de todos os lados do campo experimental. A seguir, a área restante é dividida em pequenas parcelas, todas com as mesmas dimensões. A produção destas parcelas é colhida e registrada separadamente. Com os dados de produção dos ensaios de uniformidade, o pesquisador pode ter então uma idéia de heterogeneidade do solo, detectando também manchas boas ou ruins do campo experimental. O principal método de controlar a heterogeneidade é o de adequar a área experimental escolhida ao delineamento experimental, tamanho e forma de parcela, número de repetições e número de tratamentos com a precisão requerida para o experimento.

**ENSAIO ABSOLUTO:** É aquele em que existe apenas um (1) tratamento a ser testado.

**ENSAIO COMPARATIVO:** É aquele em que mais de um tratamento são testados.

**EXPERIMENTOS ALEATÓRIOS:** São aqueles em que no seu planejamento entra ou é utilizado o princípio básico da experimentação denominado de casualização. Como exemplo desta categoria pode-se citar o delineamento em blocos completos casualizados e o delineamento em quadrado latino.

**EXPERIMENTOS SISTÊMICOS OU SISTEMÁTICOS:** São aqueles em que no seu planejamento não entra ou não é empregado o princípio básico da casualização. Este tipo de experimento é empregado nos estudos de população de plantas e espaçamento no cultivo em consórcio.

**EXPERIMENTOS PRELIMINARES:** É aquele conduzido dentro de estações experimentais para a obtenção de novos fatos. É científico, mas apresenta baixa precisão. O pesquisador avalia um grande número de tratamentos e, em geral, usa poucas repetições, apenas alguns tratamentos são repetidos para estimar o erro experimental com objetivo de descartar os tratamentos menos promissores para trabalhos futuros. Isto ocorre, com frequência nos programas de melhoramento genético de espécies cultivadas, nas quais um grande número de progênies são obtidas e avaliadas, para selecionar apenas as melhores. Exemplo: experimento de introdução de variedades.

**EXPERIMENTOS CRÍTICO, CRITERIOSO OU INTENSIVO:** É aquele que visa negar ou confirmar uma hipótese obtida no experimento preliminar e é conduzido dentro ou fora das fronteiras das estações experimentais. É científico e apresenta boa precisão. O número de tratamentos é menor e o número de repetições é maior. Há um maior controle do erro experimental, permitindo identificar as diferenças entre os tratamentos. Um aspecto importante, neste caso, é que os resultados obtidos num experimento criterioso são válidos apenas para aquele local ou estação experimental, ano, época, manejo, dentre outros, em que se conduziu o experimento. Faz-se necessário, por isto, repetir o mesmo experimento em vários locais de uma região onde os resultados devem ser aplicados.

**EXPERIMENTOS DEMONSTRATIVOS OU EXTENSIVO:** É aquele lançado pela rede de extensão rural. É de cunho demonstrativo. Os experimentos demonstrativos possuem poucas repetições e poucos tratamentos e são conduzidos, geralmente, nas propriedades agrícolas com fins de divulgação de resultados de experimentos intensivos pois é uma técnica extensionista. Neste tipo de experimento comparam-se os tratamentos ou procedimentos tradicionais com as propostas pela pesquisa. Além disso, é realizada uma análise econômica da viabilidade de substituir um procedimento usual por um novo, e verificada a aceitação dos proprietários e/ou consumidores. É importante destacar a seguinte observação: Em geral, as categorias preliminares e demonstrativas constituem-se de experimentos que não seguem o rigor dos princípios da experimentação e, por isso, geralmente, não são analisados estatisticamente.

**ESPERANÇA CONDICIONAL:** É o melhor estimador possível de  $Z(x_0)$  dados  $\{Z_a, a=1, n\}$ : onde:  $EnZ_0$  é a projeção do valor desconhecido  $Z(X_0)$  em  $H_n$ , que é o espaço vetorial de todas as funções  $f(Z_1, Z_2, \dots, Z_n)$ , onde a função  $f(Z_1, Z_2, \dots, Z_n)$  é a forma mais geral de um estimador de  $Z(x_0)$  a partir dos  $n$  dados disponíveis. O estimador  $EnZ_0$  é melhor, pois resulta da pesquisa em um espaço vetorial maior, que contém inclusive o subespaço vetorial  $En$  gerado por todas as combinações lineares. Embora  $EnZ_0$  seja o melhor estimador de  $Z(x_0)$ , o seu cálculo é quase sempre impossível, pois a distribuição das  $(n+1)$  variáveis  $\{Z(x_0), Z(x_1), \dots, Z(x_n)\}$  deve ser conhecida. Tem-se geralmente a informação limitada a uma única realização da função aleatória  $Z(x)$ , o que torna impossível a inferência e conseqüentemente o cálculo da esperança condicional.

**ESPERANÇA CONDICIONAL:** É a média de uma variável aleatória  $X$  dado uma variável  $Y$ , isto é  $E[X/Y]$ . conditional expectation.

**ESPERANÇA DE VIDA:** É o termo técnico utilizado em estatística vital para designar o número médio de anos que ainda restam para serem vividos pelos indivíduos que sobrevivem até a idade considerada, pressupondo-se que as probabilidades de morte que serviram para o cálculo continuem as mesmas.



**ESPERANÇA MATEMÁTICA DA VARIÁVEL ALEATÓRIA Y:**  $E(y)$ .

**ESTACIONARIDADE:** É uma característica da variável regionalizada, que se manifesta pela variabilidade constante a partir de uma dada distância, refletindo a independência entre essas amostras. A estacionaridade é identificada no variograma experimental, quando este apresenta um patamar ou tendência a estabilização a partir de uma certa distância e, por isso, denominada estacionaridade de segunda ordem.

**ESTANINO:** Escore derivado expresso numa escala de 1 a 9: a média de distribuição é igual a 5 e o desvio-padrão igual a 1,96. A palavra inglesa stanine deriva de standard e nine. As proporções de cada estanino são aproximadamente as seguintes:

<b>Escore Estanino</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>
<b>Porcentagem de Escores</b>	<b>4</b>	<b>7</b>	<b>12</b>	<b>17</b>	<b>20</b>	<b>17</b>	<b>12</b>	<b>7</b>	<b>4</b>

Em inglês stanine.

**ESTANINOS:** Uma escala que combina a facilidade interpretativa de porcentagens com as propriedades da curva normal de probabilidade. A escala tem nove intervalos, cada intervalo representando a metade de um desvio-padrão. O intervalo de meio (5º percentil) abrange os 20% do meio. O quarto, terceiro, segundo e primeiro estaninos, indo do centro para o extremo inferior da distribuição, em passo de meio d- padrão, abrange respectivamente 17, 12, 7 e 4 por cento dos casos. Correspondentemente, o 6º, o 7º, o 8º e o 9º estaninos, indo do centro para o extremo superior da distribuição, também abrangem respectivamente 17, 12, 7 e 4 por cento dos casos. O uso da curva normal é necessário.

**ETA QUADRADO:** É uma medida do efeito tamanho que é igual ao quociente entre a soma dos quadrados dos grupos entre, pela soma dos quadrados somados para todos os efeitos principais, de interação e de erro. Neste caso os efeitos das covariáveis não são computados. Esta medida pode ser interpretada como o percentual da variância da variável dependente que são explicadas pelos fatores. Quando existem relações curvilíneas entre a variável dependente e os fatores o Eta Quadrado será maior que o correspondente coeficiente de correlação múltiplo ou  $R^2$ . Em inglês eta-square.

**ESQUEMA SEQUENCIAL ABERTO:** É aquele em que não se prefixou o tamanho máximo de amostra a ser atingido. Opõem-se ao esquema seqüencial fechado.

**ESQUEMA SEQUENCIAL FECHADO:** É dito o plano de análise seqüencial em que a probabilidade de que o processo de seleção de amostra terminará é igual à unidade. Distingue de esquema seqüencial truncado e opõe-se a esquema seqüencial aberto.

**ESQUEMA SEQUENCIAL TRUNCADO:** É dito o plano de análise seqüencial no qual se impõe um extremo superior para o tamanho da amostra. Distingue de esquema sequencial fechado.

**ESTABILIDADE:** Falta de mudanças nos resultados de testes para retestes.

**ESTADO ABSORVENTE:** Estado de um processo estocástico que, uma vez atingido, não pode ser deixado. Do inglês Absorving state.

**ESTADO ABSORVENTE:** É o que, por si só, forma um conjunto fechado de estados.

**ESTADO ERGÓDICO:** É o persistente, não periódico, e de tempo médio de recorrência finito.

**ESTADO ESSENCIAL:** O mesmo que estado transitório, ou passageiro.

**ESTADO NÃO ESSENCIAL:** É aquele cuja probabilidade total de recorrência é nula.

**ESTADO NULO:** É o recorrente cujo tempo médio de recorrência é infinito.

**ESTADO PERIÓDICO:** De período  $t$ , é dito o estado recorrente, ou transitório, se um retorno a ele é impossível, exceto, talvez, em  $t$ ,  $2t$ ,  $3t$ , ... passos, sendo  $t > 1$  o maior inteiro com tal propriedade.

**ESTADO PERSISTENTE:** É o estado cuja probabilidade de recorrência é igual à unidade; repete-se infinitas vezes com probabilidade um.

**ESTADO RECORRENTE:** É aquele que é capaz de repetir-se. O mesmo que estado persistente.

**ESTADOS COMUNICANTES:**  $E_i$  e  $E_j$ , de um processo estocástico, são aqueles para os quais tanto a probabilidade de passagem de  $E_i$  para  $E_j$ , como a de passagem de  $E_j$  para  $E_i$ , são maiores que zero.

**ESTADO TRANSITÓRIO:** É o estado cuja probabilidade total de recorrência é menor que a unidade; com probabilidade um, repete-se apenas um número finito de vezes.

**ESTATÍSTICA:** Qualquer número usado para descrever um aspecto da amostra. Assim, a média, o desvio padrão, o coeficiente de correlação entre dois conjuntos de medidas dos membros de uma amostra são exemplos de estatísticas. A cada estatística na amostra correspondente um parâmetro na população. As estatísticas são representadas por letras do nosso alfabeto ao passo que os parâmetros são simbolizados pelas letras gregas correspondentes. Em inglês statistic.

**ESTATÍSTICA:** É a ciência que coleta, analisa e interpreta dados de maneira mais significativa possível. É particularmente útil em situações sujeitas a erros experimentais e, assim, pode ser definida como a ciência de tomada de decisões em face da incerteza.

**ESTATÍSTICA:** Termo que vem do latim status, estado, ou, talvez, situação. É o conjunto dos processos que tem por objeto a observação, a classificação formal, e a análise dos fenômenos coletivos, ou de massa, e, por fim, a indução das leis a que eles obedecem. Apresentação numérica tabular ou gráfica dos resultados da observação de fenômenos de massa. O mesmo que elemento típico. Modelo probabilístico de um fenômeno físico. Por exemplo: a estatística de Fermi-Dirac, a estatística de Bose-Einstein.

**ESTATÍSTICA:** Estudo da análise de dados; compreende a estatística descritiva e a inferência estatística; uma grandeza calculada com base nos elementos de uma mostra, como a média amostral  $\bar{X}$ ; costuma-se usar uma estatística como estimador de um parâmetro populacional desconhecido. Também chamada medida estatística.

**ESTATÍSTICA:** Palavra originária do latim e significa estado. Este termo provém do primeiro uso da estatística eu tinha como função o registro de dados, tais como, número de habitantes da população, número de casamentos e a elaboração de tabelas e gráficos para descrever resumidamente um determinado país em números.

**ESTATÍSTICA:** Uma boa definição de estatística é a de ser um conjunto de métodos especialmente apropriados à coleta, à apresentação, ou seja, a organização, resumo e descrição, à análise e à interpretação de dados de observação, tendo como objetivo a compreensão de uma realidade específica para a tomada da decisão.



**ESTATÍSTICA:** É o método que ensina a recolher, classificar, apresentar e interpretar um conjunto de dados numéricos.

**ESTATÍSTICA:** É a ciência que utiliza-se das teorias probabilísticas para explicar a frequência da ocorrência de eventos, tanto em estudos observacionais quanto experimentais, modelar a aleatoriedade e a incerteza de forma a estimar ou possibilitar a previsão de fenômenos futuros, conforme o caso.

**ESTATÍSTICA:** Um valor de resumo calculado a partir de uma amostra observada, geralmente, mas não necessariamente, como um estimador de algum parâmetro de população; uma função de valores amostrais. Em inglês *Statistic*.

**ESTATÍSTICA:** Um número ou valor. Na teoria da estimação é utilizada também como sinônimo de estimador, isto é, uma função dos elementos da amostra. Este termo foi introduzido por Sir Ronald Aylmer Fisher em 1922. Em inglês *statistics*.

**ESTATÍSTICA:** O termo estatística surgiu da expressão em Latim *statisticum collegium* palestra sobre os assuntos do Estado, de onde surgiu a palavra em língua italiana *statista*, que significa homem de estado, ou político, e a palavra alemã *Statistik*, designando a análise de dados sobre o Estado. A palavra foi proposta pela primeira vez no século XVII, em latim, por Schmeitzel na Universidade de Lena e adotada pelo acadêmico alemão Godofredo Achenwall. Aparece como vocabulário na Enciclopédia Britânica em 1797, e adquiriu um significado de coleta e classificação de dados, no início do século 19.

**ESTATÍSTICA:** Apesar de este termo ser usado como referência geral a dados numéricos, ele tem um significado mais específico em relação à amostragem de survey: uma descrição numérica de uma amostra, como 45% de homens. Corresponde a um parâmetro, que é uma descrição numérica de toda uma população. Em amostragens, estatísticas de amostras são calculadas como estimativas de parâmetros da população.

**ESTATÍSTICA:** Estatística é a ciência que trata da coleta, da organização, da sintetização e a da apresentação de dados; da medição da variação nos dados e levantamento de dados; da estimativa dos parâmetros da população e da determinação da precisão das estimativas; da aplicação dos testes de hipótese em relação aos parâmetros; da análise da relação entre duas ou mais variáveis. A estatística trabalha com dois conjuntos de dados: o universo e a amostra. Apesar de a estatística se preocupar em obter informações sobre a população, dificilmente ela estuda todos os componentes da mesma ou o censo.

**ESTATÍSTICA:** A arte e a ciência de coletar, analisar, apresentar e interpretar dados. Um valor de resumo calculado a partir de uma amostra observada, geralmente, mas não necessariamente, como um estimador de algum parâmetro de população; uma função de valores amostrais. Em inglês *Statistics*.

**ESTATÍSTICA:** É a ciência que lida com a coleta, o processamento e a disposição de dados ou informação, atuando como ferramenta fundamental nos processos de solução de problemas. Trata da coleta de dados informativos e da interpretação destes dados, facilitando o estabelecimento de conclusões confiáveis sobre algum fenômeno que esteja sendo estudado. A estatística é muito importante em um programa para a melhoria da qualidade, pois, suas técnicas podem ser utilizadas para descrever e interpretar a variabilidade, que é a causa da fabricação de um produto, bem ou serviço defeituoso. Ainda pode ser definido como o valor calculado a partir de uma amostra usado para fornecer informações sobre o valor de um parâmetro populacional de interesse.

**ESTATÍSTICA:** Característica medida de uma amostra.

**ESTATÍSTICA:** É toda a função dos valores amostrais não envolvendo parâmetros desconhecidos da distribuição.

**ESTATÍSTICA ANCILAR:** Estatística que, embora não dê informações diretas sobre o parâmetro que se deseja estimar, fornece indicações sobre a precisão da estimativa usada.

**ESTATÍSTICA ANCILAR:** De acordo com a teoria de inferência associada com o nome de Ronald Aylmer Fisher, a função de verossimilhança contém todas as informações dadas pela amostra sobre os parâmetros desconhecidos da população que se possa querer estimar. Se não há estimadores suficientes a perda de alguma informação será inevitável; mas essa perda pode ser reduzida considerando-se funções adicionais das variáveis que podem ser combinadas com o estimador de máxima verossimilhança. Tais funções são chamadas estatísticas ancilares. Termo em inglês: *Ancillary statistic*.

**ESTATÍSTICA NÃO OFICIAL:** Termo usado para denominar estatísticas e dados coletados, organizados e tornados disponíveis por instituições privadas ou por instituições públicas para uso próprio. Por exemplo: resultados de pesquisas de intenção de voto e de opinião pública, estatísticas de dados de sindicatos, etc.

**ESTATÍSTICA NÃO PARAMÉTRICA:** Conjunto de técnicas utilizadas quando não se pode fazer nenhuma suposição acerca dos parâmetros da distribuição de probabilidade dos dados. Por exemplo, não se pode afirmar que os dados possuam uma distribuição normal ou quando os níveis de mensuração são nominais ou ordinais.

**ESTATÍSTICA OFICIAL:** Estatísticas de dados coletados, organizados e publicados por órgãos governamentais. Por exemplo: os dados oriundos dos censos populacionais, estatísticas ministeriais, registros de nascimentos, de acidentes de trabalho, de óbitos, índice ou taxa de ocorrência de doenças, etc.

**ESTATÍSTICA APLICADA:** É qualquer sistema de investigação científica que proceda por intermédio da metodologia estatística. Compreende a Demografia, a Biometria, a Econometria, a Psicometria, a Mecânica Estatística, dentre outros campos de estudo. Opõe-se a estatística teórica.

**ESTATÍSTICA CIRCULAR:** Tipo de análise estatística denominada de circular ou direcional a qual é aplicada a dados que estão dispostos ao redor de uma circunferência. Os exemplos de dados desse tipo são diversos, e podem incluir as direções de voo de um pássaro, o número de pacientes chegando a um hospital ao longo do dia, a orientação do pólo magnético da Terra, entre outros. É importante ressaltar que o emprego de métodos estatísticos convencionais para essas amostras não gera resultados com significado, daí a necessidade de se utilizar fórmulas e testes específicos. Alguns exemplos de tipos de análise circular podem ser o tamanho da amostra, a média dos cossenos, a média dos senos, a média angular, o comprimento do vetor médio, a variância angular, o desvio padrão angular, o intervalo de confiança, o teste de Rayleigt para uma amostra, o teste V de Rayleigt para uma amostra, o teste de Watson-Williams para duas amostras independentes, o teste de Watson-Williams para três amostras independentes, o teste paramétrico para duas amostras relacionadas, a correlação circular, dentre outras.

**ESTATÍSTICA CONDICIONADA:** É aquela estatística amostral cuja distribuição depende de um parâmetro.

**ESTATÍSTICA DESCRITA:** Parte da estatística que tem como objetivo resumir as principais características em um conjunto de dados fazendo uso de tabelas, gráficos e resumos numéricos. Descrever os dados pode ser comparado ao ato de tirar uma fotografia da realidade. Caso a câmera fotográfica utilizada não seja

adequada, ou esteja sem foco, o resultado, no caso a foto, pode sair distorcido. Portanto, quem faz uso da estatística deve ter extremo cuidado em escolher os métodos e técnicas corretas para resumir os dados.

**ESTATÍSTICA INFERENCIAL (INDUTIVA):** A estatística inferencial utiliza informações incompletas para tomar decisões e tirar conclusões satisfatórias. O alicerce das técnicas de estatística inferencial está no cálculo de probabilidades. As duas técnicas de estatística inferencial são: estimação e teste de hipóteses.

**ESTATÍSTICA CONJECTURAL:** É o sistema de investigação científica no qual se procede por meio de levantamentos indiretos, nos quais a relação existente entre o fenômeno observado e o que se tem em vista conhecer é imperfeitamente definida.

**ESTATÍSTICA COM DISTRIBUIÇÃO QUI QUADRADO:**  $Y \sim \chi^2$  ou  $Y \sim \chi^2$ , significa que a variável aleatória contínua Y possui distribuição de teórica de probabilidade qui quadrado.

**ESTATÍSTICA PARAMÉTRICA:** Conjunto de técnicas estatísticas inferenciais utilizadas para dados cujos parâmetros sejam conhecidos suas distribuições de probabilidades, como por exemplo, a distribuição normal e com dados em níveis de mensuração intervalar ou racional.

**ESTATÍSTICA  $C_p$  DE MALLOWS:** Critério de seleção de um modelo de regressão onde adota-se o seguinte procedimento: Suponha que temos K variáveis regressoras disponíveis e ajustamos um modelo de regressão linear múltipla com (p-1) variáveis regressoras, além de termo constante. Seja a SQEp a soma de quadrados do erro obtida para este modelo. Seja  $\hat{\sigma}^2$  o estimador de  $\sigma^2$ , correspondente ao ajuste de modelo com todas as variáveis regressoras. A estatística  $C_p$  de Mallows, calculada para o modelo com (p-1) variáveis regressoras, é definida como se segue.  $C_p = \frac{SQE_p}{\hat{\sigma}^2} - (n - 2p)$ . Se o modelo com (p-1)

variáveis regressoras é adequado, isto é, o uso de apenas (p-1) regressoras equivale ao uso de todas, temos que  $\frac{SQE_p}{(n-p)}$  é um estimador de  $\sigma^2$ , assim como  $\hat{\sigma}^2$ . Então, o valor esperado de  $\left[ \frac{SQE_p}{\hat{\sigma}^2} \right]$  é aproxima-

damente igual a 1; o valor esperado de  $\left[ \frac{SQE_p}{\hat{\sigma}^2} \right]$  é aproximadamente igual a (n - p) e, conseqüentemente,

o valor esperado de  $\left[ \frac{SQE_p}{\hat{\sigma}^2} - (n - 2p) \right]$ , ou seja, de  $C_p$ , é aproximadamente igual a p. Se o valor observado

da estatística  $C_p$  de Mallows estiver muito acima de seu valor esperado, p, este modelo não parece ser adequado, por ter um valor relativamente alto da SQE. Se o modelo é adequado, esperamos observar que o valor da estatística  $C_p$  de Mallows esteja próximo de seu valor esperado. Assim temos definido um critério que nos permita avaliar a adequação do modelo ajustado.

**ESTATÍSTICA CULTURAL:** É, conforme os preceitos estabelecidos por várias publicações, como por exemplo, o Anuário Estatístico do Brasil, a estatística aplicada que se ocupa com o registro, o estudo e a apresentação dos dados relativos à situação e movimento da educação, seja esta oriunda de agentes sistemáticos e específicos, como escolas, seja a devida aos agentes de difusão cultural intencional, tais como, bibliotecas, museus, monumentos, cinema educativo, dentre outros.

**ESTATÍSTICA DA AMOSTRA:** Um valor numérico usado como medida de sumário para uma amostra, por exemplo, a média da amostra, a variância da amostra,  $s^2$ , e o desvio padrão da amostra, s.

**ESTATÍSTICA DE BOSE-EINSTEIN:** É o modelo probabilístico que descreve o conjunto D das possíveis distribuições distintas, de r objetos indiscerníveis, por n compartimentos, na hipótese de equiprobabilidade de todas elas. Nesse caso, a probabilidade de uma distribuição será dada por:

$$P(D) = \binom{n+r-1}{r}^{-1}.$$

**ESTATÍSTICA DE COOK:** Estatística utilizada como critério para análise da adequação do modelo de regressão a qual está relacionada como os resíduos e os valores h

$$\left[ h_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}} \right] \text{ ou } \left[ h_{ij} = \frac{1}{n} + \frac{(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}} \right]$$

Esta estatística mede a mudança nas estimativas dos coeficientes de regressão, quando uma determinada observação é omitida. O seu valor é dado como a seguir.

$$D_i = \frac{(\hat{\beta} - \hat{\beta}_{(i)})' X'X (\hat{\beta} - \hat{\beta}_{(i)})}{K \hat{\sigma}^2}, i = 1, 2, \dots, n. \text{ Sendo } \hat{\beta}_{(i)} \text{ o estimador de quadrados mínimos do vetor}$$

$\hat{\beta}$ , omitindo-se a i-ésima observação, e k é o número de colunas da matriz X. Um valor alto da estatística de Cook indica que tal observação afeta bastante o ajuste do modelo. A situação ideal seria aquela onde tivéssemos valores moderados desta estatística para todas as observações.

**ESTATÍSTICA DE DURBIN E WATSON:** Uma estatística de teste da independência dos erros em uma regressão de mínimos quadrados contra uma alternativa de correlação serial. Em inglês Durbin-Watson statistic.

**ESTATÍSTICA DE DURBIN-WATSON:** Estatística usada na análise de regressão para testar a presença de correlação em série.

**ESTATÍSTICA DE FERMI-DIRAC:** É o modelo probabilístico que descreve o conjunto D das possíveis distribuições distintas de  $r \leq n$  objetos indiscerníveis por n compartimentos, nenhum dos quais pode receber mais de um objeto, na hipótese de equiprobabilidade de todas estas distribuições. Neste caso, a probabilidade de uma distribuição será determinada por:

$$P(D) = \binom{n}{r}^{-1}.$$

**ESTATÍSTICA DE MAXWELL-BOLTZMANN:** É o modelo probabilístico que descreve o conjunto D das possíveis distribuições distintas de r objetos diferentes por n compartimentos, na hipótese de equiprobabilidade de todas elas Neste caso, a probabilidade de uma distribuição será calculada por:  $P(D) = n^{-r}$ .

**ESTATÍSTICA DE POSTO:** É toda estatística que se baseia exclusivamente nos postos dos elementos em causa, como, por exemplo, o coeficiente de correlação de Spearman. Não confundir com estatística ordinal em que intervém o conceito de valor além do de posto.

**ESTATÍSTICA DE PROVA:** Função dos valores de uma ou mais amostras mediante a qual se põe a prova uma hipótese relativa à especificação das respectivas populações originárias. Exemplo: a razão de Student.

**ESTATÍSTICA DE ORDEM:** Quando uma amostra de uma variável está arrumada em ordem de magnitude, estes valores ordenados são conhecidos como estatísticas de ordem. Em inglês Order statistics.

**ESTATÍSTICA D<sup>2</sup>:** Variável que exprime a distância de duas populações. Esse conceito foi dado por Mahalanobis (1936).

**ESTATÍSTICA DESCRITIVA:** Ramo da Estatística que tem por finalidade descrever certas propriedades relativas a um conjunto de dados.

**ESTATÍSTICA DESCRITIVA:** Parte da Estatística que trata do resumo e da apresentação de conjuntos de dados. Em inglês descriptive statistics.

**ESTATÍSTICA DESCRITIVA:** Estudo das maneiras de sintetizar dados.

**ESTATÍSTICA DESCRITIVA:** Estatística descritiva refere-se a um conjunto de conceitos e métodos usados na organização, resumo, tabulação, representação e descrição de uma coleção de dados. O objetivo da estatística descritiva é fornecer uma representação dos dados que descrevem em forma numérica, gráfica ou tabular os resultados de pesquisa.

**ESTATÍSTICA DESCRITIVA:** Segundo uma antiga divisão tripartida de estatística, surgida com Rümelin (1863) é a parte deste conhecimento que descreve quantitativamente os fatos notáveis da sociedade. Indica ainda, para qualquer ramo da estatística, os estudos feitos com o fim exclusivo de descrever uma coletividade dada, sem objeto de generalizar as conclusões assim obtidas para o universo de indivíduos que inclui a coletividade em causa. Opõem-se a estatística indutiva, indução ou inferência estatística.

**ESTATÍSTICA DESCRITIVA:** Parte da estatística cujo objetivo é a coleta, a organização, a classificação dos dados amostrais ou das populações, as apresentações tabular e gráfica e o cálculo de determinadas medidas: média, mediana, variância, desvio padrão, coeficiente de variação, de assimetria, de curtose e outras.

**ESTATÍSTICA DESCRITIVA:** Computações estatística descrevendo as características de uma amostra ou a relação entre variáveis numa amostra. Estatística descritiva apenas resume um conjunto de observações amostrais, enquanto estatística inferencial vai além da descrição de observações específicas para inferir sobre a população maior da qual foram retiradas às observações amostrais.

**ESTATÍSTICA DESCRITIVA:** Termo usado para denotar dados estatísticos do tipo descritivo ou os métodos de obtenção destes dados, em contraste com a estatística teórica que, ao tratar dados práticos, usualmente envolve algum processo de inferência em probabilidade para sua interpretação. Corresponde a coleta, a organização, a descrição dos dados, o cálculo e a interpretação de coeficientes. Em inglês Descriptive statistics.

**ESTATÍSTICA DESCRITIVA:** É a parte da estatística que procura os melhores métodos para coletar, ordenar e sumarizar os dados dos experimentos.

**ESTATÍSTICA DESCRITIVA:** Método usado para resumir as principais características de dados populações conhecidos.

**ESTATÍSTICA DE TESTE:** Uma função de uma amostra de observações que oferece um critério para testes de uma hipótese estatística. Em inglês Test statistic.

**ESTATÍSTICA DE TESTE:** Estatística amostral baseada em dados amostrais, usada para tomada de decisão quando à rejeição da hipótese nula.

**ESTATÍSTICA DE TESTE:** Fórmula ou algoritmo usado para um teste de significância; o valor numérico calculado por essa fórmula ou esse algoritmo, para um teste específico de significância, usando um conjunto de dados.

**ESTATÍSTICA DE TESTE:** Grandeza calculada com base em observações, e usada para testar uma hipótese nula; uma estatística de teste é construída de modo que se identifique com uma distribuição conhecida, se a hipótese nula é verdadeira; assim, a hipótese nula é rejeitada se o valor observado da estatística de teste não parece provir daquela distribuição.

**ESTATÍSTICA ECONÔMICA:** É a estatística aplicada que tem por objetivo o levantamento, o estudo e a apresentação dos dados relativos aos fenômenos de produção, circulação e consumo da riqueza e dos seus meios, bem como do trabalho, sua situação, movimento e remuneração. O Anuário Estatístico do Brasil adota o título situação econômica, dele excluindo, porém os dados relativos ao trabalho, que figuram no título situação social. Ver também, econometria.

**ESTATÍSTICA ESPACIAL:** Da mesma forma que a estatística clássica é a ciência que coleta, analisa e interpreta dados espaciais de maneira mais significativa possível. A diferença neste caso é que o valor de uma amostra é função de sua posição na mineralização do depósito e a posição relativa das amostras é levada em consideração. A similaridade entre os valores das amostras é quantificada como uma função da distância entre amostras e essa relação representa a base da estatística espacial.

**ESTATÍSTICA EXPERIMENTAL:** É a parte da estatística que fornece os métodos de análise e interpretação dos resultados dos experimentos.

**ESTATÍSTICA F:** Na análise de variância, uma estatística que testa a hipótese nula de que todos os grupos provêm de populações com a mesma média; na análise de regressão, uma estatística que testa a hipótese nula de que não há relacionamento entre as variáveis independentes e a variável dependente.

**ESTATÍSTICA FINANCEIRA:** É a estatística aplicada que tem por objeto o levantamento, o estudo e a apresentação dos dados relativos à situação e ao movimento dos valores.

**ESTATÍSTICA g:** Valores que assumem, numa amostra, os elementos típicos da mesma definição que os parâmetros  $\gamma$ .

**ESTATÍSTICA GERAL:** O mesmo que estatística metodológica

**ESTATÍSTICA INEFICIENTE:** Uma estatística com menos do que a eficiência possível no sentido de ter variância amostral mínima. Em inglês Inefficient statistic.

**ESTATÍSTICA INFERENCIAL:** Corpo de computações estatísticas relevantes à feitura de inferências baseadas em observações amostrais de uma população maior. Ver também estatística descritiva.



**ESTATÍSTICA INFERENCIAL:** Estatística inferencial refere-se a um conjunto de métodos usados para se obter inferência sobre um grande grupo de pessoas a partir de dados disponíveis sobre apenas um subconjunto representativo do grupo. Em estatística, o grande grupo de pessoas se chama população. O subconjunto do grande grupo se chama amostra.

**ESTATÍSTICA INFERENCIAL OU INDUTIVA:** Também chamada como a medida da incerteza ou métodos que se fundamentam na teoria da probabilidade. O processo de tirar conclusões sobre a natureza ou o modelo de populações a partir de amostras aleatórias retiradas destas populações.

**ESTATÍSTICA INFERENCIAL OU INDUTIVA OU ANALÍTICA:** O processo de tirar conclusões sobre a natureza ou o modelo de populações a partir de amostras aleatórias retiradas destas populações. Em inglês inferential or inductive statistic.

**ESTATÍSTICA INFERENCIAL OU INDUTIVA:** Ramo da Estatística que procura inferir propriedades da população a partir de propriedades verificadas numa amostra da mesma.

**ESTATÍSTICA INFERENCIAL OU INDUTIVA:** Métodos que envolvem o uso de dados amostrais para fazer generalizações ou inferências sobre uma população.

**ESTATÍSTICA K:** Família de funções  $K_r$  de conjunto amostral, tais que, para todo  $r$ , a esperança matemática de  $K_r$  é igual ao cumulante de ordem  $r$ , independentemente do tamanho da amostra. Dadas por Ronald Aylmer Fisher, em Moments and product-moments of sampling distributions (1929). Em função da média da amostra  $\bar{X}$ , dos momentos centrais  $m_r$  de amostra e em função de

$$S_r = \sum_{i=1}^n X_i^r$$

$$S_2 = s_2 - \frac{1}{n} s_1^2; S_3 = s_3 - \frac{3}{n} s_2 s_1 + \frac{2}{n^2} s_1^3$$

$$S_4 = s_4 - \frac{4}{n} s_3 s_1 + \frac{6}{n^2} s_2 s_1^2 - \frac{3}{n^3} s_1^4$$

Tem-se que, para a amostra  $x_1, x_2, \dots, x_n$ ,

$$K_1 = \frac{1}{n} s_1 = \bar{X}$$

$$K_3 = \frac{1}{n-1} s_2 = \frac{n}{n-1} m_2$$

$$K_3 = \frac{n}{(n-1)(n-2)} s_3 = \frac{n^2}{(n-1)(n-2)} m_3$$

$$K_4 = \frac{n}{(n-1)(n-2)(n-3)} \left[ (n+1) s_4 - 3 \frac{n-1}{n} s_2^2 \right] = \frac{n^2}{(n-1)(n-2)(n-3)} \left[ (n+1) m_4 + 3(n-1) m_2^2 \right]$$

**ESTATÍSTICA MATEMÁTICA:** É a parte da Estatística Metodológica que tem por objeto o estudo das propriedades matemáticas dos fenômenos de massa, e, por fim, a dedução e a demonstração dos processos e fórmulas usadas naquela. Também se diz Estatística Teórica. A fronteira que separa a estatística matemática do Cálculo de Probabilidades sempre foi vaga e contestada. Modernamente existe uma tendência para reservar o nome de Estatística Matemática apenas ao estudo do conjunto dos processos em

que intervém a estimação e os testes de hipóteses. Para muitos, no entanto, o Cálculo de Probabilidades é a teoria da Estatística, devendo ocupar-se não só com os problemas cuja solução se baseia no conceito clássico de probabilidade, mas também com os que requerem a introdução do conceito de verossimilhança de Ronald Aylmer Fisher, por exemplo. Nesse caso, a Estatística perde o caráter abstrato e teórico, passando a ser apenas aplicação.

**ESTATÍSTICA METODOLÓGICA:** É a exposição racional e sistemática dos processos aplicáveis a todas as fases do estudo dos fenômenos de massa em geral, desde a coleta dos dados para isso necessários, até a apresentação e a interpretação dos resultados elaborados. É o mesmo que estatística geral.

**ESTATÍSTICA MORAL:** É de acordo com a praxe estabelecida por algumas publicações estatísticas, a estatística aplicada que se ocupa com o registro dos fatos sociais relativos à religião ou cultos, criminalidade, suicídio, divórcio e, às vezes educação, não havendo nenhuma definição rigorosa e satisfatória do seu campo. O Anuário Estatístico do Brasil não consigna este título. Ver também, Estatística social.

**ESTATÍSTICA ORDINAL:** De ordem  $K$  ( $1 \leq K \leq n$ ) da amostra  $S_n: x_1, x_2, \dots, x_n$  de  $n$  valores da variável  $x$ , dispostos por ordem não decrescente, é o valor de  $x_k$ .

**ESTATÍSTICA OU ESTIMADOR:** É uma função dos valores da amostra, ou seja uma variável aleatória, pois seu resultado depende dos elementos selecionados naquela amostra. São utilizados para estimar os parâmetros populacionais, para isto é preciso conhecer sua distribuição de probabilidades, que via de regra, pressupõe normalidade ou amostras grandes. Por exemplo: a média amostral, a proporção amostral, a variância amostral, dentre outros.

**ESTATÍSTICAS P:** Nome dado a um conjunto de estatísticas usado no problema da discriminação de duas populações normais  $p$ -dimensionais, quanto às suas matrizes de covariâncias, dadas por Roy (1939).

**ESTATÍSTICAS PERMANENTES:** São as estatísticas que são objeto de levantamento, contínuo ou periódico, excluindo-se, portanto, dessa categoria, as que são objeto de levantamento ocasional.

**ESTATÍSTICA PIVOTAL:** Seja  $x_1, x_2, x_3, \dots, x_n$  uma amostra aleatória de  $p(x; \theta)$ . Seja  $Q = q(x_1, \dots, x_n; \theta)$ ; isto é,  $Q$  é função de  $x_1, x_2, x_3, \dots, x_n$  e  $\theta$ . Se  $Q$  tem distribuição que não depende de  $\theta$ , então  $Q$  é chamada de estatística pivotal. Em inglês Pivotal statistic.

**ESTATÍSTICA PRESS:** Medida de validação obtida eliminando-se cada observação, uma por vez, e prevendo-se esse valor dependente com o modelo de regressão estimado a partir das demais observações.

**ESTATÍSTICA Q DE PRESS:** Medida do poder classificatório da função discriminante quando comparada com os resultados esperados de um modelo de chances. O valor calculado é comparado com um valor crítico baseado na distribuição qui-quadrado. Se o valor calculado exceder o valor crítico, os resultados da classificação serão significativamente melhores do que se esperaria do acaso.

**ESTATÍSTICA ROBUSTA:** É o conjunto de técnicas utilizadas para atenuar o efeito de outliers e preservar a forma de uma distribuição tão aderente quanto possível aos dados empíricos. Em inglês statistics.

**ESTATÍSTICA SUFICIENTE:** Qualquer estatística que possui a propriedade de suficiência. Em inglês Sufficient statistic.

**ESTATÍSTICA t:** Teste estatístico que avalia a significância estatística da diferença entre dois grupos em uma única variável dependente. Ver teste t.

**ESTATÍSTICA T:** Estatística usada para testar a hipótese nula de que o verdadeiro valor de um coeficiente na análise de regressão é sempre zero.

**ESTATÍSTICA TEÓRICA:** O mesmo que estatística matemática.

**ESTATÍSTICA TESTE:** É o valor amostral da estatística utilizada para testar um parâmetro no teste de hipóteses. Em inglês Test statistic.

**ESTATÍSTICA TERRITORIAL:** É, de acordo com a praxe estabelecida por grande número de publicações, a estatística aplicada que têm por objeto a descrição do território em geral politicamente definido, sob seus aspectos físico, político e administrativo, ainda mesmo que alguns dos fatos aí incluídos não correspondam a fenômenos estatísticos, no rigor da definição destes. O Anuário Estatístico do Brasil adota o título situação física.

**ESTATÍSTICA TESTE:** É o valor amostral da estatística utilizada para testar um parâmetro no teste de hipóteses.

**ESTATÍSTICA U:** Ver lambda de Wilks.

**ESTATÍSTICA V:** Generalização, para mais populações, da estatística  $D^2$  de Mahalanobis que traduz a distância de duas populações.

**ESTATÍSTICA VITAL:** O mesmo que bioestatística.

**ESTATÍSTICA WALD:** Teste usado em regressão logística para a significância do coeficiente logístico. Sua interpretação é semelhante aos valores F ou t usados para o teste de significância de coeficientes de regressão.

**ESTATÍSTICA Z:** A conversão de desvios padrão calculados em unidades padrão de distância em relação à média em distribuições normais.

**ESTATÍSTICA  $\chi^2$ :** Valor que a variável qui quadrado  $\chi^2$  assume para uma dada amostra.

**ESTATÍSTICAS:** Conjunto dos métodos para planejar experimentos, obter dados, organizar, resumir, apresentar, analisar, interpretar dados e tirar conclusões neles baseados.

**ESTATÍSTICAS:** Valores numéricos das amostras: média, mediana, variância, coeficiente de assimetria, dentre outros, constituindo nas amostras probabilísticas estimativas não enviesadas dos parâmetros, conforme é demonstrado pelo Teorema do Limite Central.

**ESTATÍSTICO:** Que diz respeito à estatística. Especialista em estatística geral, especial, ou, ainda, aplicada. A literatura científica nacional registra também os termos estaticista e estatisticista com o fim evidente de distinguir a coisa da pessoa, conforme se dá, para outros idiomas, como exemplo podemos citar optical e optician, physique e physicien, dentre outros; expressões estas, que também não possuímos.

**ESTATISTICAMENTE SIGNIFICANTE:** Dizer que um resultado é estatisticamente significativo significa que as diferenças encontradas são grandes o suficiente para não serem atribuídas ao acaso. Uma diferença

estatisticamente significativa pode não ser clinicamente importante; a importância em termos biológicos não deve ser julgada pelos estatísticos, mas sim pelos profissionais da área em que a pesquisa está sendo feita.

**ESTEREOGRAMA:** Representação a três dimensões quer, ou se constrói por meio de um material rígido, ou se desenha à custa de um processo qualquer de redução ao plano, mais comumente por perspectiva isométrica. Termo do grego στερεος sólido, e γραμμη linha, desenho.

**ESTEREOGRAMAS:** São gráficos geométricos dispostos em três dimensões, pois representam volume. São usados nas representações gráficas das tabelas de dupla entrada. Em alguns casos este tipo de gráfico fica difícil de ser interpretado dado à pequena precisão que oferecem.

**ESTIMAÇÃO:** Estimação está relacionada com a inferência sobre o valor numérico de valores desconhecidos da população baseados em dados incompletos como uma amostra. Se um único valor for calculado para cada parâmetro desconhecido, o processo é chamado estimação pontual. O processo de estimação por intervalos consiste em calcular um intervalo de forma que, com certo grau de certeza, ele contenha o parâmetro de interesse. Em inglês Estimation.

**ESTIMAÇÃO:** Parte da inferência estatística que trata do processo de estimação e das propriedades dos estimadores. Em inglês estimation.

**ESTIMAÇÃO:** Determinação do valor de um parâmetro populacional ou dos parâmetros de um modelo probabilístico, com base num conjunto de observações ou dados. Exemplos de estimação: altura média de um povoamento, coeficientes de regressão de um modelo linear e não linear, como por exemplo um modelo de relação hipsométrica, de equação de volume, equação de produção ou modelos de crescimento. Ver também predição.

**ESTIMAÇÃO:** Ato de se obter uma estimativa.

**ESTIMAÇÃO DE BAYES:** A estimação de parâmetros da população pelo uso de métodos Bayesianos.

**ESTIMAÇÃO DE BAYES (PROCESSO DE):** De um parâmetro incógnito, é o que consiste em, perante uma amostra de valores, oriundos da população por ele caracterizada, adotar como seu valor o que tornar máxima a probabilidade a posteriori desse parâmetro.

**ESTIMAÇÃO DE MÁXIMA VEROSSIMILHANÇA (MLE):** Método de estimação comumente empregado em modelos de equações estruturais, incluindo LISREL e EQS. Urna alternativa aos usuais mínimos quadrados em regressão múltipla, MLE é um procedimento que melhora por iterações as estimativas de parâmetros para minimizar uma função de ajuste especificada. Esta técnica é considerada a técnica mais robusta de estimação de parâmetros. Em inglês o termo é conhecido como maximum likelihood estimation.

**ESTIMAÇÃO DE PARÂMETRO:** Procedimento estatístico utilizado para aproximar o valor verdadeiro de um parâmetro proveniente de uma população. Utiliza dados amostrais. Em inglês Parameter estimation.

**ESTIMAÇÃO DE PARÂMETRO:** Parte da inferência estatística, cujo procedimento indutivo consiste em generalizar os valores numéricos amostrais para o universo investigado.

**ESTIMAÇÃO DIRETA:** Processo no qual um modelo é diretamente estimado com um procedimento de estimação selecionado e o intervalo de confiança, bem como o erro padrão de cada estimativa de parâmetro é baseado no erro amostral.

**ESTIMAÇÃO DO INTERVALO:** Uma estimativa de um parâmetro da população que fornece um intervalo que se acredita contenha o valor do parâmetro.

**ESTIMAÇÃO POR INTERVALOS:** Estimação de um parâmetro populacional mediante a especificação de um intervalo cujos limites superior e inferior contêm o parâmetro. Em inglês Interval estimation.

**ESTIMAÇÃO POR INTERVALO:** É todo processo de estimação de um ou mais parâmetros que resulta em determinar intervalos de mais de um ponto, dos quais cada um dos parâmetros respectivamente, é elemento. Distingue de estimação por ponto.

**ESTIMAÇÃO POR PONTO:** É todo processo de estimação de um ou mais parâmetros que resulta em atribuir um só valor para cada um deles. Distingue de estimação por intervalo.

**ESTIMAÇÃO POR PONTOS:** É todo processo de estimação de um ou mais parâmetros que resulta em atribuir um só valor para cada um deles. Distingue de estimação por intervalo.

**ESTIMAÇÃO SEQUENCIAL:** Processo de estimação baseado em amostragem sequencial.

**ESTIMAÇÃO SIMULTÂNEA:** De dois ou mais parâmetros de uma mesma população é a que se baseia em uma só amostra dessa população.

**ESTIMAÇÃO SIMULTÂNEA:** Estimação da(s) função(ões) discriminante(s) ou do modelo de regressão logística em um único passo, no qual pesos para todas as variáveis independentes são calculados simultaneamente. Contrasta com a estimação stepwise, na qual as variáveis independentes entram sequencialmente de acordo com o poder discriminante.

**ESTIMAÇÃO STEPWISE:** Processo de estimação de função(ões) discriminante(s) ou do modelo de regressão logística no qual variáveis independentes entram sequencialmente de acordo com o poder discriminatório que elas acrescentam à previsão de pertinência no grupo.

**ESTIMAÇÃO STEPWISE:** Método de seleção de variáveis para inclusão no modelo de regressão que começa selecionando o melhor preditor da variável dependente. Variáveis independentes adicionais são selecionadas em termos do poder explicativo incremental que podem acrescentar ao modelo de regressão. Variáveis independentes são acrescentadas à medida que seus coeficientes de correlação parcial são estatisticamente significantes. Variáveis independentes também podem ser eliminadas se seu poder preditivo cair para um nível não significativo quando uma outra variável independente for acrescentada ao modelo.

**ESTIMADOR:** Um estimador é uma regra ou método de estimar uma constante de uma população. É geralmente expressado como uma função de valores de uma amostra e, conseqüentemente, uma variável cuja distribuição é de grande importância no cálculo da confiabilidade da estimativa a ser calculada a partir dele. Em inglês Estimator.

**ESTIMADOR:** É a estatística que permite inferir o parâmetro desconhecido da população, a partir dos dados de amostragem.

**ESTIMADOR:** Estatística amostral, como a média amostral  $\bar{X}$ , utilizado para aproximar um parâmetro populacional.

**ESTIMADOR:** Grandeza baseada em observações amostrais, cujo valor serve de indicador do valor de um parâmetro populacional desconhecido, por exemplo, a média amostral  $\bar{X}$  é usada como estimador da média populacional desconhecida  $\mu$ .

**ESTIMADOR:** Expressão algébrica utilizada para se obter a estimativa.

**ESTIMADOR:** Função dos membros de uma amostra que permite avaliar ou calcular um parâmetro da população de onde ela se originou. Geralmente, a esta função está associada uma medida da confiança que se pode atribuir à validade da avaliação ou cálculo realizado. A ideia é que o verdadeiro valor está sendo obtido a partir do valor calculado, dentro dos limites de variação da amostragem.

**ESTIMADOR:** De um parâmetro de dada população, é toda função de elementos de amostra oriunda dessa população que mantém para com o parâmetro uma certa relação probabilística.

**ESTIMADOR:** A estatística, fórmula ou expressão utilizada para avaliar o valor de um parâmetro. Um estimador é uma variável aleatória. Também conhecido como estatística amostral ou simplesmente estatística. Em inglês estimator.

**ESTIMADOR:** Procedimento utilizado para se estimar parâmetros populacionais ou os parâmetros de um modelo probabilístico. Exemplos: a média amostral é um estimador da média populacional; o procedimento padrão em regressão linear é utilizar os estimadores de quadrados mínimos para se obter estimativas dos coeficientes de regressão do modelo linear. Uma vez que os valores gerados por um estimador depende dos dados observados ou amostra, os estimadores devem ser considerados variáveis aleatórias. Um estimador é chamado de enviesado ou viciado quando a esperança matemática do estimador não for matematicamente igual ao parâmetro que ele busca estimar. A diferença entre o valor do parâmetro a ser estimado e a esperança matemática do estimador é chamada de viés ou vício.

**ESTIMADOR ABSOLUTAMENTE NÃO-VIESADO:** É o estimador não viesado seja qual for a distribuição da variável cujo parâmetro ele pretende estimar.

**ESTIMADOR ASSINTOTICAMENTE NÃO-VIESADO:** Baseado numa amostra de tamanho  $n$  é aquele cujo viés tende a zero para  $n$  tendente a infinito.

**ESTIMADOR CONSISTENTE:** Um estimador é dito consistente se ele converge em probabilidade, quando o tamanho de amostra cresce, ao parâmetro do qual é estimador. Em inglês Consistent estimator.

**ESTIMADOR CONSISTENTE:** Um estimador diz-se consistente quando converge em probabilidade para o estimador ao fazer-se aumentar o tamanho da amostra, ou seja, assume valores concentrados tão perto quanto se quiser do parâmetro.

**ESTIMADOR CONSISTENTE:** Estimador que tende a convergir para o verdadeiro valor quando o tamanho da amostra aumenta.

**ESTIMADOR DE BAYES EMPÍRICO:** Um estimador derivado de um procedimento de Bayes empírico. Em inglês Empirical Bayes estimator.

**ESTIMADOR DE MÁXIMA VEROSSIMILHANÇA GENERALIZADA:** Uma aproximação para a estimação do parâmetro envolvendo o conceito de eficiência assintótica proposto por Weiss e Wolfowitz



(1966) para remover alguns dos problemas associados com os estimadores de máxima verossimilhança clássicos. Em inglês Generalized maximum likelihood estimator.

**ESTIMADOR DE MÁXIMA VEROSSIMILHANÇA:** Qualquer estimador de um parâmetro obtido pelo método de máxima verossimilhança. Em inglês Maximum likelihood estimator (MLE).

**ESTIMADOR DE MÁXIMA VEROSSIMILHANÇA:** Um estimador com a seguinte propriedade: se o verdadeiro parâmetro desconhecido tem este valor, então a probabilidade de obter a amostra efetivamente observada é máxima.

**ESTIMADOR DE MÍNIMOS QUADRADOS: É aquele estimador que torna mínima a soma dos quadrados dos erros ou resíduos. Em inglês least square estimator.**

**ESTIMADOR DE MOMENTOS:** Estimador de um parâmetro desconhecido de uma distribuição obtido pelo método dos momentos. Em inglês Moment estimator.

**ESTIMADOR DE RAZÃO:** Designa, na teoria de amostragem, o estimador que estima a média populacional da variável de interesse por meio de da razão da média amostral da variável de interesse pela média amostral de uma variável auxiliar. A estimativa da média populacional da variável de interesse é dada pelo produto da razão pela média populacional da variável auxiliar, a qual é conhecida sem erro amostral ou com erro desprezível. O exemplo clássico do estimador de razão em levantamentos florestais é a árvore média, isto é, a razão do volume de madeira numa parcela pelo número de árvores na parcela. Conhecendo-se o número de árvores num povoamento, por meio de censo, o volume do povoamento pode ser obtido fazendo-se produto da árvore média que é o estimador de razão e o número de árvores no povoamento. O estimador de razão é um estimador enviesado, sendo que o limite superior do viés relativo é o coeficiente de variação da média amostral da variável auxiliar. Ver também: estimador.

**ESTIMADOR DE MÍNIMOS QUADRADOS GENERALIZADOS:** Método proposto por Aitken (1934) para estimar os  $k$  parâmetros do vetor  $\pi$  em uma equação linear quando os distúrbios não são independentes, porém, sua matriz de variância-covariância é conhecida. Em inglês Generalized least squares estimator.

**ESTIMADOR DE REGRESSÃO:** Em amostragem, designa o estimador que estima a média populacional de uma variável de interesse a partir de um modelo linear simples da média populacional de uma variável auxiliar. Os coeficientes de regressão do modelo são obtidos a partir das observações amostrais da variável de interesse e da variável auxiliar, que devem ser registradas de modo pareado nas mesmas unidades amostrais. A aplicação do estimador de regressão assume o conhecimento da média populacional da variável auxiliar e é um estimador enviesado ou viciado.

**ESTIMADOR EFICIENTE:** Um estimador é dito eficiente se ele é de variância mínima. Em inglês Efficient estimator.

**ESTIMADOR EQUIVARIANTE AFIM:** Um estimador é dito afim equivariante se:  $T\{(x_i, y_i); i = 1, 2, \dots, n\} = A^{-1}T\{(x_i, y_i); i = 1, 2, \dots, n\}$ , onde  $A$  é uma matriz quadrada, não singular.

**ESTIMADOR EQUIVARIANTE DE REGRESSÃO:** Um estimador de regressão  $T$  é dito equivariante de regressão se:  $T\{(x_i, y_i + x_i v); i = 1, 2, \dots, n\} = T\{(x_i, y_i); i = 1, 2, \dots, n\} + v$ , onde  $v$  é um vetor qualquer.

**ESTIMADOR EQUIVARIANTE DE ESCALA:** Um estimador é dito equivariante de escala se:  $T\{(x_i, c y_i); i = 1, 2, \dots, n\} = c T\{(x_i, y_i); i = 1, 2, \dots, n\}$ , onde  $A$  é uma matriz quadrada, não singular.

**ESTIMADOR GRÁFICO:** Constante escolhida por tentativa e erro da representação geométrica, geralmente para manter um desenho linear de observações dentro de um papel pautado específico. Em inglês Graphical estimator.

**ESTIMADOR LINEAR:** Estimador que é função linear das observações. Em inglês Linear estimator.

**ESTIMADOR MAIS EFICIENTE:** Um estimador não viciado cuja variância não é maior que a de qualquer outro estimador não viciado é chamado estimador mais eficiente. Em inglês Most efficient estimator.

**ESTIMADOR NÃO VICIADO:** Estimador cujo vício é igual a zero. Em inglês Unbiased estimator.

**ESTIMADOR NÃO TENDENCIOSO:** Estatística amostral que tende a visar o parâmetro populacional estimado.

**ESTIMADOR NÃO TENDENCIOSO:** Um estimador cujo valor esperado é igual ao verdadeiro valor do parâmetro que procura estimar.

**ESTIMADOR OU ESTIMATIVA DO PARÂMETRO  $a$ :**  $\hat{a}$ .

**ESTIMADORES:** São parâmetros calculados a partir de uma amostra, que servem como aproximações de parâmetros desconhecidos da população. A notação de praxe faz uso do símbolo com um acento circunflexo ( $\hat{\cdot}$ ). Exemplos:  $\sigma$  = desvio-padrão da população e  $\hat{\sigma}$  = desvio padrão estimado da população. A média da amostra ( $\bar{X}$ ) é o estimador comum para a média  $\mu$  da população. Portanto,

$$\hat{\mu} = \bar{X} = \frac{\sum_{i=1}^n X_i}{n}, \text{ onde } x_i \text{ são os valores e } n \text{ é o número de elementos da amostra.}$$

**ESTIMADORES DE MÍNIMOS QUADRADOS:** Estimadores obtidos pelo método de mínimos quadrados. Em inglês Least squares estimators.

**ESTIMADORES DE MÍNIMOS QUADRADOS:** Estimativas para os coeficientes da análise de regressão; estas estimativas são calculadas porque minimizam o quadrado do erro entre os valores efetivos da variável dependente e os valores preditos pela equação de regressão.

**ESTIMADORES  $L_1$ :** São estimadores que se baseiam na minimização da soma dos valores absolutos dos resíduos, isto é, minimizam a função dos resíduos  $L_1 = \sum_{i=1}^n |r_i|$ , onde  $n$  é o número de observações e  $r_i = Y_i - X_i \beta$  é o  $i$ -ésimo resíduo.

**ESTIMADORES ROBUSTOS:** Denominam-se estimadores robustos, os estimadores de regressão que possuem certas propriedades que os permitem estimações confiáveis mesmo quando os dados possuem um determinado percentual de contaminação, isto é, são estimadores robustos a desvios dos pressupostos básicos do método de mínimos quadrados.

**ESTIMATIVA:** De forma restrita, uma estimativa é um valor particular obtido para um estimador em determinada amostra. Em inglês Estimate.

**ESTIMATIVA:** Valor aproximado de um parâmetro populacional desconhecido, calculado a partir de uma amostra.

**ESTIMATIVA:** Valor específico ou intervalo de valores utilizado para aproximar um parâmetro populacional.

**ESTIMATIVA:** É o valor particular de um estimador, isto é, é o resultado de algum cálculo realizado sobre valores amostrais. Em inglês estimate.

**ESTIMATIVA:** Valor de um estimador calculado a partir de uma amostra.

**ESTIMATIVA:** Valor de um estimador em uma circunstância particular.

**ESTIMATIVA:** É a avaliação de uma determinada grandeza feita com base em dados de amostragem, incluindo-se neste processo a estimativa ou interpolação de valores em pontos não amostrados.

**ESTIMATIVA (OU ESTIMATIVA PONTUAL):** É o valor que a estatística ou estimador toma em uma amostra determinada.

**ESTIMATIVA COMBINADA DE :** Estimativa da variância comum a duas populações, obtida calculando-se a média ponderada das duas variâncias amostrais.

**ESTIMATIVA COMBINADA DE  $p_1$  e  $p_2$ :** Probabilidade obtida pela combinação dos dados amostrais, dividindo-se o número total de sucessos pelo número total de observações.

**ESTIMATIVA DE PONTO:** Inferência relativa ao erro de amostragem associado a uma determinada estimativa de valor de uma população.

**ESTIMATIVA INTERVALAR:** Estimativa de um parâmetro que dá um intervalo de valores possíveis para o mesmo, em vez de um único valor. Ver também intervalo de confiança.

**ESTIMATIVA PONTUAL:** Valor único utilizado como estimador de um parâmetro desconhecido. Em contraste, ver estimativa intervalar.

**ESTIMATIVA PONTUAL:** Valor isolado que serve como estimativa de um parâmetro populacional.

**ESTIMATIVA PONTUAL:** Fornece um único valor, calculado com os dados da amostra, que representa o parâmetro populacional; por exemplo, a prevalência de hipertensão, na amostra, ser de 10%.

**ESTIMATIVA POR INTERVALO (OU INTERVALAR):** Fornece um intervalo de possíveis valores, no qual admite-se que, provavelmente, o parâmetro populacional esteja contido; por exemplo, a proporção de hipertensos estar entre 8% e 12%. Ver erro de amostragem e intervalo de confiança.

**ESTIMATIVA POR INTERVALO DE CONFIANÇA:** A estimativa por intervalo do valor médio de  $y$  para um dado valor de  $x$ .

**ESTIMATIVA POR INTERVALO DE PREVISÃO:** A estimativa por intervalo de um valor individual de  $y$  para um dado valor de  $x$ .

**ESTIMATIVA POR PONTO:** O valor da estatística amostral que corresponde ao parâmetro populacional. Em inglês point estimate.

**ESTIMATIVAS TRANSGRESSORAS:** Qualquer valor que excede seus limites teóricos. As ocorrências mais comuns são os casos Heywood com variâncias negativas de erro, em que o valor mínimo deveria ser zero, indicando nenhum erro de mensuração, ou erros padrão muito grandes. O pesquisador deve corrigir a estimativa transgressora com uma, entre várias ações corretivas, antes que os resultados possam ser

interpretados para o ajuste geral do modelo e os coeficientes individuais possam ser examinados quanto à significância estatística.

**ESTRATÉGIA DE PESQUISA:** Termo também conhecido como delineamento de pesquisa, refere-se ao conjunto de decisões que o pesquisador tem que tomar ao longo do período da realização da pesquisa, em função das características particulares do fenômeno que está investigando e de outros fatores que afetam seu trabalho, como por exemplo, o tempo programado para sua realização ou o cronograma de execução da pesquisa, o orçamento, etc.

**ESTÍMULO:** Em experimento, evento cujo impacto está sendo testado. Em pesquisa médica, pode ser uma nova droga. Em criminologia, um programa de aconselhamento. O estímulo experimental corresponde à variável independente da análise de survey.

**ESTÍMULO:** Conjunto específico de níveis, sendo um por fator avaliado por respondentes o qual também conhecido como um tratamento. Um método de definir estímulos em um experimento fatorial é obtido considerando-se todas as combinações de todos os níveis entre os diversos fatores. Por exemplo, três fatores com dois níveis cada criariam oito ( $2 \times 2 \times 2$ ) estímulos ou tratamentos. No entanto, em muitas análises conjuntas, o número total de combinações é muito grande para um respondente avaliar todas. Nesses casos, alguns subconjuntos de estímulos são criados de acordo com um plano sistemático, geralmente um delineamento fatorial fracional.

**ESTÍMULOS DE VALIDAÇÃO:** Conjunto de estímulos que não são usados na estimação de utilidades parciais. As utilidades parciais estimadas são então usadas para prever a preferência pelos estímulos de validação para avaliar a validade e a confiabilidade das estimativas originais. É semelhante em conceito à amostra de validação de respondentes em análise discriminante.

**ESTOCÁSTICA:** Diz-se da união da estatística com a probabilidade.

**ESTOCÁSTICO:** Ao acaso. Não é determinado de antemão. Em inglês stochastic.

**ESTRATIFICAR:** Termo usado para designar a divisão de uma população em estratos, subgrupos homogêneos, amostras, subpopulações mais ou menos parecidas, etc.

**ESTRATO:** Termo usado para denominar os subgrupos homogêneos em que se subdivide uma população heterogênea; o mesmo que amostra.

**ESTRATÉGIA DE DESENVOLVIMENTO DE MODELO:** Estratégia de modelagem estrutural que incorpora reespecificação de modelo como um método teoricamente orientado de melhoria de um modelo especificado empiricamente. Isso permite exploração de formulações alternativas de modelos que podem ser apoiadas por teoria. Não corresponde a uma abordagem exploratória na qual reespecificações de modelo são feitas sem base teórica.

**ESTRATÉGIA DE MARKETING:** Orientar o uso a longo prazo dos recursos de uma empresa com base em suas capacidades existentes e projetadas e nas mudanças projetadas no ambiente externo.

**ESTRATÉGIA DE MODELAGEM CONFIRMATÓRIA:** Estratégia ou metodologia que avalia estatisticamente um único modelo quanto ao seu ajuste aos dados observados. Essa abordagem é realmente menos rigorosa do que a estratégia de modelos concorrentes, pois ela não considera modelos alternativos que possam se ajustar melhor ou tão bem como o modelo proposto.

**ESTRATÉGIA DE MODELOS CONCORRENTES:** Estratégia que compara o modelo proposto com alguns modelos alternativos em uma tentativa de demonstrar que não existe melhor modelo de ajuste. Isso é particularmente relevante em modelagem de equações estruturais porque um modelo pode ter apenas ajuste aceitável, mas somente ajuste aceitável não garante que um outro modelo não se ajustará de melhor forma ou tão bem quanto.

**ESTRATÉGIA DE COLETA DE DADOS:** Geralmente, as pesquisas possuem duas técnicas de coleta de dados: a primeira refere-se ao local onde os dados são coletados e, neste item, existem duas situações que são o laboratório ou o campo. Quando os dados são coletados em situação controlada, o ambiente de coleta é chamado de laboratório. Quando não ocorre este controle, a pesquisa é conhecida com o de campo. A segunda situação refere-se à fonte dos dados, as quais podem ser documental ou de campo. Sempre que uma pesquisa se utiliza apenas de documentos tais como, revistas, livros, arquivos, arquivos em mídia eletrônica, documentos oficiais, etc., diz-se que a pesquisa possui estratégia documental. Quando a pesquisa não se restringe à utilização de documentos, mas também se utiliza de pessoas ou não, diz-se que a pesquisa possui situação de campo.

**ESTRATIFICAÇÃO:** Agrupamento de unidades que formam uma população em grupos homogêneos ou estratos antes da amostragem. Este procedimento, que pode ser usado juntamente com amostragens aleatória simples, sistemática ou por conglomerado, melhora a representatividade da amostra, pelo menos em termos das variáveis de estratificação.

**ESTRATIFICAÇÃO:** O processo de divisão da população ou da amostra em estratos ou subgrupos. Em uma investigação, ela pode ser feita, a priori, na fase de recrutamento dos participantes, ou a posteriori, durante a fase de análise de dados; em geral, com a finalidade de neutralizar variáveis de confusão.

**ESTRATIFICAÇÃO:** Processo de classificar os dados em subgrupos baseados em características ou categorias. Estratificar é dividir as informações ou dados em grupos ou estratos, constituindo-se numa ferramenta para a busca das causas ou origens de um problema. Os dados devem ser agrupados por tempo, local, tipo, sintoma e outros fatores. A estratificação é fundamental para a construção do gráfico de Pareto, sendo uma das sete ferramentas do controle da qualidade.

**ESTRATIFICAÇÃO PROGNÓSTICA:** O mesmo que alocação estratificada.

**ESTRUTURA DE AMOSTRAGEM:** Relação dos elementos da população dos quais se selecionam unidades a serem amostradas.

**ESTRUTURA DE PREFERÊNCIA:** Representação da importância ou utilidade relativa de cada fator e do impacto de níveis individuais que afetam a utilidade.

**ESTUDO ANALÍTICO:** Investigação que tem por objetivo examinar associação de eventos. Por exemplo, entre obesidade e arteriosclerose. Em geral, formula-se uma hipótese que relaciona os dois eventos e procura-se esclarecer se a associação é casual. Comparar com estudo descritivo.

**ESTUDO ANTES-DEPOIS (OU ANTES E DEPOIS):** Investigação na qual os indivíduos, submetidos a determinado tratamento ou exposição, são comparados com os de um período anterior, cujo tratamento ou exposição é diferente. Por exemplo, a análise das estatísticas de mortalidade por acidentes de trânsito, antes e depois de uma lei para limitar a velocidade dos veículos; os pacientes com diagnóstico comprovado de raiva, e sob um novo tratamento, são comparados com os resultados obtidos anteriormente, quando

a letalidade era invariavelmente de 100%. Se o tratamento não é mortal e o tratamento não cura definitivamente, pode-se estudar os mesmos indivíduos em dois períodos, anterior e posterior; a vantagem é a maior comparabilidade dos grupos, em confronto com estudos em que os indivíduos são diferentes nos dois períodos.

**ESTUDO CASO-CONTROLE:** Estudo no qual os fatores de risco de pessoas com uma doença são comparados com aqueles sem a doença.

**ESTUDO CASO-CONTROLE:** Comparação entre um grupo de doentes ou casos e um grupo de pessoas não doentes ou controles. O objetivo é verificar se os casos diferem significativamente dos controles, em relação à exposição a um dado fator de risco.

**ESTUDO COHORTE:** Também chamado de follow-up, tem por essência identificar um grupo de indivíduos que serão acompanhados para ver o que acontece com eles. Em inglês Cohort study.

**ESTUDO COMPARATIVO:** Estudo que envolve dois ou mais grupos de pacientes para comparar e julgar a influência de algum fator, condição, característica, ou procedimento, presente ou aplicado a um dos grupos, mas não ao outro. Sinônimo de ensaio clínico se o estudo exige a comparação de tratamentos diferentes que envolvam pacientes tratados no mesmo período de tempo.

**ESTUDO CONTEXTUAL:** Survey no qual se coletam dados para descrever as condições nas quais os respondentes funcionam. Pode-se fazer um survey com estudantes e também colher dados sobre suas escolas.

**ESTUDO COORTE:** Estudo no qual um grupo específico é estudado ao longo do tempo, embora se possa coletar dados de membros diferentes do grupo em cada conjunto de observações. Exemplo: um estudo da história ocupacional da classe de 2000, com questionários enviados a cada quinquênio ou cinco anos.

**ESTUDO COORTE:** Estudo que envolve a identificação de um grande número de pessoas denominado de coorte, algumas expostas a um fator causal suspeito, outras não expostas a esse fator. Essas pessoas são acompanhadas durante um período de tempo relativamente longo para verificar se ocorreu ou não um resultado ou condição de interesse, normalmente doença ou morte. Depois se comparam as proporções de ocorrências de doença ou morte nos dois grupos, isto é nas pessoas expostas ao fator causal suspeito e nas não expostas. Também chamado estudo prospectivo.

**ESTUDO COMPOSTO:** Combinação de eventos simples.

**ESTUDO CONTROLADO:** Investigação em que são tomados cuidados para neutralizar fatores de confundimento; em que há grupo-controle formado durante o desenrolar da pesquisa, ou seja, com formação de grupo interno para comparação de resultados. Exemplo: estudo de caso-controle.

**ESTUDO CROSS-SECTIONAL:** É o estudo em que os indivíduos são observados uma única vez. Em inglês Cross-sectional study.

**ESTUDO CRUZADO (OU DE CRUZAMENTO; CROSSOVER DESIGN):** Método para comparar dois ou mais tratamentos, no qual os participantes, após completar um tratamento, passam a receber o outro tratamento. A alocação aleatória, em geral, é empregada, assim como a avaliação duplo-cega. Quando os participantes mudam de um para outro tratamento, há uma fase de limpeza ou wash out.



**ESTUDO DE CASO:** O mesmo que relato de caso; investigação aprofundada da doença, em um paciente; utiliza-se a expressão estudo de casos quando estão incluídos poucos pacientes, ou para alguns elementos, menos de uma dezena de casos. Ver série de casos.

**ESTUDO DE CASO-CONTROLE:** Neste tipo de estudo identifica-se um grupo de indivíduos ou caso com uma condição de interesse, e um grupo não afetado ou controle, sendo feitas comparações à exposição a um fator ou fatores em estudo. Em inglês Case control study.

**ESTUDO DE CASO:** Estudo conduzido com um único elemento que pode ser uma pessoa, uma cidade, uma empresa, um evento, etc.. Embora frequentemente exista a pressuposição de que este sujeito único possa ser considerado típico, raramente os resultados obtidos a partir desse tipo de estudo podem ser generalizados. A escolha do caso deve ter como base justificativas estabelecidas de forma bem consolidada e o processo de coleta de dados deve se utilizar de múltiplas estratégias, como por exemplo, de documentos, entrevistas, filmes, etc.

**ESTUDO DE CASO-CONTROLE:** Estudo que envolve a identificação de pessoas com uma doença ou condição de interesse ou casos e de um grupo comparável de pessoas sem a doença ou condição de interesse ou controles. Casos e controles são comparados com respeito a algum atributo existente, passado ou de exposição que se acredita esteja relacionado à doença ou condição. Também chamado estudo retrospectivo.

**ESTUDO DE CASO-CONTROLE:** Particular forma de investigação etiológica, de cunho retrospectivo; parte-se do efeito em busca das causas. Grupos comparáveis de indivíduos, com e sem um determinado agravo à saúde, são comparados com respeito a exposições que sofreram no passado, de modo a testar a hipótese de que tais fatores de risco são causas contribuintes da doença. Por exemplo, pacientes acometidos de câncer de estômago são comparados com igual número de indivíduos ou grupo controle, de mesmo sexo e idade, mas sem câncer de estômago; o passado de alcoolismo, o fator de risco em avaliação, é determinado nos dois grupos. Comparar com estudo de coorte.

**ESTUDO DE COORTE:** Estudo realizado com um mesmo grupo de pessoas ou coorte ao longo do tempo, embora não necessariamente acompanhando as mesmas pessoas.

**ESTUDO EPIDEMOLÓGICO:** Estudo que tem por finalidade estudar a existência e as características das doenças nas populações.

**ESTUDO CASO-CONTROLE:** Estudo capaz de abordar hipóteses etiológicas, produzindo medidas de incidências e, por conseguinte, medidas diretas de risco. A maioria dos estudos de coorte parte da observação de grupos comprovadamente expostos a um fator de risco suposto como causa de doença a ser detectada no futuro que é a prospectoanálise. Assim, a coorte constitui-se em um grupo de pessoas consideradas sadias quanto à doença sob investigação e que se caracteriza pela composição homogênea por vários fatores que não a variável independente investigada. Sinônimos: estudo prospectivo, de seguimento ou follow-up.

**ESTUDO DE CASO-CONTROLE ANINHADO:** É o estudo de caso-controle realizado dentro de um outro tipo de investigação, por exemplo, de um estudo de coorte.

**ESTUDO DE COORTE:** Comparação entre um grupo exposto a um fator de risco e outro grupo não exposto. Visa verificar se indivíduos expostos ao fator de risco desenvolvem a doença em questão, em maior ou menor proporção, do que um grupo de indivíduos não expostos.

**ESTUDO DE COORTE:** Estudo no qual um grupo claramente identificado é caracterizado pela exposição e é seguido para verificação do desfecho.

**ESTUDO DE COORTE:** Um estudo longitudinal com o mesmo conjunto ou o coorte de pessoas ao longo de um determinado tempo. Em inglês cohort study.

**ESTUDO DE COORTE:** Particular forma de investigação de fatores etiológicos; parte-se da causa em busca dos efeitos. Um grupo de pessoas é identificado, e a informação pertinente sobre a exposição de interesse é coletada, de modo que o grupo pode ser seguido, no tempo, verificado os que desenvolvem e os que não desenvolvem a doença em foco, e se esta exposição prévia está relacionada à ocorrência da doença. No caso mais simples de comparação, o subgrupo de indivíduos expostos a um fator de risco é comparado como o subgrupo de não-expostos a este mesmo fator. A incidência do agravo à saúde, nos dois grupos, é comparada. Por exemplo, indivíduos expostos a um acidente radioativo são comparados com controles não-expostos à radiação; a incidência de leucemia é determinada para cada grupo. Existem dois tipos de estudos de coorte: prospectivo e retrospectivo ou histórico. Comparar com estudo de caso-controle.

**ESTUDO DE COORTE CONCORRENTE:** Refere-se ao estudo no qual a coorte é acompanhada desde o momento de exposição, procedendo-se ao monitoramento e ao registro dos casos de doença ou de óbito à medida que esses ocorrem até a data prevista para encerramento das observações. Nesse tipo de estudo o encaminhamento da pesquisa e o fenômeno pesquisado que pode ser a doença progridem em paralelo, concomitantemente.

**ESTUDO DE COORTE NÃO-CONCORRENTE:** Refere-se ao estudo que é efetivado com a identificação de coortes em algum ponto do passado, com a seleção e a classificação dos seus elementos nesse ponto e com o início e o fim do acompanhamento antes do momento da pesquisa.

**ESTUDO DE COORTE PROSPECTIVO:** O investigador reúne os grupos de estudo no presente com base na exposição, coleta a base de dados a eles referente e acompanha os sujeitos ao longo do tempo para verificar a ocorrência do desfecho de interesse.

**ESTUDO DE COORTE RETROSPECTIVO:** As coortes são identificadas com base na exposição passada e acompanhada até o presente para verificar a ocorrência do desfecho de interesse.

**ESTUDO DE COORTE RETROSPECTIVO (OU HISTÓRICO):** Trata-se de uma pesquisa sobre eventos passados, conservando-se o princípio do estudo de coorte, isto é, da causa em direção ao efeito e com grupo-controle; o efeito já ocorreu quando a investigação é realizada.

**ESTUDO DE COORTE TRANSVERSAL:** Ver estudo de prevalência.

**ESTUDO DE FOLLOW-UP:** Ver estudo de coorte.

**ESTUDO DE OBSERVAÇÃO:** Investigação de situações que ocorrem naturalmente, sem a interferência do investigador; observação opõe-se a experimentação.

**ESTUDO EXPLORATÓRIO:** Termo usado para designar o tipo de estudo que tem por finalidade aumentar o conhecimento do comportamento de um fenômeno pouco conhecido, ou de um problema de pesquisa ainda não estudado ou pouco explorado. Pode ser chamado também de estudo preliminar, estudo prospectivo.

**ESTUDO LONGITUDINAL:** Estudo no qual se acompanha o comportamento de uma ou mais variáveis num mesmo grupo de indivíduos ao longo de um período de tempo. É o contrário de estudo transversal.

**ESTUDO DE PAINEL:** Tipo de estudo longitudinal no qual dados são coletados da mesma amostra ou painel em diversos pontos do tempo.

**ESTUDO DE PREVALÊNCIA:** Estudo epidemiológico baseado em investigações que produzem instantâneos da situação de saúde de um grupo ou comunidade, observando fator e efeito no mesmo momento histórico. Utiliza amostras representativas de população de referência precisamente delimitada, produzindo medidas de prevalência de doenças. Sinônimos: estudo transversal; estudo de corte transversal, estudo seccional, inquérito.

**ESTUDO DE SEGUIMENTO:** Ver estudo de coorte.

**ESTUDO DE TENDÊNCIAS:** Tipo de estudo longitudinal no qual uma característica de uma população é monitorada por algum tempo. Exemplo: as séries de pesquisas da Gallup mostrando as preferências do eleitorado político durante uma campanha, embora amostras diferentes sejam entrevistadas em cada momento.

**ESTUDO DESCRITIVO:** Investigação que tem o propósito de informar a distribuição de freqüências, sem a preocupação de testar hipóteses. Por exemplo, inquérito sobre a prevalência de parasitoses intestinais em escolares, especificando as freqüências por sexo e faixa etária. Comparar com estudo analítico.

**ESTUDO DUPLO-CEGO:** Estudo no qual nem os sujeitos nem os investigadores têm conhecimento da designação do tratamento que é o agente ativo ou placebo até que o ensaio esteja concluído.

**ESTUDO ECOLÓGICO:** Estudo caracterizado pelo desenho agregado, observacional e transversal. Aborda área geográfica bem delimitada, analisando, comparativamente, variáveis, quase sempre por meio de correlação entre indicadores de saúde.

**ESTUDO ECOLÓGICO:** Estudo que aborda áreas geográficas, analisando comparativamente indicadores globais, quase sempre por meio de correlação entre variáveis ambientais ou sócio- econômicas e indicadores de saúde. Nesse estudo, a área ecológica pode sistematizar um conjunto enorme de variáveis, aproximado mais este tipo de estudo da realidade social concreta.

**ESTUDO ECOLÓGICO:** Pesquisa realizada com estatística; a unidade de observação e análise não é constituída de indivíduos, mas de grupos de indivíduos; daí sua sinonímia: estudo de grupos, de agregados, de conglomerados, estatísticos ou comunitários. Por exemplo, a investigação sobre a variação dos coeficientes de mortalidade por cirrose e o consumo per capita de álcool em diversos países.

**ESTUDO EXPERIMENTAL:** Em um estudo experimental o pesquisador deliberadamente influencia os eventos e pesquisa o efeito da intervenção. Em inglês Experimental study.

**ESTUDO EXPERIMENTAL:** Investigação na qual as condições estão sob direito controle do investigador.

**ESTUDO EXPERIMENTAL ALEATORIZADO:** O mesmo que estudo randomizado.

**ESTUDO EXPLICATIVO:** Estudo projetado para revelar relações causais entre variáveis.

**ESTUDO INDIVIDUALIZADO RANDOMIZADO (OU n DE 1, n DE NÚMERO OU TAMANHO DA AMOSTRA):** Investigação de um único paciente, em que partes de tratamentos, por exemplo, uma droga e um placebo são aplicados, em ordem aleatória, em geral, duplo-cega. Os pares de tratamentos são continuados até que a efetividade da droga seja evidenciada ou refutada.

**ESTUDO HORIZONTAL (O MESMO QUE CROSS-SECTIONAL E O OPOSTO DE ESTUDO LONGITUDINAL):** Estudo de uma comunidade, que pode ser estratificado por idade, sexo, etnicidade, dentre outros, porém em um momento dado ou em um pequeno intervalo de tempo. Embora seja um estudo instantâneo, como se fosse uma fotografia da comunidade tirada num momento de tempo, inquéritos horizontais de prevalência e intensidade dentro de diferentes faixas etárias de uma população podem, todavia, dar valiosa informação sobre a taxa de infecção ou infestação, dado que as populações de hospedeiros e parasitas permanecem aproximadamente estáveis por um período de tempo que chama-se endemicidade estável.

**ESTUDO INTER-SECCIONAL:** Estudo baseado em observações que representam um só ponto temporal, ao contrário de um survey longitudinal.

**ESTUDO LONGITUDINAL:** É aquele no qual se coletam informações sobre os indivíduos selecionados ao longo de um intervalo de tempo especificado.

**ESTUDO LONGITUDINAL:** Investigação na qual cada individuo é observado em mais de uma ocasião; as observações referem-se a, pelo menos, dois momentos na vida das pessoas. Comparar com estudo transversal.

**ESFERICIDADE:** O teste de esfericidade checa se a matriz de correlação é igual a matriz identidade, ou seja, na diagonal formados por uns e zero fora da diagonal. Multicolinearidade. Este é um pressuposto importante na análise regressão múltipla, pois se uma variável independente for uma combinação linear de outras variáveis independentes, coloca em risco toda a análise. Não adianta tentar modelar uma variável em função de várias variáveis correlacionadas, essas não incrementam o poder explicativo do modelo. Neste caso, use o modelo de regressão stepwise, backward ou forward.

**ESTUDO LONGITUDINAL:** São aqueles que estudam mudanças ao longo do tempo, possivelmente com relação a uma intervenção. Em inglês Longitudinal study.

**ESTUDO LONGITUDINAL:** Estudo realizado na população ao longo do tempo. Pode ser prospectivos ou retrospectivos. Se um grupo é selecionado para ser acompanhados, este é um estudo de coorte longitudinal. Se não são indivíduos que são acompanhados, mas classes geralmente classes etárias, este será um estudo transversal longitudinal. O contrário é um estudo horizontal.

**ESTUDO LONGITUDINAL:** Refere-se ao estudo de temporalidade serial na investigação epidemiológica, havendo qualquer tipo de segmento, em uma escada temporal.

**ESTUDO LONGITUDINAL:** Estudo no qual alguns respondentes são reamostrados ao longo do tempo.

**ESTUDOS LONGITUDINAIS ECOLÓGICOS:** Uso da vigilância contínua ou de estudos transversais freqüentes para medir a tendência das taxas das doenças ao longo dos anos em uma população definida. Forma benéfica de estresse que deve estar presente para que o indivíduo goze de boa saúde.

**ESTUDO PILOTO:** Estudo preliminar em uma pesquisa conduzida com a finalidade de obter informações antecipadas para servir de suporte para o pesquisador.

**ESTUDO PROSPECTIVO:** Também chamado de estudo piloto e de estudo preliminar. Neste caso a mensuração é realizada de forma intencional, isto é, um estudo em que a mensuração foi prevista e a análise dos dados se fará de acordo com um projeto previamente planejado.

**ESTUDO N DE 1:** Ver estudo individualizado randomizado.

**ESTUDO NÃO-CONTROLADO:** Investigação em que não há grupo-controle formado durante o desenrolar da investigação; ou seja, não há grupo interno para comparação de resultados, como por exemplo, a série de casos; investigação em que as variáveis de confundimento não foram neutralizadas. Ver estudo controlado.

**ESTUDO NÃO-RANDOMIZADO:** Investigação na qual a locação de indivíduos, para formar os grupos de estudo, não é feita ao acaso; ela é determinada pela conveniência do pesquisador ou por outra razão, que não se enquadra em seleção estatística ao acaso. Ver estudo randomizado.

**ESTUDO OBSERVACIONAL:** Em um estudo observacional o pesquisador colhe a informação sobre os atributos ou faz as medições necessárias sem influenciar os eventos. Em inglês Observational study.

**ESTUDO PILOTO:** Estudo preliminar, projetado para indicar se um estudo maior é viável. Também usado para estabelecer o tamanho da amostra.

**ESTUDO PILOTO:** Estudo completo miniaturizado, para testar todos os aspectos do desenho do estudo.

**ESTUDO PILOTO:** Um estudo feito antecipadamente ao principal estudo ou pesquisa e que pode ser entendido como um modelo em menor escala, conduzindo com o propósito de ganhar informação adicional útil ao melhoramento do estudo propriamente dito.

**ESTUDO PROSPECTIVO:** É um tipo de estudo em que os dados são coletados a partir do início do mesmo, sem levar em conta dados históricos. Em inglês Prospective study.

**ESTUDO PROSPECTIVO:** Estudo no qual pessoas com uma característica ou um atributo específico são identificadas e observadas por um período de tempo para verificar se ocorreu ou não um resultado ou condição de interesse, normalmente doença ou morte.

**ESTUDO PROSPECTIVO:** Estudo longitudinal no qual os indivíduos são observados a partir de um dado momento, prosseguindo-se ao longo do tempo previamente fixado.

**ESTUDO PROSPECTIVO:** Ver estudo de coorte.

**ESTUDO QUASE-EXPERIMENTAL:** O mesmo que estudo experimental não-randomizado.

**ESTUDO RANDOMIZADO (OU ALEATORIZADO):** Investigação na qual a alocação de indivíduos, para a formação de grupos de estudo, é feita ao acaso, como, por exemplo, com o uso da tabela de números aleatórios. Conforme a unidade a alocar, os dois tipos principais de estudo randomizado são: o estudo clínico randomizado ou alocação casualizada de indivíduos ou estudo comunitário randomizado ou alocação casualizada de grupos de indivíduos.

**ESTUDO RETROSPECTIVO:** É o tipo de estudo em que os dados se referem a eventos passados, sendo obtidos de recursos já existentes. Em inglês Retrospective study.

**ESTUDO RETROSPECTIVO:** Estudo longitudinal no qual as informações de interesse estão contidas em registros anteriores, em arquivos de dados como, por exemplo, em prontuários hospitalares.

**ESTUDO RETROSPECTIVO:** Classe de estudo no qual se analisam os efeitos de um fenômeno, observando o passado para averiguar as causas do mesmo. Neste tipo de estudo, o processo de medição não prevê a realização de uma pesquisa futura que utilizasse os dados obtidos anteriormente.

**ESTUDO RETROSPECTIVO:** Estudo no qual pessoas com uma característica ou uma doença são identificadas e questionadas para saber se foram ou não expostas a determinado fator.

**ESTUDO SECCIONAL:** Ver estudo de prevalência.

**ESTUDO SOCIOMÉTRICO:** Exame de redes de relações, tais como amizades, tomada de decisões dentre outros.

**ESTUDO TRANSVERSAL:** Dados coletados de um grupo de indivíduos em um momento definido, avaliando-se, sobretudo a prevalência de uma determinada afecção.

**ESTUDO TRANSVERSAL:** Levantamento sobre uma população realizado em um único momento.

**ESTUDO TRANSVERSAL:** Ver estudo de prevalência.

**ESTUDO TRANSVERSAL (OU SECCIONAL):** Investigação para determinar prevalência; para examinar a relação entre eventos, tais como, exposição, doença e outras variáveis de interesse, em um dado momento. Por exemplo, a associação entre hábito de fumar e bronquite crônica, entre os participantes de um congresso.

**ESTUDO TRANSVERSAL ECOLÓGICO:** Estudo da frequência em que ocorrem determinada característica, por exemplo, o hábito de fumar e determinado efeito de interesse, por exemplo, câncer de pulmão na mesma população, geograficamente definida, e em um momento determinado.

**ESTUDO TRANSMETODOLÓGICO:** É o termo utilizado para a classe de estudo no qual se emprega mais de um método para a realização da pesquisa ou para a análise dos dados.

**ESTUDOS CASUAIS:** Estes estudos examinam se uma variável causa ou determina o valor de outra variável.

**ESTUDOS CORRELACIONAIS:** Consistem de observações relevantes a duas ou mais variáveis pertinentes a cada amostra de pessoas ou grupos. Por meio de métodos estatísticos se determina o grau e direção do relacionamento entre as observações. Não há manipulação de variáveis. Em vez de criar condições e determinar seus efeitos, o pesquisador observa os fenômenos e os relaciona um ao outro. A grande desvantagem do método correlacional está no fato de que não permite inferências conclusivas sobre o relacionamento causal entre as variáveis. Raramente é possível determinar por meio de estudos correlacionais de uma associação estatística entre duas variáveis, A e B, é devido ao fato de que A cause B, B cause A, ou se A e B são causados por uma terceira variável C.

**ESTUDOS DESCRITIVOS:** Estudos que respondem às perguntas quem, o que, quando, onde e como.

**ESTUDO TRANSVERSAL:** Estudo no qual se acompanha o comportamento de uma ou mais variáveis de um grupo de indivíduos que se encontram em períodos de tempo ou estratos diferentes. É o contrário de estudo longitudinal.



**EVENTO DEPENDENTE:** Termo usado na teoria das probabilidades, para indicar a situação da ocorrência de um evento quando afeta a probabilidade de ocorrência de um segundo evento no mesmo espaço amostral. Dois eventos A e B são mutuamente dependentes quando a ocorrência de um deles provoca mudança na probabilidade de ocorrência do outro. Pode ser chamado também de eventos não mutuamente exclusivos.

**ETAC:** Entrevistas por telefone assistidas por computador.

**ETA QUADRADO:** É uma medida do efeito tamanho que é igual ao quociente entre a soma dos quadrados dos grupos entre pela soma dos quadrados somados para todos os efeitos principais, de interação e de erro. Os efeitos das covariáveis não são computados. Esta medida pode ser interpretada como o percentual da variância da variável dependente que são explicadas pelos fatores. Quando existem relações curvilíneas entre a variável dependente e os fatores o Eta Quadrado será maior que o correspondente coeficiente de correlação múltiplo ou  $R^2$ .

**EVENNESS:** Ver equabilidade.

**EVENTO:** Para um espaço amostral S ou  $\Omega$  definido num experimento aleatório, um evento é qualquer subconjunto do espaço amostral. Em inglês Event.

**EVENTO (n):** É um dos subconjunto de um espaço amostral S, e assim é escolhido um dos possíveis eventos dentro do espaço amostral. Por exemplo: Um dado é lançado, e é observada a face de cima. O espaço amostral é {1, 2, 3, 4, 5, 6}, e são exemplos de eventos os seguintes acontecimentos.

i) Ocorrência de um número par. Assim o evento é {2, 4, 6}.

ii) Ocorrência de um número primo. Assim o evento é {2, 3, 5}.

iii) Ocorrência de um número ímpar maior que 3. Assim o evento é {5}.

iv) Ocorrência de um número primo maior que 5. Assim o evento é { }.

**EVENTO:** Um subconjunto de um espaço amostra é denominado de evento. Se o espaço amostral for enumerável então todo subconjunto é um evento. Se for não-enumerável então podem existir subconjuntos que não são eventos. Em inglês event.

**EVENTO:** Resultado de um experimento.

**EVENTO ALEATÓRIO:** É aquele que pode ser re-executado várias vezes, sempre nas mesmas condições, e se obtém resultados diferentes, que estão previstos dentro dos possíveis resultados para um determinado experimento, isto ocorre devido ao acaso, não podemos ter a absoluta certeza do resultado de cada destes eventos. Por exemplo: i) lançar uma moeda para cima e observar a face que irá ficar virada para cima após a queda; ii) escolhermos um aluno dentre os 30 alunos de uma classe.

**EVENTO CERTO:** Um evento que deve ocorrer. Ele tem probabilidade um. Em inglês certain event.

**EVENTO COMPOSTO:** Um evento que consiste de dois ou mais eventos simples. Por exemplo, AB, AB e A-B. Em inglês compound event.

**EVENTO INDEPENDENTE:** Na teoria de probabilidades, refere-se a um evento cuja possibilidade de ocorrer ou não, não está condicionada à ocorrência de outro evento, ou seja, a probabilidade condicional de um ocorrer dado que o outro já ocorreu é igual a probabilidade marginal simples dele ocorrer.

**EVENTOS MUTUAMENTE EXCLUSIVOS:** Diz-se que dois eventos são mutuamente exclusivos quando eles não podem ocorrer simultaneamente. As probabilidades desses eventos podem ser somadas para se obter a probabilidade de que pelo menos um, ao menos um, poderá ocorrer ou a probabilidade da reunião dos dois eventos em um dado experimento aleatório numa coleção dos eventos de um dado espaço amostral.

**EVENTO DE RISCO:** Morte, doença ou lesão.

**EVENTO ELEMENTAR:** Um evento que contém um único ponto amostral. Em inglês elementary event.

**EVENTO IMPOSSÍVEL:** Um evento que não pode ocorrer. Neste caso o acontecimento possui probabilidade nula ou zero em qualquer realização do experimento aleatório. Em inglês impossible event.

**EVENTO SIMPLES:** Resultado experimental que não comporta mais qualquer subdivisão.

**EVENTOS ASSOCIADOS:** Guardam dependência entre si.

**EVENTOS DEPENDENTES:** Eventos tais que a ocorrência de um deles afeta as probabilidades de ocorrências dos outros.

**EVENTOS DEPENDENTES:** Dois eventos são ditos dependentes se a probabilidade de um ocorrer altera a probabilidade do outro ocorrer, isto é,  $P(A/B) \neq P(A)$ . Em inglês dependent events.

**EVENTOS DEPENDENTES:** Dois eventos são dependentes se a ocorrência de um dos eventos afeta a probabilidade de ocorrência do segundo evento.

**EVENTOS DISJUNTOS:** Eventos que não podem ocorrer simultaneamente. Conhecidos também como incompatíveis, mutuamente exclusivos ou excludentes.

**EVENTOS INDEPENDENTES:** Eventos em que a ocorrência de qualquer um deles não afeta as probabilidades de ocorrência dos outros.

**EVENTOS INDEPENDENTES:** Dois ou mais eventos que não se afetam mutuamente.

**EVENTOS INDEPENDENTES:** Eventos não associados entre si.

**EVENTOS INDEPENDENTES:** Dois eventos de um espaço amostral S, são denominados de independentes se a probabilidade de um deles ocorrer não afeta ou não altera a probabilidade de o outro ocorrer. Em inglês independent events.

**EVENTOS MUTUAMENTE EXCLUSIVOS:** Dois eventos são mutuamente exclusivos quando a probabilidade deles acontecerem simultaneamente é identicamente igual a zero. Em inglês Mutually exclusive events.

**EVENTOS MUTUAMENTE EXCLUSIVOS:** Se temos um conjunto de eventos mutuamente exclusivos, ou seja, somente um dos eventos pode ocorrer em qualquer situação ou lance, as probabilidades desses eventos podem ser somadas para obter a probabilidade de que pelo menos um de uma dada coleção dos eventos ocorrerá.

**EVENTOS MUTUAMENTE EXCLUSIVOS OU EXCLUDENTES:** Dois eventos são mutuamente excludentes ou exclusivos ou exaustivos, se não puderem ocorrer simultaneamente, ou seja quando ocorrer o evento  $A \mid B = \phi$ . Em inglês mutually exclusive events.

**EVIDÊNCIA CONFIRMATÓRIA:** A evidência que dá apoio a uma hipótese é grandemente fortalecida, quando vários pesquisadores observaram os mesmos resultados e fizeram interpretações similares dos dados.

**EVOLUTA:** A evoluta de uma curva é o envelope das normais da curva. Ela pode ainda ser pensada como o locus ou lugar geométrico dos centros de curvatura. Em inglês evolute.

**EXATIDÃO:** Mesmo que validade.

**EXATIDÃO:** Define o fato de um resultado ser a representação real ou correta da grandeza medida. Uma medida exata é tão próxima quanto possível da abstração ou idealização que é o valor real. É governada pelos fatores metrológicos e uma medição de alta precisão não é necessariamente uma medição exata.

**EXISTE:** Símbolo matemático dado por  $\exists$ .

**EXCEL:** Software que pode ser utilizado para programação, organização, tabulação e cálculos matemáticos e estatísticos.

**EXPERIMENTIA CRUCIS:** Vocábulo latino que significa, experimento último ou experimento crucial.

**EXPANSÃO DOS VETORES:** Vetor escalonado no qual o vetor original é modificado para representar a razão F correspondente. Usado para representar graficamente as cargas da função discriminante de uma maneira combinada com os centróides de grupo.

**Exp ( $\beta$ ):** Distribuição exponencial.

**EXPECTÂNCIA:** É um conceito associado a uma distribuição de probabilidade. A expectância é definida como sendo a soma do produto dos valores da variável pelas probabilidades se a variável for discreta. Se a variável for contínua então a soma dá lugar a integral. A expectância é também denominada de valor esperado, média de uma variável aleatória ou esperança matemática. Ou seja  $E[X] = \sum_{i=1}^n X_i \cdot P(X = x_i)$  e  $E[X] = \int_{-\infty}^{+\infty} X_i f(X_i) dx$  respectivamente. Em inglês expectation.

**EXPECTATIVA DE VIDA:** O mesmo que longevidade, o tempo médio de vida dos indivíduos em uma população.

**EXPECTATIVA DE VIDA:** Tradicionalmente definida como o número médio de anos de vida restantes em uma determinada idade.

**EXPERIMENTAÇÃO:** Mecanismo ou processo de gerar dados brutos ou experimentais. Metodologia científica usada para verificar hipóteses e obter conclusões válidas para a população.

**EXPERIMENTAÇÃO LABORATORIAL:** Termo usado para definir o tipo de experimentação na qual o cientista utiliza um rigoroso controle das condições experimentais.

**EXPERIÊNCIA:** Obviamente, o ponto de partida da ciência em seu nível mais elementar é a experiência; seja o fenômeno do trovão, uma tempestade de neve, um vaso quebrado como resultado da expansão da água quando se congela, um eclipse, ou simplesmente a regularidade do dia e da noite. A ciência começa com uma observação a que se adicionam outras observações similares ou diferentes. Eventualmente deriva-se um sistema de princípios básicos que responde pela ocorrência e não ocorrência de um dado conjunto de experiências. O objetivo da ciência é a aquisição e sistematização de conhecimento relativo a fenômenos experienciados.

**EXPERIMENTO:** Mecanismo de pesquisa em que uma ou mais variáveis independentes são sistematicamente controladas, segundo um plano previamente estabelecido, com o fim de determinar os efeitos dessa manipulação.

**EXPERIMENTAÇÃO:** Técnica cientificamente sofisticada capaz de oferecer respostas precisas a problemas precisos. Deve ser usada em estágios mais avançados da investigação de um problema de pesquisa.

**EXPERIMENTAÇÃO LABORATORIAL:** É qualquer situação que permita a pesquisadores um estrito controle das condições em que são feitas as observações. Por meio de do experimento laboratorial se pode contrastar os efeitos das variáveis experimentais com suas condições antecedentes.

**EXPERIMENTAÇÃO NÃO-CONTROLADA:** Consiste na coleta de dados em uma ocasião e na comparação desses dados com outros coletados subsequentemente. A primeira mensuração serve de ponto de referência para se conhecer a mudança. Entre as duas mensurações se insere uma variável experimental. Pressupõe-se que qualquer diferença significativa nas observações é devida à introdução de uma nova variável.

**EXPERIMENTO:** Uma situação de pesquisa em que uma ou mais variáveis independentes são sistematicamente variadas, segundo um plano preconcebido, a fim de determinar os efeitos dessa variação.

**EXPERIMENTO:** Processo de coleta de observações.

**EXPERIMENTO (OU ENSAIO):** É um trabalho planejado previamente, que segue determinados princípios básicos e no qual se faz a comparação dos efeitos dos tratamentos.

**EXPERIMENTO:** Estudo analítico no qual há intervenção do pesquisador nos indivíduos submetidos à pesquisa, com controle efetivo dos fatores casuais e dos respectivos efeitos.

**EXPERIMENTO:** Trabalho experimental planejado e realizado de propósito.

**EXPERIMENTO:** Trabalho científico que se destina a verificar um fenômeno físico; ensaio; tentativa.

**EXPERIMENTO:** Método de pesquisa com manipulação controlada da variável independente que é o estímulo para determinar seu impacto na variável dependente. Normalmente compreende o uso de um grupo de controle e pré e pós-teste.

**EXPERIMENTO:** Uma abordagem de pesquisa na qual uma variável é manipulada e o efeito sobre outra variável é observado.

**EXPERIMENTO:** É uma pesquisa de forma ordenada e controlada que visa gerar informações, obtendo novos dados ou confirmações para hipóteses já existentes.

**EXPERIMENTO:** Situação provocada com o objetivo de observar a reação de determinado fenômeno.

**EXPERIMENTO:** Qualquer processo que gere dados brutos, ou ainda uma ação que apresenta mais de um resultado que ocorrem de forma imprevisível e que pode ser repetida indefinidamente sob as mesmas condições. Conhecido também como ensaio. Em inglês experiment.

**EXPERIMENTO BINOMIAL:** Experimento com um número fixo de provas independentes, em que cada resultado se enquadra exatamente em uma de duas categorias.

**EXPERIMENTO CEGO:** Processo utilizado em experimentos, segundo o qual o indivíduo não sabe se está recebendo um tratamento ou um placebo.

**EXPERIMENTO CEGO:** Tipo de experimento no qual o pesquisador não tem informações acerca dos indivíduos ou do tipo de procedimento experimental ao qual foram submetidos.

**EXPERIMENTO DE BERNOULLI:** Qualquer experimento que apresente unicamente dois resultados exclusivos entre si. Em inglês Bernoulli trial.

**EXPERIMENTO DE MULTI-RESPOSTA INCOMPLETO:** É um tipo de planejamento de experimento introduzido por Srivastava (1964) para tratar de situações onde nem todas as variáveis são medidas para cada unidade experimental. Isto pode acontecer quando existe impossibilidade física proveniente de inconvenientes ou de outras considerações como custo. Em inglês Incomplete multiresponse design.

**EXPERIMENTO EX POST FACTO:** Este tipo de experimento envolve situações em que um pesquisador poderia ter manipulado uma variável, mas não o fez, ou situações em que uma variável não pode ser manipulada, por considerações éticas ou práticas.

**EXPERIMENTO FATORIAL:** Experimento planejado para examinar o efeito de dois ou mais fatores, cada um sendo aplicado em pelo menos dois níveis de forma que o efeito diferencial pode ser observado. Em inglês Factorial design.

**EXPERIMENTO COMPARATIVO:** Tipo de ensaio no qual trabalha-se pelo menos três grupos experimentais, um grupo experimental, um grupo controle e um grupo controle positivo, como exemplo pode ser citado, um experimento que tem por objetivo comparar a eficácia de dois tratamentos, como dois tipos de drogas diferentes para uma mesma doença, onde estabelece-se um grupo testemunha ou controle, que recebe um tratamento placebo, um grupo controle positivo, que recebe o tratamento convencional, e um grupo experimental, que submete-se ao tratamento propriamente dito.

**EXPERIMENTO IDEAL:** Um experimento ideal pode ser descrito como aquele em que se mantêm constantes todas as variáveis relevantes, exceto a que se está estudando. Os efeitos desta são então observados.

**EXPERIMENTO INTEIRAMENTE CASUALIZADO:** Experimento em que cada parcela pode receber, por sorteio, qualquer tratamento, sem nenhuma outra restrição além do número desejado de repetições.

**EXPERIMENTO MULTINOMIAL:** Experimento com um número fixo de provas independentes, em que cada resultado se enquadra exatamente em uma dentre diversas categorias.

**EXPERIMENTO DUPLO-CEGO:** Definição usada para indicar o ensaio no qual nem os pesquisadores nem os indivíduos têm conhecimento acerca do tipo de tratamento que estão sendo submetidos. Num ensaio na área médica, por exemplo, dois grupos de pessoas são selecionados para testar uma nova droga para determinada doença. Um dos grupos chamado controle receberá um medicamento placebo e o outro grupo conhecido como experimental será submetido com a nova droga.

**EXPERIMENTO NATURAL:** Conjunto de circunstâncias que ocorrem naturalmente e em que as pessoas estão sujeitas a diferentes graus de exposição a um fator, simulando, assim, uma verdadeira experiência planejada com esta finalidade.

**EXPERIMENTO NATURAL:** Experimento que teria por substrato algum processo de massa, de caráter patológico ou de melhoria de índices vitais com desenvolvimento espontâneo e natural em certos grupos, deixando indene outros segmentos da população. O grupo afetado seria tomado como grupo experimental, e o outro seria usado como controle.

**EXPERIMENTO QUADRADO LATINO:** Um dos mais comuns planejamentos de experimentos em que  $k$  tratamentos são alocados em um quadrado  $k \times k$  de forma que cada tratamento aparece exatamente uma única vez em cada linha e coluna. Em inglês Latin square design.

**EXPERIMENTOS COM TESTE DE HIPÓTESE:** Estudos científicos que buscam descobrir se alguma variável independente influencia ou não alguma variável dependente.

**EXPERIMENTOS DE CAMPO:** Testes conduzidos fora do laboratório em ambiente real de mercado.

**EXPERIMENTOS DE LABORATÓRIO:** Experimentos conduzidos em ambiente controlado.

**EXPERIMENTOS FATORIAS:** São experimentos que incluem dois ou mais tipos distintos de tratamentos também chamados de fatores, cada fator em vários níveis. Exemplo: fatorial de 3 níveis de  $N \times 4$  níveis de  $P$ , com 12 combinações possíveis, se o fatorial for completo.

**EXPERIMENTOS MULTICÊNTRICOS:** Experimentos conduzidos em dois ou mais centros, sempre com um protocolo comum, mas com uma administração central e um centro único para receber e processar os dados.

**EXPERIMENTOS MONTE CARLO:** Simulações usando geração aleatória de números.

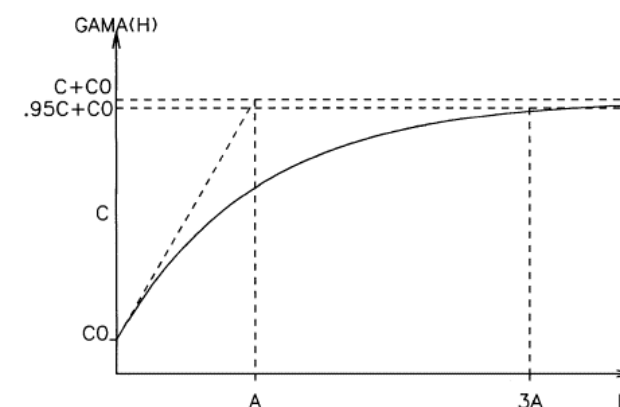
**EXPERIMENTO ROTACIONAL:** Qualquer tipo de ensaio no qual se alterna a sequência de procedimentos aos quais as unidades experimentais serão expostas.

**EXPERIMENTOS PARAMÉTRICOS:** São aqueles que produzem descrições quantitativas pormenorizadas da maneira em que as variáveis estão relacionadas, após determinação da existência do relacionamento.

**EXPLANAÇÃO:** Processo usado para demonstrar quais são as causas de um evento.

**EXPLICAÇÃO:** Resultado de um modelo de elaboração em que a relação entre duas variáveis é explicada pelo controle de uma variável antecedente. Neste caso, dizemos que a relação original era espúria e não genuína.

**EXPONENCIAL:** Modelo de variograma teórico, descrito pela seguinte equação: para  $h > 0$





Teoricamente este modelo nunca alcança o patamar, porém na prática o patamar é aceito em 3a, onde a é a intersecção entre a reta tangente e o patamar teórico.

**EXPOSIÇÃO:** É a quantidade ou intensidade de um fator ao qual o indivíduo ou grupo está ou esteve sujeito. Em epidemiologia de doenças infecciosas, significa proximidade ou contágio com agentes biológicos, potencialmente infectantes.

**EXPOSIÇÃO (RAZÃO DE CHANCE OU ODDS RATIO):** Ver razão de chance.

**EXPOSIÇÃO PRINCIPAL:** É a variável causal, ou hipoteticamente causal de um estudo; é uma das duas variáveis centrais dos estudos epidemiológicos analíticos, a outra é a doença. Por exemplo, na investigação dos efeitos da poluição atmosférica sobre a bronquite crônica, a poluição atmosférica é a exposição principal.

**EXPOSTO:** Indivíduo exposto é o submetido à ação de determinado fator; grupo exposto é aquele cujos membros estão ou estiveram sob a ação de um fator de risco ou intervenção; os desfechos clínicos, no grupo de expostos, são comparados com os do grupo de não-expostos, para inferir riscos.

**EXTRAPOLAÇÃO:** Procedimento de estimação de valores que estão fora da amplitude dos dados. Em séries temporais equivale a uma previsão, e em dados espaciais consiste em obter valores estimados para locais onde não existem observações. Em inglês Extrapolation.

**EXTRAPOLAÇÃO:** Previsão do que ocorrerá em uma situação que está fora do âmbito de dados previamente observados.

**E(X):** Valor esperado da variável aleatória X.

**ESPAÇO AMOSTRAL:** Conjunto de possibilidades, ou seja, os possíveis resultados associados a um experimento aleatório.

**EXPERIMENTO ALEATÓRIO:** Qualquer processo que venha a gerar um resultado incerto ou casual.

**ESTIMADOR:** É uma função ou fórmula que permite estimar o valor de um parâmetro que é a estimativa, baseando-se nas observações de uma amostra.

**EVENTO:** Qualquer subconjunto de um espaço amostral. Conhecido também como acontecimento aleatório, casual ou estocástico.

**ESCALAS DE MEDIÇÃO:** O assunto sobre escalas de medição está fortemente relacionado ao de classificação de tipos de variáveis que se discutiu na seção anterior, mas possui aspectos diferentes. As escalas de medição se classificam em nominal, ordinal, intervalo e de razão. As escalas nominal e ordinal estão associadas às variáveis qualitativas de mesmo nome. As escalas de intervalo e de razão estão associadas a variáveis quantitativas, tanto discretas quanto contínuas. Sendo assim as escalas podem ser descritas como a seguir. i) escala nominal: A escala nominal está associada a variáveis nominais, aquelas para as quais as categorias não têm nenhum significado numérico e nem apresentam uma ordem natural. Vamos ver em seções subseqüentes que as análises estatísticas que se podem fazer com variáveis cuja escala é nominal estão limitadas a um número bastante pequeno. No caso univariado isso se resume praticamente a construir tabelas de frequências e fazer gráficos de barras, e gráficos tipo estrela, mas estes gráficos devem ser interpretados com muito cuidado já que nem o espaçamento entre as barras nem a ordem das barras tem significado. Outro tipo de gráfico que pode ser feito é o tipo pizza conhecido também como pie chart em

inglês. Mas, este é um tipo de gráfico que deverá ser evitado devido a suas péssimas características em termos de interpretação e também do ponto de vista estético, especialmente se for feito com efeito tridimensional ou com uma fatia destacada das demais, mas tudo no seu devido tempo; ii) escala ordinal: A escala ordinal está associada a variáveis qualitativas do mesmo nome, ou seja, a variáveis ordinais. Neste caso já se amplia mais o leque de opções de análise e de apresentação. E no caso de gráficos as barras o espaçamento entre elas ainda não tem significado, mas as mesmas deverão ser apresentadas em sua ordem natural; iii) escala de intervalo: O importante nesta definição é entender o que significam as escalas de intervalo e de razão, que se usam com variáveis quantitativas. Uma escala é de intervalo se os valores da variável que pode ser discreta ou contínua não estão limitados por um valor zero que tenha significado real. O melhor exemplo e talvez o mais fácil de entender se tem com temperatura medida em graus Celsius. Temos tanto temperaturas acima como abaixo de zero. Não faz sentido dizer que a temperatura de 20 °C é duas vezes maior do que a temperatura de 10 °C. Esta escala se chama de intervalo porque diferenças entre valores tem significado, mas não os quocientes deles. Existem poucos exemplos de variáveis que se medem com escala de intervalo nas áreas de ciências sociais incluída aí a área de ciências econômicas, mas, no entanto é comum nas ciências agrárias e biológicas. Mas, considerando, o caso de renda domiciliar em reais onde esta pode ter valores negativos, seria correto dizer que esta é uma variável com escala de intervalo. Ainda assim em geral os economistas consideram apenas casos de renda não negativa e trabalham com escala de razão para esta variável, o que não acontece com outros tipos de variáveis como a temperatura média diária em graus Celsius, por exemplo. Na realidade é muito raro encontrar exemplos de escalas de intervalo fora das ciências exatas, onde elas existem em alguns casos nos quais se criaram escalas para problemas específicos surgidos em laboratório e iv) escala de razão: Uma escala é de razão quando o zero é uma barreira, um limite natural para os valores de uma variável, que pode ser tanto discreta quanto contínua. É o caso de idade em anos de uma pessoa, número de anos de escolaridade, peso, e muitas outras. Podemos agora não apenas trabalhar com diferenças entre valores, mas também com quocientes (razões) entre eles. É perfeitamente legítimo dizer que um indivíduo que esteve 10 anos na escola tem o dobro de escolaridade que outro que esteve apenas 5 anos na escola. É importante destacar que para uma escala ordinal a única operação que tem sentido é a de comparação de posição. Podemos dizer, por exemplo, que o nível socioeconômico alto está acima do nível baixo e este é menor do que o nível médio. Para uma escala de intervalo podemos efetuar operações de diferença de valores. Pode-se dizer que a diferença de temperatura entre 30 °C e 20 °C é a igual à diferença entre 25 °C e 15 °C. É claro que também podemos fazer comparações, 30 °C é mais ou maior do que 20 °C. Já para escala de razão pode-se efetuar a operação de razão, divisão ou quociente dizendo por exemplo que João que pesa 100 quilos tem o dobro do peso de Maria que tem apenas 50 quilos. Vale salientar que para a escala de razão obviamente também se podem efetuar operações de comparação e de diferença. Estas classificações de escalas de medição e a classificação de tipos de variáveis formam uma taxonomia útil, mas, como sempre, existem casos que se situam em áreas nebulosas, e as classificações neste caso devem servir principalmente para que se reflexione com cuidado ao usar técnicas que pressupõem certas propriedades das variáveis.

**FATORES QUE INFLUENCIAM O TAMANHO E FORMA DE PARCELAS EXPERIMENTAIS:** O tamanho e a forma ideais de parcelas em experimentos agrônomicos estão em função dos seguintes fatores tais como: i) tipo de experimentos; ii) heterogeneidade do solo; iii) efeito bordadura; iv) disponibilidade de recursos; v) características avaliadas; vi) precisão; vii) número de tratamentos em estudos; viii) uso de máquinas agrícolas; ix) custo, tempo e mão-de-obra; x) considerações e bom senso do pesquisador.

**ESCALA ABSOLUTA:** A escala absoluta é referente à contagem ou enumeração de ocorrências. O ponto zero absoluto nesta escala pressupõe que todas as operações matemáticas, tal como na escala de razão podem ser aplicadas. A escala absoluta poderia ser tomada como um mero caso especial de escala de razão, onde a única transformação matemática permitida para a escala absoluta é a transformação identidade, isto, é, transformá-la nela mesma. Porém vamos considera-la uma categoria de escala a parte, para que se possa mantê-la destacada dos outros tipos de escalas. Num certo sentido, a escala absoluta é a escala mais informativa possível já que ela é única. Seria uma espécie de escala fundamental. A quantidade de pessoas alocadas em uma equipe de trabalho tipifica uma variável na escala absoluta de mensuração, assim como a quantidade de ovos posto por uma ave poedeira. Como exemplo ainda podem ser citados o número de insetos numa planta que é enumerável. Observe que o número total de insetos não pode ser transformado em nenhum outro número e permanecer sendo discreto, com o significado de número de insetos. É por isso que a escala absoluta é diferente da escala de razão. Sendo assim verifica-se que as escalas de medição são hierárquicas.

# F

**F:** variável estatística de Fisher, a qual é uma função definida pelo quociente entre duas estimativas de variâncias supostas independentes. Variância é o quadrado do desvio padrão, onde  $F > 1$ , admitindo que a população 1 tem sempre maior variância que a população 2.

**F:** Função cumulativa ou função de distribuição acumulada.

**$F_i$ :** Frequência acumulada relativa até à classe  $i$ .

**F1:** A primeira geração de filhos; primeira geração resultante de um dado cruzamento.

**F2:** Segunda geração de filhos; geração resultante do acasalamento entre si ou autofecundação de indivíduos F1.

**FALÁCIA ECOLÓGICA:** Conclusões errôneas sobre indivíduos baseadas somente em observação de grupos.

**FALÁCIA ECOLÓGICA:** Uso de dados ecológicos para fazer inferências sobre relações causais em indivíduos. Se a frequência de uma exposição a frequência de um desfecho forem determinadas em uma população, mas não for fornecida nenhuma informação sobre a ocorrência da exposição e do desfecho no mesmo indivíduo, os dados não podem ser interpretados para o estabelecimento de causalidade.

**FALÁCIA ECOLÓGICA:** Falsa conclusão tirada de uma análise, devido ao exame de unidades de análise inapropriadas.

**FALÁCIA ECOLÓGICA:** É a suposição de que associações produzidas em nível do agregado, correlacionando médias, aplicam-se ao nível indivíduo.

**FALÁCIA ECOLÓGICA:** Interpretação enganosa, decorrente de atribuir-se a um indivíduo o que se observou em estudos com base em estatísticas.

**FALSO POSITIVO:** Em experimentos clínicos, quando um teste acusa a presença de um fator que não existe. Em inglês False positive.

**FAMÍLIA CONJUGADA DE DISTRIBUIÇÕES NATURAIS:** Se para qualquer verossimilhança  $l(\theta;x)$  obtida da família  $F$ , existe uma constante  $k$  que define uma ensidade  $p$  de forma que  $p(\theta)=kl(\theta;x)$ , então a família  $P$  de todas as densidades  $p$  é chamada de família conjugada natural com relação ao modelo de amostragem que tem função de verossimilhança  $l$ . Em inglês Natural conjugate family of distributions

**FAMÍLIAS CONJUGADAS:** A construção da distribuição à priori é importante pois usualmente não é fácil obter a distribuição à posteriori de forma fechada como é feito com as distribuições Beta e Binomial. Observando o teorema de Bayes e seu aspecto sequencial, percebe-se que a posteriori se origina da multiplicação da função de verossimilhança com uma priori. Procura-se, então, misturar as duas para que a distribuição à posteriori esteja na mesma família da distribuição à priori. Para isso, é necessário que  $l(\theta)$



e  $P(\theta)$  tenham o mesmo núcleo. É importante também que tais famílias sejam suficientemente amplas para acomodar várias opiniões sobre  $\theta$ . Ou seja, a partir do conhecimento que se tem sobre  $\theta$ , pode-se definir uma família paramétrica de densidades, e neste caso, a distribuição a priori é representada por uma forma funcional, cujos parâmetros devem ser especificados de acordo com este conhecimento. Estes parâmetros indexadores da família de distribuições a priori são chamados de hiperparâmetros para poder distingui-los dos parâmetros de interesse  $\theta$ . Esta abordagem em geral facilita a análise, e o caso mais importante é o de distribuições conjugadas. A ideia é que as distribuições a priori e a posteriori pertençam a mesma classe de distribuições, e assim a atualização do conhecimento que se tem de  $\theta$  envolve apenas uma mudança nos hiperparâmetros. Sendo assim a definição deste tipo de distribuição é a seguinte: se  $F = \{p(x/\theta), \theta \in \Theta\}$  é uma classe de distribuições amostrais, então uma classe de distribuições  $P$  é conjugada a  $F$  se  $\forall p(x/\theta) \in F$  e  $p(\theta) \in P \Rightarrow$ .

**FAMÍLIA CONJUGADA DE DISTRIBUIÇÕES:** Seja  $F = \{p(x|\theta), \theta \in \Theta\}$  uma família de distribuições amostrais ou observacionais. A classe  $P$  de distribuições é chamada de família conjugada com relação a  $F$  se para toda  $p \in F$  e  $p(\theta) \in P$  tem-se que  $p(\theta|x) \in P$ . Em inglês Conjugate family of distributions.

**FAMÍLIA EXPONENCIAL:** A família de distribuições com função densidade de probabilidade  $p(x|\theta)$  pertence à família exponencial com  $r$  parâmetros se  $p(x|\theta)$  pode ser escrita como  $p(x|\theta) = a(x) \exp\{\sum_j U_j(x) \phi_j(\theta) + b(\theta)\}$ ,  $x \in \mathcal{X}$  e  $\mathcal{X}$  não depende de  $\theta$ . Em inglês Exponential family.

**FAMÍLIAS EXPONENCIAIS UNIDIMENSIONAIS:** Dizemos que a distribuição da variável aleatória  $X$  pertence à família exponencial unidimensional de distribuições, se pudermos escrever sua função de probabilidade ou função densidade de probabilidade como

$$f(x/\theta) = e^{c(\theta)T(x)+d(\theta)+S(x)}, \quad x \in A$$

Onde  $c, d$  são funções reais de  $\theta$ ;  $T, S$  são funções reais de  $x$  e  $A$  não depende de  $\theta$ . Notemos que no caso em que  $X$  é uma variável aleatória contínua, para que  $f(x/\theta)$  em

$$f(x/\theta) = e^{c(\theta)T(x)+d(\theta)+S(x)}, \quad x \in A$$

seja uma função densidade de probabilidade, é necessário que

$$\int_A e^{c(\theta)T(x)+d(\theta)+S(x)} dx = 1.$$

**FAMÍLIAS EXPONENCIAIS MULTIDIMENSIONAIS:** Dizemos que a distribuição da variável aleatória ou de um vetor aleatório  $X$  pertence à família exponencial de dimensão  $K$  se a função de densidade ou de probabilidade de  $X$  é dada por

$$f(x/\theta) = e^{\sum_{j=1}^k c_j(\theta)T_j(x)+d(\theta)+S(x)}, \quad x \in A$$

Onde  $c_j, T_j, d$  e  $S$  são funções reais,  $j = 1, 2, \dots, k$ , e como no caso unidimensional,  $d(\theta)$  está associado à constante de normalização de

$$f(x/\theta) = e^{\sum_{j=1}^k c_j(\theta)T_j(x)+d(\theta)+S(x)}, \quad x \in A$$

e  $A$  não depende de  $\theta$ .

Também, no caso de dimensão  $k$ , amostras de famílias exponenciais de dimensão  $k$  têm distribuições que são membros da família exponencial de dimensão  $k$ . para uma amostra  $X_1, X_2, \dots, X_n$  de uma variável aleatória com função de densidade ou de probabilidade dada por

$$f(x/\theta) = e^{\sum_{j=1}^k c_j(\theta)T_j(x)+d(\theta)+S(x)}, \quad x \in A$$

temos que a função de densidade (ou de probabilidade)

conjunta de  $X_1, X_2, \dots, X_n$  é dada por

$$f(x_1, x_2, \dots, x_n/\theta) = e^{\sum_{j=1}^k c_j^*(\theta) \sum_{i=1}^n T_j(x_i)+d^*(\theta)+S^*(x)}$$

$$\text{Onde } T_j^*(x) = \sum_{i=1}^n T_j(x_i), \quad c_j^*(\theta) = c_j(\theta) \text{ e } S^*(x) = \sum_{i=1}^n S(x_i), \quad d^*(\theta) = nd(\theta)$$

Neste caso,  $(T_1^*, \dots, T_k^*)$  é conjuntamente suficiente para  $\theta$ .

**FAMÍLIAS COMPLETAS DE DISTRIBUIÇÃO:** Se, para uma família paramétrica de distribuições univariadas ou multivariadas  $f(x|\theta)$ , que depende de um vetor de parâmetros  $\theta$ ,  $h(x)$  é uma estatística independente de  $\theta$ , e, para todo  $\theta$ :  $E[h(x)] = \int h(x) f(x|\theta) dx = 0$  implica que  $h(x) = 0$  identicamente (a não ser possivelmente por uma medida igual a zero), então a família  $f(x|\theta)$  é chamada completa. Em inglês Complete families of distributions.

**FASE:** Diz-se que dois termos  $y_i = (t_i)$  e  $y_i = f(t_i)$  de uma sequência, se acham na mesma fase quando: 1º são iguais, em grandeza e sinal e 2º a função  $F(t)$  é, ou crescente para ambos os pontos  $t_i$  e  $t_j$ , ou decrescente para ambos, ou, ainda, estacionária para ambos.

**FATOR:** Esta palavra é usada no contexto estatístico de diversas formas: primeiro na forma matemática ordinal, um fator de uma expressão matemática, por exemplo; segundo para denotar uma quantidade sob exame em um experimento como causa possível de variação; terceiro em análise multivariada, para denotar uma função das variáveis observadas, geralmente linear, que pode ser considerada como parte das variáveis; e assim como um fator da variação; e em quarto para denotar um item constituinte de um promédio ou número índice. Em inglês Factor.

**FATOR:** Em análise da variância, propriedade ou característica que nos permite distinguir diferentes populações umas das outras.

**FATOR:** Combinação linear ou variável estatística das variáveis originais. Os fatores também representam as dimensões latentes ou constructos que resumem ou explicam o conjunto original de variáveis observadas.

**FATOR:** Variável independente não-métrica, também chamada de tratamento ou variável experimental.

**FATOR:** Variável que o pesquisador manipula e que representa um atributo específico. Na análise conjunta, os fatores ou variáveis independentes são não-métricos. Os fatores devem ser representados por dois ou mais valores, também conhecidos como níveis, os quais também são especificados pelo pesquisador.

**FATOR:** É aquilo que se aplica em um ensaio de forma não homogênea, por exemplo, cultivar, quando se testam várias delas; adubação, ao se compararem diversas formulações; etc.

**FATOR:** Um fator é uma denominação alternativa, utilizada na análise de variância, para uma variável. Assim se num determinado experimento consiste em determinar o efeito da dosagem de um remédio, a dosagem do remédio é um fator. Se neste experimento for levado em conta não apenas a dosagem mas também o tempo em que o remédio será tomado, então o tempo será um segundo fator. Cada valor que

o fator ou variável assume é denominado nível do fator. Assim se a dosagem testado for de: 0, 100, 150 e 200 t/ha, este fator terá quatro níveis. Um fator deve ter pelo menos dois níveis. Em inglês fator.

**FATOR AUTOREGRESSIVO:** Um dos fatores considerados dentro de um modelo autoregressivo e que estão associados a uma das variáveis defasadas no tempo. Em inglês Autoregressive factor.

**FATOR ALEATÓRIO:** Em uma análise de variância, um fator cujos níveis sejam escolhidos aleatoriamente a partir de alguma população de níveis de fator.

**FATOR DE CORREÇÃO:** Um termo usado para a expressão  $\left(\frac{1}{n}\right)\left(\sum_{i=1}^n X_i\right)^2$  para corrigir a soma dos quadrados, definida como  $\left(\frac{1}{n}\right)\left(\sum_{i=1}^n (X_i - \bar{X})^2\right)$ . O fator de correção pode também ser escrito como  $n\bar{X}^2$ .

**FATOR DE CORREÇÃO PARA POPULAÇÃO FINITA:** Um termo na fórmula para a variância de uma variável aleatória hipergeométrica.

**FATOR DE RUÍDO:** Um fator que influencia provavelmente a variável de resposta, porém não tem qualquer interesse no presente estudo. Quando os níveis do fator de ruído podem ser controlados, blocagem é a técnica de planejamento costumeiramente usada para remover seu efeito.

**FATOR FIXO (OU EFEITO FIXO):** Na análise de variância, um fator ou efeito é considerado fixo se todos os níveis de interesse para aquele fator são incluídos no experimento. Conclusões são então válidas somente acerca desse conjunto de níveis, embora quando o fator for quantitativo, é costume ajustar o modelo aos dados para interpolar entre esses níveis.

**FATORES CRUZADOS:** Outro nome para fatores que interagem em um planejamento de experimentos.

**FATORES DE AUMENTO DE VARIÂNCIA:** Grandezas usadas em regressão múltipla de modo a avaliar a extensão da multicolinearidade ou dependência linear próxima em regressores. O fator de aumento de variância para o  $i$ -ésimo regressor,  $FAV_i$ , pode ser definido como  $FAV_i = \left[\frac{1}{(1-R_i^2)}\right]$ , em que  $R_i^2$  é o coeficiente de determinação obtido quando  $x_i$  é regredido sobre as outras variáveis regressoras. Assim, quando  $x_i$  é aproximadamente linearmente independente em um subconjunto dos outros regressores,  $R_i^2$  estará perto da unidade e o valor correspondente do fator de aumento de variância será grande. Valores dos fatores de aumento de variância que excedem 10 são geralmente tomados como um sinal de que a multicolinearidade está presente.

**FATOR (FACTOR):** Em delineamentos de análise de variância (ANOVA), um conjunto de níveis de um único tratamento experimental.

**FATOR DE BAYES (BAYES' FACTOR):** A probabilidade a posteriori em relação da hipótese nula com a alternativa. Se as probabilidades a priori das duas hipóteses forem iguais, ele pode ser calculado como a razão das verossimilhanças das duas hipóteses.

**FATOR DE PARCELA COMPLETA (WHOLE PLOT FACTOR):** Em um delineamento de parcelas subdivididas, o tratamento que é aplicado a um bloco ou parcela inteira. Comparar com fator de subparcela.

**FATOR DE SUBPARCELA (SUBPLOT FACTOR):** Em um delineamento de parcelas subdivididas, é o tratamento que é aplicado dentro de um único bloco ou parcela. Comparar com fator de parcela completa.

**FRAÇÃO DEFEITUOSA:** Em controle estatístico da qualidade, aquela porção de um número de unidades ou a saída de um processo que é defeituosa.

**FREQUÊNCIA RELATIVA:** A frequência relativa de um evento é a proporção de vezes que o evento ocorreu em uma série de tentativas de um experimento aleatório.

**FUNÇÃO DE DISTRIBUIÇÃO:** Outro nome para a função distribuição cumulativa.

**FUNÇÃO DE PROBABILIDADE:** Uma função que fornece probabilidades para os valores na faixa de uma variável aleatória discreta.

**FUNÇÃO DE PROBABILIDADE CONJUNTA:** Uma função usada para calcular probabilidades para duas ou mais variáveis aleatórias discretas.

**FUNÇÃO DE PROBABILIDADE MARGINAL:** A função de probabilidade de uma variável aleatória discreta obtida por meio de distribuição de probabilidades conjuntas de duas ou mais variáveis.

**FUNÇÃO DE PROBABILIDADES CONDICIONAIS:** A função de probabilidade da distribuição de probabilidades condicionais de uma variável aleatória discreta.

**FUNÇÃO DENSIDADE:** Outro nome para a função densidade de probabilidade.

**FUNÇÃO DENSIDADE DE PROBABILIDADE CONDICIONAL:** A função densidade de probabilidade da distribuição de probabilidades condicionais de uma variável aleatória contínua.

**FUNÇÃO DENSIDADE DE PROBABILIDADE CONJUNTA:** Uma função usada para calcular as probabilidades para duas ou mais variáveis aleatórias contínuas.

**FUNÇÃO DENSIDADE DE PROBABILIDADE MARGINAL:** A função densidade de probabilidade de uma variável aleatória contínua obtida por meio de distribuição de probabilidades conjuntas de duas ou mais variáveis aleatórias.

**FUNÇÃO DE RISCO:** A função de risco também chamada taxa de falha específica da idade ou força de mortalidade, representa a taxa instantânea de morte de um indivíduo no instante  $t$ , sabendo que sobreviveu até esse instante. Em inglês hazard function.

**FUNÇÃO DENSIDADE DE PROBABILIDADE:** Chama-se função densidade de probabilidade,  $f(t)$ , à função definida por:  $f(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t < T \leq t + \Delta t)}{\Delta t}$ ,  $0 \leq t < \infty$ . No contexto da análise de sobrevivência, a

função  $f(t)$  pode ser interpretada como a probabilidade de um indivíduo sofrer o acontecimento de interesse no intervalo  $(t, t + \Delta t)$  em que  $\Delta t \rightarrow 0$ .

**FUNÇÃO DE DISTRIBUIÇÃO:** A função de distribuição  $F(t)$  é definida da seguinte forma:  $F(t) = P(T \leq t)$ ,  $0 \leq t < \infty$ . Representa, portanto, a probabilidade de ocorrência do acontecimento de interesse até ao instante  $t$ .

**FUNÇÃO DE SOBREVIVÊNCIA:** Chama-se função de sobrevivência,  $S(t)$ , à probabilidade de ocorrência do acontecimento de interesse após o instante  $t$ . A função de sobrevivência define-se então da seguinte

forma:  $S(t) = P(T > t)$ ,  $0 \leq t < \infty$ . Em termos clínicos, a função de sobrevivência representa a probabilidade de um indivíduo sobreviver para além do instante  $t$ .

**FUNÇÃO DE RISCO:** A função de risco também chamada taxa de falha a qual é específica da idade ou força de mortalidade, representa a taxa instantânea de morte de um indivíduo no instante  $t$ , sabendo que sobreviveu até esse instante. Define-se da seguinte forma:  $H(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t < T \leq t + \Delta t | T > t)}{\Delta t}$ ,  $0 \leq t < \infty$ . Em

estatística, a função de probabilidade condicional de falha, função de risco ou função hazard descreve o potencial ou a probabilidade instantâneo de ocorrência do evento sobrevivência até um tempo arbitrário. Em outras palavras, esta função descreve a taxa condicional de ocorrência de evento num determinado instante de tempo  $t$  dado sobrevivência até imediatamente antes do tempo  $t$ . Por exemplo, se a variável aleatória  $T$  representa a vida útil de um objeto, por exemplo, a duração de vida de um componente eletrônico, a função hazard  $h_T(t)$  associada a esta variável aleatória é definida como  $H(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t < T \leq t + \Delta t | T > t)}{\Delta t}$ . Junto com a função de sobrevivência, a função hazard é muito utilizada

em trabalhos econômicos que tentam estimar o tempo médio de permanência de indivíduos no emprego ou desemprego. Por exemplo, a probabilidade de um indivíduo conseguir um emprego na quinta semana dado que esteve desempregado nas quatro semanas anteriores. Em inglês hazard function.

**FUNÇÃO DE RISCO CUMULATIVA:** A função de risco cumulativa mede o risco de ocorrência do acontecimento de interesse até ao instante  $t$ . Define-se da seguinte forma:  $H(t) = \int_0^t h(u) du$ . A função de risco

cumulativa, que é uma função monótona não decrescente, pode ser obtida a partir da função de sobrevivência, uma vez que:

$$S(t) = \exp(-H(t)) \Leftrightarrow H(t) = -\ln S(t).$$

**FORMA FUNCIONAL:** Tipo de forma de relação funcional usada em regressão e econometria. A teoria econômica pode algumas vezes sugerir a forma funcional de uma relação econômica. Por exemplo, a teoria microeconômica postula a curto prazo que a curva de custo médio tem o formato de U e a curva de custo fixo médio constantemente cai e se aproxima assintoticamente do eixo da quantidade, quando o custo fixo total se distribui sobre mais e mais unidades produzidas. O diagrama de pontos também pode sugerir a forma funcional apropriada em uma relação com duas variáveis. Quando nem a teoria nem o diagrama de pontos ajudam, a função linear é usualmente tentada primeiro, devido a sua simplicidade. Algumas das mais úteis e comuns transformações de funções não lineares em funções lineares são as funções di-log, mono-log, recíproca e a polinomial. Uma das vantagens da forma di-log é que os parâmetros de declividade representam elasticidade. A função mono-log é apropriada quando a variável dependente cresce a uma taxa constante no tempo, como no caso da fora de trabalho e da população. As funções recíprocas e polinomial são apropriadas para estimar curvas de custo médio e de custo total. A estimação de uma função transformada di-log pelo MMQ resulta em estimadores não viesados da declividade. Entretanto, é um estimador viesado, porém consistente de . O fato de ser viesado não traz grandes consequências porque a constante geralmente não é de interesse para análise. Em outras funções transformadas, também é não viesado. O modelo linear di-log é apropriado quando o log Y plotado contra log X encontra-se aproximadamente sobre uma linha reta.

**FUNÇÕES HIPERBÓLICAS:** Na pesquisa em geral e na teoria científica biológica em particular existem ocasiões em que soma ou a diferença de  $e^x$  e  $e^{-x}$  é usada em fórmulas e equações. Em tal ocasião, é costume reescrever a fórmula utilizando as seguintes funções: O seno hiperbólico, definido por  $\sinh x = \frac{1}{2}(e^x - e^{-x})$ , o coseno hiperbólico definido por  $\cosh x = \frac{1}{2}(e^x + e^{-x})$ , a tangente hiperbólica, definida

por  $\tanh x = \frac{\sinh x}{\cosh x}$ , a cotangente hiperbólica definida por  $\cotanh x = \frac{\cosh x}{\sinh x}$ . Todas as quatro funções são

chamadas funções hiperbólicas. Tanto a palavra hiperbólica, quanto seno, coseno, etc. necessitam de uma justificativa. Para este propósito calcula-se a seguinte equação  $\cosh^2 x - \sinh^2 x = \frac{1}{4}(e^{2x} + 2 + e^{-2x}) - \frac{1}{4}(e^{2x} - 2 + e^{-2x})$

que resulta em  $\cosh^2 x - \sinh^2 x = 1$ . A relação é análoga a  $\cos^2 \alpha + \sin^2 \alpha = 1$ , que já é conhecida da fórmula

$(\cos \alpha)^2 + (\sin \alpha)^2 = 1$ . A única diferença formal é o sinal de menos em  $\cos^2 \alpha + \sin^2 \alpha = 1$ . Além

disso, se  $\alpha$  é um ângulo polar variável, os pontos  $(x, y)$  com  $x = \cos \alpha$  e  $y = \sin \alpha$ , lançados em um

sistema de coordenadas retangulares, ficam localizados na circunferência do círculo unitário cuja equação é  $x^2 + y^2 = 1$ , devido a fórmula  $\cosh^2 x - \sinh^2 x = 1$ . Esta curva é uma hipérbole com assíntotas, dadas

pelas equações  $y = x$  e  $y = -x$ . Então, o seno e o coseno hiperbólicos estão relacionados a uma hipérbole, da mesma forma que o seno e coseno comuns estão relacionados com a circunferência. Mesmo sabendo-se que os resultados demonstrados anteriormente oferecem sólida e justificativa suficiente para os termos seno hiperbólico e coseno hiperbólico, porém existem analogias mais restritas. Por exemplo, pode-se estudar as derivadas do coseno e do seno através das seguintes equações

$$(\cosh x)' = \frac{d}{dx} \frac{1}{2}(e^x + e^{-x}) = \frac{1}{2}(e^x - e^{-x}) = \sinh x \quad \text{e} \quad (\sinh x)' = \frac{d}{dx} \frac{1}{2}(e^x - e^{-x}) = \frac{1}{2}(e^x + e^{-x}) = \cosh x.$$

Essas fórmulas são bastante análogas a  $(\cos \alpha)' = -\sin \alpha$  e  $(\sin \alpha)' = \cos \alpha$ . apesar de muitas analogias, existem também grandes diferenças. As funções hiperbólicas não são periódicas. Algumas fórmulas importantes pois são de uso frequente são as seguintes:

$$e^u e^v = e^{u+v},$$

$$\frac{e^u}{e^v} = e^{u-v},$$

$$(e^u)^r = e^{ur},$$

$$q^x = e^{x \ln q} \quad (q > 0),$$

$$\ln a + \ln b = \ln(ab) \quad (a > 0, b > 0),$$

$$\ln a - \ln b = \ln\left(\frac{a}{b}\right) \quad (a > 0, b > 0),$$

$$\ln a^s = s \ln a \quad (a > 0),$$

$$\ln e = 1,$$

$$\frac{d}{dx} e^x = e^x \quad (\text{todo } x)$$



$$\int e^x dx = e^x + C \quad (\text{todo } x),$$

$$\frac{d}{dx} \ln x = \frac{1}{x} \quad (x \neq 0),$$

$$\int \frac{1}{x} dx = \ln|x| + C \quad (x \neq 0) \text{ e finalmente as duas seguintes fórmulas } e^x \approx 1 + x \text{ e } \ln(1+x) \approx x \text{ ambas válidas}$$

para  $x$  suficientemente próximo a zero.

**FUNÇÃO DENSIDADE DE PROBABILIDADE:** Uma função usada para calcular probabilidades e para especificar a distribuição de probabilidades de uma variável aleatória contínua.

**FUNÇÃO DISTRIBUIÇÃO CUMULATIVA:** Para uma variável aleatória  $X$ , a função de  $X$  definida como  $P(X \leq x)$  que é usada para especificar a distribuição de probabilidades.

**FUNÇÃO DISTRIBUIÇÃO CUMULATIVA NORMAL:** A distribuição cumulativa da distribuição normal padrão, frequentemente denotada como  $\Phi(x)$  e tabelada na tabela II do apêndice.

**FUNÇÃO DE VEROSSIMILHANÇA:** É uma função de verossimilhança que formaliza a contribuição dos dados amostrais para o conhecimento sobre  $\theta$ , visto que conecta a distribuição a priori à distribuição a posteriori. A função de verossimilhança  $[\ell(\theta; y)]$  de  $\theta$  é a função que associa a cada  $\theta$ , o valor  $f(y/\theta)$ . Dessa forma  $\theta \rightarrow \ell(\theta; y) = f(y/\theta)$ . A função  $\ell(\theta; y)$  associa para um valor fixo de  $y$  a probabilidade de ser observado  $y$  a cada valor de  $\theta$ . Assim, quanto maior o valor de  $\ell$  maiores são as probabilidades atribuídas pelo particular valor de  $\theta$  considerado, ao valor fixado de  $y$ . Ao fixar um valor de  $y$  e variar os valores de  $\theta$  observa-se a plausibilidade ou verossimilhança de cada um dos valores de  $\theta$ . É interessante observar que  $\int_R f(y/\theta) dy = 1$ , mas  $\int_{\theta} \ell(\theta; y) d\theta = K \neq 1$ , ou seja, a integral da função densidade de probabilidade

equivale a 1, mas a função de verossimilhança não integra 1. É importante ressaltar que os termos probabilidade e verossimilhança são conceitos diferentes. No cômputo da verossimilhança, fixa-se a amostra ou conjunto de dados ( $y$ ) e varia-se o parâmetro  $\theta$ , procurando encontrar o parâmetro verossímil ou plausível com o conjunto de dados. Por outro lado, no cálculo de uma probabilidade utiliza-se uma distribuição com parâmetro  $\theta$  conhecido e calcula-se a probabilidade de observar um determinado valor  $y = y_0$ . Em outras palavras, o problema da probabilidade é prever a chance de ocorrer  $y$  sabendo  $\theta$  e o problema de verossimilhança é fazer afirmações sobre  $\theta$  com base no valor observado  $y$ . No cálculo de probabilidades fixa-se o parâmetro  $\theta$  conhecido e varia-se  $y$ , ao passo que no cálculo da verossimilhança fixa-se  $y$  e varia-se  $\theta$ . A função de verossimilhança é a base do princípio da verossimilhança, o qual postula que toda a informação contida na amostra ou experimento encontra-se representada nesta função. Existem diferentes verossimilhanças que podem ser empregadas na estimação paramétrica, tais quais verossimilhança incondicional marginal, verossimilhança condicional e verossimilhança parcial.

**FUNÇÃO ESTIMÁVEL:** É a função linear nos parâmetros que é idêntica a uma função linear do valor esperado do vetor de observações.

**FUNÇÃO GAMA:** Uma função usada na função densidade de probabilidade de uma variável aleatória gama que pode ser considerada para estender fatoriais.

**FUNÇÃO GERADORA:** Uma função que é usada para determinar propriedades da distribuição de probabilidades de uma variável aleatória. Veja função geradora de momento.

**FUNÇÃO GERADORA DE MOMENTO:** Uma função que é usada para determinar propriedades (tais como momentos) da distribuição de probabilidades de uma variável aleatória. É o valor esperado de  $\exp(tX)$ . Veja Função geradora; Momento.

**FUNÇÃO LINEAR DE VARIÁVEIS ALEATÓRIAS:** Uma variável aleatória que é definida como uma função linear de muitas variáveis aleatórias.

**FUNÇÃO VEROSSIMILHANÇA:** Suponha que as variáveis aleatórias  $X_1, X_2, \dots, X_n$  tenham uma distribuição conjunta dada por  $f(x_1, x_2, \dots, x_n; \theta_1, \theta_2, \dots, \theta_n)$  em que os  $\theta$ s sejam parâmetros desconhecidos. Essa distribuição conjunta, considerada como uma função dos  $\theta$ s para  $x$ 's fixados, é chamada de função verossimilhança.

**FATOR COMUM:** É o que intervém em mais de um teste ou variável de uma bateria analisada fatorialmente. Quando um fator intervém em todos os testes diz-se geral e, quando somente em alguns, grupal. Opõe-se a fator específico.

**FATOR DE BAYES:** A razão das chances dos dados considerando-se duas hipóteses que competem entre si. Do inglês Bayes factor.

**FATOR DE BLOCAGEM:** Característica de respondentes em análise de variância (ANOVA) ou análise de variância multivariada (MANOVA) que é usada para reduzir a variabilidade interna do grupo. Essa característica se toma um tratamento adicional na análise. Fazendo-se isso, são formados grupos adicionais, que são mais homogêneos. Como exemplo, considere que clientes são questionados sobre intenções de compra de um produto e que a medida independente empregada é idade. O exame dos dados descobriu variação substancial devido a sexo. Logo, sexo poderia ser acrescentado como um tratamento adicional, de forma que cada categoria de idade foi dividida em grupos de homens e mulheres com maior homogeneidade interna.

**FATOR DE CONFUSÃO:** O mesmo que variável de confusão.

**FATOR DE CORREÇÃO:** Utilizado quando se está trabalhando com população finita, ou seja, no processo de amostragem sem reposição.  $\frac{(N-n)}{(N-1)}$ .

**FATOR DE CORRELAÇÃO:** Ver fator de correção para população finita.

**FATOR DE CORREÇÃO AMOSTRAL:** É utilizado para calcular a variância e o desvio padrão amostral. Em função de estar utilizando um estimador, por exemplo, a média amostral para calcular outro estimador como a variância ou o desvio padrão requer um fator de correção, esse fator é na verdade o número de observações menos um ( $n-1$ ).

**FATOR DE CORREÇÃO DA CONTINUIDADE:** Um valor de 0,5 que é adicionado e/ou subtraído de um valor  $x$  quando a distribuição normal contínua de probabilidade é usada para aproximar a distribuição binomial discreta de probabilidade.

**FATOR DE CORREÇÃO DA POPULAÇÃO FINITA:** O termo que é usado nas fórmulas para e sempre que uma população finita, em vez de população infinita, esteja sendo amostrada. A regra prática geralmente aceita é ignorar o fator de correção da população finita sempre que  $\frac{n}{N} \leq 0,05$ .

**FATOR DE CORREÇÃO DA POPULAÇÃO FINITA:** Fator que corrige o erro-padrão da média, quando o tamanho da amostra excede em 5% o tamanho de uma população finita.

**FATOR DE CORREÇÃO FINITO DA POPULAÇÃO (FPC):** Um ajuste ao tamanho da amostra necessária que é feito no caso em que se espera que a amostra seja igual a 5% ou mais da população total.

**FATOR DE CORREÇÃO PARA POPULAÇÃO FINITA:** Em amostragem sem reposição, as fórmulas para o desvio-padrão da média amostral e da proporção amostral devem ser multiplicadas por  $\sqrt{\frac{(N-n)}{(N-1)}}$ ,

onde N é o tamanho da população e n é o tamanho da amostra; este fator torna-se desnecessário no caso de a população ser muito maior do que a amostra.

**FATOR DE INFLAÇÃO DE VARIÂNCIA (VIF):** Medida do efeito das outras variáveis preditoras sobre um coeficiente de regressão. VIF é inversamente relacionada ao fator de tolerância ( $VIF_i = 1 + TOL_i$ ). Valores VIF grandes, sendo uma referência comum é 10,0, que corresponde a uma tolerância de 0,10 e indicam um alto grau de colinearidade ou multicolinearidade entre as variáveis independentes.

**FATOR DE INFLAÇÃO DE VARIÂNCIA (VIF):** Indicador do efeito que as outras variáveis independentes têm sobre o erro padrão de um coeficiente de regressão. O fator de inflação de variância está diretamente relacionado ao valor de tolerância  $VIF_i = \frac{1}{TOL_i}$ . Valores VIF altos também indicam um alto

grau de colinearidade ou multicolinearidade entre as variáveis independentes. VIF é a sigla para variance inflation factor.

**FATOR DE RISCO:** Exposição ambiental, característica pessoal ou evento que afetem a probabilidade de contrair determinada doença ou experimentar mudança no estado de saúde. Uma análise dos fatores de risco normalmente implica em algum tipo de análise estatística para apontar ou identificar fatores de risco para determinada doença ou condição.

**FATOR DE RISCO:** Um fator de exposição supostamente associado com o desenvolvimento de uma doença. Pode ser genético ou ambiental. Ao contrário dos fatores prognósticos, os fatores de risco agem antes da instalação da doença. Considera-se fator de risco de um dano toda característica ou circunstância que acompanha um aumento de probabilidade de ocorrência do fator indesejado, sem que o dito fator tenha de intervir necessariamente em sua casualidade.

**FATOR DE RISCO:** Atributo ou circunstância do ambiente ou característica do indivíduo, herdada ou adquirida, associada à maior probabilidade de este mesmo indivíduo apresentar, no futuro, um dano a saúde.

**FATOR DE RISCO:** Característica que, se presente e ativa, claramente aumenta a probabilidade de uma doença específica em um grupo de pessoas que têm o fator quando comparado com outro grupo de pessoas, semelhante em outros aspectos, que não têm esse fator. Um fator de risco não é causa necessária nem a causa suficiente de uma doença.

**FATOR DE RISCO:** Se o fator está presente, é a probabilidade do efeito deste fator na ausência de fatores intermediários ou variáveis intervenientes.

**FATOR ESPECÍFICO:** Fator que intervém em somente um teste ou variável de uma bateria analisada fatorialmente. Diz-se, também, fator único. Opõe-se a fator comum, geral ou grupal.

**FATOR EM ESTUDO:** Medida básica de um estudo epidemiológico, relacionada a um desfecho de fenômeno saúde-doença. Refere-se ao agente em investigação que é o fator de risco ou exposição, fator prognóstico ou de tratamento que, supostamente, determina o desfecho.

**FATOR GERAL:** Em análise de componentes, um componente que é comum a todas as variáveis. Em inglês General factor.

**FATOR PROGNÓSTICO:** Uma condição associada com o desenvolvimento da complicação de um doença. Difere do fator de risco por atuar no prognóstico e não na etiologia de uma doença.

**FATOR TONELAGEM:** Conceito conhecido também como densidade aparente é o parâmetro que permite fazer a conversão de volume em tonelagem.

**FATORES AGREGADOS:** São os agentes etiológicos e as condições propiciatórias, ambos constituintes na geração da doença, que, até então inexistentes na área sob estudo, são trazidos de outros lugares ou são gerados na própria área, por modificações da estrutura epidemiológica.

**FATORES (TRATAMENTOS) DE CLASSIFICAÇÃO CRUZADA:** Quando os fatores são todos cruzados e é possível estudar as interações uma vez que ao se repetir o mesmo nível de um fator em cada um dos níveis do outro fator é possível estudar o comportamento diferencial de um fator quando o outro muda de nível.

**FATORES (TRATAMENTOS) DE CLASSIFICAÇÃO HIERÁRQUICA:** Quando ao se mudar o nível de um fator que é o de agrupamento principal mudam-se também os níveis do outro fator que é o fator de subagrupamento. Nota-se, nesses casos que existe uma hierarquia principal e o outro fator está aninhado ou dentro dele. Observa-se que o fator de subagrupamento não pode ser considerado sem que se indique em qual nível do fator principal ele está aninhado. Neste caso não é possível estudar a interação entre os fatores, pois quando muda-se de nível um fator os níveis do outro fator passam a ser outros, tirando a possibilidade de avaliar esse tipo de efeito.

**FATORES DE CLASSIFICAÇÃO MISTA:** Surge quando, no mesmo conjunto, aparecem fatores cruzados e hierárquicos. Pelo próprio conceito de fator, vê-se que em um experimento, a escolha dos fatores e seus respectivos níveis, que farão parte do experimento é basicamente um problema do pesquisador.

**FATOR DE EFEITO FIXO:** É aquele onde o pesquisador estabelece quais são os níveis a serem utilizados ou ensaiados. Como por exemplo, temos: Um pesquisador dispõe de variedades da cultura de cana-de-açúcar para realizar um experimento de competição de variedades. Neste caso diz-se que o fator variedade é fixo.

**FATOR DE EFEITO ALEATÓRIO:** É aquele cujos níveis são selecionados aleatoriamente que é a amostra de uma população de níveis para os quais as inferências deverão ser válidas, assim as inferências obtidas poderão ser estendidas para a população de níveis do fator considerado. Um pesquisador entre todas as variedades da cultura de cana-de-açúcar, sorteou 5 variedades e com elas deseja realizar um experimento de competição de variedades. Neste caso diz-se que o fator variedade é aleatório.

**FATORIAL:** O fatorial de um número inteiro  $n$  é o produto  $1.2.3. \dots n$ . Ele é representado por  $n!$ . Sua generalização para valores não-inteiros é feita por meio de função gama. Em inglês factorial.

**FATORIAL:** De um número inteiro  $n$ , o produto de todos os números inteiros consecutivos, desde 1 até  $n$ .

**FATORIAL COMPLETO:** Ensaio fatorial cujos tratamentos sejam todas as combinações possíveis entre todos os níveis de todos os fatores.

**FATORIAL FRACIONÁRIO:** Experimento fatorial que inclua somente um frção de todas as combinações possíveis entre todos os níveis de todos os fatores. Por exemplo, o fatorial fracionário  $\left(\frac{1}{5}\right) 5 \times 5 \times 5$  inclui

apenas 25 tratamentos, isto é,  $\frac{1}{5}$  das 125 combinações possíveis do fatorial de  $5 \times 5 \times 5 = 5^3$ .

**FATORIAL INCOMPLETO:** Experimento fatorial que não inclua todas as combinações possíveis entre todos os níveis de todos os fatores.

**FATORIZAÇÃO DE NEYMAN:** Essa fatorização é o chamado teorema de fatorização de Neyman enunciado assim. Seja  $X_1, X_2, \dots, X_n$  uma amostra aleatória independente e identicamente distribuída (iid) proveniente de uma população com função densidade de probabilidade  $f(x/\theta)$ , onde  $\theta$  é um parâmetro a determinar. Seja  $\hat{\theta} = G(X_1, X_2, \dots, X_n)$  um estimador de  $\theta$ . A estatística  $\hat{\theta}$  é suficiente para  $\theta$  se e somente se a função densidade de probabilidade f.d.p. (ou função de probabilidade, no caso de variáveis aleatórias discretas) conjunta de  $X_1, X_2, \dots, X_n$  que é dada por  $f_{x_1 x_2 \dots x_n}(x_1, x_2, \dots, x_n | \theta) = \prod_{i=1}^n f(x_i | \theta)$  puder ser fatorizada do seguinte modo:

$f_{x_1 x_2 \dots x_n}(x_1, x_2, \dots, x_n | \theta) = g(G(x_1, x_2, \dots, x_n) | \theta) h(x_1, x_2, \dots, x_n)$ . Onde  $h(x_1, x_2, \dots, x_n)$  é uma função real não negativa e independente do parâmetro  $\theta$  e

$g(G(x_1, x_2, \dots, x_n) | \theta)$  é outra função real não negativa que depende de  $x_1, x_2, \dots, x_n$  apenas por meio de função  $G$ .

**FECHO DE UM CONJUNTO DE ESTADOS:** É o menor conjunto fechado de estados que contém aquele.

**FECUNDIDADE:** Capacidade de produzir descendência; taxa de produção de descendência por fêmea.

**FLUTUAÇÕES DE AMOSTRAGEM:** Variações casuais das estatísticas das diversas amostras extraídas de uma população.

**FREQUÊNCIA:** Número de ocorrências de determinado fenômeno.

**FREQUÊNCIA ACUMULADA:** O número que, numa distribuição de frequências de valores, mostra, para cada intervalo, quantos valores acham-se situados entre a classe extrema e determinado intervalo, inclusive.

**FREQUÊNCIA DE CLASSE:** Número de casos incluídos em determinada classe ou intervalo de uma distribuição de frequências.

**FAIXA, REGIÃO OU INTERVALO DE CONFIANÇA PARA A LINHA:** A região ao redor da linha de regressão linear a qual supõe-se ser a linha verdadeira, com um grau de certeza fixado.

**FATOR ENTRE SUJEITOS (BETWEEN-SUBJECTS FACTOR):** Na análise de variância (ANOVA), o termo que leva em conta a variação entre grupos de tratamentos. Em um delineamento de medidas

repetidas, o fator entre sujeitos é equivalente ao de parcela completa de um delineamento de parcelas subdivididas. Em inglês split-plot design.

**FATOR INTRASSUJEITOS (WITHIN-SUBJECTS FACTOR):** Em uma análise de variância (ANOVA), é o termo que explica a variância dentro dos grupos de tratamentos. Em um delineamento de medidas repetidas, o fator intrassujeitos é equivalente ao de subparcelas do delineamento de parcelas subdivididas.

**FATORES COMUNS (COMMON FACTORS):** Os parâmetros gerados em uma análise fatorial. Cada variável original pode ser reexpressa como uma combinação linear dos fatores comuns, mais um fator adicional que é único àquela variável. Ver também análise fatorial e análise de componentes principais.

**FATORIAL (FACTORIAL):** A operação matemática indicada por um ponto de exclamação(!). Para um dado valor de  $n$ , o valor do  $n!$ , ou  $n$  fatorial, é calculado como:  $n \times (n-1) \times (n-2) \times \dots \times (3) \times (2) \times (1)$ . Portanto,  $5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$ . Por definição,  $0! = 1$ , e fatoriais não podem ser calculados para números negativos e que não sejam inteiros.

**FDP:** Ver função de densidade de probabilidade.

**FECHAMENTO (CLOSURE):** Uma propriedade matemática dos conjuntos. Os conjuntos são considerados fechados para uma operação matemática em particular, por exemplo, a adição ou a multiplicação, se o resultado de aplicar uma operação a um membro ou a membros do conjunto também for um membro do conjunto. Por exemplo, o conjunto de números pares  $\{2, 4, \dots\}$  é fechado para a multiplicação, pois o produto de dois ou mais números inteiros também será um número par. Em contraste, o conjunto dos números naturais  $\{0, 1, 2, \dots\}$  não é fechado para a subtração, pois é possível subtrair um número natural de outro e obter um número negativo, por exemplo,  $3 - 5 = -2$ , e números negativos não são números naturais.

**FORMA QUADRÁTICA (QUADRATIC FORM):** Uma matriz  $1 \times 1$  igual ao produto de  $x^T A x$ , onde  $x$  é uma matriz coluna  $n \times 1$  e  $A$  é uma matriz  $n \times n$ .

**FATORES:** Uma variável com uma ou mais categorias ou níveis nas quais os indivíduos podem ser classificados.

**FOREST PLOT:** Um diagrama utilizado em uma metanálise, que mostra o efeito estimado em cada estudo, bem como suas médias e os intervalos de confiança associados.

**FRAME:** Uma lista completa de unidades da amostra.

**FREQUÊNCIA:** O número de vezes que ocorre um valor específico ou o número de repetições de valores de uma variável.

**FUNÇÃO DE DENSIDADE DA PROBABILIDADE:** A curva, definida por uma fórmula matemática, que descreve a distribuição de frequência relativa da população. A área total sob a curva é homogênea e a proporção de observações entre quaisquer dois limites é a área sob a curva entre esses limites.

**FUNÇÕES DE LIGAÇÃO:** Uma transformação particular, por exemplo, a transformação logística na regressão logística ou a transformação log na regressão de Poisson do valor médio da variável resposta em um modelo linear generalizado; é exibida como uma combinação linear das variáveis explicativas.

**F.D.P.:** Função densidade de probabilidade.



**FENÔMENO ESTATÍSTICO:** É qualquer evento que se pretenda analisar, cujo estudo seja possível da aplicação do método estatístico. São divididos em três grupos.

**FENÔMENO DE TRANSIÇÃO:** Ocorre na borda de lentes superpostas, que constituem uma rede de descontinuidades, detectada pela variável regionalizada.

**FENÔMENO NÃO ESTACIONÁRIO:** É aquele que apresenta tendência de aumento da variabilidade com a distância, ou seja, não apresenta a estacionaridade de 2ª ordem. Isto deve-se à própria natureza da variável regionalizada ou então devido ao campo geométrico total não ter sido amostrado adequadamente.

**FENÔMENO, PROPRIEDADES:** Toda pesquisa é vista como direcionada para a avaliação de três propriedades de afirmativas sobre fenômenos: i) sua verificabilidade empírica por métodos aceitáveis; ii) sua consistência lógica com outros fatos aceitos ou conhecidos; iii) sua utilidade social. A maioria das pesquisas disciplinadas busca avaliar propriedade em vários graus.

**FENÔMENO, REPETIÇÃO DE:** É de natureza das coisas surgirem em certas circunstâncias e de certas maneiras. Quando quer que se apresentem as mesmas circunstâncias, surge sempre o mesmo fenômeno e não outro.

**FENÔMENO, UNIVERSALIDADE:** Podem-se identificar três aspectos da universalidade de um fenômeno: i) generalidade por meio de tempo, por exemplo, será o fenômeno, um texto, um auto-conceito, dentre outros, de interesse daqui a 80 anos?; ii) generalidade por meio de espaço será que é o fenômeno de interesse a pessoas da outra cidade, do outro estado, do outro lado do oceano?; iii) aplicabilidade a um número de exemplos específicos do fenômeno geral, por exemplo, o fenômeno está sendo estudado ou este é o único?.

**FENOMENOLOGIA:** É o estudo de fenômenos, e enfatiza a descrição cuidadosa de fenômenos em todas as áreas de experiência. Os fenomenologistas não assumem que sabem o que significam as coisas às pessoas que estão sendo observadas. Eles enfatizam os aspectos subjetivos dos comportamentos das pessoas.

**FENÔMENOS DE MASSA OU COLETIVO:** São aqueles que não podem ser definidos por uma simples observação. A estatística dedica-se ao estudo desses fenômenos.

**FENÔMENOS DE MULTIDÃO:** Quando as características observadas para a massa não se verificam para o particular.

**FENÔMENOS INDEPENDENTES:** São fenômenos respeitantes à mesma variável que não têm qualquer ligação um com o outro.

**FENÔMENOS INDIVIDUAIS:** São aqueles que irão compor os fenômenos de massa.

**FENÓTIPO:** Característica de um indivíduo resultante do produto dos genes e expressada de diversas maneiras. As pessoas do tipo sanguíneo ABO, por exemplo, são classificadas em fenótipos do tipo A, B, AB ou O.

**FENÓTIPO:** Expressão visível ou mensurável de um caráter; por exemplo, peso a desmama, ao abate, reprodução, dentre outros.

**FREQUÊNCIA (FREQUENCY):** O número de observações em uma categoria em particular. Os valores do eixo-y ou ordenada de um histograma são frequências de observações em cada grupo identificadas no eixo-x ou abscissa.

**FREQUÊNCIA RELATIVA (RELATIVE FREQUENCY):** Ver proporção.

**FUNÇÃO (FUNCTION):** Uma regra matemática para atribuir um valor numérico a outro valor numérico. Em geral, as funções envolvem a aplicação de operações matemáticas, como adição ou multiplicação ao valor inicial, para obter o resultado da função. As funções muitas vezes são escritas com letras em itálico: a função  $f(x) = \beta_0 + \beta_1 x$ , significa pegar o x, multiplicá-lo por  $\beta_1$  e somar o  $\beta_0$  ao produto. Isto dá um novo valor, que é relacionado ao valor original, x, pela operação da função f(x).

**FUNÇÃO DENSIDADE DE PROBABILIDADE (FDP) (PROBABILITY DENSITY FUNCTION):** A função matemática que atribui a probabilidade de observar um resultado  $X_i$  a cada valor de uma variável aleatória X.

**FUNÇÃO DE DISTRIBUIÇÃO (OU DE REPARTIÇÃO F(x) OU CUMULATIVA (FDC) ) (CUMULATIVE DISTRIBUTION FUNCTION):** A área, acima de um ponto X, sob a curva que descreve uma variável aleatória. Formalmente, se X for uma variável aleatória com função de distribuição de probabilidade f(X), então a função de distribuição cumulativa F(X) é igual a:  $P(x < X) = \int_{-\infty}^x f(x) dx$ . A função de distribuição

cumulativa é usada para calcular a probabilidade em cauda de uma hipótese, ou cotas de probabilidades.

**FUNÇÃO DE INFLUÊNCIA (INFLUENCE FUNCTION):** Um gráfico de diagnóstico para avaliar o impacto de cada observação na inclinação e no intercepto de um modelo de regressão. Neste gráfico, os interceptos obtidos por jackknife no eixo-y são plotados contra as inclinações obtidas por jackknife no eixo-x. A inclinação desse gráfico sempre deve ser negativa, mas dados discrepantes em observações com efeitos indevidos ou influência no modelo de regressão.

**FUNÇÃO LOG DA VEROSSIMILHANÇA (LOG-LIKELIHOOD FUNCTION):** O logaritmo da função de verossimilhança. Maximizá-la é uma forma de determinar quais parâmetros de uma fórmula, ou hipótese, melhor explicam os dados observados. Ver também verossimilhança.

**F.G.M.:** Função geradora de momentos.

**F<sub>i</sub>:** Uma medida de associação que é uma variação do V de Cramer, usada somente para tabelas contingenciais em que pelo menos uma das variáveis contém somente duas categorias.

**f<sub>i</sub> :** frequência absoluta da classe i.



**FISHER:** Sir Ronald Aylmer Fisher, FRS foi um matemático, estatístico, biólogo evolutivo e geneticista inglês. Foi descrito por Anders Hald como um gênio que criou praticamente sozinho as fundações para a moderna ciência. Os relatos sobre a vida de Fisher foram baseados em documentos assentados em Londres Inglaterra, afirmam que os pais de Fisher eram George e Katie Fisher. Seu pai era um famoso homem de negócios, ligado às artes finas, leiloeiro da firma Robinson & Fisher, em Londres, Londres, e sua mãe era uma dedicada dona do lar. George e Katie tiveram sete filhos: quatro meninos (Geoffrey, Alan, Ronald e Alwin) e três meninas (Evelyn,

Sybyl e Phyllis). Depois do nascimento do primogênito Geoffrey, em 1876 e Evelyn, em 1877, o casal deu o nome de Alan ao terceiro filho, que nasceu no ano seguinte, em 1878, mas que morreu com apenas três anos de idade. Katie, que era uma pessoa muito supersticiosa, decidiu que todos os seus filhos a partir de então teriam a letra y em seus nomes, já que ela julgava que Alan teria morrido, prematuramente, por não possuir esta letra em seu nome. Em 1890, mesmo antes do nascimento de Fisher, sua chegada ao mundo já era uma grande vitória e surpreendeu a todos os familiares. Katie, ao ser internada no hospital para dar a luz ao seu quarto filho, o qual receberia o nome de Isaac, novamente teve uma má notícia: seu pequeno tinha nascido morto. Entretanto, esta notícia foi rapidamente substituída por uma boa nova, quando o obstetra anunciou que Katie tinha mais uma criança em seu ventre. Naquela época os recursos da medicina eram limitados e não permitiam, na maioria das vezes, diagnosticar a gravidez de gêmeos. Dessa forma, as esperanças reascenderam e o segundo gêmeo nasceu forte e saudável: era Fisher. Por esse fato, desde seu nascimento, ele era uma criança preciosa e muito amada pelos seus pais e irmãos, sendo chamado carinhosamente por Ron ou Ronnie. Ainda na infância, passou a sofrer de hipermetropia tendo de usar óculos com lentes grossas. Desde então, Fisher internalizou sua visão, desenvolvendo grande habilidade na solução de problemas através do raciocínio com imagens, por meio da trigonometria. A maioria de suas teorias teve como base a formulação através de imagens para só, num passo a seguir, ser desenvolvida sob a forma escrita. Aos 10 anos, em 1900, Fisher foi matriculado na escola preparatória. Desde então, já se destacava, terminando seu curso, em 1903, em primeiro lugar em matemática, latim e ciências e em último lugar em francês, dentre 11 alunos. Joan Fisher Box, filha de Fisher, em seu livro sobre a vida do pai, se referindo à péssima classificação dele em francês, escreveu que ele nunca teve muita paciência com irrelevantes. Fisher sofreu sua primeira grande perda na vida aos 14 anos quando sua mãe morreu, em 1904. Neste mesmo ano, ingressou no Harrow College, conquistando, em 1906, sua primeira medalha numa competição de matemática, a Neeld Medal. Um ano e meio após a morte da mãe, seu pai perdeu toda a fortuna da família em negócios arriscados e, diante da crise financeira eminente, ele teria poucas oportunidades de continuar seus estudos em excelentes centros de ensino. Dessa forma, ainda em Harrow, pleiteou uma bolsa de estudos para ingressar no Caius & Goinville College, em Cambridge, e graças a seu brilhante desempenho e destaque, obteve uma resposta positiva. Em 2010 comemora-se o 120º ano de nascimento do mais importante estatístico do século XX: Ronald Aylmer Fisher. Ele, também chamado carinhosamente de Ron ou Ronnie, nasceu em 17 de fevereiro de 1890, em Londres. Foi o quarto filho do casal George e Katie Fisher. Desde a infância passou a sofrer de hipermetropia e, por isso, desenvolveu grande habilidade com imagens geométricas em sua mente. Destacou-se desde os primeiros anos de escola por ser um brilhante aluno. Em 1909, obteve uma bolsa de estudos, ingressando no Caius & Goinville College, Cambridge, para estudar matemática e astronomia, graduando-se em 1913. Porém, já naquela época, demonstrava grande interesse pela biologia. Casou-se com Ruthie Eileen em 26 de abril de 1917, tendo dois filhos e seis filhas. Em 1919, assumiu o cargo de estatístico chefe da Rothamsted Experimental

Station, onde foi capaz de desenvolver suas ideias ligadas à eugenia e à experimentação. Assumiu o posto de Galton Professor of Eugenics, na University College of London no lugar de K. Pearson, em 1933, e a cadeira de Arthur Balfour Professor of Genetics, na Cambridge University, em 1943. Em 1959, foi convidado a assumir um cargo de pesquisador na Commonwealth Scientific and Industrial Research Organisation, junto à University of Adelaide, Austrália. Fisher publicou em seus livros e artigos novas ideias e conceitos sobre pequenas amostras, delineamentos experimentais tais como em blocos e fatorial, casualização, análise de variância (ANOVA), estudos relacionados às distribuições teóricas de probabilidade t,  $\chi^2$ , N, z e F, método da máxima verossimilhança, teste de significância, análise multivariada, regressões múltiplas, dentre outros.

**FICHAMENTO:** São as anotações de coletas de dados registradas em fichas para posterior consulta.

**FIDEDIGNIDADE:** O mesmo que reprodutibilidade.

**FIDEDIGNIDADE:** Refere-se a estabilidade dos resultados de um teste, ou seja, ao grau de consistência dos escores. Veja validade e fidedignidade.

**FILTRO:** Qualquer método para isolar componentes harmônicos em uma série temporal; o análogo matemático do filtro de um raio de luz ou som pela remoção de efeitos não sistemáticos e evidenciando os harmônicos que o constituem. Em inglês Filter.

**FLUTUAÇÃO:** Pequenas quantias de ruído aleatório acrescentado aos valores usados no treinamento de uma rede neural que suavizam a função de erro e auxiliam na obtenção de uma solução ótima global.

**FLUTUAÇÕES DE AMOSTRAGEM:** Variações casuais das estatísticas das diversas amostras extraídas de uma população. Em inglês sampling fluctuations.

**FOLHA DE REGISTRO:** Relação de registro do entrevistador com o número e os resultados de um contato.

**FUNÇÃO MONÓTONA CONTÍNUA (CONTINUOUS MONOTONIC FUNCTION):** Uma função que é definida para todos valores em um intervalo e que preserva a ordem dos operandos.

**FOLLOW-UP:** Também chamado de estudo cohorte, tem por essência identificar um grupo de indivíduos que serão acompanhados para ver o que acontece com eles. Em inglês Follow-up.

**FORÇA DA ASSOCIAÇÃO:** Grau em que a variação em uma variável explica a variação em outra. Quanto maior for a força de associação entre as variáveis, mas completamente a variação de um predirá a variação da outra.

**FORMA QUADRÁTICA:** Uma forma quadrática é uma expressão genérica com termos de segunda ordem. Em modelos com duas variáveis é normalmente escrita como:  $ax^2 + 2hxy + by^2 + 2gx + 2fy + c$ . Em inglês quadratic form.

**FÓRMULA ALTERNATIVA PARA A INFORMAÇÃO DE FISHER  $I(\theta)$ :** Fórmula dada pelo seguinte teorema. Se  $\ln f(x \setminus \theta)$  for duas vezes continuamente derivável em ordem ao parâmetro  $\theta$ , e sob condições de regularidade semelhante ao teorema de Cramer-Rao, a informação de Fisher,  $I(\theta)$ , da variável aleatória x é dada por:

$$I(\theta) = E \left[ \left( \frac{\partial}{\partial \theta} \ln f(X \setminus \theta) \right)^2 \right] = E \left( \frac{\partial^2}{\partial \theta^2} \ln f(X \setminus \theta) \right).$$

**FÓRMULA DA PROBABILIDADE BINOMIAL:** Expressão usada para o cálculo de probabilidades em um experimento binomial.

**FÓRMULA DE KELLEY:** Para o cálculo do afastamento médio, AM, de N números  $x_1$ , de média aritmética  $\bar{X}$ , sendo n o número dos  $x_1$  menores que  $\bar{X}$ :  $AM = \frac{2}{N}(n\bar{X} - \sum X_i)$ ; Onde o sinal de somatório

abrange somente os  $x_1 < X_1 < \bar{X}$ . Também se diz fórmula de Wald-Herring.

**FÓRMULA DE PEREQUAÇÃO (MECÂNICA) DE WITTSTEIN:** Equação dada pela expressão  $u'_1 = [u_{i-4} + 2u_{i-2} + 4u_{i-1} + 5u_i + 4u_{i+1} + 3u_{i+2} + 2u_{i+3} + u_{i+4}]$ , em que  $u_1$  representa os termos sucessivos de uma série de dados experimentais e  $u'$ , os valores que lhes devem ser substituídos (WITTSTEIN, 1867).

**FÓRMULA DE WALD-HERRING:** O mesmo que fórmula de Kelley.

**FÓRMULA DE WAPPÄUS:** É fórmula  $t = \frac{2(P-p)}{n(P+p)}$  da taxa de crescimento relativo de uma população,

onde p denota a população do início de um dado intervalo constituído de n períodos, geralmente anos e P a população final. Usa-se para cálculos de interpolação e extrapolação.

**FÓRMULA DE WATERS:** Para o cálculo do crescimento da população da circunscrição territorial e de dado país P. Escreve-se:  $P_t = \frac{P_1 - P_t}{P_1 - P_0} P_0 + \frac{P_t - P_0}{P_1 - P_0} P_1$ , onde,  $p_t$  denota população de e na época t;  $p_0$ , a popu-

lação de e em um determinado censo (origem);  $p_1$ , a população de e no censo imediato a  $p_0$ ;  $P_t$ ,  $P_0$  e  $P_1$  têm sentido análogo a  $P_t$ ,  $P_0$  e  $P_1$ , mas em relação à população de P (WATERS, 1907).

**FÓRMULA DE WAUGH:** Para o cálculo aproximado da moda da distribuição de frequência da variável aleatória x, é dada pela igualdade

$$M_0 = \frac{X_i + 2M(\bar{X} - X_i) - (m-1)h_i}{2(m-1)}$$

Onde  $M_0$  denota moda,  $x_1$  o ponto-médio da classe extrema inferior, m o número de classes da distribuição,  $\bar{X}$  sua média aritmética e h a amplitude uniforme das classes, dada por Waugh (1943).

**FÓRMULA QUADRÁTICA:** Uma função quadrática homogênea da forma  $Q = \sum_i \sum_j a_{ij} x_i x_j = x'Ax$  onde A é a matriz da forma quadrática e geralmente assumida como simétrica. A forma quadrática é importante em análises multivariadas e, em particular, análises de variância. Em inglês Quadratic form.

**FIDEDIGNIDADE TESTE-RETESTE:** Coeficiente de correlação entre duas variáveis obtidas nos mesmos indivíduos em dois distintos períodos.

**FÓRMULAS TRIGONOMÉTRICAS:** Para qualquer triângulo com ângulo direito.

$$\text{Sen } \theta = \frac{\text{Cateto oposto}}{\text{Hipotenusa}};$$

$$\text{Cos } \theta = \frac{\text{Cateto adjacente}}{\text{Hipotenusa}};$$

$$\text{Tan } \theta = \frac{\text{Cateto oposto}}{\text{Cateto adjacente}} = \frac{\text{Sen } \theta}{\text{Cos } \theta};$$

$$\text{Cosec } \theta = \frac{1}{\text{Sen } \theta};$$

$$\text{Sec } \theta = \frac{1}{\text{Cos } \theta};$$

$$\text{Cot } \theta = \frac{1}{\text{Tan } \theta}.$$

$f_{r_i}$ : Frequência relativa da classe i.

**FRAÇÃO AMOSTRADA:** Ver intensidade de amostragem.

**FRAÇÃO ATRIBUÍVEL:** Ver risco atribuível.

**FRAÇÃO DE AMOSTRAGEM:** A proporção do número total de unidades amostrais na população, estrato, ou unidade de estágio mais elevado em que a amostragem aleatória simples com unidades amostrais de contagem múltipla, quando amostrada com reposição, é feita. Em inglês Sampling fraction.

**FRAÇÃO CONSTANTE DE AMOSTRAGEM:** Propriedade da amostra estratificada em que as frações de amostragem de todos os estratos é a mesma.

**FRAÇÃO DE AMOSTRAGEM:** De uma amostra, é o quociente da divisão do tamanho da amostra pelo tamanho da população, ou seja,  $\frac{n}{N}$ . Já de um estrato de uma amostra, é o quociente da divisão do tamanho

do estrato amostral pelo tamanho do estrato populacional de que aquele é oriundo.

**FRAÇÃO (PORCENTAGEM) DO RISCO ATRIBUÍVEL NA POPULAÇÃO (RAP %):** A RAP % responde à seguinte questão: na população geral, que porcentagem do risco total para a doença em pauta ocorre devido ao fator de risco ou exposição de interesse? Valor preditivo positivo. Proporção dos indivíduos com resultado de teste positivo que realmente têm a doença de interesse.

**FRAÇÃO ETIOLÓGICA (OU ATRIBUÍVEL):** O mesmo que risco atribuível.

**FRAÇÃO ÓTIMA DE AMOSTRAGEM:** É a que conduz à partilha ótima. É o valor da fração de amostragem que leva a estimativa com a precisão desejada e o custo mínimo.

**FRAÇÃO VARIÁVEL DE AMOSTRAGEM:** É o oposto de fração constante de amostragem.

**FRAÇÕES CONTÍNUAS:** Uma sequência de frações encadeadas. A fração contínua de um número r é uma expressão semelhante a figura ao lado. Se r é um número racional a expansão termina. Como pode ser visto abaixo. Em inglês fraction continued.

$$r = \frac{1}{a + \frac{1}{b + \frac{1}{c + \frac{1}{d + \dots}}}}$$

**FRACTIS:** Números que dividem a distribuição dos dados em partes aproximadamente de mesmo tamanho.

**FREQUÊNCIA:** Número de ocorrências de um dado tipo de evento, ou o número de membros de uma população em uma classe específica. Em inglês Frequency.

**FREQUENCIA:** Número de ocorrências de determinado fenômeno. Em inglês frequency.



**FREQUÊNCIA:** É um termo genérico utilizado em epidemiologia para descrever a frequência de uma doença ou de outro atributo ou evento identificado na população, sem fazer distinção entre incidência ou prevalência. Sinônimo: ocorrência.

**FREQUÊNCIA (DE ESPÉCIES ARBÓREAS):** Número de ocorrências de uma espécie em relação ao número de unidades amostrais utilizadas, geralmente expresso em porcentagem.

**FREQUÊNCIA ABSOLUTA:** É a frequência de um valor da variável ou ainda é o número de vezes que esse valor aparece na lista de dados.

**FREQUÊNCIA ABSOLUTA:** A frequência absoluta, ou apenas frequência, de um valor é o número de vezes que uma determinada variável assume esse valor. Ao conjunto das frequências dos diferentes valores da variável dá-se o nome de distribuição da frequência ou apenas distribuição.

**FREQUÊNCIA ABSOLUTA ( $f_i$ ):** É o número de vezes que o valor de determinada variável é observado.

**FREQUÊNCIA ABSOLUTA ACUMULADA ( $F_i$ ):** É a soma das frequências absolutas anteriores com a frequência absoluta deste valor.

**FREQUÊNCIA ABSOLUTA ACUMULADA:** De um valor da variável é o somatório da sua frequência absoluta com todas as frequências dos valores anteriores.

**FREQUÊNCIA ACUMULADA:** Soma das frequências de uma classe e de todas as classes que a precedem.

**FREQUÊNCIA ACUMULADA:** A frequência acumulada de um valor é o número de vezes ou de repetições que uma variável assume um valor inferior ou igual a esse valor.

**FREQUÊNCIA ACUMULADA:** O número que, numa distribuição de frequências de escores, mostra, para cada intervalo, quantos escores se acham situados entre a classe extrema mais baixa ou mais alta e determinado intervalo inclusive. Em inglês cumulative frequency.

**FREQUÊNCIA DE CLASSE:** Número de casos incluídos em determinada classe ou intervalo de uma distribuição de frequência. Em inglês class frequency.

**FREQUÊNCIA ESPERADA:** Número de vezes que um valor da variável deve se repetir se a hipótese nula for verdadeira ou ainda número de vezes que um valor da variável deve se repetir de acordo com um determinado modelo. Em inglês expected frequency.

**FREQUÊNCIA ESPERADA:** Frequência teórica para uma célula de uma tabela de contingência ou tabela multinomial.

**FREQUÊNCIA GÊNICA:** Proporção de um alelo de um par ou série presente em uma população ou amostra, ou seja, o número de loci que um determinado gene ocorre, dividido pelo número de loci que ele poderia ocorrer, expresso como um número puro entre 0 e 1.

**FREQUÊNCIA INDEPENDENTE:** Em uma tabela de contingência, a frequência que deveria ser achada em uma cela particular, caso os atributos que a definem sejam independentes. Em inglês Independence frequency.

**FREQUÊNCIA OBSERVADA:** Contagem efetiva registrada em uma célula de uma tabela de contingência ou tabela multinomial.

**FREQUÊNCIA OBSERVADA:** Número de vezes que o valor de uma variável se repete. Em inglês observed frequency.

**FREQUÊNCIA RELATIVA:** De um valor da variável é o quociente entre a sua frequência absoluta e o número total de dados.

**FREQUÊNCIA RELATIVA:** A frequência relativa é a porcentagem relativa à frequência.

**FREQUÊNCIA RELATIVA (F<sub>ri</sub>):** Frequência de uma classe, dividida pela frequência total.

**FREQUÊNCIA RELATIVA (F<sub>ri</sub>):** É o quociente entre a frequência absoluta do valor da variável e o número total de observações.

**FREQUÊNCIA RELATIVA ACUMULADA:** De um valor da variável é o somatório da sua frequência relativa com todas as frequências relativas dos valores anteriores.

**FREQUÊNCIA RELATIVA ACUMULADA:** A frequência relativa acumulada, é a porcentagem relativa à frequência acumulada.

**FREQUÊNCIA RELATIVA ACUMULADA (F<sub>ri</sub>):** É a soma das frequências relativas anteriores com a frequência relativa desse valor.

**FREQUÊNCIA RELATIVA DE UM EVENTO:** A frequência relativa de um evento A é definida como sendo o número de vezes que A ocorre,  $n(A)$  nas  $n(S)$  repetições de um experimento com um espaço amostra associado S, isto é,  $f_{r_i} = \frac{n(A)}{n(S)}$ . Em inglês event relative frequency.

**FREQUÊNCIAS RELATIVAS:** São os valores das razões entre as frequências absolutas de cada classe e a frequência total da distribuição. A soma das frequências relativas é igual a 1 ou 100 %.

**FREQUÊNCIA SIMPLES ACUMULADA DE UMA CLASSE:** é o total das frequências de todos os valores inferiores ao limite superior do intervalo de uma determinada classe.

**FREQUÊNCIAS SIMPLES OU ABSOLUTAS:** São os valores que realmente representam o número de dados de cada classe. A soma das frequências simples é igual ao número total dos dados da distribuição.

**FIDEDIGNIDADE:** Grau fiducial que determinado conhecimento científico pode assumir. Quando se faz repetidas medidas de um evento obtendo-se resultados semelhantes, diz-se que o aparelho de medida é confiável.

**FRONTEIRA DE CLASSE:** Valor obtido em uma distribuição de frequências aumentando-se os limites superiores de classe e diminuindo-se os limites inferiores de classe da mesma quantidade, de forma que não haja lacunas entre classes consecutivas.

**FUNÇÃO:** De um modo geral, dados dois conjuntos A e B, e uma relação entre eles, dizemos que essa relação é uma função de A em B se e somente se, para todo  $X \in A$  existe um único  $Y \in B$  de modo que X se relacione com Y.

**FUNÇÃO:** Relação matemática entre um fator ou efeito e sua causa. Por exemplo, pode-se analisar o comportamento de uma variável em função de outra, ou seja,  $g = f(z)$ : rendimento de uma cultura agrícola

em função de doses crescentes de adubação nitrogenada, o ataque de insetos em função da temperatura ambiente, etc.

**FUNÇÃO AFIM:** É a função da forma  $Y(X) = aX + b$ ,  $a, b \neq 0$ . Conhecida também como função linear.

**FUNÇÃO ALEATÓRIA:** É a denominação dada ao conjunto de teores  $Z(x)$ , para todos os pontos  $x$  dentro do depósito.

**FUNÇÃO ESCORE:** É a derivada primeira da função Log L que é o logaritmo da função de máxima verossimilhança e, considerando  $\theta$  um parâmetro escala, sendo dada por,  $S(\theta) = \frac{\partial}{\partial \theta} \text{Log}(\theta)$  em que  $\partial$

refere-se ao operador de derivação ou diferenciação parcial da primeira ordem.

**FUNÇÃO BETA:** A função beta  $B(m, n)$  é definida pela seguinte integral.  $B_{(m,n)} = \int_0^1 x^{m-1} (1-x)^{n-1} dx$ . Ela

também pode ser definida em termos da função gama como:  $B_{(m,n)} = \frac{\Gamma(m)\Gamma(n)}{\Gamma(m+n)}$ . Muitas integrais podem

ser avaliadas por meio da função Beta. Em inglês beta function.

**FUNÇÃO BIJETORA:** Uma função é bijetora se ela é injetora e sobrejetora. Por exemplo, a função  $f: A \rightarrow B$ , tal que  $f(x) = 5x + 4$ . Note que ela é injetora, pois  $x_1 \neq x_2$  implica em  $f(x_1) \neq f(x_2)$ . É sobrejetora, pois para cada elemento em  $B$  existe pelo menos um em  $A$ , tal que  $f(x) = y$ .

**FUNÇÃO BINOMIAL DE PROBABILIDADE:** A função usada para calcular probabilidades em um experimento binomial.

**FUNÇÃO BIUNÍVOCA:** Dizemos que uma função é bijetiva, bijetora, biunívoca ou um a um quando ela é ao mesmo tempo injetiva ou injetora e sobrejetiva ou sobrejetora.

**FUNÇÃO COVARIÓGRAMA:** Ou função covariância, como na estatística clássica, tem por objetivo medir a correlação ou associação entre pontos separados por uma distância ou amplitude  $h$ :

$$C(h) = E[Z(X+h) \cdot Z(X)] - m^2.$$

**FUNÇÃO COVARIÓGRAMA CRUZADO:** Mede a covariância cruzada:

$$C_{ab}(h) = E\{Z_a(X+h)Z_b(X)\} - m_a m_b$$

**FUNÇÃO CUMULATIVA:** Função que indica para cada valor real  $x$  a frequência absoluta ou relativa de observações com intensidade menor ou igual a  $x$ . A representação gráfica desta função é em forma de escada.

**FUNÇÃO DE AUTOCORRELAÇÃO:** A função de autocorrelação de um processo estocástico estacionário é a autocovariância dividida pela variância. Do inglês Autocorrelation function.

**FUNÇÃO DE AUTOCOVARIÂNCIA:** Para qualquer processo estocástico estacionário a função  $\gamma(k) = C(x_{t+k}, x_t)$ , onde  $C$  representa a covariância dos termos entre parêntese é conhecida como função de autocovariância. Do inglês Autocovariance function.

**FUNÇÃO DE DENSIDADE:** Ver função de densidade de probabilidade.

**FUNÇÃO DE DENSIDADE CONDICIONAL:** Função de densidade de uma variável aleatória, quando se sabe que outra variável aleatória tem um valor especificado.

**FUNÇÃO DE DENSIDADE DE PROBABILIDADE:** Para uma variável aleatória discreta, a função de densidade de probabilidade para um valor específico é a probabilidade de uma variável aleatória tomar ou assumir esse valor; já para uma variável aleatória contínua, a função de densidade de probabilidade é representada por uma curva tal que a área sob e entre dois números é a probabilidade de a variável aleatória estar entre esses números.

**FUNÇÃO COBB-DOUGLAS:** Função potencial da forma  $Y = \alpha X^\beta \mu_i$  admitindo um erro multiplicativo, cujo modelo com transformação é dado por:  $\text{Log} Y = \text{Log} \alpha + \beta \text{Log} X + \text{Log} \mu_i$ .

**FUNÇÃO CONSTANTE:** É a função da forma  $Y = Y(X) = K$ , sendo  $K$  constante.

**FUNÇÃO CÚBICA:** É a função da forma  $Y(X) = aX^3 + bX^2 + cX + d$ ,  $a \neq 0$ .

**FUNÇÃO DE CLASSIFICAÇÃO:** Método de classificação no qual uma função linear é definida para cada grupo. A classificação é realizada calculando um escore para cada observação na função de classificação de cada grupo e então designando a observação ao grupo com o maior escore. É diferente do cálculo do escore  $Z$  discriminante, que é calculado para cada função discriminante.

**FUNÇÃO DE COVARIÂNCIA:** Uma forma coloquial para o termo função de autocovariância. Ver coeficiente de autocorrelação e função. Em inglês Covariance function.

**FUNÇÃO DE DECISÃO:** Uma função de decisão é uma regra de conduta que, em qualquer estágio de uma pesquisa amostral, indica ao estatístico se é necessário obter mais informação ou se já foi coletada suficiente informação, e em último caso, que decisão tomar sobre esta. Em inglês Decision function.

**FUNÇÃO DE DENSIDADE:** Ver Função de frequência. Em inglês Density function.

**FUNÇÃO DENSIDADE DE PROBABILIDADE CONJUNTA:** É também conhecida como f.d.p. conjunta. Se for conhecida uma variável aleatória bidimensional contínua  $(X, Y)$ , então existe sempre uma função não negativa  $f(x, y)$  designada por função densidade de probabilidade conjunta se e somente se as condições seguintes forem satisfeitas:

$$\text{i) } f(x, y) \geq 0, \quad \forall_{(x,y) \in \mathbb{R}^2};$$

$$\text{ii) } \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = 1;$$

$$\text{iii) } P(X = x, Y = y) = 0 \neq f(x, y);$$

$$\text{iv) } P[(X, Y) \in A] = \iint_A f(x, y) dx dy, \quad \forall_{A \subset \mathbb{R}^2}.$$

**FUNÇÃO DE DENSIDADE CONDICIONAL:** Se um conjunto de variáveis  $x_1, \dots, x_p, x_{p+1}, \dots, x_q$  têm densidade conjunta  $f(x_1, \dots, x_p, x_{p+1}, \dots, x_q)$ , a sub-densidade obtida mantendo constantes algumas das variáveis é chamada de condicional; assim, a densidade de  $x_1, \dots, x_p$  para  $x_{p+1}, \dots, x_q$  fixos é comumente representada por  $f(x_1, \dots, x_p | x_{p+1}, \dots, x_q)$ . Em inglês Conditional density function.

**FUNÇÃO DE DENSIDADE DE PROBABILIDADE (f.d.p.):** Nome alternativo para a função de frequências quando a variável de estudo é contínua. Em inglês Probability density function.

**FUNÇÃO DE DENSIDADE MARGINAL:** Para um vetor de variáveis aleatórias contínuas, a função de densidade de uma variável independente das outras. Em inglês Marginal density function.

**FUNÇÃO DENSIDADE DE PROBABILIDADE MARGINAL DE X:** É a função dada por

$$f_1(x) = \int_{-\infty}^{+\infty} f(x, y) dy.$$

**FUNÇÃO DENSIDADE DE PROBABILIDADE MARGINAL DE Y:** É a função dada por

$$f_2(y) = \int_{-\infty}^{+\infty} f(x, y) dx.$$

**FUNÇÃO DE DEVIANCE:** A função de deviance, definida como o logaritmo da razão de verossimilhanças avaliadas em  $y$  e  $\mu$ , é uma função que mede a discrepância entre os valores observados e os ajustados por um modelo, levando em conta a parametrização. Em inglês Deviance function.

**FUNÇÃO DE DISCREPÂNCIA:** Função que avalia a proximidade, ou não, de dois conjuntos de valores gerados por dois processos. Também utilizada para comparar as estimativas feitas por um modelo e os dados reais. Em inglês Discrepancy function.

**FUNÇÃO CANÔNICA:** Relação correlacional entre duas composições lineares que são variáveis estatísticas canônicas. Cada função canônica tem duas variáveis estatísticas canônicas, uma para o conjunto de variáveis dependentes e outra para o conjunto de variáveis independentes. A força da relação é dada pela correlação canônica.

**FUNÇÃO COMPOSTA:** Situação que aparece na prática por meio de um encadeamento de funções como, por exemplo,  $Y = Y(\mu)$  e  $\mu = \mu(X)$ , então,  $Y = Y(\mu(X)) = \varphi(X)$ . A função  $\varphi(X)$  denomina-se função composta das funções  $Y(\mu)$  e  $\mu(X)$ . Um exemplo que aparece nas aplicações estatísticas é dado pela função densidade de probabilidade do tipo normal ou Gaussiana, ou seja,  $\varphi(X) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}X^2}$ ,  $\mu = \mu(X) = -\frac{1}{2}X^2$  e  $Y = Y(\mu) = \frac{1}{\sqrt{2\pi}} e^{\mu}$ . E  $x$  é igual a  $(X) = Y[\mu(X)] = \frac{1}{2\pi} e^{-\frac{1}{2}X^2}$ . Os exemplos mais importantes de funções compostas são dados pelas funções  $\alpha(X) = [\mu(X)]^a$  sendo  $\alpha$  constante,  $\alpha(X) = e^{\mu(X)}$  e  $\alpha(X) = \ln[\mu(X)]$ .

**FUNÇÃO CONTÍNUA:** Para muitas funções de grande importância prática, além de possuírem o limite quando  $x$  tende para um dado valor  $a$ , isto é além de existir  $\lim_{x \rightarrow a} Y(X)$ , esse limite é igual a  $Y(a)$ . Em outras palavras, existe o limite e é igual ao valor da função no ponto  $a$ , isto é,  $\lim_{x \rightarrow a} Y(X) = Y(a)$ . Funções com essas características são chamadas de contínuas em  $a$ , e contínuas em um intervalo se forem contínuas em todos os pontos desse intervalo.

**FUNÇÃO DISCRIMINANTE:** Uma variável estatística das variáveis independentes selecionadas por seu poder discriminatório usado na previsão de pertinência ao grupo. O valor previsto da função discriminante é o escore  $Z$  discriminante, o qual é calculado para cada objeto, pessoa, empresa ou produto na análise. Ela toma a forma da equação linear

$$Z_{jk} = a + W_1 X_{1k} + W_2 X_{2k} + \dots + W_n X_{nk}, \text{ onde}$$

$Z_{jk}$  = escore  $Z$  discriminante da função discriminante  $j$  para o objeto  $k$

$a$  = intercepto

$W_i$  = peso discriminante para a variável independente  $i$ .

$X_{ik}$  = variável independente  $i$  para o objeto  $k$

**FUNÇÃO DISCRIMINANTE:** Dimensão de diferença ou discriminação entre os grupos na análise de variância multivariada (MANOVA). A função discriminante é uma variável estatística das variáveis dependentes.

**FUNÇÃO DISCRIMINANTE LINEAR DE FISHER:** Ver função de classificação.

**FUNÇÃO DE ATIVAÇÃO:** Função matemática com a interconexão que traduz o escore múltiplo dos valores ponderados de entrada em um único valor de saída. Apesar de a função poder ser de qualquer tipo, a forma mais comum é uma função sigmóide.

**FUNÇÃO DE BASE RADIAL:** Forma alternativa de rede neural que funciona muito semelhantemente ao perceptron multicamadas, no sentido de ser apropriada para problemas de previsão e classificação.

**FUNÇÃO DE DENSIDADE DE PROBABILIDADE CONTÍNUA:** Se  $f_{x,y}(x, y)$  é uma função de densidade de probabilidade conjunta de duas variáveis aleatórias  $X$  e  $Y$ , então  $f_{x,y}(x, y) = \Pr[(X = x) e (Y = y)]$

**FUNÇÃO DENSIDADE DE PROBABILIDADE CONDICIONADA DE X DADO Y = y FIXO:** É a função dada por  $f(x|Y = y) = \frac{f(x, y)}{f_2(y)}$ , (sendo  $f_2(y) \neq 0$ ).

**FUNÇÃO DENSIDADE DE PROBABILIDADE CONDICIONADA DE Y DADO X = x FIXO:** É a função dada por  $f(y|X = x) = \frac{f(x, y)}{f_1(x)}$ , (sendo  $f_1(x) \neq 0$ ).

**FUNÇÃO DE DISTRIBUIÇÃO:** A função de distribuição também denominada de função de repartição ou função de distribuição acumulada, representada por  $F(x)$ , é definida como  $F(x) = P(X \leq x)$ . Em inglês distribution function.

**FUNÇÃO DE DISTRIBUIÇÃO:** A função de distribuição acumulada ou simplesmente função de distribuição de uma variável aleatória é definida em cada valor  $x$  real como sendo  $F(x) = P(X \leq x)$ . Em inglês cumulative distribution function.

**FUNÇÃO DE DISTRIBUIÇÃO:** A função de distribuição  $F(x)$  de uma variável  $x$  é a frequência total dos membros com valores da variável menores ou iguais a  $x$ . Como uma regra geral, a frequência total é igual a um, neste caso a função de distribuição é a proporção dos membros que têm valores menores ou iguais a  $x$ . Em inglês Distribution function.

**FUNÇÃO DE DISTRIBUIÇÃO CONDICIONADA DE X DADO Y = y FIXO:** É a função de probabilidade dada por  $F(x|Y = y) = F(x|y) = P(X \leq x|Y = y) = \int_{-\infty}^x f(u|Y = y) du$ ,  $\forall_{x \in \mathbb{R}}$ .

**FUNÇÃO DE DISTRIBUIÇÃO CONDICIONADA DE Y DADO X = x FIXO:** É a função de probabilidade dada por  $F(y|X = x) = F(y|x) = P(Y \leq y|X = x) = \int_{-\infty}^y f(v|X = x) dv$ ,  $\forall_{y \in \mathbb{R}}$ .



**FUNÇÃO DE DISTRIBUIÇÃO CONJUNTA:** É também conhecida como f.d. conjunta. Se for conhecida uma variável aleatória bidimensional contínua (X,Y), então existe sempre uma função F(x,y) designada por função de distribuição conjunta se e somente se as condições seguintes forem satisfeitas:

$$i) F(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) du dv, \quad \forall_{(x,y) \in \mathbb{R}^2}; \quad ii) 0 \leq F(x, y) \leq 1, \quad \forall_{(x,y) \in \mathbb{R}^2}.$$

**FUNÇÃO DE DISTRIBUIÇÃO EMPÍRICA:** Dada uma amostra ordenada de n observações independentes  $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ , a função  $S_n(x)$  definida como: 0 para  $x \leq x_{(1)}$ ,  $k/n$  para  $x_{(k)} < x < x_{(k+1)}$ , 1 para  $x_{(n)} < x$  é chamada função de distribuição empírica. Em inglês Empirical distribution function.

**FUNÇÃO DE DISTRIBUIÇÃO CONDICIONADA DE X DADO Y = y<sub>j</sub> FIXO:** É a função dada por  $F(x|Y = y_j) = P(X \leq x_i | Y = y_j) = \sum_{x_i \leq x} f(x_i | Y = y_j), \quad \forall_{x \in \mathbb{R}}$ .

**FUNÇÃO DE DISTRIBUIÇÃO CONDICIONADA DE Y DADO X = x<sub>i</sub> FIXO:** É a função dada por  $F(y|X = x_i) = P(Y \leq y_j | X = x_i) = \sum_{y_j \leq y} f(y_j | X = x_i), \quad \forall_{y \in \mathbb{R}}$ .

**FUNÇÃO DE DISTRIBUIÇÃO CONJUNTA ASSOCIADA A UMA VARIÁVEL ALEATÓRIA BIDI-MENSIONAL DISCRETA (X,Y):** É a função dada por

$$F(x, y) = P(X \leq x, Y \leq y) = \sum_{x_i \leq x} \sum_{y_j \leq y} f(x_i, y_j), \quad \forall_{(x,y) \in \mathbb{R}^2}.$$

**FUNÇÃO DE DISTRIBUIÇÃO MARGINAL DE X:** É a função dada por

$$F_1(x) = P(X \leq x) = \sum_{x_i \leq x} f_1(x_i), \quad \forall_{x \in \mathbb{R}}.$$

**FUNÇÃO DE DISTRIBUIÇÃO MARGINAL DE Y:** É a função dada por

$$F_2(y) = P(Y \leq y) = \sum_{y_j \leq y} f_2(y_j), \quad \forall_{y \in \mathbb{R}}.$$

**FUNÇÃO DE DISTRIBUIÇÃO MARGINAL DE X:** É a função dada por

$$F_1(x) = P(X \leq x) = \int_{-\infty}^x f_1(u) du, \quad \forall_{x \in \mathbb{R}}.$$

**FUNÇÃO DE DISTRIBUIÇÃO MARGINAL DE Y:** É a função dada por

$$F_2(y) = P(Y \leq y) = \int_{-\infty}^y f_2(v) dv, \quad \forall_{y \in \mathbb{R}}.$$

**FUNÇÃO DE ESCORE:** A função de escore de X, denotada por U(X;θ), é definida como U(X;θ) = ∂logp(X|θ)/∂θ, sendo p(X|θ) a distribuição de X. Em inglês Score function.

**FUNÇÃO DE FREQUÊNCIA:** Expressão que fornece a frequência de uma variável x como função de x; ou, para variáveis contínuas, a frequência elementar em uma amplitude dx. Em inglês Frequency function.

**FUNÇÃO DE LIGAÇÃO:** Em modelos lineares generalizados, é a função que relaciona a média da parte aleatória com o preditor linear. Em inglês Link function.

**FUNÇÃO DE PERDA:** No processo de decisão tendo como base a variável x, a desvantagem pode aparecer por meio de desconhecimento da verdadeira distribuição de x. A extensão da desvantagem é geralmente uma função da verdadeira distribuição e da decisão que é realmente feita. Esta função é chamada de função de perda. Em inglês Loss function.

**FUNÇÃO DE PREVISIBILIDADE:** Especificar como usar pesquisa descritiva e diagnóstica para prever os resultados de uma decisão planejada de marketing.

**FUNÇÃO DE PROBABILIDADE:** Para uma variável aleatória discreta, a função de probabilidade de a variável tomar esse valor.

**FUNÇÃO DE PROBABILIDADE:** Uma função f é dita de probabilidade se associa a cada evento A do espaço amostra S, um número real que satisfaz as seguintes condições: i) f(x) ≥ 0 e ii) Σf(x) = 1, para todo x. A função de probabilidade é também denominada, em alguns textos, de função massa de probabilidade. Observação: por função de probabilidade entende-se a função associada a uma variável aleatória discreta. No caso contínuo a função é denominada de densidade de probabilidade. Em inglês probability function.

**FUNÇÃO DE PROBABILIDADE DE POISSON:** A função usada para calcular as probabilidades de Poisson.

**DISTRIBUIÇÃO DE PROBABILIDADE DE UMA FUNÇÃO DE UMA VARIÁVEL ALEATÓRIA:**

Seja (X,Y) uma variável aleatória bidimensional contínua com função densidade de probabilidade conjunta f(x,y), e φ uma função definida sobre o domínio da variável aleatória (X,Y), definida por φ(x,y) = (φ<sub>1</sub>(x,y), φ<sub>2</sub>(x,y)) = (z,w). Seja φ a função inversa, isto é,

$$\varphi^{-1} = \psi : (z, w) \rightarrow (\psi_1(x, y), \psi_2(x, y)) = (z, w). \text{ Sendo o jacobiano da aplicação inversa diferente de zero, isto é, } J_\psi = \begin{vmatrix} \frac{\partial x}{\partial z} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial z} & \frac{\partial y}{\partial w} \end{vmatrix} \neq 0, \text{ sendo que a função densidade de probabilidade conjunta de (Z,W) é dada por}$$

$$g(z, w) = f(\psi_1(z, w), \psi_2(z, w)) |J_\psi|.$$

**FUNÇÃO DE PROBABILIDADE CONDICIONADA DE X DADO Y = y<sub>j</sub> FIXO:** É a função dada por

$$f(x_i | Y = y_j) = f(x_i | y_j) = P(X = x_i | Y = y_j) = \frac{f(x_i, y_j)}{f_2(y_j)}, \quad \forall_{x_i} \text{ (sendo } f_2(y_j) \neq 0).$$

**FUNÇÃO DE PROBABILIDADE CONDICIONADA DE Y DADO X = x<sub>i</sub> FIXO:** É a função dada por

$$f(y_j | X = x_i) = f(y_j | x_i) = P(Y = y_j | X = x_i) = \frac{f(x_i, y_j)}{f_1(x_i)}, \quad \forall_{y_j} \text{ (sendo } f_1(x_i) \neq 0).$$

**FUNÇÃO DE PROBABILIDADE MARGINAL:** Para um vetor de variáveis aleatórias discretas, a função de probabilidade de uma variável independente das outras. Em inglês Marginal probability function.

**FUNÇÃO DE PROBABILIDADE MARGINAL:** Função de probabilidade para uma variável aleatória particular, calculada a partir de uma função de probabilidade conjunta.

**FUNÇÃO DE PROBABILIDADE MARGINAL DE X:** É a função dada por

$$f_1(x_i) = P(X = x_i) = P(X = x_i, Y \text{ qualquer}) = \sum_j f(x_i, y_j), \quad \forall x_i.$$

**FUNÇÃO DE PROBABILIDADE MARGINAL DE Y:** É a função dada por

$$f_2(y_j) = P(Y = y_j) = P(X = \text{qualquer}, Y = y_j) = \sum_i f(x_i, y_j), \quad \forall y_j.$$

**FUNÇÃO DE PROBABILIDADE:** A função usada para definir as probabilidades iguais da distribuição discreta uniforme de probabilidade.

**FUNÇÃO DE PROBABILIDADE:** Uma função, denotada por  $f(x)$ , que fornece a probabilidade de que  $x$  assumo um determinado valor para uma variável aleatória discreta.

**FUNÇÃO DE PROBABILIDADE ACUMULADA:** É a função  $F^*(A; z)$  estimada por meio da krigagem indicadora, fornece a exata proporção de teores  $Z(x)$  abaixo do teor de corte  $z$ , dentro de qualquer área  $A$ :  $\phi^*(A; Z) = \sum_{j=1}^n \lambda_j \cdot i(X_j; Z)$ , com  $x_j$  e  $A$ , onde:  $\lambda_j$  são os ponderadores da krigagem indicadora;  $i(x_j; z)$

é a função indicadora de  $z$ .

**FUNÇÃO DE PROBABILIDADE CONJUNTA ASSOCIADA A UMA VARIÁVEL ALEATÓRIA BIDIMENSIONAL DISCRETA (X,Y):** É a função dada por

$$f(x_i, y_j) = P(X = x_i, Y = y_j) = P(x_i, y_j), \quad \forall (x_i, y_j).$$

**FUNÇÃO DE RISCO:** Probabilidade que um item, funcionando no tempo  $t_0$  falhe no intervalo  $(t, t + \delta t)$  e é dado por  $h(t) = f(t) / [1 - F(t)]$ . Se esta taxa instantânea é calculada em um número de tempos de sobrevivência, as taxas de risco resultantes serão conhecidas como função de risco. Em inglês Hazard function.

**FUNÇÃO DE VALOR ABSOLUTO:** Função matemática cujo resultado é o próprio valor sem levar em consideração o signo. Do inglês Absolute value function.

**FUNÇÃO DE VEROSSIMILHANÇA:** Se a função de distribuição de variáveis aleatórias  $x_1, \dots, x_n$ , que dependem dos parâmetros  $\theta_1, \dots, \theta_k$ , é representada como  $dF = f(x_1, \dots, x_n; \theta_1, \dots, \theta_k) dx_1 \dots dx_n$ , a função  $f(x_1, \dots, x_n; \theta_1, \dots, \theta_k)$  considerada como função dos  $\theta$ 's para valores fixos dos  $x$ 's é chamada função de verossimilhança. Em inglês Likelihood function.

**FUNÇÃO DE VEROSSIMILHANÇA DE  $\theta$ , A QUAL ASSOCIA A CADA VALOR DE  $\theta$ , O VALOR F:**

$$(Y|\theta).$$

**FUNÇÃO DESCRITIVA:** Coleta e apresentação de afirmações de fato.

**FUNÇÃO DIAGNÓSTICA:** Explicação de dados ou ações.

**FUNÇÃO DISTÂNCIA NORMALIZADA:** Processo que converte cada escore de dados originais em uma variável estatística padronizada com uma média de 0 e um desvio-padrão de 1, para remover o viés introduzido por diferenças em escalas de diversas variáveis.

**FUNÇÃO ELÍPTICA:** Uma função elíptica é uma função analítica de  $C$  para  $C$  que tem duplo período, isto é, para dois valores independentes de números complexos  $z$  e  $w$  o valor em  $f(z)$  e  $f(w + z)$  são os mesmos.

Elas também podem ser definidas como a função inversa de certas integrais (denominadas de integrais elípticas) da forma  $\int \frac{dz}{\sqrt{R(z)}}$  onde "R" é um polinômio de grau três ou quatro. Em inglês elliptic function.

**FUNÇÃO ERRO:** De uma variável  $t$  é dada pela integral definida entre 0 e  $t$ :  $\text{erf}(t) = \frac{2}{\sqrt{\pi}} \int_0^t e^{-y^2} dy$ . A

probabilidade que um erro cai entre  $\pm a$  é  $\text{erf}(ha)$ , onde  $h$  é um índice de precisão, para dados que tenham uma distribuição de Gauss.

**FUNÇÃO EXPLÍCITA:** A forma como se usa uma função, explicita uma variável em relação a outra, e por isso é chamada de forma explícita. Toda função escrita na forma  $Y = Y(X)$  é chamada de explícita.

**FUNÇÃO EXPONENCIAL COM ERRO MULTIPLICATIVO:** Função da forma  $Y = \alpha \beta^{X_i} \mu_i$ , cujo modelo com transformação é dado por:  $\text{Log } Y = \text{Log } \alpha + X_i \text{Log } \beta + \text{Log } \mu_i$ .

**FUNÇÃO EXPONENCIAL (SEMILOGARÍTMICA):** Função cuja forma original é dada pela equação  $Y = \beta_0 b^X$ , e cuja forma linearizada é dada por  $\text{Ln } Y = \text{Ln } \beta_0 + \beta_1 X$ , e há restrições de variáveis na forma transformada dada por  $X > 0$ .

**FUNÇÃO EXPONENCIAL:** É a função da forma  $Y(X) = a^X$ ,  $a > 0$  e  $a \neq 1$ . Um caso especial de função exponencial, de grande aplicação prática, é dado por:  $Y(X) = \beta e^{\alpha X}$ ,  $\alpha, \beta > 0$ , onde  $e$  é um número aproximadamente igual a 2,71818.

**FUNÇÃO GAMA:** Função definida como  $\Gamma(c) = \int_0^\infty x^{c-1} e^{-x} dx$ , para  $c > 0$  e  $0 \leq x < \infty$ . Em inglês Gamma function.

**FUNÇÃO GAMA:** A função gama é uma generalização da função fatorial para a linha real ou o plano complexo. Ela é definida por:  $\Gamma(n+1) = \int_0^\infty x^n e^{-x} dx$ . Se  $n$  é um número inteiro então  $G(n+1) = n!$ . Em inglês gama function.

**FUNÇÃO GERADORA:** Função de um parâmetro  $t$  que, quando expandido como uma série de potências de  $t$ , produz os valores de algumas quantidades de interesse estatístico em forma de coeficientes como a probabilidade de eventos ou os momentos de uma distribuição de frequência. Em inglês Generating function.

**FUNÇÃO GERADORA DE MOMENTOS:** Função de uma variável  $t$  que, quando expandida como uma série de potências em  $t$  fornece os momentos de uma distribuição de frequências como coeficientes das potências. Em inglês Moment generating function.

**FUNÇÃO GERADORA DE MOMENTOS:** Introduzida por Laplace, é uma transformada de Fourier exponencial de argumento imaginário ou uma transformada de Laplace bilateral em que o parâmetro infelizmente é chamado  $t$  na maior parte da literatura e aparece com sinal positivo.

$$G_X(t) = E(e^{tX}) = \int_{-\infty}^{+\infty} e^{tX} f_X(X) dx.$$

É portanto invertível por um integral no domínio complexo, tal como as transformadas de Fourier e de Laplace. Para distribuições discretas, usa-se preferivelmente o parâmetro  $s = e^t$ .

**FUNÇÃO GLOBAL:** Veja método de interpolação.

**FUNÇÃO HIPERBÓLICA:** É a função da forma  $Y(X) = \frac{1}{X}$ ,  $X \neq 0$ .

**FUNÇÃO HIPERBÓLICA:** Função cuja forma original é dada pela equação  $Y = \frac{1}{(\beta_0 + \beta_1 X)}$ , e cuja

forma linearizada é dada por  $\frac{1}{Y} = \beta_0 + \beta_1 X$ , e há restrições de variáveis na forma transformada dada por Y diferente de zero.

**FUNÇÃO HIPERBÓLICA:** Função da forma  $y = \frac{[X]}{[aX - \beta]}$ , cujo modelo com transformação é dado por:  $\frac{1}{Y} = a - \left(\frac{\beta}{X}\right)$ .

**FUNÇÃO HIPERBÓLICA (RECÍPROCA I):** Função cuja forma original é dada pela equação  $Y = \beta_0 + \beta_1 \left(\frac{1}{X}\right)$ ,

e cuja forma linearizada é dada pela transformação  $\frac{1}{X}$ , e há restrições de variáveis na forma transformada dada por X diferente de zero.

**FUNÇÃO HIPERGEOMÉTRICA:** A função hipergeométrica, introduzida por Gauss, é a soma da série hipergeométrica:  $F_{(a,b;c;x)} = 1 + \frac{a \cdot b}{1 \cdot c} x + \frac{a(a+1)b(b+1)}{1 \cdot 2 \cdot c(c+1)} x^2 + \dots$ . Ela é a solução da equação hipergeométrica:

$x(x-1) \frac{d^2 y}{dx^2} + \{c - (a+b+1)x\} \frac{dy}{dx} - aby = 0$ . Muitas outras funções comuns podem ser escritas como

funções hipergeométricas. Em inglês hypergeometric function.

**FUNÇÃO HIPERGEOMÉTRICA DE PROBABILIDADE:** A função usada para calcular a probabilidade de x sucessos em n ensaios quando os ensaios são dependentes.

**FUNÇÃO IDENTIDADE:** Seja x uma variável, a função f é a função identidade se f(x)=x. Em inglês Identity function.

**FUNÇÃO ÍMPAR:** Uma função f é denominada ímpar quando f(x)=-f(-x), para todo x do domínio f.

**FUNÇÃO IMPLÍCITA:** A forma como se usa uma função, explicita uma variável em relação a outra, e por isso é chamada de forma explícita. Há, porém, em algumas situações, casos em que esse destaque não aparece, e por isso a função é representada ou simbolizada na forma chamada implícita F(X,Y)=0. Toda função na forma explícita Y=Y(X) pode ser escrita na forma implícita F(X,Y)=Y-Y(X)=0. Mas, da forma implícita para explícita, algumas condições devem ser satisfeitas.

**FUNÇÃO INDICADORA:** Dada uma classe de eventos A, sobre uma amostra S, a função indicadora é definida para todos os pontos s em S como I(A;S) igual a 1 ou 0 dependendo se s pertence ou não a A. Em inglês Indicator function.

**FUNÇÃO INJETORA:** Uma função é injetora se os elementos distintos do domínio tiverem imagens distintas. Por exemplo, dada a função f: A→B, tal que f(x) = 3x.

**FUNÇÃO INJETORA (INJETIVA):** Uma função f é dita injetora ou injetiva se dados dois pontos x e y do seu domínio, com  $X \neq Y$  então, necessariamente,  $f(X) \neq f(Y)$ .

**FUNÇÃO INVERSA:** Seja uma função f(x), a função inversa g=f<sup>-1</sup> tal que g(f(x))=x. Esta função nem sempre existe. Em inglês Inverse function.

**FUNÇÃO INVERSA:** uma função será inversa se ela for bijetora. Se f: A→B é considerada bijetora então ela admite inversa f: B→A. Por exemplo, a função y = 3x-5 possui inversa y = (x+5)/3. Em algumas oportunidades se por exemplo Y = Y(X), temos necessidade de obter X = X(Y), que é chamada função inversa de Y(X). na maioria dos casos, isso é facilmente obtido, porém, de um modo geral, isto se faz mediante algumas condições. Por exemplo, uma função afim  $Y = aX + b, a \neq 0$ , pode ser invertida assim.  $X = \frac{1}{a}Y - \frac{b}{a}, a \neq 0$ . Um outro exemplo é dado por.  $Y = e^X$  por emio de função  $\ln Y = X$ .

**FUNÇÃO LINEAR:** É a função da forma  $Y(X) = aX, a \neq 0$ .

**FUNÇÃO LINEAR:** Função cuja forma original é dada pela equação  $Y = \beta_0 + \beta_1 X$ , e cuja forma linearizada não existe e não há nenhuma restrição de variáveis na forma transformada.

**FUNÇÃO LOCAL:** Veja método de interpolação.

**FUNÇÃO LOGARÍTMICA:** É a função  $Y(X) = \log_a X, X > 0, a > 0 e a \neq 1$ . m particular, se  $\alpha$  for igual a e, então, a função logarítmica é representada por  $Y = \ln X$ .

**FUNÇÃO LOGÍSTICA:** É a função da forma  $Y(X) = \frac{\delta}{1 + \beta e^{-\alpha t}}, \alpha, \beta, \delta > 0 e t > 0$ .

**FUNÇÃO LOGÍSTICA:** Função da forma  $Y = \frac{N}{D}$ , onde  $N = e^{\alpha + \beta X}$  e  $D = 1 + e^{\alpha + \beta X}$ , cujo modelo com transformação é dado por:  $Ln = \frac{[Y]}{[1-Y]} = \alpha + \beta X$ .

**FUNÇÃO LOGÍSTICA:** Função cuja forma original é dada pela equação  $Y = \frac{M}{(1 + \beta_1 e^{-mX})}$ , e cuja forma linearizada é dada pela equação  $Ln\left(\frac{M}{Y-1}\right) = Ln\beta_1 - mX$ , cujas restrições de variáveis na forma trans-

formada se refere ao fato de que M é dado, M > Y, e M > 0.

**FUNÇÃO MONOTÔNICA:** Função f(x) de forma que, para xi < xj, ou f(xi) ≤ f(xj) ou f(xi) ≥ f(xj), isto é, ela é crescente ou decrescente. Em inglês Monotonic function.

**FUNÇÃO MONOTÔNICA:** Uma função f é dita monotônica, ou monótona, se puder ser classificada como crescente, estritamente crescente, decrescente ou estritamente decrescente, segundo as definições abaixo: Crescente, se:  $X < Y f(X) \leq F(Y)$ ; Estritamente crescente, se:  $X < Y f(X) < F(Y)$ . Decrescente, se:  $X < Y f(X) \geq F(Y)$ . Estritamente decrescente, se:  $X < Y f(X) > F(Y)$ .

**FUNÇÃO OPERATIVA-CARACTERÍSTICA:** Na teoria de decisões, e especialmente em controle de qualidade e análise seqüencial, uma descrição da regra de decisão que fornece a probabilidade de aceitar hipóteses alternativas quando uma hipótese nula é verdadeira. Em inglês Operating characteristic function.

**FUNÇÃO PAR:** Uma função f é denominada par quando f(x)=f(-x), para todo x do domínio f.

**FUNÇÃO PARCIAL:** Trata-se de uma função em que nenhum elemento de A que é o domínio admite mais de um representante em B que é o contradomínio.

**FUNÇÃO PODER:** Quando as hipóteses alternativas para uma hipótese nula formam uma classe que pode ser especificada por um parâmetro θ o poder de um teste de hipótese nula considerado como uma



função de  $\theta$  é chamado de função poder. Comparações entre o número de diferentes testes são feitos superpondo os gráficos das correspondentes funções poder. Em inglês Power function.

**FUNÇÃO POLÍVOCA:** são funções que, para um valor dado da variável  $z$ , associam dois ou mais números  $w=f(z)$  distintos. Conhecidas também como funções plurívocas, multiformes ou multivalentes.

**FUNÇÃO POTÊNCIA:** É a função matemática da forma  $Y(X) = \alpha X_i^\beta$ .

**FUNÇÃO POTENCIAL (LOGARÍTMICA):** Função cuja forma original é dada pela equação  $Y = \beta_0 X^{\beta_1}$ , e cuja forma linearizada é dada por  $LnY = Ln\beta_0 + \beta_1 LnX$ , e há restrições de variáveis na forma transformada dada por  $Y > 0$ ; e  $X > 0$ . Onde  $Ln$  = logaritmo neperiano e  $e$  = base dos logaritmos neperianos  $Ln$  a qual é igual a 2,27828... .

**FUNÇÃO QUADRÁTICA:** Função da forma  $Y = \alpha + \beta X + \phi X^2$ , que é uma parábola de 2º grau com curvatura sem ponto de inflexão; e de 3º grau com ponto de inflexão, dentre outros. cujo modelo com transformação não existe.

**FUNÇÃO QUADRÁTICA:** Função cuja forma original é dada pela equação  $Y = \beta_0 + \beta_1 X + \beta_2 X^2$ , e cuja forma linearizada é dada pelo uso de  $X^2$  além de  $X$ , e não há nenhuma restrição de variáveis na forma transformada.

**FUNÇÃO QUADRÁTICA:** É a função da forma  $Y(X) = aX^2 + bX + C$ ,  $a \neq 0$ .

**FUNÇÃO REAL:** Se a cada elemento  $x$  pertencente a um conveniente subconjunto de números reais, associa, por algum processo, um único número real  $Y$ . Neste caso, a correspondência é chamada de função real, pois  $y$  é um número real, de variável real devido a  $X$  ser real, representam-se as funções reais de variável real por  $Y = f(X)$  ou  $Y = Y(X)$ . Esta notação significa que a cada  $X$  corresponde um único  $Y$ , sem, contudo, dizer como se realiza essa associação.

**FUNÇÃO REPARTIÇÃO:** É uma função  $F(x)$  para V.A.D. ou V.A.C. definida para todo o valor de  $x$  de  $S$ , e que para cada  $x$  dá a proporção de elementos da amostra menores ou iguais a  $x$ .

**FUNÇÃO SEMILOGARÍTMICA:** É a função matemática da forma  $e^{Y_i} = \alpha X_i^\beta$ . Em particular, se  $\alpha$  for igual a  $e$ , então, a função logarítmica é representada por  $Y = \ln X$ .

**FUNÇÃO SEMILOGARÍTMICA:** Função da forma  $Y = \alpha e^{\beta X}$ , onde  $e$  é a base dos logaritmos neperianos, cujo modelo com transformação é dado por:  $LnY = Ln\alpha + \beta X$ .

**FUNÇÃO SEMILOGARÍTMICA:** Função cuja forma original é dada pela equação  $e^Y = \beta_0 X^{\beta_1}$ , e cuja forma linearizada é dada por  $Y = Ln\beta_0 + \beta_1 LnX$ , e há restrições de variáveis na forma transformada dada por  $X > 1$ .

**FUNÇÃO SIGMÓIDE:** Função não linear com uma distribuição geral em forma de S. Um exemplo comum é a função logística.

**FUNÇÃO SOBREJETORA:** uma função é sobrejetora se, e somente se, o seu conjunto imagem é especificadamente igual ao contradomínio,  $Im = B$ . Por exemplo, se temos uma função  $f: Z \rightarrow Z$  definida por  $y = x + 1$  é sobrejetora, pois  $Im = Z$ .

**FUNÇÃO TOTAL:** Uma função total é uma relação binária na qual cada elemento do conjunto de partida  $A$  que é o domínio está associado a um único elemento do conjunto de chegada  $B$  que é o contradomínio.

**FUNÇÃO UNÍVOCA:** Uma função  $w = f(z)$  é uniforme ou unívoca se a cada valor de  $z$ , corresponde um único valor de  $w$ .

**FUNÇÃO VARIOGRAMA:** É definida como sendo a esperança matemática dos quadrados das diferenças entre os valores de pontos no espaço, separados por uma distância  $h$ :

$$2\gamma(h) = E \left\{ [z(x+h) - z(x)]^2 \right\} \text{ ou em termos computacionais: } 2\gamma(h) = \frac{1}{n} \sum_{i=1}^n [Z(x+h) - Z(x)]^2, \text{ onde:}$$

$2\gamma(h)$  é a função variograma;  $n$  é o número de pares de pontos separados por uma distância  $h$ ;  $Z(x)$  e  $Z(x+h)$  são os valores da variável regionalizada nos pontos  $(x)$  e  $(x+h)$ , respectivamente. Esta é a função variograma absoluto, que toma os valores absolutos das diferenças ao quadrado entre os valores de dados. Contudo, dependendo da natureza dos dados, principalmente para dados com alta variação, é interessante calcular outros tipos de variogramas como segue.

**FUNÇÃO VARIOGRAMA DA VARIÁVEL INDICADORA:** É aquela obtida pela média aritmética dos quadrados das diferenças entre os valores  $(0,1)$  da variável indicadora.

**FUNÇÃO VARIOGRAMA GERAL RELATIVO:** É obtida pela divisão da soma absoluta das diferenças ao quadrado pelo quadrado da média dos pares dentro da abertura  $h$  e na direção considerada.

$$2\gamma(h) = \frac{1}{n} \cdot \frac{\sum_{i=1}^n [(Z+h) - Z(X)]^2}{\left\{ \frac{1}{n} \sum_{i=1}^n [Z(X+h) + Z(X)] \right\}^2}$$

**FUNÇÃO VARIOGRAMA LOCAL RELATIVO:** É obtida pelo quadrado da razão entre a diferença do par pela média do mesmo, moderando o efeito da deriva de segunda ordem:

$$2\gamma(h) = \frac{1}{n} \sum_{i=1}^n \frac{[(Z+h) - Z(X)]^2}{\left[ \frac{(Z+h) - Z(X)}{2} \right]^2}$$

**FUNÇÃO VARIOGRAMA LOGARÍTMICO:** É dada pela média aritmética dos quadrados das diferenças dos logaritmos dos valores das variáveis regionalizadas:

$$2\gamma(h) = \frac{1}{n} \sum_{i=1}^n \left\{ \log[Z(X+h)] - \log[Z(X)] \right\}^2$$

**FURNIVAL, ÍNDICE DE:** Índice utilizado para se comparar erro padrão da estimativa entre modelos de regressão linear com diferentes transformações da variável resposta. Ao se transformar a variável resposta num modelo linear, produz-se uma transformação na escala de análise e, conseqüentemente, o erro padrão do modelo transformado está na nova escala. O índice de Furnival é definido como um fator de correção vezes a raiz quadrada do quadrado médio do resíduo do modelo ajustado na escala transformada. O fator de correção é o inverso da média geométrica da primeira derivada da função de transformação de escala.

**FONTE PRIMÁRIA:** Fonte que dá origem direta a informação obtida num levantamento.

**FONTE SECUNDÁRIA:** Fonte que origina indiretamente a informação em uma pesquisa.

**FREQUÊNCIA ABSOLUTA (fa):** Corresponde ao numero de observações que temos em uma determinada classe ou em um determinado atributo de uma variável qualitativa.

**FREQUÊNCIA RELATIVA (fr):** Corresponde à proporção do número de observações em uma determinada classe em relação ao total de observações que temos. Esta frequência pode ser expressa em termos percentuais. Para isto, basta multiplicar a frequência relativa obtida por 100.

**FREQUÊNCIA ACUMULADA:** Frequência acumulada corresponde à soma da frequência daquela classe às frequências de todas as classes abaixo dela.

$F_{(v_1, v_2)}$ : Distribuição F de Fisher-Snedecor.

$f_X$ : Função densidade de probabilidade da variável aleatória X.

$F_X$ : Função de distribuição acumulada da variável aleatória X.

$f_{X/Y=y}$ : Função densidade de probabilidade condicional da variável aleatória X dado Y = y.

$F_{X/Y=y}$ : Função de distribuição acumulada condicional da variável aleatória X dado Y = y.

$f_{XY}$ : Função densidade de probabilidade do vetor aleatório (X,Y).

$F_{XY}$ : Função de distribuição acumulada do vetor aleatório (X,Y).

$f_\alpha$ : Quantil 1-  $\alpha$  da distribuição F.

# G

**GAMA ( $\alpha, \beta$ ):** Distribuição gama.

**GAMMA:** Medida de associação apropriada a variáveis ordinais.

**GANHO MÉDIO DIÁRIO:** Medida de mudança diária de peso corporal do animal em um teste de alimentação. Normalmente duram entre 140 ou 160 dias.

**GARIMPAGEM DE DADOS:** O uso de software avançado ou estatístico para descobrir padrões não óbvios em um banco de dados.

**GAUSSIANO:** Modelo de variograma teórico descrito pela seguinte equação:

$$\lambda(h) = C_0 + C \left[ 1 - \exp\left(-\left(\frac{h}{a}\right)^2\right) \right], \text{ para } h > 0. \text{ Da mesma forma que o modelo exponencial, este modelo nunca}$$

alcança o patamar, sendo aceito na prática o patamar em  $a$ , onde  $a$  é a distância na qual ocorre a interseção entre a reta tangente e o patamar teórico.

**GEOCODIFICAÇÃO:** Definição da posição de elementos geográficos referenciada a um sistema de coordenadas padrão. Normalmente é feita por meio da definição de um centróide.

**GEOM ( $p$ ):** Distribuição geométrica.

**GEOMETRIA DIFERENCIAL:** A geometria diferencial estuda as propriedades das variedades que são diferenciáveis. Alguns exemplos das propriedades estudadas são a curvatura e a torção para curvas e curvaturas de superfícies. Em inglês differential geometry.

**GEOMETRIA NÃO-EUCLIDIANA:** É uma geometria na qual o postulado das paralelas de Euclides é substituído pela idéia de que não existe uma única linha ou então de que existe mais de uma linha paralela a linha dada. Se existe mais de uma linha então a geometria é dita hiperbólica e se a hipótese é de que não existem paralelas então a geometria é dita elíptica. Em inglês non-Euclidean geometry.

**GEOPROCESSAMENTO:** Tecnologia que abrange o conjunto de procedimentos de entrada, manipulação, armazenamento e análise de dados espacialmente referenciados.

**GENÓTIPO:** Conjunto de todos os genes que determinado indivíduo possui. No caso, por exemplo, do sistema sanguíneo ABO, as pessoas do grupo A possuem o genótipo AA ou AO; do grupo B, genótipo BB ou BO; do grupo AB, genótipo AB; e do grupo O, genótipo OO.

**GENÉTICA:** Ciência que estuda a hereditariedade e a variação.

**GENÉTICA QUALITATIVA:** Parte da genética que estuda caracteres que apresentam distribuição descontínua.



**GENÉTICA QUANTITATIVA:** Porção da genética que estuda caracteres que apresentam distribuição contínua.

**GENERALIZAÇÃO:** A nível de pesquisa em ciências sociais, a generalização é conhecida como validade externa. Os resultados de uma pesquisa são generalizáveis ou têm validade externa, quando eles podem ser aplicados em outras situações que não aquela em que a pesquisa foi realizada. As ameaças a validade externa ou generalização de uma pesquisa são: i) efeito interativo ou reativo do teste. Não podemos generalizar os efeitos do tratamento, quando o preteste sensibiliza os indivíduos ao tratamento; ii) interação entre seleção e variável experimental. Quando os indivíduos selecionados são muito suscetíveis aos efeitos da variável experimental; iii) efeitos reativos de arranjos experimentais. Refere-se à situação e ambiente da pesquisa. Observações feitas em laboratórios podem não ser verificáveis na vida real; iv) interferência devida a tratamentos múltiplos. Quando tratamentos múltiplos são aplicados em sequência, há o risco de que os indivíduos sintam os efeitos da experiência cumulativa.

**GENÓTIPO X FENÓTIPO:** Genótipo é uma descrição exata da constituição genética de um indivíduo. Fenótipo é a característica observada em um indivíduo, resultante da interação entre genes e fatores de ambiente.

**GEORREFERÊNCIA:** Relação entre as coordenadas de uma planta e as coordenadas do mundo real conhecidas.

**GERENCIAMENTO DA UTILIZAÇÃO:** O mesmo que Assistência gerenciada.

**GERADOR:** Efeitos em um planejamento fatorial fracionário que são usados para construir os testes experimentais usados no experimento. Os geradores também definem os pares associados.

**GRADIENTE ASCENDENTE (DESCENDENTE):** Uma estratégia para uma série de testes de modo a otimizar uma resposta usada juntamente com os modelos de superfície de resposta.

**GRÁFICOS:** Os gráficos são figuras geométricas destinadas a sintetizar os dados apresentados em tabelas e facilitar a análise da informação, revelando certas relações menos evidentes no material tabulado, ou seja, revelar detalhes da intensidade de variações do fenômeno estudado de forma mais clara, elegante, objetiva e concisa, onde nas tabelas não é possível ou é mais difícil de enxergar, despertando no leitor uma maior atenção e curiosidade. Algumas vezes a finalidade de um gráfico é simplesmente demonstrar a evolução de um determinado fenômeno, uma vez que, enquanto a interpretação de uma tabela requer certos conhecimentos, qualquer pessoa pode compreender que uma linha ascendente traduz um aumento do fenômeno estudado e uma linha descendente significa uma diminuição. Os princípios mais importantes na construção de um gráfico são: i) os gráficos mais simples são os mais eficientes. Não devem ser usados mais linhas ou símbolos, em um gráfico, de que aqueles que a vista possa seguir facilmente; ii) cada gráfico deve ser autoexplicativo e iii) os gráficos, assim como as tabelas, devem possuir título, fonte e notas explicativas. Certas normas devem ser seguidas na construção de um gráfico, entre elas: i) o título pode ser colocado acima ou abaixo do gráfico, que deve ser também numerado, para referência no texto; ii) quando mais de uma variável é mostrada, cada uma delas deve ser claramente diferenciada por meio de legenda. O uso de tracejados diferentes é preferível ao uso de cores, pois permite a reprodução do material; iii) as linhas do gráfico devem ser mais fortes que as das coordenadas; iv) as divisões da escala devem ser claramente indicadas, bem como as unidades nas quais a escala está dividida. Na escala em que se representam frequências, as divisões devem ser sempre uniformes, a não ser que se trate de um gráfico

logarítmico; v) o comprimento da abscissa e o da ordenada devem guardar entre si uma relação de 1:1 ou 1:1,5, pois a abscissa deve ser mais longa que a ordenada, para não produzir distorções no que se quer representar e vi) a escala vertical deve sempre começar do zero, mas, quando os valores representados forem elevados e apresentarem pouca flutuação a escala poderá ser quebrada, ou melhor, como a escala vertical deve ter início em zero, as distribuições que apresentarem somente frequências elevadas, com pequenas flutuações, têm sua representação dificultada por falta de espaço. Esse inconveniente pode ser contornado pela amputação do diagrama, interrompe-se a escala após a origem e cria-se uma nova escala mais favorável à representação do fenômeno estudado. Em algumas circunstâncias, a escala aritmética habitualmente usada não permite apresentar adequadamente o fenômeno estudado. Uma alternativa para a escala aritmética é a escala logarítmica, ou seja, a ordenada e/ou a abscissa apresentam graduações logarítmica. Existe no comércio um papel apropriado para a construção de gráficos semilogarítmicos denominados monolog, e que apresenta a escala vertical em ciclos de tal modo que as distâncias 1 a 10, 10 a 100, e 100 a 1000 são as mesmas. A escala horizontal é aritmética. Esse tipo de gráfico pode ser utilizado para: i) evidenciar o crescimento geométrico de um fenômeno, caso em que a trajetória no papel será uma linha reta; ii) representar em um mesmo gráfico duas ou mais séries com grandezas muito diferentes entre si e iii) comparar, em duas ou mais séries, as mudanças relativas em vez das variações absolutas. Exemplos de gráficos: diagramas, cartogramas, pictogramas, histogramas, polígono de frequência, gráfico em linhas, gráfico em hastes ou bastão, etc.

**GRÁFICOS-OBJETIVOS:** O objetivo de um gráfico é o mesmo de uma tabela, ou seja, tentar mostrar o essencial do fenômeno estudado em um conjunto de dados de modo que ele seja rápido, conciso e facilmente compreendido. Ao agrupar os dados em qualquer tipo de tabela, alguns detalhes se perdem para se mostrar melhor os aspectos principais. Construindo um gráfico o pesquisador estará realmente traçando um quadro da situação, e neste caso se perde mais os detalhes. Se o gráfico é bem feito, em geral, é mais fácil de lê-lo e interpretá-lo que a tabela.

**GRÁFICO HISTOGRAMA DE FREQUÊNCIAS:** Tipo de gráfico usado para representar o comportamento de variáveis contínuas ou variáveis discretas quando agrupadas em intervalos de classes. Se os intervalos de classe na distribuição de frequência são iguais pode-se traçar o histograma diretamente, usando frequências absolutas ou frequências relativas. Inicia-se o traçado com duas linhas em ângulos reto. Um eixo horizontal, o das abscissas ou eixo de x e um eixo vertical, o das ordenadas ou eixo de y. Marca-se uma escala conveniente em cada eixo. No eixo horizontal representam-se os intervalos de classes numa escala contínua, não sendo necessário que a escala comece de zero. A escala vertical deve começar de zero e nela se representam as frequências. Há ampla liberdade na escolha das escalas para os dois eixos, pois elas podem mesmo ser escolhidas de modo que o histograma pareça muito alto e estreito, ou muito largo e achatado, mas o ideal é evitar qualquer extremo. Escolhida a escala, marcam-se os intervalos de classe no eixo horizontal e levanta-se uma barra vertical que é o retângulo sobre cada intervalo de classe o que resulta numa série de retângulos justapostos. A altura de cada retângulo é proporcional à frequência, absoluta ou relativa, de cada classe. Depois de traçar os retângulos, resta tornar o histograma autoexplicativo, dando-lhe um título adequado e indicando a fonte das informações. A escala de cada eixo deve ser marcada e deve-se indicar adequadamente o que se representa em cada eixo. É necessário destacar um aspecto importante, que é o de que a comparação entre histogramas ou entre classes do mesmo histograma se faz realmente em termos de áreas, em vez de se considerar apenas as alturas dos retângulos. A importância dessa observação será mais facilmente percebida quando se for traçar um histograma a partir de classe com intervalos desiguais ou de amplitudes diferentes. Para traçar um histograma a partir

de um conjunto de dados agrupados em intervalos de classe desiguais, deve-se primeiro ajustar as frequências em função do tamanho dos intervalos, de modo a evitar um gráfico que dê uma impressão errada. Esse ajuste é feito dividindo-se a frequência pela amplitude do intervalo de classe de maneira a obter a média de ocorrência em cada intervalo. Esses valores médios é que serão representados na escala vertical do gráfico. Na escala horizontal as classes serão representadas por segmentos da reta proporcionais às amplitudes dos intervalos. O histograma de frequência permite evidenciar mais claramente as diferentes frequências das várias classes.

**GRÁFICO POLÍGONO DE FREQUÊNCIA:** Tipo de gráfico usado para representar dados quantitativos contínuos e também quantitativos discretos agrupados em intervalos de classes. O polígono de frequência substitui muitas vezes o histograma na representação de uma distribuição de frequência. É traçado do mesmo modo, exceto que, em vez de um retângulo de altura apropriada sobre cada intervalo de classe, coloca-se um ponto na mesma altura, sobre o ponto médio do intervalo. Os pontos são ligados por linhas retas para formar o polígono de frequência. De modo geral, o polígono é fechado em cada extremidade, traçando-se uma linha reta a partir dos pontos que representam as frequências da 1ª e da última classe, até o eixo horizontal. Os pontos do eixo horizontal que são escolhidos para fechar o polígono de frequência são os pontos médios do intervalo imediatamente anterior ao 1º e do imediatamente seguinte ao último da tabela, ambos com frequência zero. Isto se deve para que a área sob o polígono continue representando 100% da distribuição como no histograma, é necessário fechar o polígono nos dois extremos, conectando o primeiro e o último pontos à abscissa. Além de ser de traçado mais fácil que o histograma, o polígono de frequência tem a seu favor o fato de que é mais fácil comparar dois polígonos de frequência do que dois histogramas de frequência. O polígono permite uma melhor visualização do conjunto de dados, por isso mesmo, ele é mais adequado quando se deseja comparar duas ou mais distribuições no mesmo gráfico.

**GRÁFICO POLÍGONO DE FREQUÊNCIAS ACUMULADAS:** Tipo de gráfico usado para representar dados quantitativos contínuos e também quantitativos discretos agrupados em intervalos de classes usando as frequências acumuladas tanto absolutas como percentuais. Esse gráfico é usado para representar distribuições de frequência de variáveis quantitativas contínuas, com fins analíticos, ou, para resumir certas séries cronológicas quando se está interessado no efeito acumulado durante o período considerado, mais do que nas flutuações ocorridas nesse período. As frequências acumuladas são obtidas somando-se a frequência da 2ª classe à da 1ª, e assim sucessivamente, até a última classe, cuja frequência é somada a todas as anteriores. O gráfico será construído com esses valores representando-se cada frequência acumulada por um ponto cuja abscissa é o limite superior da classe correspondente. O primeiro ponto será conectado à abscissa no ponto correspondente ao início da primeira classe, e o último ponto será ligado à abscissa no ponto correspondente ao final da última classe, por uma linha que poderá ser graduada em porcentagens.

**GRÁFICO HISTOGRAMA DE FREQUÊNCIAS:** Tipo de gráfico usado para organizar dados de distribuições de frequências simples, o qual é um gráfico formado por um conjunto de retângulos justapostos, de forma que a área de cada retângulo seja proporcional à frequência da classe que ele representa. Sendo assim a soma dos valores correspondentes às áreas dos retângulos será sempre igual à frequência total. O histograma é construído tomando-se como referência dois eixos coordenados. No eixo horizontal, ou eixo das abscissas, são escritos os valores dos limites das classes. Por conseguinte, a dimensão horizontal de cada retângulo representará a classe. No eixo vertical, ou eixo das ordenadas, será construída a escala onde serão lidos os valores relativos ao número de observações ou frequências da classe. A área de cada retângulo do histograma corresponde à frequência da classe que o retângulo representa.

**GRÁFICO POLÍGONO DE FREQUÊNCIAS:** Tipo de gráfico usado para organizar dados de distribuições de frequências simples o qual é construído unindo-se os pontos médios das bases superiores dos retângulos do histograma de frequências. O polígono de frequências pode se referir às frequências absolutas ou às frequências relativas, conforme a escala utilizada no eixo vertical. As escalas podem aparecer simultaneamente.

**GRÁFICO HISTOGRAMA DE FREQUÊNCIAS ACUMULADAS E POLÍGONO DE FREQUÊNCIAS ACUMULADAS OU OGIVA DE GALTON CRESCENTE:** A ogiva de Galton, em homenagem ao estatístico inglês Francis Galton (1822-1911), ou polígono de frequências acumuladas tem por finalidade a representação gráfica das tabelas de frequências acumuladas. Esse tipo de gráfico pode ser utilizado para representar as frequências acumuladas abaixo de e acima de. O histograma de frequências acumuladas e o polígono de frequências acumuladas ou ogiva de Galton crescente, são construídos utilizando-se a mesma metodologia e procedimentos do histograma e do polígono de frequências para frequências simples, no entanto as escalas devem ser construídas para representar e se fazer leituras das frequências acumuladas abaixo de e acima de. O histograma é construído através de retângulos justapostos tendo como base os limites de classes e como altura as frequências acumuladas. Sendo que o polígono é obtido unindo-se os pontos referentes ao limite superior de cada classe quando tratar-se de frequência acumulada abaixo de, e unindo-se os limites inferiores quando tratar-se de frequência acumulada acima de. Quando o polígono de frequências acumuladas se refere às frequências relativas, usa-se a denominação de ogiva percentual ou polígono de frequências relativas acumuladas. A representação das frequências relativas pode ser feita utilizando-se o mesmo gráfico das frequências absolutas, alterando-se apenas a escala de leitura. Pode-se representar a escala das frequências relativas à direita do gráfico.

**GRÁFICO HISTOGRAMA DE FREQUÊNCIAS ACUMULADAS E POLÍGONO DE FREQUÊNCIAS ACUMULADAS OU OGIVA DE GALTON DECRESCENTE:** Ver histograma de frequências acumuladas e polígono de frequências acumuladas ou ogiva de Galton crescente.

**GRÁFICO OU DIAGRAMA DE RAMOS E FOLHAS:** Tipo de gráfico usado para representar pequenos conjuntos de dados, um gráfico de ramos e folhas criado por John Wilder Tukey (1915-2000) é mais rápido de fazer e apresenta informação mais detalhada. Para construir um ramo e folhas, separe cada observação em um ramo consistindo em todos os algarismos exceto o final que é o mais à direita e uma folha, o algarismo final. Uma vantagem desse diagrama sobre o histograma é que não se perde ou perde-se pouca informação sobre os dados em si. Ou melhor, para dados quantitativos, uma técnica alternativa e uma forma engenhosa para organização e apresentação de dados observados é o diagrama de ramos e folhas. A obtenção deste diagrama, além de ser fácil e rápida, tem a vantagem de preservar as informações individuais de cada dado. No diagrama de ramos e folhas, cada dado é dividido ou separado em duas partes: a parte da esquerda é denominada ramo e a da direita, folha. A seguir, as folhas são colocadas nos seus respectivos ramos. Um gráfico ramo e folhas é um gráfico de frequências construído para pontos amostrais de dois ou três algarismos, utilizando-se o último algarismo como um indicador para a classe de frequência correspondente aos algarismos iniciais. Em geral, os gráficos ramos e folhas são traçados com o ramo que são os algarismos iniciais dispostos verticalmente, e os últimos algarismos que formam as folhas dispostos horizontalmente. Muitas vezes, o comprimento de cada folha é registrada à esquerda do ramo. A vantagem do sistema ramo e folhas consiste em sua capacidade de proporcionar a informação desejada de modo rápido e fácil, sem a necessidade de fazer cálculos; uma amostra de 150 dados pode ser analisada em uns poucos minutos. Se houver mais de dois

ou três Algarismos nos valores dos dados, provavelmente deveremos arredondá-los antes de prosseguir. Pode ainda acontecer que o ramo e folhas, tal como descrito acima, resulte no agrupamento dos dados em um número muito pequeno de folhas. A escolha do número de linhas do ramo e folhas é equivalente à escolha do número de classes de um histograma.

**GRÁFICO EM BARRAS:** Tipo gráfico usado para representar as distribuições de frequências de dados qualitativos. Os gráficos em barras têm por finalidade comparar, por meio de retângulos de igual largura e alturas proporcionais às respectivas grandezas. Cada barra representa a intensidade de uma modalidade do atributo. As magnitudes das barras são representadas pelos respectivos comprimentos e seu traçado é feito tendo-se como referência uma escala horizontal. Em geral, as divisões da escala se prolongam em traços verticais por todo o gráfico, facilitando assim a leitura do comprimento de cada barra. A identificação da barra é inscrita à esquerda do gráfico. As barras só diferem em comprimento, e não em largura, a qual é arbitrária, e elas devem vir separadas uma das outras pelo mesmo espaço, como regra prática pode-se tomar o espaço entre as barras como aproximadamente a metade ou dois terços de suas larguras.

**GRÁFICO EM COLUNAS:** Tipo gráfico usado para representar as distribuições de frequências de dados qualitativos. Os gráficos em colunas ou gráficos em barras verticais prestam-se à mesma finalidade que os gráficos em barras horizontais, sendo, entretanto, preferíveis a esses últimos quando as legendas a se inscreverem sob os retângulos forem breves. Caso contrário, o emprego do gráfico em barras é mais adequado.

**GRÁFICO CIRCULAR OU EM SETORES OU SETOGRAMAS OU CARTOGRAMAS EM SETORES:** Tipo gráfico usado para representar as distribuições de frequências de dados qualitativos. Os gráficos em setores ou setogramas são usados para representar valores absolutos ou porcentagens complementares pode ser feita através de diferentes tipos de gráficos. Para construir esse tipo de gráfico parte-se do fato de que o número total de graus de um arco de circunferência é  $360^\circ$ . Assim, o número total de valores analisados que pode ser 100%, se quiser representar as porcentagens complementares corresponderá a  $360^\circ$ . Cada uma das parcelas componentes do total dos valores poderá, então, ser expressa em graus, e a correspondência se fará através de uma regra de três simples. O gráfico pode ser feito a partir de um círculo de circunferência ( $360^\circ$ ) como já comentado anteriormente ou de um semicírculo ( $180^\circ$ ) ou de um quadrante ( $90^\circ$ ) conforme se utilize o semicírculo ou o quadrante para a representação completa do fenômeno.

**GRÁFICO BOX-PLOT:** O box-plot é um gráfico que mostra a posição central, dispersão e simetria dos dados de uma amostra, comprimento de caudas e dados discrepantes. É utilizado para resumir as informações de um conjunto de dados.

**GRÁFICO RAMO E FOLHAS:** Gráfico cujo objetivo é resumir os dados de tal forma que se possa ter uma noção da forma da distribuição dos dados se simétrica ou assimétrica, a frequência de observações, vazios entre os dados e possíveis dados discrepantes ou Outlier's. As observações devem ser divididas em duas partes. A primeira, chamada de ramo, é colocada à direita de uma linha vertical, e a segunda, chamada de folhas, é colocada à esquerda.

**GRÁFICOS DE SETORES OU CIRCULARES:** O gráfico de setores ou circular, também conhecido como pizza pode ser utilizado para representar a frequência de observações de diferentes categorias. O tamanho de cada setor é proporcional ao número de observações em cada categoria. O tamanho pode ser definido em números percentuais ou absolutos. Em geral, um gráfico de setores não é um bom modo de

representar dados porque o olho humano tem dificuldades para comparar áreas relativas. em comparação com medidas lineares.

**GRÁFICO DE BARRAS:** Tipo de gráfico que assim como o gráfico de setores é utilizado para representar a frequência absoluta ou percentual de diferentes categorias. Basicamente, são construídas barras proporcionais as frequências.

**GRÁFICO HISTOGRAMA:** Um histograma é utilizado para representar a distribuição de uma variável contínua. Basicamente, as frequências observadas são representadas por classes de ocorrência. As frequências absolutas podem ser substituídas por frequências relativas ou proporcionais. Um dos principais aspectos na construção de um histograma é a definição do número de classes.

**GRÁFICOS DE DISPERSÃO OU SCATTER PLOTS:** Gráficos que mostram a dispersão de dados são úteis para identificar muitas características de dados. Além da dispersão, Outlier's, tendências entre outros aspectos, podem ser explorados com gráficos de dispersão ou Scatter plots.

**GRÁFICO NORMAL DE PROBABILIDADE:** Em muitas análises estatísticas, um dos pressupostos mais comuns é de que a variável aleatória em estudo tenha distribuição Normal. Além de vários testes de hipóteses existentes para realizar essa avaliação, no entanto existem, também, maneiras gráficas de se avaliar se uma variável aleatória possui ou não distribuição Normal que é através do gráfico normal de probabilidades. Os dados são plotados em um sistema de eixos Fazendo um gráfico para todos os possíveis valores de probabilidades acumuladas sob a curva normal ou de Gauss versus os valores da variável  $x$  e assim obtém-se a curva normal acumulada, que tem a característica sigmóide em forma de S. Um gráfico das observações ordenadas versus os valores percentuais  $P_i$  indica o comportamento da variável. No caso da avaliação da Normalidade de uma variável, espera-se que a maioria dos pontos esteja sobre a reta e concentrados na região central.

**GRÁFICO DE SUPERFÍCIE DE RESPOSTA:** Este tipo de gráfico é gerado a partir de modelos que expressam o comportamento de uma variável resposta em função de dois ou mais fatores. Para construir o gráfico de superfície de resposta são necessárias três variáveis, ou seja,  $x$  e  $y$ , representando os níveis dos fatores e uma variável  $z$ , representando o comportamento da variável resposta.

**GRÁFICO DE BARRAS:** O gráfico de barras é composto por duas linhas ou eixos, um vertical e outro horizontal. No eixo vertical são construídas as barras que representam a variação de um fenômeno ou de um processo de acordo com sua intensidade. Essa intensidade é indicada pela altura da barra. No eixo horizontal especificam-se as categorias da variável. As barras devem sempre possuir a mesma largura e a distância entre elas deve ser constante.

**GRÁFICO DE SETOR:** Os gráficos de setor ou pizza são representados por círculos divididos proporcionalmente de acordo com os dados do fenômeno ou do processo a ser representado. Os valores são expressos em números ou em percentuais (%).

**GRÁFICO HISTOGRAMA:** Na estatística, um histograma é uma representação gráfica da distribuição de frequências de uma massa de medições, normalmente um gráfico de barras verticais. Tal gráfico é composto por retângulos justapostos em que a base de cada um deles corresponde ao intervalo de classe e a sua altura à respectiva frequência.



**GRÁFICO BOXPLOT:** O boxplot é um gráfico que possibilita representar a distribuição de um conjunto de dados com base em alguns de seus parâmetros descritivos, quais sejam: a mediana ( $q_2$ ), o quartil inferior ( $q_1$ ), o quartil superior ( $q_3$ ) e do intervalo interquartil ( $IQR = q_3 - q_1$ ).

**GRÁFICO DE DISPERSÃO:** Os diagramas de dispersão são representações de duas variáveis que são organizadas em um gráfico, para observar o padrão de relacionamento entre as mesmas. É um método gráfico que permite verificar a existência ou não de relação entre duas variáveis de natureza quantitativa, ou seja, variáveis que podem ser medidas ou contadas.

**GRÁFICO DE LINHAS:** Este gráfico é útil para plotar uma série de dados ao longo do tempo ou para ligar pontos por linhas.

**GRÁFICO OU DIAGRAMA DE LINHA:** O número de ocorrências de uma variável hidrológica discreta pode ser convenientemente representado pelo chamado diagrama de linha, o qual dispõe os valores possíveis da variável em um eixo horizontal, enquanto os correspondentes números de ocorrências são representados pelas alturas das linhas verticais.

**GRÁFICO OU DIAGRAMA UNIAXIAL DE PONTOS:** O diagrama uniaxial de pontos é uma representação gráfica apropriada para amostras pequenas, de tamanho arbitrado como menor ou igual a 25 ou 30 observações, de variáveis contínuas. Os dados são inicialmente classificados em ordem crescente e, em seguida, grafados como pontos em um único eixo horizontal.

**GRÁFICO TIPO HISTOGRAMA:** Tipo de gráfico usado em hidrologia para as amostras médias e grandes, nas quais é conveniente classificá-las ou agrupá-las em subconjuntos, de modo a se ter uma melhor compreensão do padrão de variabilidade da variável em questão. Esse expediente dá origem a diversos tipos de gráficos, entre os quais se destaca o histograma. Para se construir um histograma, é necessário, primeiramente, agrupar as observações em classes, definidas por intervalos de largura fixa ou variável, e, em seguida, contar o número de ocorrências, ou seja, a frequência absoluta em cada classe.

**GRÁFICO EM POLÍGONO DE FREQUÊNCIAS:** O polígono de frequências é outra representação gráfica da tabela de frequências, sendo muito útil para diagnosticar o padrão de distribuição de uma variável. Esse polígono é aquele formado pela junção dos pontos médios dos topos dos retângulos do histograma, depois de estendê-lo por uma classe adicional de cada um de seus lados.

**GRÁFICO OU DIAGRAMA DE FREQUÊNCIA RELATIVA ACUMULADA:** O diagrama de frequências relativas acumuladas resulta da união, por linhas contínuas, dos pares formados pelos limites superiores dos intervalos de classe e pelas ordenadas consecutivamente acumuladas do histograma, desde a menor até a maior. No eixo das ordenadas, o diagrama fornece a frequência de não superação do valor correspondente da variável, lido no eixo das abscissas. De modo alternativo, o diagrama de frequências relativas acumuladas pode também ser elaborado sem a prévia construção do histograma ou da tabela de frequências.

**GRÁFICO DE CURVA DE PERMANÊNCIA:** A chamada curva de permanência é uma variação do diagrama de frequências relativas acumuladas, na qual a frequência de não superação é substituída pela porcentagem de um intervalo de tempo específico em que o valor da variável, indicado em abscissas, foi igualado ou superado. Em hidrologia, a curva de permanência é muito usada para ilustrar o padrão de variação de vazões, assim como o é para indicadores de qualidade da água, tais como turbidez de um trecho fluvial, dureza da água e concentrações de sedimento em suspensão, entre outros.

**GRÁFICO OU DIAGRAMA BOX PLOT:** O diagrama box plot consiste em um retângulo definido pelo primeiro e pelo terceiro quartis, contendo a mediana em seu interior, traça-se uma linha até o ponto que não exceda ( $Q_3 + 1,5AIQ$ ), considerado limite superior para a identificação de Ouliers. De modo análogo, traça-se outra linha a partir do lado inferior do retângulo até o limite dado por ( $Q_1 - 1,5AIQ$ ). As observações que estiverem acima ou abaixo desses limites são identificadas no diagrama e consideradas Outlier's ou valores atípicos. Para a construção dos diagramas do tipo box plot, existem outras alternativas, tais como estender as linhas verticais até os pontos de máximo e mínimo, os quais são assinalados no gráfico por barras horizontais; nesse caso, o diagrama recebe a denominação de box & whisker. Os diagramas do tipo box plot são muito úteis por permitirem uma visão geral do valor central, da dispersão, da assimetria, das caudas e de eventuais pontos amostrais discordantes. O valor central é dado pela mediana e a dispersão pela amplitude interquartil. A simetria ou assimetria da distribuição pode ser visualizada pelas posições relativas de  $Q_1$ ,  $Q_2$  e  $Q_3$ . Pode-se ter uma ideia das caudas superior e inferior por meio dos comprimentos das linhas verticais que saem do retângulo de quartis. Os diagramas do tipo box plot são particularmente úteis para comparar as características de duas ou mais amostras diferentes.

**GRÁFICO OU DIAGRAMA RAMO-E-FOLHA (STEM-AND-LEAF):** Para amostras de tamanho médio a grande, o histograma é um procedimento gráfico eficaz para ilustrar a forma da distribuição de frequências de uma variável. Para amostras menores, uma interessante alternativa ao histograma é dada pelo diagrama ramo-e-folha. De fato, esse diagrama agrupa os dados de tal modo, que há nenhuma ou pouca perda da informação contida em cada elemento amostral, realçando a presença de pontos extremos.

**GRÁFICO DIAGRAMA DE DISPERSÃO:** Um diagrama de dispersão consiste em um gráfico onde são lançados em coordenadas cartesianas os pares  $\{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$  de observações simultâneas das variáveis X e Y.

**GRÁFICO OU DIAGRAMA QUANTIS-QUANTIS (Q-Q OU Q-Q PLOT):** O diagrama quantis-quantis, ou diagrama Q-Q, é uma representação gráfica que permite visualizar a associação entre duas variáveis X e Y. Diferentemente do diagrama de dispersão entre observações simultâneas das variáveis, o gráfico Q-Q é uma representação dos dados ordenados ou quantis do conjunto  $\{x_1, x_2, \dots, x_N\}$  contra os dados ordenados ou quantis da amostra de mesmo tamanho  $\{y_1, y_2, \dots, y_N\}$ . Para elaborar um diagrama Q-Q, é necessário: i) classificar os dados de X (e Y) em ordem crescente; ii) associar aos dados classificados os seus respectivos números de ordem da classificação m, com  $1 \leq m \leq N$ ; e iii) associar aos dados classificados as correspondentes frequências ou probabilidades empíricas de não superação. Em seguida, os dados de X e Y, com igual frequência ou probabilidade empírica de não superação, são lançados em coordenadas cartesianas, formando, assim, o diagrama Q-Q. De modo diverso de um diagrama de dispersão, o qual estabelece uma associação global entre as variáveis, o gráfico Q-Q demonstra se os valores mais baixos, médios e mais altos de X estão relacionados aos seus correspondentes de Y. Em um caso limite, se as distribuições dos dois conjuntos de dados fossem idênticas, a menos de suas medidas de posição e escala ou dispersão, os pontos estariam sobre a reta  $y=x$ . O modo como os pontos se afastam dessa linearidade revelam as diferenças entre as distribuições de frequências de X e Y.

**GRÁFICOS DIAGRAMAS:** São gráficos em que a magnitude das frequências é representada através de figuras geométricas. Existem diagramas de pontos, de linhas, de áreas, de volumes, sendo este último pouco usado.

**GRÁFICOS CARTOGRAMAS:** São gráficos onde as frequências são projetadas em mapas geográficos.

**GRÁFICO CARTOGRAMA:** São mapas geográficos onde as frequências das categorias da variável estudada são projetadas nas áreas específicas mediante cores ou traçados, cujos significados constam de legenda que acompanha os mapas. Esse tipo de gráfico é usado para representar séries geográficas, ou seja, conjuntos de dados classificados de acordo com sua localização geográfica, num dado período de tempo. Por exemplo, um gráfico mostrando a variação da densidade populacional na região metropolitana do Rio de Janeiro, em 2014.

**GRÁFICO DIAGRAMA DE BARRAS:** O diagrama de barras é um gráfico em que as frequências do fenômeno estudado são representadas por meio da altura de retângulos não justapostos. É utilizado para representar distribuições de variáveis qualitativas e distribuições de variáveis quantitativas discretas. Cada barra deve ter a mesma largura e estar separada da seguinte por um espaço igual ou menor que sua largura. As barras não devem ser justapostas já que as classes da variável representada são descontínuas. Se as classes representadas não têm uma sequência especial que deva ser mantida, as barras podem ser ordenadas em sentido crescente ou decrescente que é o mais comum, para facilitar a compreensão. De modo geral, usam-se as barras no sentido vertical, mas também podem ser desenhadas barras horizontais quando os nomes das classes forem muito longos e não houver espaço suficiente para escrevê-los em baixo das barras verticais. Neste caso, a escala de frequência aparece no alto do gráfico, horizontalmente, e em espaço suficiente para a designação das classes deve ser deixado à esquerda do gráfico. No gráfico de barras podem ser utilizadas barras simples, barras compostas ou múltiplas, tais como duplas, triplas, etc.

**GRÁFICO DIAGRAMA DE BARRAS SIMPLES:** Tipo de gráfico que é usado para representar dados qualitativos e para dados quantitativos discretos. Pode ser usado para séries cronológicas quando há poucos valores a serem representados ou quando o fenômeno é estudado em períodos.

**GRÁFICO DIAGRAMA DE BARRAS COMPOSTAS:** Tipo de gráfico que é utilizado para comparar a distribuição percentual de duas variáveis qualitativas, como os dados de associação.

**GRÁFICO DIAGRAMA DE BARRAS DUPLAS:** Tipo de gráfico que é utilizado para comparar duas distribuições de variáveis qualitativas ou para comparar duas séries cronológicas ou temporais. As barras duplas podem ser parcialmente sobrepostas ou simplesmente justapostas, e diferenciadas através de um tracejado diferente que será explicado em legenda. A comparação pode ser estendida para três ou mais variáveis qualitativas, com o uso de barras triplas e assim por diante.

**GRÁFICO DIAGRAMA DE SETORES:** O diagrama de setores é u gráfico em que as frequências da variável estudada são representadas pelos setores de um círculo. Esse diagrama é ideal para ilustrar a divisão do todo em segmentos. Pode ser usado para variáveis qualitativas e para variáveis quantitativas discretas. Como os 360 ° do círculo correspondem a 100 % da distribuição, cada 1% de frequência corresponderá a 3,6 °. Assim, multiplicando-se as frequências relativas ou porcentagens por 3,6° será obtida a medida em graus do setor correspondente a cada classe da variável. A convenção é iniciar na posição correspondente a 12 horas pelo maior segmento e dispor os demais em ordem decrescente no sentido horário.

**GRÁFICO DIAGRAMA LINEAR:** Diagrama linear, ou gráfico de linhas quebradas, é o gráfico em que as frequências da variável estudada são representadas por pontos, unidos por linhas retas. É usado para mostrar a variação de um fenômeno em relação ao tempo, ou seja, uma série cronológica. Em epidemiologia, por exemplo, é usado para mostrar a variação sazonal ou distribuição mensal do número de casos ou óbitos de uma doença. Quando a variável idade é tratada como variável qualitativa, isto é, cada classe de idade é vista como uma categoria isolada, o diagrama linear pode ser utilizado.

**GRÁFICO DIAGRAMA DE CORRELAÇÃO:** O diagrama de correlação ou gráfico de dispersão é u tipo especial de gráfico utilizado para apontar relações ou associações entre duas variáveis quantitativas. Para estudo dessa relação são feitas, em cada indivíduo, duas medidas concomitantes como, por exemplo, peso e altura, peso e idade, pressão arterial, antes e depois da aplicação de um medicamento. Se uma dessas variáveis é apresentada no eixo de x e outra no eixo de y, cada indivíduo é representado no gráfico por um ponto colocado na intersecção de duas linhas imaginárias que passam pelos valores correspondentes na abscissa e na ordenada. O padrão mostrado pelo agrupamento dos pontos é indicativo de uma possível relação. Assim, se eles tendem a seguir uma linha reta, a relação é linear. Se a inclinação da reta for ascendente, reflete uma relação direta entre as variáveis estudadas, e se for descendente, reflete uma relação inversa. Se o padrão não seguir uma linha reta, a relação pode ser curvilínea como parábola, exponencial, etc.. Se o padrão revelar uma dispersão dos pontos, sem tender a uma linha definida, pode indicar ausência de correlação. Para construir o gráfico deve-se traçar os eixos do mesmo tamanho, não sendo obrigatório neste caso, iniciar nenhuma das escalas a partir de zero. Uma escala para notas de uma prova de estatística, por exemplo, será marcada no eixo horizontal ou eixo das abscissas, outra para notas de exercício, no eixo vertical ou eixo das ordenadas. Cada indivíduo será representado por um ponto resultante do cruzamento das linhas que partem das duas cotas correspondentes. Outro exemplo poderá ser os pesos ao desmame y e os pesos ao nascer ambos em quilogramas de bezerros machos da raça bovina guzerá. Nestes casos não precisa identificar uma variável dependente e outra independente, pois ambas deveram variar de forma conjunta aos pares conforme um modelo de uma distribuição teórica ou especial de probabilidade normal bivariada, sendo assim as variáveis estudadas podem ser denominadas livremente como  $x_1$  e  $x_2$ ,  $y_1$  e  $y_2$ , w e z, u e v, etc..

**GRÁFICO DE CONTROLE:** Uma disposição gráfica usada para monitorar um processo. Geralmente, ele consiste em uma linha central horizontal, correspondendo ao valor de controle do parâmetro que está sendo monitorado, e de limites inferior e superior de controle. Os limites de controle são determinados por critérios estatísticos e não são arbitrários nem são relativos a limites de especificação. Se os pontos amostrais caem dentro dos limites de controle; isto é, causas atribuídas estão provavelmente presentes. Isso sinaliza a necessidade de encontrar e remover as causas atribuídas.

**GRÁFICO C:** Um gráfico de controle para atributo que expressa o número total de defeitos por unidade em um subgrupo. Similar ao gráfico de defeitos por unidade ou gráfico U.

**GRÁFICO DE BLOCOS (OU GRÁFICOS DE BLOCOS E LINHAS):** Uma disposição gráfica de dados em que o bloco contém os 50% de dados intermediários, que é a faixa de interquartil com a mediana dividindo-a e as linhas se estendendo para os valores menores e maiores ou algum limite inferior e limite superior definidos.

**GRÁFICO DE CONTROLE DA SOMA CUMULATIVA (CUSUM):** Um gráfico de controle em que o ponto plotado no tempo t é a soma dos desvios medidos em relação a um valor-alvo para todas as estatísticas até o tempo t.

**GRÁFICO DE CONTROLE DE DEFEITOS POR UNIDADE:** Veja gráfico U.

**GRÁFICO DE CONTROLE DE SHEWHART:** Um tipo específico de gráfico de controle, desenvolvido por Shewhart (1931). Tipicamente, cada ponto plotado é um resumo estatístico calculado a partir de dados em um subgrupo racional. Os gráficos de controle foram originalmente propostos em 1924, por Walter Shewhart, que trabalhou no Bell Telephone Laboratories, com a intenção de eliminar variações anormais

pela diferença entre variações devidas às causas assinaláveis e aquelas devidas às causas aleatórias. Veja Gráfico de controle.

**GRÁFICO DE CONTROLE PARA A FRAÇÃO DEFEITUOSA:** Veja gráfico P.

**GRÁFICO DE CONTROLE PARA ATRIBUTOS:** Qualquer gráfico de controle para uma variável aleatória discreta. Veja gráfico de controle para variáveis.

**GRÁFICO DE CONTROLE PARA VARIÁVEIS:** Qualquer gráfico para uma variável aleatória contínua. Veja gráfico de controle para atributos.

**GRÁFICO DE PARETO:** Um gráfico de barras usado para classificar as causas de um problema.

**GRÁFICO DE PROBABILIDADE:** Um gráfico de dispersão usado para julgar se os dados seguem razoavelmente uma particular distribuição de probabilidades. Um gráfico de probabilidade normal é frequentemente usado para avaliar a suposição de normalidade dos dados ou dos resíduos.

**GRÁFICO DE PROBABILIDADE NORMAL:** Um gráfico construído especialmente para uma variável  $x$  geralmente na abcissa, no qual a escala de  $y$  geralmente na ordenada é feita de modo que o gráfico da distribuição normal cumulativa seja uma linha reta.

**GRÁFICO NP:** Um gráfico de controle para atributo que plota o total de unidades defeituosas em um subgrupo. Também chamado de gráfico de controle de fração defeituosa ou gráfico P.

**GRÁFICO P:** Um gráfico de controle para atributo que plota a proporção de unidades defeituosas em um subgrupo. Também chamado de gráfico de controle de fração defeituosa. Similar ao gráfico NP.

**GRÁFICO (DE CONTROLE) PARA AMPLITUDE:** Um gráfico de controle usado para monitorar a variabilidade ou dispersão em um processo. Veja Gráfico de controle.

**GRÁFICO R:** Um gráfico de controle que plota a amplitude das medidas em um subgrupo que é usado para monitorar a variância do processo.

**GRÁFICO S:** Um gráfico de controle que plota o desvio-padrão das medidas em um subgrupo de modo a monitorar a variância do processo.

**GRÁFICO U:** Um gráfico de controle para atributo que plota o número médio de defeitos por unidade em um subgrupo. Também chamado de gráfico de controle de defeitos por unidade. Similar ao gráfico C.

**GRÁFICO X:** Um gráfico de controle que plota a média das medidas em um subgrupo que é usado para monitorar a média do processo.

**GRÁFICOS INDIVIDUAIS DE CONTROLE:** É um gráfico de controle de Shewart, em que cada ponto plotado é uma medida individual, em vez de uma estatística resumida. Veja gráfico de controle e gráfico de controle de Shewart.

**GAUSSIANA:** Nome dado à distribuição normal de probabilidades, simétrica, em forma de sino, em forma campanular, de chapéu de napoleão, ou de Gauss, deduzida pelo matemático alemão Johann Carl Friedrich Gauss.

**GENERALIZAÇÃO:** Processo indutivo em que os resultados obtidos em um determinado número de indivíduos denominados de amostra são válidos também para toda a população da qual foram selecionados esses elementos.

**GRAUS DE LIBERDADE:** O número de comparações independentes que podem ser feitas entre os elementos de uma amostra. O termo é análogo ao número de graus de liberdade para um objeto em um sistema dinâmico, que é o número de coordenadas independentes requeridas para determinar o movimento do objeto.

**GL:** Em estatística é o número  $n$  de observações em que se baseia o cálculo do desvio padrão  $s$  quando se conhece a média verdadeira  $m$ .  $n$  dá, pois, uma indicação sobre a precisão da estimativa  $s$  obtida, e constitui o seu número de graus de liberdade (GL).

**GL:** Graus de liberdade, correção de Bessel ou liberdade de escolha.

**GLC:** Gerador linear congruencial.

**GLM:** Modelos lineares generalizados.

**GLS:** Quadrados mínimos generalizados.

**GOODNESS OF FIT:** Teste estatístico de modelo de distribuição de probabilidades, no qual as proporções observadas se ajustam às proporções esperadas, deduzidas matematicamente ou estabelecidas de acordo com alguma teoria. É também denominado de teste de aderência.

**GRADE:** Malha retangular em um plano formada por dois conjuntos de linhas ortogonais. Cada linha de cada conjunto é colocada a intervalos constantes da linha adjacente. Utilizada em algumas formas de amostragem de áreas. Em inglês Grid.

**GRÁFICO:** Qualquer representação geométrica de um fenômeno quantitativo. Em inglês graph.

**GRÁFICO:** Uma apresentação figurativa de variação de uma quantidade com respeito à outra, ou uma comparação de duas ou mais quantidades. Os tipos mais comuns são: i) gráfico de linha, em que a variável independente fica no eixo horizontal e a dependente, no vertical; ii) gráfico de barras vertical, útil para demonstrações de percentuais; iii) gráfico de barra horizontal às vezes preferido ao de barra vertical.

**GRÁFICO:** É a representação gráfica das escalas quantitativas recolhidas durante o trabalho de pesquisa.

**GRÁFICO BOX-PLOT (MEDIANA E QUARTIS):** Mostra a mediana, o primeiro e o terceiro quartis e, ainda, o maior e o menor escores da amostra.

**GRÁFICO BOX-PLOT (MÉDIA E DESVIO):** Mostra a média, o desvio ou erro padrão, o maior e o menor escores da amostra.

**GRÁFICO CAULE-E-FOLHA (STEM AND LEAF):** Organiza os escores em ordem crescente e separa cada valor em dois componentes: o caule, constituído por todos os valores de cada variável, exceto o último à direita, o qual representa a folha.

**GRÁFICO CIRCULAR:** É representado por um círculo que está dividido em sectores cujas amplitudes são proporcionais à frequência que lhe corresponde.



**GRÁFICO CIRCULAR (PIE CHART):** Representação gráfica da frequência relativa ou percentagem de cada categoria de uma variável. Este gráfico é utilizado para variáveis nominais e ordinais. É uma opção ao gráfico de barras quando se pretende dar ênfase à comparação das percentagens de cada categoria. Suas principais características são: a área do gráfico equivale à totalidade de casos (100 %) e cada fatia representa a percentagem de cada categoria.

**GRÁFICO DE ÁREA:** É a representação que consiste em construir figuras cujas áreas sejam proporcionais às magnitudes a serem representadas. Prestam-se à representação gráfica de valores que crescem muito rapidamente.

**GRÁFICO DE ÁREA:** Compara as áreas abrangidas pelos escores de duas ou mais amostras.

**GRÁFICO DE BARRAS:** É constituído por barras, horizontal ou vertical, de comprimento proporcional à frequência.

**GRÁFICO DE BARRAS:** Representação dos escores de determinados fenômenos por meios de barras de comprimentos proporcionais às respectivas frequências. Quando as barras são verticais, chama-se gráficos de colunas. Em inglês bar chart.

**GRÁFICO DE BARRAS:** Representação gráfica da distribuição de frequências de variáveis categóricas nominais e ordinais. Suas principais características são: todas as barras devem ter a mesma largura e devem existir espaços entre as barras.

**GRÁFICO DE CAIXAS (BOXPLOT):** Método para representar a distribuição de uma variável. Uma caixa representa a maior parte da distribuição, e as extensões chamadas de whiskers atingem os pontos extremos da distribuição. Muito útil para fazer comparações de uma ou mais variáveis em grupos.

**GRÁFICO DE COLUNAS:** Compara os escores de cada categoria em retângulos verticais ou horizontais, com a mesma largura e cuja altura é proporcional às grandezas que representam. Podem ser classificadas em: i) simples; ii) justapostas; e iii) superpostas.

**GRÁFICO DE COMPOSIÇÃO EM SETORES:** Destinado a representar a composição, centesimal ou não, de um dado todo, consiste numa circunferência de raio arbitrário, toda como representado esse todo, em que o círculo é dividido em setores, tomando-se arcos proporcionais às magnitudes das diversas parcelas do total em causa; também se diz setorgrama.

**GRÁFICO DE CONTROLE:** Um dentre diversos tipos de gráficos que ilustram alguma característica de um processo a fim de determinar se há estabilidade estatística.

**GRÁFICO DE DISPERSÃO:** Gráfico de duas variáveis, cada uma representada em um eixo, que mostra a dispersão e associação existente entre elas. Em inglês Scatterplot.

**GRÁFICO DE DISPERSÃO:** Utilizado para duas variáveis, X e Y, retiradas da mesma unidade, como ocorre nas análises de correlação e regressão.

**GRÁFICO DE DISTRIBUIÇÃO CONJUNTA:** Traçado de duas variáveis contínuas que pode ser utilizada para visualizar a relação entre ambas caso ela exista e determinar a direção positiva ou negativa e a linearidade dessa relação.

**GRÁFICOS ESTATÍSTICOS:** São formas de figuras usadas na apresentação dos dados estatísticos, cujo objetivo é o de produzir, no investigador ou no público em geral, uma impressão mais rápida e viva do fenômeno em estudo. A representação gráfica de um fenômeno deve obedecer a certos requisitos fundamentais para ser realmente útil: i) Simplicidade: O gráfico deve ser destituído de detalhes de importância secundária, assim como de traços desnecessários que possam levar o observador a uma análise com erros. ii) Clareza: O gráfico deve possibilitar uma correta interpretação dos valores representativos do fenômeno em estudo e iii) Veracidade: O gráfico deve expressar a verdade sobre o fenômeno em estudo. Por exemplo, gráficos como, o histograma e o diagrama de Pareto, o boxplot e o dotplot.

**GRÁFICO DE LINHA:** É todo gráfico em que o elemento básico de construção é a linha. Usa-se para distinguir dos gráficos de barras, faixas, setores, dentre outros. e inclui os diagramas cartesianos e polares.

**GRÁFICO DE LINHA (CURVA):** Utilizado, sobretudo, nas séries temporais, considerando-se a disposição dos dados nos eixos das abscissas (X) e das ordenadas (Y).

**GRÁFICO DE MÁXIMO E MÍNIMO:** Gráfico e linhas verticais permitindo comparar a amplitude e os valores máximo e mínimo de cada amostra.

**GRÁFICO DE np:** Carta de controle em que os números de defeitos são marcados de forma que se possa monitorar um processo.

**GRÁFICO DE p:** Carta de controle usada para monitorar a proporção  $p$ .

**GRÁFICO DE PARETO:** Gráfico em barras para dados qualitativos em que as barras são ordenadas de acordo com as frequências.

**GRÁFICO DE PARETO:** O gráfico de Pareto é um histograma onde se pode visualizar, com facilidade, a importância de cada classe de uma variável expressa no eixo X.

**GRÁFICO DE PROBABILIDADE NORMAL:** Comparação gráfica da forma da distribuição com a distribuição normal. No gráfico de probabilidade normal, a distribuição normal é representada por uma linha reta inclinada em 45 graus. A distribuição real é colocada no gráfico em contraste com essa reta, de modo que quaisquer diferenças são mostradas como desvios da linha reta, tornando a identificação de diferenças bastante visível e interpretável.

**GRÁFICO DE PROBABILIDADE NORMAL:** Comparação gráfica da forma da distribuição da amostra em relação à distribuição normal. No gráfico, a distribuição normal é representada por uma reta com inclinação de 45°. A verdadeira distribuição é representada em contraste com essa reta, de modo que quaisquer diferenças são mostradas como desvios da reta, tornando a identificação de diferenças bastante simples.

**GRÁFICO DE R:** Carta de controle baseada em intervalos amostrais, usada para monitorar variação em um processo.

**GRÁFICO DE REGRESSÃO PARCIAL:** Representação gráfica da relação entre a variável dependente e uma única variável independente. O diagrama de dispersão de pontos representa a correlação parcial entre as duas variáveis, com os efeitos de outras variáveis independentes mantidas constantes. Essa representação é particularmente útil na avaliação da forma da relação linear versus não linear e na identificação de observações influentes. Ver coeficiente de correlação parcial.

**GRÁFICO DE REPETIÇÕES:** Gráfico seqüencial de valores individuais de dados ao longo do tempo, onde um eixo em geral o vertical é usado para os valores dos dados, e o outro eixo em geral o horizontal é usado para sequencia do tempo.

**GRÁFICO DE s:** Carta de controle, baseada em desvios-padrão amostrais, usada para monitorar a variação em um processo.

**GRÁFICO DE SETOR:** Representa parcelas, em um círculo, de variáveis categóricas, comparando cada valor com o todo, dividindo-se a circunferência em setores, cada um representando a respectiva categoria. É recomendável que o número de setores não deve ultrapassar sete (7) variáveis.

**GRÁFICO DE VOLUMES:** Processo de representação gráfica que consiste em construir figuras planas que são a perspectiva de sólidos cujos volumes são proporcionais às magnitudes a serem representadas.

**GRÁFICO DE  $\bar{X}$ :** Gráfico de controle usado para monitorar a média de um processo.

**GRÁFICOS EM SETORES:** Gráfico em que o círculo representa o total de observações, e um setor do mesmo indica a fração de cada categoria.

**GRÁFICO MÁXIMO-MÍNIMO:** Em cada ponto  $x_i$  da abscissa marcam-se duas ordenadas  $F_m(x_{i-1})$  e  $F_m(x_{i+1})$ , valores no ponto anterior, de uma função  $f(x)$ , do tempo, tal como preço, temperatura, dentre outros. Traça-se uma poligonal dos máximos e outra dos mínimos, ou reúnem-se aqueles dois pontos de ordenada por um bastonete

**GRÁFICO NULO:** Gráfico de resíduos versus valores previstos que exhibe um padrão aleatório. Um gráfico nulo é indicativo da ausência de violações identificáveis das suposições inerentes à análise de regressão.

**GRÁFICO POLAR:** Representa períodos cíclicos tais como meses, semanas, horas, graus, dentre outros, com representação circular, indicado nas séries temporais.

**GRÁFICOS DE ANÁLISE:** São gráficos que se prestam melhor ao trabalho estatístico, fornecendo elementos úteis à fase de análise dos dados, sem deixar de ser também informativos. Os gráficos de análise freqüentemente vêm acompanhados de uma tabela estatística. Inclui-se, muitas vezes um texto explicativo, chamando a atenção do leitor para os pontos principais revelados pelo gráfico.

**GRÁFICOS DE INFORMAÇÃO:** São gráficos destinados principalmente ao público em geral, objetivando proporcionar uma visualização rápida e clara. São gráficos tipicamente expositivos, dispensando comentários explicativos adicionais. As legendas podem ser omitidas, desde que as informações desejadas estejam presentes.

**GRÁFICOS DE LIBERDADE:** Número de valores que podem variar após terem sido impostas certas restrições a todos os valores.

**GRÁFICO DE PROBABILIDADE NORMAL:** É o gráfico que é construído considerando-se como abscissa a função de distribuição acumulada empírica e como ordenada a função de distribuição acumulada da normas correspondente.

**GRÁFICO DE SETORES:** Representação das freqüências de um fenômeno pelas áreas de setores de um círculo. Os setores correspondem a certa percentagem. Em inglês pie chart.

**GRÁFICOS EM BARRAS COMPOSTAS, GRÁFICOS EM COLUNAS SUPERPOSTAS:** Eles diferem dos gráficos em barras ou colunas convencionais apenas pelo fato de apresentar cada barra ou coluna segmentada em partes componentes. Servem para representar comparativamente dois ou mais atributos.

**GRÁFICOS EM BARRAS HORIZONTAIS, GRÁFICOS EM BARRAS VERTICAIS (COLUNAS):** Quando as legendas não são breves usam-se de preferência os gráficos em barras horizontais. Nesses gráficos os retângulos têm a mesma base e as alturas são proporcionais aos respectivos dados. A ordem a ser observada é a cronológica, se a série for histórica, e a decrescente, se for geográfica ou categórica.

**GRÁFICOS EM CAIXAS:** Representação gráfica de um conjunto de dados.

**GRÁFICOS EM LINHAS OU LINEARES:** São freqüentemente usados para representação de séries cronológicas com um grande número de períodos de tempo. As linhas são mais eficientes do que as colunas, quando existem intensas flutuações nas séries ou quando há necessidade de se representarem várias séries em um mesmo gráfico.

**GRÁFICOS EM PONTOS:** Gráfico em que cada valor é marcado como um ponto ao longo de uma escala de valores.

**GRÁFICOS EM SETORES:** Este gráfico é construído com base em um círculo, e é empregado sempre que desejamos ressaltar a participação do dado no total. O total é representado pelo círculo, que fica dividido em tantos setores quantas são as partes. Os setores são tais que suas áreas são respectivamente proporcionais aos dados da série. O gráfico em setores só deve ser empregado quando há, no máximo, sete dados. Observação: as séries temporais geralmente não são representadas por este tipo de gráfico.

**GRÁFICOS EM SETORES:** Representação gráfica de dados em forma de um círculo dividido em setores.

**GRÁFICOS ESTATÍSTICOS:** São representações visuais dos dados estatísticos que devem corresponder, mas nunca substituir as tabelas estatísticas.

**GRÁFICOS RAMO-E-FOLHAS:** Método de disposição dos dados em forma de um círculo dividido em setores.

**GRÁFICOS RESIDUAIS:** Representações gráficas dos resíduos que podem ser usadas para determinar se as suposições feitas sobre o modelo de regressão parecem ser válidas.

**GRÁFICO DE BARRAS:** Representação gráfica que proporciona uma comparação visual das freqüências das variáveis. O comprimento das barras ou colunas representa a freqüência das variáveis. As barras podem ser dispostas vertical ou horizontalmente.

**GRÁFICO DE SETORES:** Representação gráfica circular na qual cada fatia do círculo simboliza a freqüência relativa de uma categoria de elementos componentes de uma variável. Este tipo de gráfico é muito útil para evidenciar a importância relativa dos componentes que estão sendo comparados.

**GARANTIA DE QUALIDADE E CONTROLE DE QUALIDADE (GQ/CQ) (QUALITY ASSURANCE AND QUALITY CONTROL-QA/QC):** O processo pelo qual os dados, ou qualquer atividade são monitorados para determinar que eles têm acurácia e precisão.

**GRÁFICO DE CAULE-E-FOLHA (STEM AND LEAF PLOT):** Um histograma rico em informação em que o valor real de cada dado é ilustrado.

**GRÁFICO DE DISPERSÃO (SCATTERPLOT):** Um gráfico de duas dimensões usado para ilustrar dados bivariados. A variável preditora, ou independente, geralmente é colocada no eixo x e a variável resposta, ou dependente, no eixo y. Cada ponto representa as medições dessas duas variáveis em cada observação.

**GRÁFICO DE RESÍDUOS (RESIDUAL PLOT):** Um gráfico de diagnóstico usado para determinar se as premissas da regressão ou análise de variância (ANOVA) foram cumpridas. Mais comumente, os resíduos são plotados contra os valores preditos.

**GRÁFICO MOSAICO (MOSAIC PLOT):** Um gráfico destinado à ilustração de tabelas de contingência. O tamanho dos retângulos do mosaico é proporcional à frequência relativa de cada observação.

**GRÃO (GRAIN):** O tamanho espacial ou a extensão temporal da menor unidade amostral em um estudo experimental. Comparar com extensão.

**GRAUS DE LIBERDADE (GL) (DEGREES OF FREEDOM-DF):** Refere-se ao número ou a quantas observações independentes tem-se para uma certa variável que podem ser utilizadas para estimar um parâmetro estatístico.

**GRAU DE LIBERDADE:** Número de maneiras em que os dados sejam livres para variar; número de observações menos o número de restrições impostas aos dados. O grau de liberdade (gl) informa-nos quantos valores foram utilizados para o cálculo de determinada estatística. É um valor necessário para o cálculo do qui quadrado e dos valores t, por exemplo.

**GRUPO DE CONTROLE:** i) Numa pesquisa experimental, refere-se ao grupo de sujeitos que não se submeterá às condições experimentais, sendo utilizado apenas como parâmetro de comparação. Por exemplo: numa pesquisa sobre a eficácia de um novo medicamento, os sujeitos podem ser divididos em dois grupos: experimental e controle. O primeiro será tratado com o medicamento em teste, enquanto o segundo receberá um tratamento inócuo ou placebo. Ao final do experimento, comparam-se os efeitos do procedimento entre o grupo experimental e o grupo de controle; ii) Qualquer unidade experimental utilizada para efeito de comparação, por exemplo, o olho esquerdo em relação ao olho direito, um mesmo sujeito em dois momentos diferentes de uma pesquisa, dentre outros.

**GRUPO CONTROLE POSITIVO:** Tipo de grupo controle numa pesquisa experimental que, por não poder se submeter a um placebo, recebe um tratamento convencional. Esse tipo de grupo é muito comum na área médica, por exemplo, na qual seria eticamente questionável submeter os sujeitos de um estudo a um tratamento sem efeito, testemunha, controle ou placebo.

**GRUPO DE COMPARAÇÃO:** Termo que pode ser utilizado como sinônimo de grupo controle. Frequentemente, porém, refere-se ao grupo de sujeitos num dado experimento que é submetido a procedimentos alternativos ao invés de não ser submetido a nenhum procedimento, como ocorre no caso de grupo controle.

**GRUPO (GROUP):** Um conjunto que possui quatro propriedades. Primeiro, o conjunto é fechado sob alguma operação matemática  $\oplus$  (como adição ou multiplicação). Segundo, a operação é associativa, de modo que quaisquer três elementos a, b, c do conjunto  $a \oplus (b \oplus c) = (a \oplus b) \oplus c$  (onde  $\oplus$  indica qualquer operação matemática). Terceiro, o conjunto tem um elemento identidade I tal que para qualquer elemento a do conjunto  $a \oplus I = a$ . Quarto, o conjunto tem inversos, de maneira que para cada elemento a do conjunto existe outro elemento b para o qual  $ab = ba = I$ . Ver também fechamento.

**GRUPO EXPERIMENTAL:** Numa pesquisa experimental, refere-se ao conjunto de sujeitos que será submetido às condições experimentais. Os dados obtidos pela manipulação das variáveis deste grupo serão comparados àqueles obtidos pelo grupo controle.

**GRAU DE CONFIANÇA:** Probabilidade de um parâmetro populacional estar contido em determinado intervalo de confiança; também chamado nível de confiança.

**GRAU DE EFICIÊNCIA:** De um plano, ou delineamento, experimental é um número, calculado segundo critérios vários, que tem por fim medir a precisão dos estimadores baseados nesse plano.

**GRAU DE LIBERDADE:** Este termo é usado de formas diferentes em estatística. Os graus de liberdade de um conjunto de observações é o número de valores que podem ser arbitrariamente alocados dentro da especificação do sistema. De um ponto de vista diferente, a expressão graus de liberdade também é usada para denotar o número de comparações independentes que podem ser feitas entre os membros de uma amostra. Em inglês Degree of freedom (D.F.).

**GRAUS DE LIBERDADE:** É um conceito ligado ao número de dados disponíveis (livres) para o cálculo da estatística. Por exemplo, ao estimarmos a média populacional, com a média amostral perdemos um grau de liberdade, assim a estatística t student terá n-1 graus de liberdade. No caso da tabela de análise de variância (ANOVA), os graus de liberdade do grupo será igual ao número de grupos menos 1, os graus de liberdade total será igual a n-1 e os graus de liberdade do resíduo, a diferença entre esses dois.

**GRAUS DE LIBERDADE:** É um conceito ligado ao número de dados disponíveis ou livres para o cálculo da estatística.

**GRAUS DE LIBERDADE:** É a quantidade de informações ou variáveis livres que serão utilizadas para o cálculo de uma estatística que é uma fórmula. O número de valores independentes que serão utilizados na estimativa de um parâmetro. Em geral, o número de graus de liberdade de uma estimativa é igual ao número de valores utilizados no seu cálculo menos o número de parâmetros estimados no cálculo intermediário para a sua obtenção. Assim para calcular a média de uma amostra de tamanho n, são necessários as n observações fazendo com que esta estatística tenha n graus de liberdade. Já a estimativa da variância por meio de uma amostra de tamanho n terá n-1 graus de liberdade, pois para a obtenção da variância amostral é necessário antes o cálculo da média amostral. Em inglês degree of freedom.

**GRAUS DE LIBERDADE:** São parâmetros indexadores estatísticos correspondentes ao número de observações independentes, como se observam nas distribuições t de Student, F da análise de variância (ANOVA), Qui quadrado e r da correlação linear de Pearson.

**GRAUS DE LIBERDADE:** Número de observações em um conjunto de dados que são livres para variar, uma vez que os parâmetros do conjunto de dados, por exemplo, a média tenham sido estabelecidos.

**GRAUS DE LIBERDADE:** O número de maneiras em que os dados sejam livres para variar; o número de observações menos o número de restrições impostas aos dados.

**GRAUS DE LIBERDADE:** Uma quantidade requerida para avaliar a significação de uma relação estatística. Num conjunto de valores, é o número de valores que precisamos conhecer para calcular o restante. Sabendo que duas pessoas juntas têm dez reais, ao saber quanto uma tem, sabemos automaticamente



quanto a outra tem. Neste caso, dizemos que há um grau de liberdade, porque só um valor pode variar. Se neste exemplo houvesse três pessoas, teríamos dois graus de liberdade.

**GRAUS DE LIBERDADE:** O número de observações em um problema estatístico que não estão restritas, ou seja, que estejam livres para variar.

**GRAUS DE LIBERDADE:** Número de células que estão livres para variar. Uma vez conhecidos os valores dessas células e também os totais de linhas e colunas, os valores de todas as outras células são fixados.

**GRAUS DE LIBERDADE:** Conceito associado a uma distribuição de probabilidade que significa o número de variáveis livres ou informações amostrais que foram necessárias para caracterizar a distribuição. Em inglês degrees of freedom.

**GRAUS DE LIBERDADE:** O número de bits de informação disponível para estimar a distribuição amostral dos dados depois que todos os parâmetros do modelo tenham sido estimados. Em termos práticos, graus de liberdade são o número de correlações ou covariâncias não redundantes na matriz de entrada menos o número de coeficientes estimados. O pesquisador tenta maximizar os graus de liberdade disponíveis enquanto ainda obtém o modelo de melhor ajuste. Cada coeficiente estimado consome completamente um grau de liberdade. Um modelo jamais pode estimar mais coeficientes do que o número de correlações ou covariâncias não redundantes, o que significa que zero é o limite inferior para os graus de liberdade de qualquer modelo.

**GRAUS DE LIBERDADE (DF, DEGREES OF FREEDOM):** Valor calculado a partir do número total de observações menos o número de parâmetros estimados. Essas estimativas de parâmetros são restrições sobre os dados porque, uma vez calculadas, elas definem a população da qual se supõe que os dados tenham sido extraídos. Por exemplo, ao estimar um modelo de regressão com uma única variável independente, estimamos dois parâmetros, o intercepto ( $b_0$ ) e um coeficiente de regressão para a variável independente ( $b_1$ ). Ao estimar o erro aleatório, definido como a soma dos erros de previsão que são os valores dependentes reais menos os previstos para todos os casos, encontraríamos  $(n-2)$  graus de liberdade. Os graus de liberdade nos dão uma medida de quão restritos estão os dados para alcançar um certo nível de previsão. Se o número de graus de liberdade é pequeno, a previsão resultante pode ser menos generalizável porque todas, exceto algumas observações, foram incorporadas na previsão. Do mesmo modo, um valor alto no número de graus de liberdade indica que a previsão é bastante robusta, no sentido de ser representativa de toda a amostra de respondentes.

**GRAUS DE LIBERDADE DO DENOMINADOR:** número de graus de liberdade correspondente ao numerador da estatística de teste F.

**GRAUS DE LIBERDADE DO NUMERADOR:** Número de graus de liberdade correspondente ao numerador da estatística de teste F.

**GRAU OU COEFICIENTE DE CONFIANÇA:** Probabilidade de um intervalo de confiança conter o valor do parâmetro de uma população representado por  $1 - \alpha$ .

**GRID ESPECÍFICO:** Similar ao grid geral, utilizado apenas para determinados testes estatístico.

**GRID GERAL:** É uma espécie de planilha eletrônica onde os dados, da maior parte dos testes estatísticos, devem ser introduzidos. Sinônimo: Matriz de Dados.

**GRUPO:** Um grupo é um conjunto combinado com uma operação que deve satisfazer certas propriedades. Assim se  $G$  é o grupo e  $*$  a operação, e  $a$  e  $b$  são dois elementos pertencentes a  $G$ , os seguintes resultados são válidos: i)  $a*b = c$ , onde  $c$  também pertence a  $G$  (a operação é fechada em  $G$ ); ii)  $(a*b)*c = a*(b*c)$  (a operação é associativa); iii) Existe  $e$  pertencente a  $G$ , tal que:  $a*e=a$  (existência de um elemento neutro); iv) Para cada  $a$  de  $G$ , existe  $a'$  tal que:  $a*a' = e$  (cada elemento do grupo apresenta um oposto). Os grupos são aplicados em várias áreas da própria matemática e na física teórica. Alguns exemplos de grupos incluem: o das simetrias, o das matrizes e o das permutações. Em inglês group.

**GRUPO:** Conjunto de elementos, indivíduos ou observações, todos os quais possuem uma, ou mais, características em comum. Em inglês Group.

**GRUPO ABELIANO:** Um grupo abeliano é um grupo em que a operação do grupo é comutativa, isto é, satisfaz a relação  $a*b = b*a$ . Em inglês abelian group.

**GRUPO-CONTROLE:** O mesmo que grupo-testemunha; utilizado para comparação de resultados; pode ser interno e externo. Ver controle.

**GRUPO CONTROLE:** Em um ensaio clínico, grupo de pacientes designados para o tratamento controle. Serve como base de comparação para o grupo que recebe o tratamento em teste.

**GRUPO DE CONTROLE:** Grupo de indivíduos em um experimento que não recebem um tratamento particular.

**GRUPO DE CONTROLE:** Em experimentação, um grupo de participantes a que nenhum estímulo experimental é dado e que, exceto por isto, se assemelha ao grupo experimental em todos os demais aspectos. A comparação do grupo de controle com o grupo experimental no final do experimento mostra o efeito do estímulo experimental.

**GRUPO DE COMPARAÇÃO:** A categoria de uma variável não-métrica que sempre recebe zeros quando a codificação indicadora é usada, ou sempre menos 1 (-1) quando a codificação dos efeitos é empregada na criação de variáveis dicotômicas.

**GRUPO DE CONTROLE:** Grupo de pessoas selecionado para comparação com o grupo de estudo. Idealmente, o grupo de controle é idêntico ao de estudo, exceto por não possuir a característica estudada ou não ter sido exposto ao tratamento investigado. Sinônimo: grupo de referência.

**GRUPO DE ENTROPIA:** Grupo de objetos independentes de qualquer agrupamento, ou seja, eles não se ajustam a agrupamento nenhum que podem ser considerados atípicos e possivelmente eliminados da análise de agrupamentos.

**GRUPO DE ESTUDO:** Em um estudo de coorte ou em um ensaio clínico controlado, esse é o grupo de indivíduos que possui as características ou está exposto aos fatores estudados. Nos estudos de casos e controles ou nos transversais, corresponde ao grupo de indivíduos que apresentam a doença investigada.

**GRUPO FOCAL:** Termo utilizado para denominar uma técnica de pesquisa qualitativa muito utilizada em pesquisa de marketing.

**GRUPO DE FOCO:** Um grupo de oito a 12 participantes liderados por um moderador em discussões profundas sobre determinado tópico ou conceito.

**GRUPO DE FOCO:** Uma discussão semi-estruturada entre indivíduos que supostamente têm algum conhecimento das questões associadas ao estudo ou interesse por elas.

**GRUPO DE TRATAMENTO:** Grupo de indivíduos a quem é ministrado determinado tratamento em um experimento.

**GRUPO EXPERIMENTAL:** Em um ensaio clínico, é o grupo de pacientes designados ao tratamento em teste. É contrastado com o grupo controle para chegar a uma conclusão sobre um fator, condição, ou tratamento.

**GRUPOS FOCAIS:** Pequenos grupos reunidos para discussões orientadas de um determinado assunto, como um novo produto de consumo, um tema político dentre outros. Podem fornecer um exame profundo de atitudes e preferências, mas em geral não têm uma representatividade rigorosa.

**GRUPOS DE NOTÍCIAS:** Sites da Internet dedicados a um tópico específico em que as pessoas podem ler e colocar mensagens.

**GUIA DE DISCUSSÃO:** Esboço escrito de tópicos a serem cobertos durante uma discussão de grupos de enfoque.

**GRÁFICO:** Qualquer representação geométrica de um fenômeno quantitativo.

**GRÁFICO EM BARRAS:** Representação dos valores de determinado fenômeno por meio de barras de comprimentos proporcionais às respectivas frequências. Quando as barras são verticais, chama-se gráfico de colunas.

**GRÁFICO POLAR:** Representação de uma série por meio de um polígono. Geralmente presta-se para apresentação de séries temporais. Para construí-lo, divide-se uma circunferência em tantos arcos iguais quantos forem os dados a representar e traçam-se raios pelas divisas.

**GRÁFICO EM SETORES:** Representação das frequências de um fenômeno pelas áreas de setores de um círculo. Os setores correspondem a certa porcentagem.

**GRAUS DE LIBERDADE:** Tecnicamente, referem-se à liberdade de variação num conjunto de escores. Se tivermos uma amostra de 6 escores, então 5 deles têm liberdade de variar, enquanto 1, e apenas 1, é fixo no que diz respeito ao valor. Portanto, numa amostra simples de tamanho 6, por exemplo, teremos 5 graus de liberdade.

**GOODNESS-OF-FIT:** Tipo de teste não paramétrico que avalia a concordância entre um conjunto de valores observados e outro conjunto de valores derivados de alguma hipótese ou teoria específica.

**GRÁFICO DE BARRAS:** Um diagrama em que todas as categorias da variável são representadas pelo tamanho de uma barra que corresponde ao número ou porcentagem de indivíduos que pertencem àquela categoria.

**GRÁFICO NORMAL:** Um diagrama graduado de tal maneira que os dados com distribuição Normal são exibidos como uma linha reta. Os desvios em relação à linha reta indicam que os dados não apresentam distribuição Normal.

**GRÁFICO PIZZA:** Um diagrama utilizado para mostrar a distribuição de um conjunto de dados categóricos. É um círculo dividido em segmentos, cada um deles retratando uma categoria diferente da variável qualitativa. A área de um segmento é proporcional à porcentagem de indivíduos naquela categoria.

**GRAUS DE LIBERDADE (GL):** O número de observações independentes que contribuem para o valor da estatística, ou seja, o número de observações disponíveis para avaliar a estatística, menos o número de restrições nessas observações.

**GRUPO:** Uma coleção de animais ou unidades experimentais.

**GRUPO CONTROLE:** Um grupo de indivíduos de um estudo experimental que recebem um tratamento padrão ou um tratamento não ativo; é utilizado como base de comparação para o grupo de indivíduos que receberam tratamento(s) ativo(s).

**GRUPO MODAL OU CLASSE MODAL:** O grupo ou classe de uma distribuição de frequência que contém mais observações do que qualquer outra classe.







**HEPERENDEMIAS:** São conjecturas com as quais se procuram explicar, por tentativa, fenômenos ocorridos ou ocorrentes. São repostas dadas e problemas postos pela ciência ou pelo senso comum. Deverão ser testadas para comprovação ou refutação.

**HERDABILIDADE (HERITABILIDADE):** Proporção das diferenças observadas ou medidas nos animais que é transmitida à descendência; proporção da variância fenotípica de um dado caráter atribuída aos efeitos médios dos genes. A herdabilidade varia de zero a um. Se uma característica tem alta herdabilidade, mais acuradamente o desempenho fenotípico individual prediz o valor genético do indivíduo e mais rápida deve ser a resposta à seleção para o dado caráter.

**HERDABILIDADE NO SENTIDO AMPLO ( $h_g^2$ ):** Quadrado da correlação entre o valor genotípico e o valor fenotípico, ou regressão do valor genotípico sobre o valor fenotípico, ou coeficiente de determinação do valor genotípico pelo fenótipo.

**HERDABILIDADE NO SENTIDO AMPLO:** Proporção da variação fenotípica observada atribuída à causas genéticas.

**HERDABILIDADE NO SENTIDO RESTRITO ( $h_a^2$ ):** É o quadrado da correlação entre o valor genético aditivo e o valor fenotípico, ou regressão do valor genético aditivo sobre o valor fenotípico, ou coeficiente de determinação do valor genético aditivo pelo fenótipo.

**HERDABILIDADE NO SENTIDO RESTRITO:** Proporção da variação fenotípica ou variação observada atribuída à variância genética aditiva.

**HEREDITARIEDADE:** Fenômeno de continuidade biológica pelo qual as formas vivas se repetem nas gerações que se sucedem.

**HEREDITARIEDADE:** Transmissão de características genéticas ou físicas dos pais para seus descendentes.

**HEREDITARIEDADE:** Grau em que a aquisição de uma doença é influenciada pela disposição genética.

**HETEROSE (VIGOR HÍBRIDO):** Termo proposto por Shull (1914), para descrever a superioridade de desempenho dos F1's cruzados sobre o desempenho médio dos parentais, para uma dada característica.  $H = \text{Média dos cruzados} - \frac{1}{2} (\text{média do parental 1} + \text{média do parental 2})$ .

**HETEROGENEIDADE DAS AMOSTRAS:** É um modelo probabilístico que indica se as amostras investigadas não são oriundas da mesma população, sendo utilizado nos teste G, do Qui-Quadrado, da correlação linear e da regressão linear.

**HETEROGENEIDADE DAS VARIÂNCIAS:** É um modelo probabilístico que revela se as variâncias dos subconjuntos testados são desiguais, não sendo oriundas da mesma população, e é empregado no F teste da análise de variância (ANOVA), no t teste de Student para duas amostras independentes e na regressão linear. Sinônimo: Heterocedasticidade.

**HETEROCEDASTICIDADE:** Vide heterogeneidade das variâncias.

**HETEROSCEDASTICIDADE:** Ver homoscedasticidade.

**HETEROCEDASTICIDADE:** Situação que ocorre na análise de variância onde não se verifica a hipótese básica de homogeneidade de variância, isto é, não existe homoscedasticidade, ou seja, a variância não é a mesma nos diferentes tratamentos.

**HETEROSCEDASTICIDADE:** Situação que ocorre na análise da variância, quando os termos erro não têm todos a mesma variância.

**HETEROCEDASTICIDADE IRREGULAR:** É aquela que ocorre quando certos tratamentos apresentam maior variabilidade que outros, como por exemplo, nos experimentos com inseticidas, nos quais é considerado um grupo de parcelas não tratadas ou testemunha. De um modo geral, verificamos que o número de insetos nas parcelas tratadas é menor e mais homogêneo do que o da testemunha, que apresenta uma maior variabilidade.

**HETEROCEDASTICIDADE REGULAR:** É aquela que ocorre devido à falta de normalidade dos dados experimentais, existindo, frequentemente, uma certa relação entre a média e a variância dos diversos tratamentos testados. Se a distribuição de probabilidade dos dados for conhecida, a relação  $\frac{\hat{m}}{s^2}$  dos tra-

tamentos também o será e, os dados poderão ser transformados de forma que passem a ter uma distribuição aproximadamente normal e as médias e variâncias se tornem independentes, permitindo estruturar a análise de variância. Neste caso devemos buscar uma transformação tal que os dados passem a apresentar uma distribuição aproximadamente normal.

**HETEROCEDASTICIDADE:** Termo usado na análise de regressão em que se a hipótese do método dos mínimos quadrados de que a variância do termo erro é constante para todos os valores das variáveis independentes não se mantém, deparamo-nos com o problema da heterocedasticidade. Isto conduz a estimativas viesadas e ineficientes isto é, maior que a variância mínima dos erros padrões e, assim, a testes estatísticos e intervalo de confiança incorretos. Um teste de heterocedasticidade envolve ordenar os dados para pequenos e grandes valores da variável independente, X, e fazer duas regressões, uma para pequenos valores de X e uma para grandes valores de X, omitindo, digamos, um quinto das observações com valores médios. Então, testamos se a razão da soma dos quadrados dos erros (VR) da segunda para a primeira regressão é significativamente diferente de zero, usando a tabela da distribuição teórica de probabilidade F com  $\frac{(n-d-2k)}{2}$  graus de liberdade, onde n é o número total de observações, d é o número de obser-

vações omitidas e K é o número de parâmetros estimados.

**HETEROCEDASTICIDADE:** A heterocedasticidade é um conceito usado em análise de regressão que refere-se ao caso em que a variância do termo erro não é constante para todos os valores da variável independente. Isto é,  $E(X_i u_i) \neq 0$ , assim  $(E(u_i))^2 \neq$ . Isto viola a terceira hipótese do modelo de regressão pelo método dos mínimos quadrados (MMQ). Isto ocorre principalmente em dados de cross-section. Por exemplo, a variância do erro associada com as despesas das famílias de baixa renda é geralmente menor do que para famílias de alta renda, pois a maioria das despesas das famílias de baixa renda são para atender suas necessidades básicas, com pouca folga para liberdade de ação. A presença da heterocedasticidade pode ser testada pelo reagrupamento dos dados do menor para o maior valor da variável independente,,

e então fazendo duas regressões separadas, uma para os menores valores de e outra para os maiores valores de , omitindo por exemplo, um quinto das observações de valores médios. Então, testa-se a razão entre a soma dos quadrados dos erros da segunda regressão e a soma dos quadrados dos erros da primeira regressão isto é,  $\left(\frac{VR_2}{VR_1}\right)$ , para ver se é significativamente diferente de zero. Utiliza-se a distribuição de

probabilidade de F para este teste, com  $\frac{(n-d-2k)}{2}$  graus de liberdade, onde n é o número total de

observações, d é o número de observações omitidas, e o k é o número de parâmetros estimados. Este é o teste de Goldfeld-Quandt para heterocedasticidade, e é mais apropriado para amostras grandes, isto é, para  $n \geq 30$ . Se nenhuma das observações com valores médios são omitidas, o teste é ainda correto, mas ele teria um poder reduzido para detectar a heterocedasticidade. Ou ainda se o adotado como é frequente o caso que var, onde C é uma constante diferente de zero, pode-se corrigir a heterocedasticidade pela divisão isto é, ponderando de cada termo da regressão por , e então reestimando a regressão, usando variáveis transformadas. No caso de duas variáveis, tem-se que: . O termo erro transformado é agora homoscedástico: var. Observe que o intercepto original torna-se uma variável, enquanto o parâmetro declividade original, , é agora o novo intercepto. Porém, deve-se tomar cuidado para interpretar corretamente os resultados da regressão transformada ou ponderada. Como os erros são homoscedásticos, as estimativas de mínimos quadrados simples são, além de não viesadas e consistentes, também eficientes. No caso de uma regressão múltipla, cada termo da regressão é dividido isto é, ponderado pela variável independente, por exemplo, que se julga estar associada com o termo erro, se tem que: . Nessa equação, o intercepto original, , torna-se uma variável, enquanto torna-se o novo intercepto. Podemos visualmente determinar se é ou que é relacionado com , pela colocação de e contra os resíduos da regressão, .

**HETEROGENEIDADE DOS ERROS:** Ver heterocedastidade.

**HETEROZIGOTOS:** Genes de um par específico ou alelos são diferentes em um animal.

**HIERARQUIA:** Sistema de organização de objetos em uma ordem ascendente ou descendente, representada por uma série de agrupamentos ordenados tais como aqueles usados para classificação de plantas ou animais, como por exemplo, classes, ordens, famílias.

**HIPERPARÂMETRO:** Em uma análise bayesiana, hiperparâmetros são os parâmetros da distribuição a priori. Em inglês Hyperparameter.

**HIPGEOM (N, K, n):** Distribuição de probabilidade teórica discreta hipergeométrica.

**HIPÓTESE NULA (NULL HYPOTHESIS):** A hipótese mais simples possível. Na ecologia e nas ciências ambientais, a hipótese nula geralmente é que a variabilidade observada nos dados pode ser atribuída inteiramente à aleatoriedade ou a erros de medida. Comparar com hipótese alternativa. Ver também hipótese nula estatística.

**HIPÓTESE NULA ESTATÍSTICA (STATISTICAL NULL HYPOTHESIS):** A hipótese formal de que não existe relação matemática entre duas variáveis, ou a forma estatística da hipótese científica de que qualquer variabilidade observada nos dados pode ser atribuída inteiramente à aleatoriedade ou a erro de medidas. Ver também hipótese nula.

**HISTOGRAMA (HISTOGRAM):** Um gráfico de barras que representa a distribuição de frequências. A altura de cada barra é igual à frequência da observação identificada no eixo dos x's.

**HOMOCEDESTICO (HOMOCEDESTIC):** A propriedade de um conjunto de dados cuja variância residual de todos os tratamentos é igual. Comparar com heterocedástico.

**HIPÓTESE:** Suposição ou conjectura sobre a relação entre dois ou mais eventos.

**HIPÓTESE:** Uma hipótese é uma declaração de crença relativa a uma fenômeno, fato ou relacionamento entre várias variáveis. Uma hipótese deve sempre ser testada contra fatos antes de ser aceita ou rejeitada.

**HIPÓTESE:** É uma afirmação, mentalmente elaborada, sobre as possíveis causas ou natureza de um problema, fato observacional ou fenômeno que se deseja investigar. Basicamente consiste em supor conhecida a resposta ou solução para o problema. A hipótese pode ser vista como uma solução provisória para o problema. Se necessário, elaboram-se soluções provisórias complementares. Hipóteses são elaboradas a partir de conhecimentos próprios do pesquisador sobre o problema e de revisão, a mais ampla possível, do que já é sabido sobre o mesmo. Formalmente, a hipótese deve ser formulada de tal modo que sua verificação possa ser feita por observação direta, por meio de procedimentos experimentais ou que deduções nestes fundamentadas e que conduzam a predições verificáveis. Uma hipótese pode consistir de afirmações como: ... numa determinada situação um fenômeno se manifesta, ... possui tais características ou natureza, ... que se relaciona ou não com outro fenômeno. Essas afirmações devem ser experimentalmente verificadas o que equivale a uma confrontação com a realidade empírica. somente os resultados experimentais autorizam a aceitação ou rejeição de uma hipótese. É conveniente, lembrar agora a opinião do cientista Isaac Newton: ... não se inventam hipóteses. A hipótese exerce importante papel na definição da metodologia experimental a ser utilizada, composta pelos seguintes elementos: delineamentos experimentais, técnicas de observação e obtenção de dados, variáveis a serem observadas, dentre outros. A hipótese, portanto, tem por função orientar o pesquisador na direção da provável causa do fenômeno ou problema observado e orientar o pesquisador no estabelecimento da metodologia.

**HIPÓTESE:** É uma explicação provisória a um problema e ela deve ser submetida à verificação para sua comprovação ou não. No entanto, a comprovação direta de certas hipóteses é empiricamente inviável. Nestas situações as hipóteses se firmam por meio de comprovações indiretas, isto é, elas são verificadas pelas suas conseqüências. As hipóteses são as proposições desencadeadoras do processo científico. A investigação não deve produzir mais do que aquilo expresso na formulação antecipada pela hipótese. Deve ficar claro que as hipóteses são sentenças afirmativas ainda que as afirmações sejam prováveis e representam, de alguma forma, relações entre variáveis. portanto, a hipótese se refere a fatos e fenômenos na tentativa de explica-los e de verificar o interrelacionamento das variáveis envolvidas. As investigações executadas para comprovação de hipóteses se acham limitadas pelo próprio conteúdo da hipótese que é, assim, a delimitadora da área de observação e de experimentação. As hipóteses podem ser explicativas ou preditivas. São explicativas quando representam o resultado de gradativas generalizações de proposições existentes na teoria, sendo formuladas após a observação de um fato. São preditivas quando formuladas precedendo a observação empírica, com base na teoria existente e utilizando uma inferência de tipo dedutiva, ao contrário da hipótese explicativa que usa a inferência indutiva.

**HIPÓTESE:** Uma hipótese é um enunciado formal das relações esperadas entre pelo menos uma variável independente e uma variável dependente. Nas pesquisas exploratórias, as hipóteses podem se tornar questões de pesquisa. Estas questões pela sua especificidade, devem dar testemunho do trabalho conceitual efetuado pelo pesquisador e, pela sua clareza, permitir uma resposta interpretável.

**HIPÓTESE:** É a suposição de uma resposta para o problema formulado em relação ao tema. A hipótese pode ser confirmada ou negada.

**HIPÓTESE:** Afirmação, ou alegação, sobre uma propriedade de uma população.

**HIPÓTESE:** Expectativa sobre a natureza das coisas derivada de uma teoria. Afirmação de que algo deve ser observado no mundo real se a teoria for correta. Ver dedução.

**HIPÓTESE:** Afirmação conjectural sobre o relacionamento entre duas ou mais variáveis que podem ser testadas com dados empíricos. Pressuposição ou teoria que um pesquisador ou gerente faz sobre uma característica da população sob estudo.

**HIPÓTESE:** Uma hipótese é um enunciado formal das relações esperadas entre pelo menos uma variável independente e uma variável dependente. Nas pesquisas exploratórias, as hipóteses podem se tornar questões de pesquisa. Estas questões pela sua especificidade, devem dar testemunho do trabalho conceitual efetuado pelo pesquisador e, pela sua clareza, permitir uma resposta interpretável.

**HIPÓTESE:** É uma suposição fundamentada sobre o porquê de um fato, ou seja, uma suposição sobre causas. Refere-se a um conjunto estruturado de argumentos e explicações que possivelmente justificam dados e informações, porém ainda carente de confirmação. Pode ser didaticamente classificada em i) positiva, tendo como exemplo: há risco de contaminação por gripe aviária ao se ingerir carne crua de ave; ii) negativa: não há risco de contaminação por gripe aviária sem ingestão de carne crua de aves e; iii) condicional: na ausência do contato humano direto com aves, não há risco de contaminação por gripe aviária.

**HIPÓTESE, UTILIDADES:** As hipóteses podem ser úteis para: i) explicar novos relacionamentos; ii) estimular novas pesquisas; iii) como fontes metodológicas; iv) avaliar técnicas de pesquisa; v) levar a novos princípios.

**HIPÓTESE ALTERNATIVA:** Alternativa para a hipótese da nulidade, que postula haver diferença entre as populações ou grupos em comparação, com relação ao fator, à característica ou à condição de interesse. Veja hipótese de nulidade.

**HIPÓTESE ALTERNATIVA:** É a hipótese de que existe de fato diferença real ou verdadeira entre médias ou proporções de grupos que estão sendo comparados. Ou de que existe uma associação real entre duas variáveis. Comparar com Hipótese nula.

**HIPÓTESE ALTERNATIVA:** Representada por  $H_1$  ou  $H_a$  é a afirmativa contrária ou complementar à hipótese nula. Ela pode ter uma das seguintes três formas: i) a média da população não é igual a um valor especificado ( $H_1 : \mu \neq 5,0$ ), também chamada hipótese alternativa não direcional; ii) a média da população é menor que um valor especificado ( $H_1 : \mu < 5,0$ ); ou iii) a média da população é maior que um valor especificado ( $H_1 : \mu > 5,0$ ). Estas duas últimas formas são também chamadas hipóteses alternativas direcionais.

**HIPÓTESE ALTERNATIVA:** Afirmação equivalente à negação da hipótese nula; denotada por  $H_1$ .

**HIPÓTESE ALTERNATIVA:** É a hipótese que afirma que a hipótese nula é falsa.

**HIPÓTESE ALTERNATIVA:** Suposição de que, no resultado final da investigação, há diferenças significativas entre grupos; é alternativa em relação à hipótese nula.



**HIPÓTESE ALTERNATIVA ( $H_1$ ):** As hipóteses de uma pesquisa devem enunciar-se por propostas tão claras e específicas quanto possível via de regra, é o que você quer mostrar, por exemplo: as atitudes em relação a matemática interferem diretamente na formação das atitudes em relação à estatística .

**HIPÓTESE ALTERNATIVA:** É a hipótese que contraria a de nulidade, no sentido de afirmar que há diferença entre os grupos específicos objetos da pesquisa, como, por exemplo:  $(\mu_1 \neq \mu_2)$

**HIPÓTESE ALTERNATIVA:** No teste de hipóteses é a hipótese que se quer provar. É geralmente uma desigualdade.

**HIPÓTESE ALTERNATIVA:** No teste de hipóteses é a hipótese que se quer provar. É geralmente uma desigualdade. Em inglês alternative hypothesis.

**HIPÓTESE ALTERNATIVA ( $H_1$ ):** No teste de hipóteses, qualquer hipótese alternativa admissível para aquela que está sendo testada. Ou ainda é aquela que difere de uma hipótese pré-fixada. Em inglês Alternative hypotheses.

**HIPÓTESE ALTERNATIVA:** Hipótese que será considerada como aceitável, caso a hipótese nula seja rejeitada.

**HIPÓTESE ALTERNATIVA ( $H_1$ ):** É aquela que contradiz a hipótese de nulidade, ou seja, assegura, com certa margem de incerteza, que as diferenças encontradas na amostra são reais, podendo-se, conseqüentemente, rejeitar a hipótese de nulidade.

**HIPÓTESE CIENTÍFICA:** É uma proposição preliminar para explicar um fenômeno ou fato observacional cuja verdadeira causa é desconhecida, os cujas explicações presentes não satisfazem ao corpo do saber. A realização de qualquer pesquisa somente se justifica se tiver por objetivo produzir novos fatos ou evidências que favoreçam a confirmação ou rejeição do saber existente ou que permitam o estabelecimento de novos conceitos sobre a verdadeira causa do fenômeno investigado.

**HIPÓTESE CIENTÍFICA:** É aquela passível de ser testada. Logo, uma hipótese científica deve ser testada e, confirmada ou negada. Uma hipótese que não pode ser testada é uma mera especulação. Convém lembrar que o teste das hipóteses requer modelos matemáticos.

**HIPÓTESE COMPOSTA:** Hipótese estatística que define um intervalo de valores possíveis para o parâmetro de interesse. Em inglês Composite hypothesis.

**HIPÓTESE DA NULIDADE:** Hipótese que postula não haver diferença entre as populações ou grupos em comparação, com relação ao fator, à característica ou à condição de interesse. Veja hipótese nula.

**HIPÓTESE DA NULIDADE ( $H_0$ ):** É a hipótese que testa considerando-se não haver diferenças entre os grupos específicos objetos do estudo, representada, por exemplo:  $(\mu_1 = \mu_2)$  .

**HIPÓTESE DE NULIDADE:** Hipótese usada nas análises estatísticas, em que se consideram nulos certos parâmetros.

**HIPÓTESE DE NULIDADE ( $H_0$ ):** É aquela na qual se indica que as diferenças encontradas na amostra são devidas a erro amostral, não tendo, portanto, significância estatística.

**HIPÓTESE DIRECIONAL:** Uma hipótese formulada de tal maneira que se dá uma direção aos resultados, isto é, hipotetiza-se que os resultados serão maiores que ou menores que.

**HIPÓTESE ESTATÍSTICA:** É a hipótese sobre os parâmetros ou a forma da distribuição de probabilidades de uma população, ou populações, ou, de forma geral, de um mecanismo probabilístico que gera as observações. Em inglês Statistical hypothesis.

**HIPÓTESE ESTATÍSTICA:** É uma suposição feita a respeito de um ou mais ou mais de um valor de parâmetros populacionais de distribuições teóricas de probabilidades de populações ou universos estatísticos de variáveis aleatórias discretas ou contínuas, ou sobre o formato ou natureza dessas distribuições de probabilidade, ou ainda baseadas em raciocínio indutivo ou em alguma lei natural.

**HIPÓTESE ESTATÍSTICA:** É uma suposição ou conjectura a respeito de valor(es) de parâmetro(s) populacional(is) tais como a média, variância dentre outros. ou até mesmo com relação a distribuição de probabilidade de uma determinada população. A qual pode ser: i) Hipótese de nulidade ( $H_0$ ): é aquela que se refere a nulidade de parâmetros, isto é que se enuncia sob a forma de uma mais iguais a zero, ou se refere a diferença de parâmetros por exemplo:  $H_0: \mu_1 = \mu_2$  ou  $\mu_2 - \mu_1 = 0$ ; ii) Hipótese alternativa ( $H_1$  ou  $H_a$ ): É qualquer hipótese que difere da hipótese prefixada  $H_0$ , ou seja  $H_1: \mu_1 \neq \mu_2$  ou  $\mu_1 >$  ou ainda  $\mu_1 < \mu_2$ .

**HIPÓTESE ESTATÍSTICA:** Uma afirmativa sobre uma ou mais distribuições da população; mais especificadamente sobre um ou mais parâmetros de tal distribuição populacional. Hipóteses estatísticas nunca são equivalentes a hipóteses científicas, que são afirmativas sobre fenômenos. É uma descrição concreta de um ou mais aspectos sumários de uma população; não há qualquer implicação do por quê a população possui as características.

**HIPÓTESE (ESTATÍSTICA) COMPOSTA:** É a hipótese estatística  $H: \theta \in \omega$ , onde  $\theta$  é um parâmetro e  $\omega$  uma região do espaço parametral, quando  $\omega$  contém mais de um ponto parametral.

**HIPÓTESE (ESTATÍSTICA) COMPOSTA:** É a hipótese estatística  $H: \theta_i = \theta^\alpha$  relativa ao valor apenas de um dos parâmetros cujos valores se desconhecem e dos quais depende a função de distribuição de uma população. Opõe-se a hipótese estatística simples.

**HIPÓTESE (ESTATÍSTICA) SIMPLES:** É a hipótese estatística  $H: \theta \in \omega$ , onde  $\theta$  é um parâmetro e  $\omega$  uma região do espaço parametral quando, reduzindo-se essa região a um só ponto  $\theta\alpha$ , a hipótese se reduz a  $H: \theta = \theta\alpha$ .

**HIPÓTESE (ESTATÍSTICA) SIMPLES:** É a hipótese estatística  $H: \theta = \theta\alpha$  relativa ao valor do único parâmetro desconhecido de que depende a função de distribuição de uma população. Opõe-se a hipótese ou estatística composta.

**HIPÓTESE EXPERIMENTAL:** É formulada de maneira declarativa, após a pergunta experimental relativa a possíveis relacionamentos entre as variáveis dependente e independente e após as definições operacionais das variáveis. É uma asserção inambígua relativa aos resultados esperados do estudo, até mesmo ao ponto de especificar quantidade e direções antecipadas.

**HIPÓTESE INTRÍNSECA:** Uma função aleatória  $Z(x)$  pode ser considerada intrínseca quando as seguintes suposições podem ser feitas.i) a esperança matemática existe e não depende do ponto  $X$ :  $E\{Z(X)\} = m$  ;

ii) para todos os vetores  $\vec{h}$ , o incremento  $[Z(X+h) - Z(X)]$  tem uma variância finita que não depende de  $X$ :

$$\text{Var}[Z(X+h) - Z(X)] = E\{[Z(X+h) - Z(X)]^2\} = 2\gamma(h), \quad \forall X.$$

**HIPÓTESE NULA:** No teste de hipóteses é a hipótese que representa o que já se conhece e é formulada com o objetivo de ser rejeitada. Em inglês null hypothesis.

**HIPÓTESE NULA:** Suposição de que, no resultado final da investigação, não há diferenças significativas entre os grupos; a diferença encontrada é trivial, em termos estatísticos, e explicada pelo fator chance.

**HIPÓTESE NULA (ESTATÍSTICA):** Uma hipótese formulada de maneira a declarar que não há diferença ou nenhum relacionamento entre as variáveis. É também chamada de hipótese não- direcional. Serve para avaliar a possibilidade de que os resultados obtidos podem ter sido devidos ao acaso e não devidos às manipulações experimentais. A hipótese declarativa de que os resultados obtidos foram casuais efeitos de variáveis não-controladas.

**HIPÓTESE NULA:** Afirmação sobre uma característica populacional, envolvendo em geral o caso de não haver nenhuma diferença; denotada por  $H_0$ .

**HIPÓTESE NULA:** Em conexão com teste de hipótese e com teste de significação estatística, é a hipótese que sugere não haver relação entre as variáveis estudadas. Pode-se concluir que duas variáveis estão relacionadas depois da rejeição estatística da hipótese nula.

**HIPÓTESE NULA:** Hipótese que é colocada a prova em teste de hipótese. Em geral indica uma igualdade a ser contestada.

**HIPÓTESE NULA:** A hipótese que está sendo testada em uma situação de teste de hipóteses; frequentemente, a hipótese nula tem a forma não há relação entre duas grandezas.

**HIPÓTESE NULA ( $H_0$ ):** Em geral, este termo está relacionado com uma hipótese em particular em teste, distinto de hipóteses alternativas que estão sob consideração. É, entretanto, a hipótese que determina a probabilidade do erro tipo I. Em alguns contextos, porém, o termo é restrito a uma hipótese em teste de sem diferença. É ainda definida como sendo aquela hipótese descrita sob a forma de uma ou mais igualdades a zero. Em inglês Null hypothesis.

**HIPÓTESE NULA:** Tradução corrente, mas equivocada, de null hypothesis, uma vez que não é a hipótese que tem a qualidade de nula, mas sim o que ela postula que é a diferença nula. Veja hipótese da nulidade.

**HIPÓTESE NULA:** Hipótese de que as amostras surgem de populações com médias iguais para uma variável dependente ou seja teste univariado ou um conjunto de variáveis dependentes que é teste multivariado. A hipótese nula pode ser aceita ou rejeitada, dependendo dos resultados de um teste de significância estatística.

**HIPÓTESE NULA ( $H_0$ ):** A hipótese nula é a negação da hipótese alternativa, por isso, via de regra, você sempre torce para que ela seja rejeitada, como no caso seguinte: as atitudes em relação a matemática não interferem na formação das atitudes em relação à estatística. Esta é a hipótese que está sendo testada por qualquer teste estatístico. A se tomar uma decisão estatística, existem duas possibilidades de erro: o erro de tipo I: rejeitar a hipótese nula ( $H_0$ ), quando ela é verdadeira e, o erro de tipo II: aceitar a hipótese nula ( $H_0$ ), quando ela é falsa. Infelizmente, quando a probabilidade de cometer um erro, a probabilidade

de cometer o outro aumenta. Assim, os testes estatísticos foram delineados para controlar o erro de tipo I, chamado de nível de significância.

**HIPÓTESE NULA:** Hipótese de que não há diferença real ou verdadeira entre as médias ou proporções dos grupos que estão sendo comparados, ou de que não há associação real entre duas variáveis contínuas. Comparar com Hipótese alternativa. Número necessário para prejudicar (NNP). Número de pacientes que precisam ser tratados, em média, para que um deles sofra os efeitos adversos de um determinado tipo de tratamento. O NNP está baseado no incremento do risco absoluto (IRA). Se, por exemplo, o risco de prejudicar é igual a 0,02 sem o tratamento e igual a 0,05 com o tratamento, o IRA = 0,03. O NNP é calculado como sendo  $\frac{1}{IRA}$ . Nesse exemplo ele seria  $\frac{1}{0,03} = 0,03$ . Comparar com Número necessário para beneficiar.

**HIPÓTESE SIMPLES:** É aquela hipótese estatística que assume um único valor para o parâmetro de estudo. Em inglês Simple hypothesis.

**HIPÓTESES:** São conjecturas com as quais se procuram explicar, por tentativa, fenômenos ocorridos ou concorrentes. São respostas dadas a problemas postos pela ciência ou pelo senso comum. Deverão ser testadas para comprovação ou refutação.

**HIPÓTESES (CARACTERÍSTICAS):** As características de uma hipótese são as seguintes: i) são formuladas de maneira declarativa; ii) devem descrever um relacionamento entre duas ou mais variáveis; iii) devem ser testáveis; iv) devem ser operacionais, no sentido de que não deve haver ambigüidade nas variáveis ou relacionamentos propostos; v) devem refletir uma preocupação com um problema, baseada em algum conhecimento, pesquisa anterior, ou necessidades identificadas.

**HIPÓTESES ALTERNATIVAS:** Há três conjecturas ou hipóteses alternativas que podem ser consideradas rivais da hipótese nula: ( $H_0 : \mu = 50; H_1 : \mu > 50; H_1 : \mu < 50; H_1 : \mu \neq 50$ ).

**HIPÓTESES ENCAIXADAS:** Sequencia de hipóteses em que a hipótese para um estágio está contida em todas as hipóteses posteriores da sequencia. em inglês nested hypothesis.

**HETEROGENEIDADE CLÍNICA:** É constatada quando os estudos de uma metanálise não são comparáveis em razão das diferenças entre as populações, definição de variáveis, dentre outros.

**HETEROGENEIDADE ESTATÍSTICA:** Ocorre quando há uma diferença estatisticamente significativa entre os efeitos de interesse em uma metanálise.

**HETEROSCEDASTICIDADE:** As variâncias das observações ou de um parâmetro estimado e grupos diferentes não são iguais.

**HIPÓTESE ALTERNATIVA:** É a proposição ou afirmação que discorda da hipótese que está sendo testada que é a hipótese nula.

**HIPÓTESE ALTERNATIVA ( $H_1$ ):** As hipóteses de uma pesquisa devem enunciar-se por propostas tão claras e específicas quanto possível, via de regra, é o que você quer mostrar, por exemplo: as atitudes em relação a matemática interferem diretamente na formação das atitudes em relação à estatística.

**HIPÓTESE NULA ( $H_0$ ):** O termo dado à proposição de que está sendo testada em um procedimento de teste de hipótese. Geralmente é expressa em termos de não influência; por exemplo, não há diferença entre as médias.

**HIPÓTESE:** Proposição provisória acerca de um fenômeno, fato ou relação entre variáveis. Suposição realizada provisoriamente com o intuito de explicar algo que se desconhece. Na atividade de pesquisa, o enunciado da(s) hipóteses(s) constitui-se na fase do método científico que vem depois da formulação do problema: sob certo aspecto, podemos afirmar que toda pesquisa científica consiste apenas em enunciar e verificar hipóteses. Por exemplo: considerando que o tabagismo exerce efeito sobre os mecanismos antioxidativos e/ou sobre o suprimento da vasculatura do sistema auditivo, pode-se formular a hipótese, por dedução, de que o tabagismo crônico compromete a audição.

**HIPÓTESE ALTERNATIVA:** Representada por ( $H_1$ ) é a afirmativa contrária ou complementar à hipótese nula. Quando a hipótese nula é rejeitada, a hipótese alternativa é verdadeira e vice-versa.

**HIPÓTESE ESTATÍSTICA:** i) Afirmção provisória acerca da igualdade ou da diferença entre dois parâmetros populacionais, com base em dados coletados de amostras dessas populações, por exemplo, duas médias amostrais; ii) Afirmção provisória acerca de uma distribuição de dados.

**HIPÓTESE EX POST FACTO:** Hipótese formulada acerca de um fenômeno que já ocorreu. Por exemplo: uma pesquisa que tenha por objetivo explicar o comportamento das bolsas de valores no ano anterior começará por elencar hipóteses ex post facto que expliquem o ocorrido. O contrário deste tipo de hipótese é a hipótese preditiva.

**HIPÓTESE NULA:** Representada por ( $H_0$ ), é a hipótese que o pesquisador deseja rejeitar. É formulada de maneira a declarar que não há diferença entre dois resultados que estão sendo comparados ou entre as variáveis investigadas. Por exemplo: numa pesquisa que visa comprovar a eficiência de determinado método de ensino, pode-se montar uma situação ou experimento na qual se compara o novo método aos métodos tradicionais. A  $H_0$  seria a de que o novo método produz os mesmos resultados que os métodos anteriores. Por outro lado, a hipótese alternativa ( $H_1$ ) seria a de que o novo método é superior aos métodos mais antigos. Toda hipótese é rejeitada ou confirmada tendo por base determinado nível de significância.

**HIPÓTESE PREDITIVA:** Hipótese que tem por objetivo prever o comportamento futuro de uma ou mais variáveis. Compara com hipótese ex post facto.

**HIPÓTESE PROBANDA:** Ver hipótese nula.

**HISTOGRAMA:** Representação gráfica dos dados de uma distribuição, consistindo em retângulos, em geral sobre uma abscissa que é o eixo inferior ou eixo x, sendo que as áreas dos retângulos são proporcionais às frequências dos dados. Difere do gráfico de barras pelo fato de a abscissa possuir necessariamente uma escala contínua, os retângulos serem justapostos e a área dos mesmos representar exatamente as frequências de cada categoria.

**HISTERESE:** Fenômeno pelo qual uma série de valores registrados à medida que aumentam de magnitude são diferentes daqueles quando que apresentam diminuição de magnitude; é verificada em alguns instrumentos.

**HISTOGRAMA:** É um diagrama bidimensional que ilustra uma distribuição de frequência de uma variável contínua. Normalmente, o eixo horizontal representa as unidades de medida da variável; os retângulos acima de cada intervalo de classe de largura constante indicam a frequência para aquela classe, de acordo com sua altura.

**HOMOSCEDASTICIDADE:** As variâncias das observações ou de um parâmetro estimado em diferentes grupos são iguais.

**HISTOGRAMA:** Representação gráfica para dados contínuos sob a forma de retângulos justapostos. A base de cada retângulo representa o intervalo de cada classe e a altura, a respectiva frequência.

**HISTOGRAMA:** É uma representação gráfica aos dados numa distribuição, consistindo de retângulos em geral sobre uma abscissa. As áreas dos retângulos devem ser proporcionais às frequências dos dados.

**HISTOGRAMA:** É a representação gráfica da distribuição de frequências simples. As observações originais são acumuladas em intervalos de classe predefinidos, conforme algum critério. As frequências numéricas das classes são transformadas em frequências percentuais e lançadas no eixo das ordenadas em função dos intervalos de classe que são representados por barras de largura proporcional ao tamanho do intervalo.

**HISTOGRAMA:** Gráfico de colunas justapostas ou de retângulos onde a base é a amplitude da classe e a altura é a densidade ou a densidade relativa da classe. É utilizado, normalmente, para representar uma variável contínua. Em inglês histogram.

**HISTOGRAMA:** Gráfico constituído por barras verticais, situadas ao longo da abscissa. A largura das colunas representa o intervalo de classe. Pelo fato de os intervalos serem iguais, as alturas das barras são proporcionais às frequências de classe. Em inglês histogram.

**HISTOGRAMA:** Diagrama em barras que ilustra uma distribuição de frequência; cada barra é construída de modo que sua área seja proporcional ao número de elementos contidos no intervalo que ela representa.

**HISTOGRAMA:** Representação gráfica da distribuição das frequências absolutas ou relativas normalmente utilizado para variáveis contínuas. Suas principais características são: as barras devem estar todas juntas, cada barra representa a frequência de um intervalo de valores e os intervalos devem ter todos a mesma amplitude.

**HISTOGRAMA:** É um gráfico de barras em que a área destas é proporcional à frequência, não havendo espaço entre as mesmas. Só se utiliza em variáveis quantitativas contínuas.

**HISTOGRAMA:** Conjunto de retângulos que têm as bases sobre o eixo x e a área proporcional às frequências de classe. Diagrama constituído por retângulos ou linhas desenhadas a partir de uma linha de base, em que a posição deles ao longo dessa linha representa o valor ou a amplitude de uma das variáveis, e a sua altura, o valor correspondente de uma segunda variável.

**HISTOGRAMA:** Um diagrama de frequências univariado cujas áreas de retângulos são proporcionais às frequências de classe, e a largura de cada seção, representando o intervalo da classe correspondente da variável. Em inglês Histogram.

**HISTOGRAMA:** Representação gráfica da distribuição de uma única variável. Ao fazer a contagem de frequência em categorias, a forma da distribuição da variável pode ser mostrada. Usando para fazer uma comparação visual com a distribuição normal.

**HIPÓTESE (COMO EM HIPÓTESE ESTATÍSTICA):** Uma afirmação acerca dos parâmetros de uma distribuição de probabilidades ou de um modelo, ou uma afirmação acerca da forma de uma distribuição de probabilidades.



**HIPÓTESE ALTERNATIVA:** Em um teste estatístico de hipóteses, essa é a hipótese diferente daquela que está sendo testada. A hipótese alternativa contém condições realizáveis, enquanto a hipótese nula especifica condições que estão sob teste.

**HIPÓTESE NULA:** Esse termo geralmente relaciona uma hipótese particular que está sob teste, diferentemente da hipótese alternativa que define outras condições que são factíveis, porém não estão sendo testadas. A hipótese nula determina a probabilidade do erro tipo I para o procedimento de teste.

**HISTOGRAMA:** Uma disposição de dados univariados que usa retângulos proporcionais em área às frequências de classes, de modo a exibir visivelmente as características dos dados, como localização, variabilidade e forma.

**HISTOGRAMA:** Representação gráfica da distribuição da frequência de níveis de cinza de uma imagem matricial de sensor remoto. Apresenta no eixo x a variação de níveis de cinza da imagem e no eixo y a frequência de cada valor digital na mesma.

**HISTOGRAMA DE FREQUÊNCIAS RELATIVAS:** Variante do histograma básico em que as frequências são substituídas por frequências relativas.

**HISTOGRAMA DE PROBABILIDADES:** Histograma com resultados dispostos ao longo do eixo horizontal e as probabilidades marcadas no eixo vertical.

**HISTÓRIA:** Coisas que acontecem ou variáveis externas que mudam entre o começo e o fim de um experimento.

**HISTÓRIA (AMEAÇA À VALIDADE INTERNA):** Eventos que são estranhos à variável independente, mas que ocorrem concomitantemente com ela, podem produzir mudanças na variável dependente. Quando tais eventos ocorrem, efeitos verdadeiros provenientes da intervenção são confundidos com história.

**HISTÓRIA (COMO PESQUISA):** História é uma narrativa integrada de eventos passados, escrita como espírito crítico. Seus métodos diferem dos das ciências naturais, pois não é uma disciplina em que se emprega observação direta e experimentação. Todavia, suas observações não podem ser repetidas.

$H_0$ : Hipótese nula.

**HOLOENDEMIA:** Ver endemia.

**HOMOCEDASTICIDADE:** Para muitas variáveis aleatórias com distribuição conjunta dada, se a variância de uma variável é a mesma para todas as outras, a distribuição é chamada homocedástica. Em inglês Homoscedasticity.

**HOMOSCEDASTICIDADE:** Condição em que a variável resposta (y) tem distribuição normal dentro de cada classe de x, e sua variância dentro de cada classe de x não está associada aos de x. Esta é uma condição necessária para a análise de regressão e análise de variância.

**HOMOCEDASTICIDADE:** Vide homogeneidade das variâncias.

**HOMOGENEIDADE:** Este termo é usado em estatística no seu sentido ordinário, mas freqüentemente ocorre em conexão com amostras de populações diferentes que podem ou não ser idênticas. Se as populações

são idênticas são chamadas homogêneas, e por extensão, os dados amostrais serão chamados homogêneos. Em inglês Homogeneity.

**HOMOSCEDASTICIDADE:** Quando a variância dos termos de erro ( $\epsilon$ ) parece constante ao longo de um domínio de variáveis preditoras, diz-se que os dados são homoscedásticos. A suposição de variância igual do erro E da população, onde E estimado a partir de  $\epsilon$ , é crítica para a aplicação correta da regressão linear. Quando os termos de erro têm variância crescente ou flutuante, diz-se que os dados são heteroscedásticos. A análise de resíduos ilustra melhor este ponto.

**HOMOSCEDASTICIDADE:** Descrição de dados para os quais a variância dos termos de erro ( $\epsilon$ ) aparece constante no intervalo de valores de uma variável independente. A suposição de igual variância do erro da população  $\epsilon$ , onde  $\epsilon$  é estimado a partir de  $e$ , é essencial à aplicação adequada de regressão linear. Quando os termos de erro têm variância crescente ou flutuante, diz-se que os dados são heterocedásticos.

**HOMOCEDASTICIDADE OU IGUALDADE DE VARIÂNCIAS:** Este pressuposto exige que o nível de dispersão da variável dentro dos grupos seja similar. Os pacotes ou softwares estatísticos SAS e SPSS automaticamente testam a hipótese de nulidade dada por:  $H_0: s_1 = s_2 = s_3 = s_4$  através do teste de Levene.

**HOMOCEDASTICIDADE:** Conceito associado a teoria da regressão e que significa que as variáveis dependentes Y, apresentam a mesma variabilidade para cada valor de  $x=x_0$ , isto é, para cada nível de x, a variável Y varia na expectância  $E(Y/x)$ , mas não na variância. Em inglês homoscedasticity.

**HOMOCEDASTICIDADE:** É uma suposição relacionada primariamente a relações de dependência entre variáveis. Refere-se à suposição de que as variáveis dependentes exibem níveis iguais de variância ao longo do domínio da (s) variável (is) preditor(a)s. A homocedasticidade é desejável porque a variância da variável dependente sendo explicada na relação de dependência não deveria se concentrar apenas em um domínio limitado dos valores independentes. Apesar de as variáveis dependentes deverem ser métricas, esse conceito de igual extensão de variância entre as variáveis independentes pode ser aplicado quando as variáveis independentes são métricas ou não. Com variáveis independentes métricas, o conceito de homocedasticidade se baseia na extensão da variância da variável dependente no domínio dos valores das variáveis independentes, o que se encontra em técnicas como a regressão múltipla. O mesmo conceito também se aplica quando as variáveis independentes são não métricas. Nestes casos, como os que ocorrem em análise de variância univariada (ANOVA) e em análise de variância multivariada (MANOVA), o objetivo neste caso é tornar a igualdade da variância uma só variável dependente ou das matrizes de variância e covariância em múltiplas variáveis independentes nos grupos formados pelas variáveis independentes não métricas. A igualdade das matrizes de variância e covariância também é vista na análise discriminante, mas nessa técnica a ênfase é na dispersão das variáveis independentes nos grupos formados pela medida dependente não métrica. Em todos esses casos, o objetivo é o mesmo, ou seja, garantir que a variância usada na explicação e previsão esteja distribuída no domínio de valores, permitindo assim um teste justo da relação entre todos os valores das variáveis não métricas. Na maioria das situações, temos muitos valores diferentes da variável dependente em cada valor da variável independente. Para que essa relação seja totalmente verificada, a dispersão ou variância dos valores da variável dependente deve ser igual em cada valor da variável preditora. A maioria dos problemas das variâncias diferentes surge de uma entre duas fontes. A primeira é o tipo de variáveis incluídas no modelo. Por exemplo, quando uma variável aumenta em valor, por exemplo unidades que variam de algo próximo a zero até milhões, ocorre um intervalo naturalmente maior de possíveis respostas para os valores maiores. A segunda fonte resulta de uma distribuição assimétrica que cria heterocedasticidade a qual

é provocada por assimetria de uma variável, ou seja para os diversos valores de uma variável há diferentes padrões de dispersão para a outra, isso faz com que as previsões sejam melhores em alguns níveis da variável independente do que em outros. Violar a suposição de homocedasticidade frequentemente torna os testes de hipóteses muito conservadores ou muito sensíveis. Muitas vezes, o efeito da heterocedasticidade também está relacionado ao tamanho da amostra, especialmente quando se examina a dispersão de variância em grupos. Por exemplo, nas análises de variância univariada (ANOVA) e multivariada (MANOVA), o impacto da heterocedasticidade sobre o teste estatístico depende do tamanho da amostra associados com os grupos de menores e maiores variâncias. Em análise de regressão múltipla, efeitos semelhantes ocorreriam em distribuições altamente assimétricas onde houvesse números desproporcionais de respondentes em certos intervalos da variável independente.

**HOMOCEDASTICIDADE (HOMOGENEIDADE DE VARIÂNCIAS):** Os erros ou desvios  $e_{ij}$ , devidos ao efeito de fatores não controlados, devem possuir uma variância comum  $\sigma^2$ , isto significa que a variabilidade das repetições de um tratamento deve ser semelhante à dos outros tratamentos, isto é, os tratamentos devem possuir variâncias homogêneas.

**HOMOGENEIDADE:** Qualidade de semelhança. Um grupo de pessoas é homogêneo na medida em que elas são semelhantes entre si.

**HOMOGENEIDADE DAS AMOSTRAS:** É um modelo probabilístico que indica se as amostras investigadas são oriundas da mesma população, sendo utilizado nos testes G, do Qui-Quadrado, da Correlação linear e da regressão linear.

**HOMOLOGIA:** É a forma de anexar grupos abelianos ou objetos algébricos mais elaborados a um espaço topológico de forma a obter invariantes algébricos. Da mesma forma ele detecta a presença de buracos de várias dimensões em um espaço. Os métodos desenvolvidos para isto levaram ao que hoje é conhecido como álgebra homológica e invariantes homológicos podem ser calculados para muitas estruturas puramente algébricas. Em inglês homology.

**HOMOZIGOTOS:** Genes de um par específico ou chamados também de alelos são semelhantes em um animal.

**HYPÉRBOLE:** Curva simétrica do segundo grau. É uma das seções cônicas. É o conjunto de pontos para os quais a diferença das distâncias de dois pontos fixos ou focos é constante. As assíntotas da hipérbole são tangentes pelo centro e tocam a hipérbole somente no infinito. A hipérbole pode ser também definida via coordenadas cartesianas como sendo o conjunto de pontos do plano que satisfazem a equação:  $ax^2 - by^2 = 1$ . Em inglês hyperbola.

**HETEROCEDÁSTICO (HETEROCEDASTIC):** A propriedade de um conjunto de dados que diz que a variância residual de todos os grupos de tratamentos são diferentes. Comparar com homocedástico.

**HIPÓTESE (HYPOTHESIS):** Uma afirmação testável de causa e efeito. Ver também hipótese alternativa e hipótese nula.

**HIPÓTESE ALTERNATIVA (ALTERNATIVE HYPOTHESIS):** A hipótese de interesse. Ela é a hipótese estatística que corresponde à questão científica que está sendo examinada com os dados. Diferente da hipótese nula, a hipótese alternativa diz que algo está acontecendo. Comparar com hipótese nula e com hipótese estatística.

**HIPÓTESE CIENTÍFICA (SCIENTIFIC HYPOTHESIS):** Uma declaração de causa e efeito. A acurácia de uma hipótese científica é inferida a partir dos resultados de testar hipóteses estatísticas nulas e alternativas.

**HIPÓTESE EXTRÍNSECA (EXTRINSIC HYPOTHESIS):** Um modelo estatístico em que os parâmetros são estimados a partir de outras fontes que não os dados. Comparar com hipótese intrínseca.

**HIPÓTESE INTRÍNSECA (INTRINSIC HYPOTHESIS):** Um modelo estatístico cujos parâmetros são estimados a partir dos próprios dados. Comparar com hipótese extrínseca.

**HISTOGRAMA:** Histogramas: são constituídos por um conjunto de retângulos, com as bases assentadas sobre um eixo horizontal, tendo o centro da mesma no ponto médio da classe que representa, e cuja altura é proporcional à frequência da classe.

**HIPÓTESE TABULAR BÁSICA:** É a hipótese que estabelece ou supõe que todas as observações contidas numa classe são consideradas numericamente iguais ao ponto médio da classe.

**H<sub>1</sub>:** É a hipótese estatística que difere de uma suposição pré fixada, também simbolizada por H<sub>a</sub> e conhecida por hipótese alternativa.





# I

**IBID OU IBDEM:** Significa na mesma obra.

**IC:** Intervalo de confiança ou intervalo fiducial.

**ICEBERG (FENÔMENO DO):** Em epidemiologia, refere-se à evidência de pesquisa porção de casos clínicos em relação à elevada proporção de infecções inaparentes de determinadas doenças. É o que ocorre, por exemplo, nas epidemias de poliomielite, meningite, difteria, febre tifóide e outras; apenas por meio de inquéritos sorológicos poder-se-ão detectar os portadores.

**IDEM OU ID:** Significa igual a anterior.

**IDENTIDADE OU IDÊNTICO A:** Símbolo matemático dado por  $\equiv$ .

**IDENTIFICABILIDADE:** Em certos sistemas de equações estocásticas pode acontecer que alguns ou todos os parâmetros não podem ser estimados de forma separada sem cometer um vício, apesar de contar com uma grande quantidade de dados, ou se o número de equações é igual ao número de variáveis endógenas desconhecidas. Em inglês Identifiability.

**IDENTIFICAÇÃO:** Grau em que existe um número suficiente de equações para resolver cada um dos coeficientes (desconhecidos) a serem estimados. Modelos podem ser subidentificados (não podem ser resolvidos), exatamente identificados (número de equações igual ao número de coeficientes estimados com nenhum grau de liberdade), ou superidentificados (mais equações do que coeficientes estimados e graus de liberdade acima de zero). O pesquisador deseja ter um modelo superidentificado para o mais rigoroso teste do modelo proposto. Ver também graus de liberdade.

**IG:** Progênie de irmãos germanos.

**IGUAL A:** Símbolo matemático dado por  $=$ .

**i.i.d.:** Independente e identicamente distribuídos.

**IME:** Índice multiefeitos.

**IMPLICA:** Símbolo matemático dado por  $\Rightarrow$ .

**IMPLEMENTAÇÃO:** Expressão usada para denotar a ação de adaptar um modelo a um conjunto de dados reais. Em inglês Implementation.

**In:** Significa em.

**INDUÇÃO:** Trata-se de um tipo especial de inferência, onde partindo-se do conhecimento de uma parte ou amostra tira-se conclusões sobre a realidade do todo que é população ou universo estatístico.

**INCERTEZA:** Característica de um fenômeno ou de uma situação em virtude da qual esta não se concretiza necessariamente da mesma maneira, ainda que se repitam as condições em que ela se realizou não se podendo sequer conhecer a probabilidade de ocorrência dos seus possíveis resultados.

**INCIDÊNCIA:** Número de novos casos de um evento, em uma população definida, em um período de tempo especificado. O coeficiente de incidência é este número expresso em unidade de população.

**INCIDÊNCIA:** Estudo epidemiológico de ocorrência de casos novos de determinada doença, constatados ao longo de um período de tempo prefixado como, por exemplo, seis (6) meses, um ano, e assim por diante. Os estudos relativos à incidência são denominados de longitudinais.

**INCIDÊNCIA:** Frequência ou número de novas ocorrências de doença, lesão ou óbito, na população em estudo durante o período de observação.

**INCIDÊNCIA:** Tempo que em epidemiologia traduz a idéia de intensidade com que acontece a morbidade em uma população. Ver coeficiente de incidência.

**INCIDÊNCIA (CASOS INCIDENTES):** Frequência ou número de ocorrências de novas doenças, acidentes ou mortes, isto é, número de transições de saudável para doente, de não-acidentado para acidentado ou de vivo para morto em um estudo populacional durante o período de tempo estudado.

**INCIDÊNCIA ACUMULADA:** Número total de casos incidentes ocorridos ao longo de um período específico de tempo. Os casos incidentes que resultam em morte devem ser incluídos na incidência acumulada, mas não na prevalência.

**INDEPENDÊNCIA:** Uma propriedade de um modelo de probabilidade e de dois ou mais eventos que permite que a probabilidade da interseção seja calculada como o produto das probabilidades.

**INFERÊNCIA:** Conclusão de uma análise estatística. Geralmente se refere à conclusão de um teste de hipóteses ou a uma estimativa de intervalo.

**INFERÊNCIA ESTATÍSTICA:** Veja Inferência.

**INSTRUMENTO:** Aparelho através do qual se obtém medidas de determinado fenômenos ou se obtém dados numa pesquisa. Nas ciências agrárias, por exemplo, instrumentos que podem ser termômetros, barômetros, microscópios, paquímetros, régua, etc.

**IID:** Representa independente e identicamente distribuído. Bons delineamentos experimentais resultam em réplicas e distribuição de erros que são independentes e identicamente distribuídos (IID). Ver também ruído branco, independente e randomização. Uma amostra casual ou aleatória simples iid é aquela em que todos os seus elementos são variáveis aleatórias independentes e identicamente distribuídas, tendo todas elas a mesma distribuição de probabilidade da população parental que deu origem a esta amostra aleatória.

**IDENTIFICAÇÃO:** Identificação refere-se à possibilidade de calcular os parâmetros estruturais de um modelo de equações simultâneas, através dos parâmetros da forma reduzida. Uma equação de um sistema é exatamente identificada se o número de variáveis exógenas excluídas da equação é igual ao número de variáveis endógenas na equação menos 1. Contudo, uma equação de um sistema superidentificada ou subidentificada se o número de variáveis exógenas excluídas da equação excede ou é menor do que o número de variáveis endógenas incluídas na equação menos 1. Embora isto seja somente uma condição

necessária, e não uma condição suficiente, para a identificação, este procedimento leva usualmente à resposta correta. Os coeficientes estruturais únicos só podem ser calculados através dos coeficientes da forma reduzida apenas no caso de uma equação exatamente identificada.

**IDENTIFICAÇÃO:** Identificação refere-se à possibilidade ou impossibilidade de obter parâmetros estruturais de um sistema de equações simultâneas através de parâmetros da forma reduzida. Uma equação de um sistema pode ser exatamente identificada, superidentificada ou subidentificada. O sistema como um todo é exatamente identificado se todas suas equações são exatamente identificadas. Uma equação e um sistema é justamente ou exatamente identificada se o número de variáveis exógenas excluídas de uma equação é igual ao número de variáveis endógenas da equação menos um (1). Para uma equação exatamente identificada, valores únicos dos parâmetros estruturais podem ser calculados através dos parâmetros da forma reduzida. Uma equação de um sistema é superidentificada se o número de variáveis exógenas excluídas da equação excede o número de variáveis endógenas na equação menos um (1). Para uma equação superidentificada, mais do que o valor numérico pode ser calculado a partir de algum dos parâmetros estruturais da equação, a partir dos parâmetros da forma reduzida. Uma equação de um sistema é subidentificada ou não identificada se o número de variáveis excluídas da equação é menor do que o número de variáveis endógenas excluídas da equação menos um (1). Neste caso, nenhum parâmetro estrutural pode ser calculado através dos parâmetros da forma reduzida. As regras precedentes para identificação chamada de condição de ordem são necessárias mas não suficientes. Contudo, desde que essas regras forneçam o resultado mais correto na maioria das aplicações, serão as únicas aqui discutidas. Uma condição suficiente para identificação é dada pela condição de rank, a qual afirma que em um sistema de  $G$  equações, alguma particular equação é identificada se e somente se for possível obter um determinante não nulo de ordem  $G-1$  dos coeficientes das variáveis excluídas daquela particular equação, mas incluídas em outra equação do modelo. Quando esta condição de rank é satisfeita, a condição de ordem é automaticamente satisfeita. Entretanto, o inverso não é verdadeiro.

**IMPUREZA (IMPURITY):** Uma medida de incerteza dos nós de uma árvore de classificação.

**INCLINAÇÃO (SLOPE):** A mudança no valor esperado de uma variável resposta para uma dada mudança na variável preditora. O parâmetro  $\beta_1$  em um modelo de regressão.

**INDUÇÃO (INDUCTION):** O processo de raciocínio científico que progride do específico, por exemplo, este cavalo é amarelo, para o geral, todos os cavalos são amarelos. Comparar com dedução.

**INTERAÇÃO:** Em experimentos fatoriais, dois fatores interagem se o efeito de uma variável é diferente em níveis diferentes das outras variáveis. Em geral, quando variáveis operam independentemente umas das outras, elas não exibem interação.

**INTERCEPTO:** O termo constante em um modelo de regressão.

**INTERFERÊNCIA:** Outro nome para controle excessivo.

**INTERVALO ESTATÍSTICO DE ESTIMAÇÃO OU DE CONFIANÇA:** Tipo de intervalo construído baseado na distribuição amostral do estimador, o qual estima através de dois limites o inferior e o superior o verdadeiro valor desconhecido do parâmetro da população sob estudo, como por exemplo, estimar a média de todos os resultados ou amostras possíveis com reposição ou sem reposição que já foram, que estão sob estudo e que serão obtidas em determinada pesquisa científica com uma probabilidade de confiança de  $[1 - \alpha]\%$ , e um erro de estimação ou de amostragem de no máximo igual a  $\alpha$ .

**INTERVALO ESTATÍSTICO DE PREDIÇÃO:** Tipo de intervalo construído baseado na distribuição amostral do estimador, o qual estima através de dois limites o inferior e o superior o verdadeiro valor desconhecido do parâmetro da população sob estudo, no que se refere a uma próxima observação. A sua diferença para o intervalo estatístico de confiança está na fórmula que calcula o erro, agora chamado de erro de predição.

**INTERVALO ESTATÍSTICO DE TOLERÂNCIA:** tipo de intervalo construído baseado na distribuição amostral do estimador, o qual estima através de dois limites o inferior e o superior a proporção de valores populacionais da variável aleatória X com um nível de probabilidade de confiança de  $[1-\alpha]\%$ .

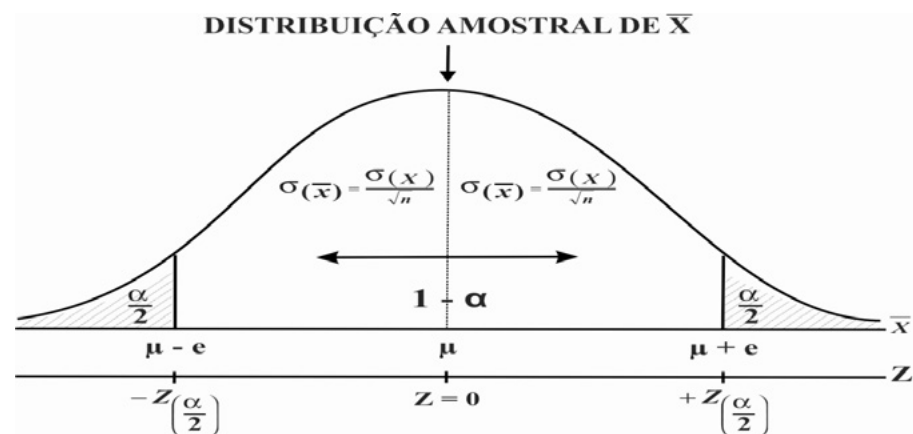
**INTERVALO DE CONFIANÇA:** Se for possível escrever um enunciado de probabilidade na forma  $P(L \leq \theta \leq U) = 1 - \alpha$ , em que L e U são funções somente dos dados da amostra e  $\theta$  é um parâmetro, então o intervalo L e U é chamado de intervalo de confiança ou um intervalo de confiança de  $100(1-\alpha)\%$ . A interpretação é que a afirmação de que o parâmetro  $\theta$  se encontra no intervalo será verdadeira em  $100(1-\alpha)\%$  das vezes que tal afirmação for feita.

**INTERVALO DE CLASSE:** Grupos em que se subdivide a amplitude total de dados numéricos com a finalidade de agrupar em uma distribuição de frequência uma massa de dados referentes a um determinado fenômeno estudado, usando determinado critério de agrupamento.

**INTERVALO DE DECISÃO:** Um conjunto de parâmetro em um algoritmo tabular de CUSUM (Cumulative Sum) o qual é um gráfico de controle das somas acumuladas, que é determinado a partir de um compromisso entre alarmes falsos e a detecção de causas atribuídas.

**INTERVALO DE CONFIANÇA PARA A MÉDIA POPULACIONAL ( $\mu$ ) DE UMA POPULAÇÃO INFINITA, NORMAL COM VARIÂNCIA ( $\sigma^2$ ) CONHECIDA OU NÃO, E COM A UTILIZAÇÃO DE AMOSTRA GRANDE [ $n \geq 30$ ]:**

É o intervalo de confiança, mostrado a seguir,



$$\mu - e \leq \bar{X} \text{ e } \bar{X} \leq \mu + e$$

$$\mu \leq \bar{X} + e \text{ e } \bar{X} - e \leq \mu$$

Ou seja, tem-se que:

$$\bar{X} - e \leq \mu \leq \bar{X} + e$$

e assim o intervalo é dado por:

$$P[\bar{X} - e \leq \mu \leq \bar{X} + e] = 1 - \alpha$$

$$Z\left(\frac{\alpha}{2}\right) = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \therefore Z\left(\frac{\alpha}{2}\right) = \frac{(\mu + e) - \mu}{\frac{\sigma}{\sqrt{n}}} \therefore Z\left(\frac{\alpha}{2}\right) = \frac{+e}{\frac{\sigma}{\sqrt{n}}} \therefore e = Z\left(\frac{\alpha}{2}\right) \cdot \frac{\sigma}{\sqrt{n}}$$

$$\bar{X} - e \leq \mu \leq \bar{X} + e$$

$$\bar{X} - Z\left(\frac{\alpha}{2}\right) \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z\left(\frac{\alpha}{2}\right) \cdot \frac{\sigma}{\sqrt{n}}$$

$$P\left[\bar{X} - Z\left(\frac{\alpha}{2}\right) \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z\left(\frac{\alpha}{2}\right) \cdot \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha$$

LSC      PARÂMETRO      LIC      COEFICIENTE DE CONFIANÇA

$$I. C.: \bar{X} \pm e \therefore \bar{X} \pm Z\left(\frac{\alpha}{2}\right) \cdot \frac{\sigma}{\sqrt{n}} \text{ OU } \bar{X} \pm Z\left(\frac{\alpha}{2}\right) \cdot \frac{S}{\sqrt{n}}$$

**Figura:** Gráfico representativo da distribuição amostral da média  $\bar{X}$  utilizado na construção do intervalo de confiança para a média populacional  $\mu$ , quando se utiliza grandes amostras ( $n \geq 30$ ).

Sendo assim o intervalo de confiança é determinado conforme a expressão abaixo.

$$P\left[\bar{X} - z\left(\frac{\alpha}{2}\right) \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z\left(\frac{\alpha}{2}\right) \cdot \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha$$

Onde  $\bar{X}$  e  $\sigma$  são, respectivamente a media e o desvio padrão da população com distribuição normal e variância conhecida, e assim o intervalo de confiança de  $100(1-\alpha)\%$  para a media populacional  $\mu$  é dada pela equação acima, onde  $Z\left[\frac{\alpha}{2}\right]$  é o valor critico ou tabelado sob a curva da distribuição normal padrão.

**INTERVALO DE CONFIANÇA PARA A MÉDIA ( $\mu$ ) DE UMA POPULAÇÃO INFINITA, OU FINITA E AMOSTRAGEM COM REPOSIÇÃO, APROXIMADAMENTE NORMAL COM VARIÂNCIA ( $\sigma^2$ ) DESCONHECIDA E AMOSTRA PEQUENA ( $n < 30$ ):** É o intervalo de confiança, dado por,

$$P\left[\bar{X} - t\left(v; \frac{\alpha}{2}\right) \cdot \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t\left(v; \frac{\alpha}{2}\right) \cdot \frac{S}{\sqrt{n}}\right] = 1 - \alpha$$



Lembrado que a variável t possui  $v = n-1$  graus de liberdade.

Onde  $\bar{X}$  e S são, respectivamente a média e o desvio padrão de uma amostra aleatória de uma população com distribuição normal e variância desconhecida, e assim o intervalo de confiança de  $100(1-\alpha)$  % para a media populacional  $\mu$  é dada pela equação acima, onde  $t_{\left[v; \frac{\alpha}{2}\right]}$  é o valor critico ou tabelado sob a curva da distribuição t de Student.

**INTERVALO DE CONFIANÇA PARA A MÉDIA DAS DIFERENÇAS ( $\mu_D$ ) DE DUAS POPULAÇÕES QUE NÃO SÃO INDEPENDENTES, ISTO É, AS VARIÁVEIS SÃO EMPARELHADAS (DADOS EMPARELHADOS). E COM A AMOSTRA DE DIFERENÇAS (n) PEQUENA ( $n < 30$ ):** É o intervalo descrito por,

$$P\left(\bar{D} - t_{\left(v; \frac{\alpha}{2}\right)} \frac{S_D}{\sqrt{n}} \leq \mu_D \leq \bar{D} + t_{\left(v; \frac{\alpha}{2}\right)} \frac{S_D}{\sqrt{n}}\right) = 1 - \alpha$$

Onde,  $v = n-1$ ;  $n =$  número de elementos da amostra de  $n$  diferenças;  $\bar{D}$  = média da amostra de  $n$  diferenças;  $S_D =$  o desvio padrão da amostra de  $n$  diferenças.

Os resultados das duas amostras constituem dados emparelhados ou pareados, quando estão relacionados dois a dois segundo algum critério que introduz uma influência marcante entre os diversos pares, que supomos, porém, influir igualmente sobre os valores de cada par.

É importante salientar que se os dados das duas amostras estão emparelhados tem sentido calcularmos as diferenças ( $d_i$ ) correspondentes a cada par de valores, reduzindo assim os dados a uma única amostra de  $n$  diferenças.

Os exemplos i e ii a seguir ilustram situações em que os dados obtidos de suas amostras são correlacionados ou dados emparelhados.

**Exemplo 1:** Quando certo caráter é medido no mesmo indivíduo, em épocas diferentes, os valores obtidos nas duas mensurações tendem a ser mais parecidos entre si do que se houvessem sido obtidos de indivíduos diferentes. Como por exemplo, a medição de taxas de crescimento de plantas de uma determinada cultura, antes e depois de se aplicar uma substância inibidora da fotossíntese.

**Exemplo 2:** As eficiências de duas rações podem ser comparadas utilizando-se vários pares de animais irmãos de uma mesma leitegada, suínos, por exemplo, ou de uma mesma ninhada, como camundongos, por exemplo.

O experimento consiste em alimentar cada membro, de cada par, com uma das rações alocada ao acaso. Indivíduos deste tipo são denominados irmãos-germanos pelas suas semelhanças genéticas, tendem a apresentar respostas correlacionadas aos estímulos a que são submetidos.

**INTERVALO DE CONFIANÇA PARA A DIFERENÇA ENTRE AS MÉDIAS ( $\mu_1 - \mu_2$ ) DE DUAS POPULAÇÕES INFINITAS, OU FINITAS E AMOSTRAGEM COM REPOSIÇÃO, COM VARIÂNCIAS ( $\sigma_1^2$  e  $\sigma_2^2$ ) CONHECIDAS [AMOSTRAS GRANDES ( $n_1 \geq 30$  E  $n_2 \geq 30$ ) OU PEQUENAS ( $n_1 < 30$  E  $n_2 < 30$ )], OU COM VARIÂNCIAS DESCONHECIDAS [AMOSTRAS GRANDES ( $n_1 \geq 30$  E  $n_2 \geq 30$ ):** É o intervalo determinado por,

$$P\left[\left(\bar{X}_1 - \bar{X}_2\right) - z_{\left(\frac{\alpha}{2}\right)} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq \left(\bar{X}_1 - \bar{X}_2\right) + z_{\left(\frac{\alpha}{2}\right)} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right] = 1 - \alpha$$

**INTERVALO DE CONFIANÇA PARA A DIFERENÇA ENTRE AS MÉDIAS ( $\mu_1 - \mu_2$ ) DE DUAS POPULAÇÕES INDEPENDENTES E APROXIMADAMENTE NORMAIS, COM VARIÂNCIAS ( $\sigma_1^2$  e  $\sigma_2^2$ ) DESCONHECIDAS E ESTATISTICAMENTE IGUAIS (HOMOCEDÁSTICAS) ( $\sigma_1^2 = \sigma_2^2 = \sigma^2$ ) E AMOSTRAS PEQUENAS E INDEPENDENTES ( $n_1 < 30$  E  $n_2 < 30$ ):** É o intervalo de confiança obtido pela expressão,

$$P\left[\left(\bar{X}_1 - \bar{X}_2\right) - t_{\left(v; \frac{\alpha}{2}\right)} S_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leq \mu_1 - \mu_2 \leq \left(\bar{X}_1 - \bar{X}_2\right) + t_{\left(v; \frac{\alpha}{2}\right)} S_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right] = 1 - \alpha$$

Em que os graus de liberdade da variável t são dados por  $v = n_1 + n_2 - 2$

Como não conhecemos  $\sigma^2$ , deveremos estimá-lo assim:

$$S_p^2 = \frac{(n_1 - 1) S_1^2 + (n_2 - 1) S_2^2}{n_1 + n_2 - 2}, \text{ Onde } S_p = \sqrt{S_p^2}$$

**INTERVALO DE CONFIANÇA PARA A DIFERENÇA ENTRE AS MÉDIAS PARAMÉTRICAS ( $\mu_1 - \mu_2$ ) DE DUAS POPULAÇÕES APROXIMADAMENTE NORMAIS, COM VARIÂNCIAS ( $\sigma_1^2$  e  $\sigma_2^2$ ) DESCONHECIDAS E ESTATISTICAMENTE DESIGUAIS ( $\sigma_1^2 \neq \sigma_2^2$ ) (HETEROCEDÁSTICAS) E COM AMOSTRAS ALEATÓRIAS PEQUENAS E INDEPENDENTES ( $n_1 < 30$  e  $n_2 < 30$ ):** É o intervalo obtido assim,

$$P\left[\left(\bar{X}_1 - \bar{X}_2\right) - t_{\left(v; \frac{\alpha}{2}\right)} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq \left(\bar{X}_1 - \bar{X}_2\right) + t_{\left(v; \frac{\alpha}{2}\right)} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right] = 1 - \alpha$$

Em que,  $v = \frac{(w_1 + w_2)^2}{\frac{w_1^2}{n_1 - 1} + \frac{w_2^2}{n_2 - 1}}$ , onde  $w_1 = \frac{S_1^2}{n_1}$  e  $w_2 = \frac{S_2^2}{n_2}$ .

**INTERVALO DE CONFIANÇA PARA A VARIÂNCIA ( $\sigma^2$ ) DE UMA POPULAÇÃO NORMAL:** É o intervalo fiducial determinado seguindo a expressão,

$$P\left(\frac{(n-1) S^2}{\chi_{sup}^2} \leq \sigma^2 \leq \frac{(n-1) S^2}{\chi_{inf}^2}\right) = 1 - \alpha$$

**Onde:**  $n$  é o tamanho ou o número de elementos da amostra e  $S^2 =$  Variância da amostra. Observação:  $v=n-1$  dá os graus de liberdade da distribuição Qui-quadrado ( $\chi^2$ ).

**INTERVALO DE CONFIANÇA PARA O DESVIO PADRÃO ( $\sigma$ ) DE UMA POPULAÇÃO NORMAL:**

Intervalo dado por,

$$P \left[ \sqrt{\frac{(n-1)S^2}{\chi_{\text{sup}}^2}} \leq \sigma \leq \sqrt{\frac{(n-1)S^2}{\chi_{\text{inf}}^2}} \right] = 1 - \alpha$$

$$\text{Onde: } \chi_{\text{inf}}^2 = \chi^2 \left( v; 1 - \frac{\alpha}{2} \right) \quad e \quad \chi_{\text{sup}}^2 = \chi^2 \left( v; \frac{\alpha}{2} \right)$$

n=número de elementos da amostra;  $S^2$  = Variância da amostra.Observação:  $v = n-1$  dá os graus de liberdade da distribuição Qui-quadrado ( $\chi^2$ ).**INTERVALO DE CONFIANÇA PARA A RAZÃO  $\left[ \frac{\sigma_1^2}{\sigma_2^2} \right]$  ENTRE AS VARIÂNCIAS OU QUADRADOS****MÉDIOS ( $\sigma_1^2$  e  $\sigma_2^2$ ), DE DUAS POPULAÇÕES NORMAIS:** Intervalo dado por,

$$P \left( \frac{S_1^2}{S_2^2} \cdot \frac{1}{F \left( v_1, v_2; \frac{\alpha}{2} \right)} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{S_1^2}{S_2^2} \cdot F \left( v_2, v_1; \frac{\alpha}{2} \right) \right) = 1 - \alpha$$

**Onde:**  $v_1 = n_1 - 1$  e  $v_2 = n_2 - 1$ .**INTERVALO DE CONFIANÇA PARA A PROPORÇÃO (P) DE UMA POPULAÇÃO INFINITA, NORMAL E COM AMOSTRA GRANDE ( $n \geq 30$ ):** Intervalo de confiança dado por,

$$P \left( f - z \left( \frac{\alpha}{2} \right) \sqrt{\frac{f(1-f)}{n}} \leq p \leq f + z \left( \frac{\alpha}{2} \right) \sqrt{\frac{f(1-f)}{n}} \right) = 1 - \alpha$$

**Onde:**  $f = \hat{p} = \frac{x}{n}$ ,  $f = \hat{p}$  = proporção ou frequência relativa de eventos sucesso na amostra; x=número

ou frequência absoluta de eventos sucesso na amostra; e n=número de elementos da amostra ou tamanho da amostra.

**INTERVALO DE CONFIANÇA PARA A DIFERENÇA ENTRE AS PROPORÇÕES OU PORCENTAGENS ( $P_1 - P_2$ ), DE DUAS POPULAÇÕES INFINITAS, NORMAIS, COM AMOSTRAS GRANDES ( $n_1 \geq 30$  e  $n_2 \geq 30$ ):** Intervalo fiducial determinado através da seguinte expressão,

$$P \left[ (\hat{p}_1 - \hat{p}_2) - z \left( \frac{\alpha}{2} \right) \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} \leq p_1 - p_2 \leq (\hat{p}_1 - \hat{p}_2) + z \left( \frac{\alpha}{2} \right) \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} \right] = 1 - \alpha$$

**Onde,**  $\hat{p}_1 = \frac{x_1}{n_1}$ ,  $\hat{p}_2 = \frac{x_2}{n_2}$ ,  $\hat{q}_1 = 1 - \hat{p}_1$ ,  $\hat{q}_2 = 1 - \hat{p}_2$ .

É importante lembrar através da seguinte observação que é necessário usar o fator de correção para população finita dado por (FCPF =  $\frac{N-n}{N-1}$ ), quando necessário. Por exemplo  $t_{\left(\frac{\alpha}{2}; v\right)} \frac{S}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$ , ou seja,

para multiplicar o valor do erro padrão do estimador que está sendo usado na construção do intervalo de confiança. Vale lembrar que a população infinita é aquela em que  $n < 0,05.N$  ou  $\frac{n}{N} < 0,05$  (ou 5%)

, e a população finita é aquela em que  $n \geq 0,05.N$  ou  $\frac{n}{N} \geq 0,05$  (ou 5%).

**INTERVALO DE CONFIANÇA:** É o intervalo em torno da estimativa por ponto, de sorte tal intervalo possua uma probabilidade conhecida de conter o verdadeiro valor do parâmetro. O valor da probabilidade, que normalmente assume os valores 90%, 95%, 98% dentre outros, é denominado nível de confiança.

**INTERVALO DE CLASSE:** Diferença entre o limite superior real e o limite inferior real de uma classe numa distribuição de frequências.

**INTERVALO SEMI-INTERQUARTIL:** Metade da diferença entre o terceiro e o primeiro quartil de uma distribuição de frequências. É, como o desvio-padrão e a amplitude, uma medida de dispersão.

**INTERVALO DE PREVISÃO:** O intervalo entre um conjunto de limites superior e inferior associado com um valor previsto designado para mostrar, em uma base de probabilidades, a faixa de erro associado com a previsão.

**INTERVALO DE TOLERÂNCIA:** Um intervalo que contém uma proporção especificada de uma população com um nível estabelecido de confiança.

**INFERÊNCIA:** Processo de raciocínio por meio de qual se derivam conclusões a partir de premissas. Inferir costuma ser utilizado como sinônimo de concluir. Há dois processos inferenciais básicos: a dedução e a indução.

**INFERÊNCIA ESTATÍSTICA:** Afirmação que se faz com base na teoria das probabilidades acerca da veracidade de um grande número de eventos cm base na observação de apenas uma parte dos mesmos. Fruto do processo de generalização, a inferência é uma conclusão válida para toda a população, embora tenha sido obtida tomando por base apenas uma amostra dessa população.

**INTERVALO DE CONFIANÇA:** Numa distribuição, refere-se a uma certa amplitude de valores que tem a probabilidade de conter o valor verdadeiro da população. Por exemplo, quando temos um conjunto de valores numa dada amostra, podemos estar interessados na média desses valores. Mas o que garante que esta média será representativa em relação a população estudada? Para resolver esse problema, ao invés de utilizar uma estimativa pontual da média da amostra que muito provavelmente não coincidirá com a média populacional real, podemos calcular certos limites inferiores e superiores para essa média, com certo grau de precisão denominado nível de confiança. O intervalo entre esses limites denomina-se intervalo de confiança.

**INDEPENDENTE (INDEPENDENT):** Dois eventos são ditos estatisticamente independentes se a probabilidade de um ocorrer não está relacionada de forma alguma à probabilidade de ocorrência do outro.

**INFERÊNCIA (INFERENCE):** Uma conclusão que, logicamente, surge de observações e premissas.

**INFERÊNCIA BAYESIANA (BAYESIAN INFERENCE):** Uma das três grandes estruturas de análises estatísticas. Com frequência, é chamada de probabilidade inversa. A inferência bayesiana é usada para estimar a probabilidade de uma hipótese estatística, de acordo ou condicionada pelos dados disponíveis. O resultado de uma inferência bayesiana é uma nova distribuição de probabilidades ou a posteriori. Comparar com as análises frequentista, paramétrica e de Monte Carlo.

**INFERÊNCIA FREQUENTISTA (FREQUENTIST INFERENCE):** Uma abordagem de inferência estatística na qual as probabilidades das observações são estimadas a partir de funções que assumem um número infinito de tentativas. Comparar com inferência bayesiana.

**INTERAÇÃO (INTERACTION):** Os efeitos conjuntos de dois ou mais fatores experimentais. Especificamente, a interação representa uma resposta que não pode ser predita apenas conhecendo o efeito principal de cada fator, de maneira isolada. As interações resultam em efeitos não aditivos. Ela é um importante componente dos experimentos multifatoriais.

**INTERCEPTO (INTERCEPT):** O ponto em que a linha de regressão cruza o eixo dos x's; ele é o valor esperado de Y quando X é igual a zero.

**INTERPOLAÇÃO (INTERPOLATION):** A estimativa de valores não observados dentro da gama de valores observados. Em geral, é mais confiável que a extrapolação.

**INCIDENTE CRÍTICO:** Procedimentos destinados à coleta de observações diretas do comportamento humano, de forma a facilitar sua utilidade potencial na solução de problemas práticos. Incidentes observados que possuem significado especial e que satisfazem critérios sistematicamente definidos. Qualquer atividade humana observável, suficientemente completa em si mesma, que permite inferências e previsões sobre a pessoa que executa a atividade. Frutífera abordagem num estudo exploratório formulativo. Incidente significa qualquer atividade humana observável que é suficientemente completa em si mesma para permitir inferências e previsões a serem feitas sobre a pessoa agente do ato. Para ser crítico, o incidente deve ocorrer numa situação em que a finalidade e intenção do ato parecem bastante claras ao observador, e em que suas conseqüências são suficientemente definidas de forma a não deixar dúvidas sobre seus efeitos.

**INCLINAÇÃO:** Distorção horizontal de uma distribuição de frequências, de modo que uma cauda do traçado é mais longa e contém mais observações do que a outra.

**INDEPENDÊNCIA:** No cálculo de probabilidades, independência é geralmente definida como referência ao princípio de probabilidades conjuntas. Dois eventos são independentes se a probabilidade de um for a mesma perante a certeza da ocorrência ou não do outro. Em inglês Independence.

**INDEPENDÊNCIA:** Em inglês independence.

**INDEPENDÊNCIA:** Suposição crítica da análise de variância univariada (ANOVA) ou da análise de variância multivariada (MANOVA) que requer que as medidas dependentes para cada respondente sejam totalmente não-correlacionadas com as respostas de outros respondentes na amostra. Uma falta de independência afeta severamente validade estatística da análise a menos que uma ação corretiva seja realizada.

**INDEPENDÊNCIA:** Dois eventos ou variáveis são independentes se o conhecimento de um não dá informação sobre o valor do outro, ou seja, as probabilidades de um não são afetadas pelas do outro, e vice-versa.

**INDEPENDÊNCIA CONDICIONAL:** Duas variáveis X e Y são condicionalmente independentes dada a variável Z se  $\Pr(X,Y|Z)=\Pr(X|Z)\Pr(Y|Z)$ . Em inglês Conditional independence.

**INDEPENDÊNCIA ENTRE VARIÁVEIS ALEATÓRIAS:** Duas variáveis aleatórias X e Y dizem-se independentes se os acontecimentos  $X \leq x$  e  $Y \leq y$  são independentes,  $\forall_{(x,y)} \in \mathfrak{R}^2$ . Ou ainda X e Y são variáveis aleatórias independentes se e somente se  $f(x_i, y_j) = f_1(x_i)f_2(y_j)$ ,  $\forall_{(x_i, y_j)}$ , se (X,Y) for uma variável aleatória bidimensional discreta, e  $f(x, y) = f_1(x)f_2(y)$ ,  $\forall_{(x,y)} \in \mathfrak{R}^2$  se (X,Y) for uma variável aleatória bidimensional contínua,

**INDETERMINÂNCIA FATORIAL:** Característica de análise de fatores comuns tal que diversos escores fatoriais diferentes podem ser calculados para um respondente, cada um se adequando ao modelo fatorial estimado. Isso significa que os escores fatoriais não são únicos para cada indivíduo.

**INTERSEÇÃO (INTERSECTION):** Na teoria de conjuntos, os elementos que são comuns a dois ou mais conjuntos. Na teoria de probabilidades, a de que dois eventos ocorram ao mesmo tempo. Em ambos os casos, a interseção é escrita com o símbolo  $\cap$ . Comparar com união.

**INTERVALO (INTERVAL):** Um conjunto de números com um mínimo e um máximo definidos.

**INTERVALO ABERTO (OPEN INTERVAL):** Um intervalo que não inclui seus extremos. Desta forma, o intervalo (0, 1) é o conjunto de todos os números entre 0 e 1, com exceção de 0 e 1. Comparar com intervalo fechado.

**INTERVALO DE CONFIANÇA (CONFIDENCE INTERVAL):** O intervalo que inclui a verdadeira média da população um determinado número de vezes. O termo intervalo de confiança sempre deve ser classificado em n %, como em intervalo de confiança de 95 %, intervalo de confiança de 50 %, etc. Os intervalos de confiança são calculados usando estatística frequentista. Comumente são mal interpretados como representando uma chance de número de que a média esteja dentro do intervalo de confiança. Comparar com intervalo de credibilidade.

**INTERVALO DE CREDIBILIDADE (CREDIBILITY INTERVAL):** Um intervalo que representa a probabilidade de  $(1 - \alpha)\%$  de que os limites deste intervalo contenha o verdadeiro valor da média da população. Como no intervalo de confiança, o de credibilidade sempre deve ser classificado: intervalo de credibilidade de 95 %, intervalo de credibilidade de 50 %, etc. Esses intervalos são determinados conforme a estatística bayesiana. Comparar com intervalo de confiança.

**INTERVALO FECHADO (CLOSED INTERVAL):** Um intervalo que inclui seus pontos finais. Assim, o intervalo fechado [0, 1] é o conjunto de todos os números entre 0 e 1, incluindo 0 e 1. Comparar com intervalo aberto.

**INTERVALO DE PREDIÇÃO INVERSA (INVERSE PREDICTION INTERVAL):** Fundamentado em um modelo de regressão, é a amplitude de valores possíveis do X (ou preditor) para um dado valor do Y (ou resposta).

**INDICADOR:** Mensuração do padrão de vida, ou nível de vida, da população com o qual esta trabalhando.

**INDICADOR:** Variável única utilizada em conjunção com uma ou mais variáveis diferentes para formar uma medida composta.



**INDICADOR:** Variável simples usada em conjunção com uma ou mais variáveis distintas para formar uma medida composta.

**INDICADOR:** Valor observado ou variável manifesta usado como uma medida de um conceito ou construto latente que não pode ser medido diretamente. O pesquisador deve especificar quais indicadores são associados com cada construto latente.

**INDICADORES DE SAÚDE:** São parâmetros utilizados internacionalmente com o objetivo de avaliar, sob o ponto de vista sanitária, a higidez de agregados humanos, bem como fornecer subsídios aos planejamentos de saúde, permitindo acompanhamento das flutuações e tendências históricas do padrão sanitário de diferentes coletividades consideradas à mesma época ou da mesma coletividade em diversos períodos de tempo.

**INDICADORES PRINCIPAIS:** Índice que consiste de 12 séries temporais econômicas diferentes que apresentam tendência ascendente ou descendente.

**INTERVALO DE PREDIÇÃO SIMULTÂNEA (SIMULTANEOUS PREDICTION INTERVAL):** Um ajuste em um intervalo de confiança de análise de regressão que é necessário caso seja feita uma estimativa de múltiplos valores.

**ÍNDICE:** É um quociente, no qual o que está expresso no denominador não está sujeito ao risco de vir a apresentar o evento que está expresso no numerador. Exemplo: Telefones/habitantes.

**ÍNDICE:** Tipo de medida composta que resume várias observações específicas e representa alguma dimensão mais geral. Contrastada com escala.

**ÍNDICE:** Número abstrato que resulta da redução das expressões numéricas de fatores diferentes a uma única expressão capaz e representar a estrutura peculiar formada pro aqueles.

**ÍNDICE:** A relação entre frequências atribuídas de determinado evento, citadas na mesma unidade, sendo que no numerador é registrada a frequência absoluta do evento que constitui subconjunto daquela que é registrada no denominador, de caráter mais abrangente. O índice é geralmente apresentado sob forma percentual. Sinônimo: proporção.

**ÍNDICE:** Pequeno número ou letra escrito próximo e ligeiramente abaixo de outra letra.

**ÍNDICE DE AJUSTE:** Índice de correlação quadrada ( $R^2$ ) que pode ser interpretado como indicativo da proporção de variância das disparidades de dados otimamente escalonados que pode ser explicada pelo procedimento de escalonamento multidimensional (MDS-Multidimensional Scaling). Ele mede o quão bem os dados iniciais se ajustam ao modelo MDS. Esse índice é uma alternativa para a medida de desajuste para determinar o número de dimensões. Semelhante a medidas de covariância em outras técnicas multivariadas, medidas de 0,60 ou mais são consideradas aceitáveis.

**ÍNDICE DE AMODALIDADE:** Corresponde ao valor particular  $X_a$ , elemento de dado conjunto  $X$ , é o valor da variável reduzida  $Z = \frac{(X_a - \bar{X})}{S_x}$ , onde  $\bar{X}$  denota a média aritmética e  $S_x$  o desvio padrão ou afastamento padrão.

**ÍNDICE DE CANIBALIZAÇÃO:** A extensão pela qual um novo produto rouba a linha de produtos existentes de uma empresa.

**ÍNDICE DE CONDIÇÃO:** Medida da quantia relativa de variância associada com um autovalor, de modo que um grande índice de condição indica um grau elevado de colinearidade.

**ÍNDICE DE REDUNDÂNCIA:** Quantidade de variância em uma variável estatística canônica, dependente ou independente, explicada pela outra variável estatística canônica na função canônica. Pode ser computado tanto para variáveis estatísticas canônicas dependentes quanto independentes em cada função canônica. Por exemplo, um índice de redundância da variável estatística dependente representa a quantidade de variância nas variáveis dependentes explicada pela variável estatística canônica independente.

**ÍNDICE DE RESPOSTA:** Porcentagem dos entrevistados em potencial contatados inicialmente que preencheram de fato o questionário.

**ÍNDICE DE SÍTIO:** Descrição quantitativa do sítio de um povoamento florestal, geralmente referida como a altura média das árvores dominantes numa idade de referência (idade índice ou idade base). A definição do índice de sítio é geralmente realizada a partir das curvas de índice de sítio, ou seja, de um conjunto de curvas do crescimento da altura média das dominantes em função da idade, onde cada curva representa o crescimento esperado num dado sítio.

**ÍNDICE DE SWAROOP & UEMURA:** Índice que significa a percentagem de pessoas que morreram com 50 anos ou mais em relação ao total de óbitos ocorridos numa determinada população. É calculado dividindo-se o número de óbitos de pessoas que faleceram com 50 ou mais anos de idade pelo total de óbitos, multiplicando-se por 100. Sinônimo: razão de mortalidade proporcional.

**ÍNDICE POTÊNCIA:** Medida composta do poder discriminatório de uma variável independente quando mais de uma função discriminante é estimada. Baseado em cargas discriminantes, é uma medida relativa usada para comparar a discriminação geral dada por cada variável independente em todas as funções discriminantes significantes.

**ÍNDICE DE SHANNON-WIENER:** índice de diversidade usado em ecologia, nos quais a contribuição de cada espécie é pesada por sua abundância relativa, que significa a proporção do número total de indivíduos numa comunidade que pertence aquela espécie. Esses índices são calculados a partir das proporções de cada espécie ( $p_i$ ) na amostra total de indivíduos. O cálculo é feito conforme a seguinte equação. Conhecido também como índice de Shannon. É uma medida de variação genética num locus específico. Se  $p_1, p_2, \dots, p_k$  representam as verdadeiras frequências populacionais dos  $K$  alelos em um locus, a diversidade genética nesse locus é.

$$h_s = H = H' = -\sum_{i=1}^k p_i \log_e p_i = -\sum_{i=1}^S p_i \log_e p_i$$

Onde  $n_i$  é o número dos indivíduos em cada espécie; a abundância de cada espécie,  $S$  é o número de espécies, chamado também de riqueza,  $N$  é o número total de todos os indivíduos, ou seja,  $N = \sum_{i=1}^S n_i$ ,  $p_i$  é a abundância relativa de cada espécie, calculada pela proporção dos indivíduos de uma espécie pelo número total dos indivíduos na comunidade, ou seja,  $\frac{n_i}{N}$ . Ao se aplicar o cálculo por meio da fórmula

acima, que pode ser demonstrado para qualquer dado número de espécies, pode se obter um valor máximo possível  $H' = H_{\max} = \ln S = \log_e S$  o qual ocorre quando todas as espécies que estão presentes ocorrem

em igual número. Isto é, de fato, a probabilidade de que dois genes, aleatoriamente escolhidos da população, sejam dissimilares no seu perfil alélico. Sua estimativa amostral,  $\hat{h}$ , é obtida pela substituição dos  $p_i$ 's por seus respectivos equivalentes amostrais,  $\hat{p}_i$ 's.

**ÍNDICE DE SIMPSON:** Índice de diversidade usado em ecologia, nos quais a contribuição de cada espécie é pesada por sua abundância relativa, que significa a proporção do número total de indivíduos numa comunidade que pertence àquela espécie. Esses índices são calculados a partir das proporções de cada espécie ( $p_i$ ) na amostra total de indivíduos. Além de ter o seu uso em estudos ecológicos esse coeficiente tem sua origem também no conceito de funções de entropia, na teoria de informação e física teórica. Tem sido também utilizado no contexto de estudos evolucionários. É importante destacar que estudos evolucionários apontam algumas desvantagens dessa medida. Primeiramente, seu sentido biológico é difuso. Em segundo lugar, mesmo sendo seu valor mínimo zero para um locus fixo, ele pode atingir valores muito grandes, ou seja, quando cada um dos  $K$  alelos é igualmente frequente na população,  $h$  ou  $h_s$  atinge seu valor máximo,  $\log_e K$ . Sua distribuição amostral depende da verdadeira distribuição de frequências. Para loci independentes, pela propriedade de aditividade dos logaritmos dos produtos, os  $h_s$ 's medidos em cada locus podem ser ponderados para obter a diversidade média por locus dentro da população. O cálculo é feito conforme a seguinte equação:

$$h = D = \frac{1}{\sum_{i=1}^k p_i^2} = 1 - \sum_{i=1}^k p_i^2$$

**ÍNDICES DE DIVERSIDADE (SHANNON E OUTROS):** Destinam-se a determinar a variabilidade de dados mensurados em escala nominal, ou seja, a distribuição de observações entre categorias.

**ÍNDICES DE MODIFICAÇÃO:** Valores calculados para cada relação não estimada possível em um modelo especificado. O Q índice de modificação para uma relação específica não estimada indica o melhoramento no ajuste geral do modelo, ou seja, a redução na estatística qui-quadrado que é possível se um coeficiente é calculado para aquela relação não testada. O pesquisador deve usar índices de modificação apenas como uma orientação para melhoramentos de modelagem daquelas relações que podem ser teoricamente justificadas como possíveis modificações.

**ÍNDICES DE MORAES:** São os índices de mortalidade proporcional dos grupos etários (em anos): <I, 1-4, 5-19, 20-49 e <50. Calcula-se tendo o denominador sempre constituído pelo total de óbitos ocorrido em certo período de tempo em uma determinada região e o numerador formado pelo número de óbitos para cada grupo etário específico.

**ÍNDICE DE PREÇO AO PRODUTOR (IPP):** Medida do nível de preços baseada em um cesta básica fixa de mercadorias utilizadas pelos produtores.

**ÍNDICES DE PREÇOS:** Uma medida do nível de preços.

**ÍNDICES DE PREÇOS AO CONSUMIDOR (IPC):** Medida do nível médio de preços em dada época, baseada no preço de aquisição de uma cesta básica para o consumidor.

**ÍNDICES DE VARIAÇÃO ESTACIONAL:** São proporções estatísticas, análogas aos números índices, mas que se calculam de mês para mês e se usam para o estudo sintético das variações estacionais da marcha global de um conjunto de fenômenos.

**INDIVÍDUO:** É igual unidade estatística.

**INDUÇÃO:** Modelo lógico em que princípios gerais são desenvolvidos a partir de observações específicas. Constatando que católicos votam mais provavelmente no partido do governo do que protestantes, pode-se concluir que minorias religiosas no Brasil se filiam mais ao partido de oposição do que os católicos e explicar por quê. Este é um exemplo de indução.

**INDUÇÃO:** Processo mental por intermédio do qual, partindo de dados particulares, suficientemente constatados, infere-se uma verdade geral ou universal, não contida nas partes examinadas.

**INDUÇÃO:** O método fundamental da ciência é generalização ou indução. O processo usado para fazer inferências sobre o todo de uma classe a partir de observações feitas em alguns dos seus membros. A abordagem indutiva leva a conclusões gerais baseadas em casos específicos. É de natureza experimental.

**INFERÊNCIA BAYESIANA:** Uma forma de inferência que trata parâmetros como variáveis aleatórias que possuem distribuições a priori, refletindo o estado acumulado de conhecimento. Por meio do teorema de Bayes são feitas atualizações das distribuições dos parâmetros. Em inglês Bayesian inference.

**INFERÊNCIA CAUSAL:** Processo de raciocínio que induz a formulação de afirmativas sobre a relação entre variáveis.

**INFERÊNCIA ESTATÍSTICA:** A utilização de amostras de uma população com o objetivo de tomar decisões sobre a população. Em inglês inferential statistics.

**INFERÊNCIA ESTATÍSTICA:** Processo que consiste em utilizar as observações de uma amostra para estimar as propriedades da população.

**INFERÊNCIA ESTATÍSTICA:** É procurar conclusões sobre um grande número de eventos com base na observação de apenas uma parte deles.

**INFERÊNCIA ESTATÍSTICA:** Processo de fazer afirmativas sobre a população a partir dos dados de amostra; envolve o uso de teste de significância, a determinação do erro amostral ou a construção de intervalo de confiança.

**INFERÊNCIA ESTATÍSTICA:** É a técnica usada para estimar parâmetros e comportamento da população por meio de dados de uma amostra dessa população.

**INFERÊNCIA ESTATÍSTICA:** A generalização dos resultados de estudos em uma amostra que represente a verdade para uma população maior. Inferência estatística de respeito só à variação ou erro aleatória na amostragem.

**INFERÊNCIA ESTATÍSTICA:** Método indutivo, a partir do qual os resultados amostrais são generalizados para a população. Infere-se, portanto, do particular que é a amostra para o geral ou seja a população.

**INFERÊNCIA ESTATÍSTICA:** Método indutivo de generalização dos valores numéricos amostrais para a população de onde os dados foram retirados. As generalizações estatísticas, diferentes das leis universais, admitem exceções, mas proporcionam conhecimentos de relevância em termos científicos.

**INFERÊNCIA FIDUCIAL:** Tipo de inferência estatística baseada na distribuição fiducial, introduzida por Fisher (1930). O objetivo da inferência é fazer afirmações probabilísticas sobre os valores de parâmetros desconhecidos. Em inglês Fiducial inference.

**INFERIR:** Significa deduzir como consequência, conclusão ou probabilidade.

**INFINITESIMAL:** É uma quantidade arbitrariamente pequena que os matemáticos antigos precisavam utilizar para incorporar nas suas teorias pela falta da teoria dos limites. Cálculo infinitesimal é o cálculo diferencial e integral. Em inglês infinitesimal.

**INFINITO:** Símbolo matemático dado por  $\infty$ .

**INFORMAÇÃO:** Dado codificado.

**INFORMAÇÃO, ALFABETIZADO EM:** Aquele que conhece as técnicas e possui habilidades para usar instrumentos gerais e específicos de informação para delinear soluções a problemas.

**INFORMAÇÃO A PRIORI:** Informação inicial existente sobre algum parâmetro de interesse. Em inglês Prior information.

**INFORMAÇÃO CIENTÍFICA (CONCEITO DE RELEVÂNCIA):** Uma informação científica relevante pode ser interpretada como uma medida de um contato eficaz entre uma fonte e o usuário dela, num processo de comunicação, em que a informação é transmitida; o conceito é eficaz quando a informação transmitida resulta em acúmulo de conhecimento do usuário da informação.

**INFORMAÇÃO DE FISHER:** Seja  $X$  um vetor de variáveis aleatórias ou estocásticas com função de densidade  $p(x|\theta)$ . A medida de informação de Fisher esperada de  $\theta$  é definida como  $I(\theta) = E[-\partial^2 \log p(x|\theta) / \partial \theta^2]$ . Em inglês Fisher information.

**INFORMAÇÕES GEORREFERENCIADAS:** Ligação de atributos não gráficos ou dados geograficamente referenciados às informações gráficas de um mapa.

**INFORMAÇÕES MANIPULATIVAS:** Informações explicativas em um questionário que pretendem prover antecedentes e perspectiva necessários, mas em vez disso servem para influenciar o entrevistado.

**INFOSCAN:** Serviço de rastreamento baseado em scanner para produtos de consumo embalados.

**INQUÉRITO:** Todo inquérito é visto como direcionado à avaliação de três propriedades de afirmativas sobre fenômenos: i) suas verificabilidades empíricas por métodos aceitos; ii) sua consistência lógica com outros fatos aceitos ou conhecidos; iii) sua utilidade social. Muitos dos inquéritos disciplinados buscam avaliar cada propriedade em variados graus.

**INQUÉRITO AVALIATIVO:** É a determinação do valor de uma coisa. Em educação ele envolve a obtenção de informação para julgar o valor de um programa, produto ou procedimento.

**INQUÉRITO ELUCIDATÓRIO:** É o processo de obtenção de conhecimento generalizável, comparando e testando afirmativas sobre relacionamentos entre variáveis ou fenômenos generalizáveis. Esse conhecimento resulta em relacionamentos funcionais ou estatísticos, modelos e, por fim, teorias. Quando os resultados de inquérito elucidatório são combinados com conhecimento de circunstância particular, obtêm-se explicações.

**INQUIRIÇÃO:** Inquirir é perguntar, interrogar em busca de fatos relevantes. Entre os instrumentos de inquirição podem ser citados: o questionário e a entrevista. Sempre que possível a inquirição deve estar conjugada a cuidadosa observação.

**INSPEÇÃO AMOSTRAL:** É a que incide apenas sobre uma amostra da partida de material a ser julgada.

**INSPEÇÃO GRÁFICA:** Método para observar algumas qualidades estudadas baseado na observação de gráficos adequados. Em inglês Graphical inspection.

**INSPEÇÃO NORMAL:** É a que incide sobre o número de exemplares previsto no plano de inspeção.

**INSTALAÇÕES PARA GRUPOS DE FOCO:** Instalação provida de sala de conferências ou ambiente de sala de estar e uma sala de observação separada. A instalação também tem aparelhos de gravação de audiovisuais.

**INSTRUÇÕES AO ENTREVISTADOR:** Instruções explícitas ao encarregado da pesquisa a respeito de como aplicar e concluir adequadamente o questionário.

**INSTRUÇÕES PARA SUPERVISORES:** Orientações por escrito para o serviço de campo sobre como conduzir um levantamento.

**INSTRUMENTAÇÃO:** É o processo de selecionar ou desenvolver instrumentos de mensuração e métodos apropriados a um dado problema de avaliação. Duas questões: i) é o instrumento fidedigno? preciso, consistente, estável?; ii) é o instrumento válido? é relevante? realmente mede o que ele é feito para medir?

**INSTRUMENTAÇÃO (AMEAÇA À VALIDADE INTERNA):** Esta ameaça à validade interna de uma pesquisa ocorre quando se emprega instrumentos inconsistentes na coleta dos dados.

**INSTRUMENTO DE PESQUISA:** Material utilizado pelo pesquisador para colher dados para a pesquisa.

**INQUÉRITO POR ENTREVISTAS:** Tipo de inquérito transversal.

**INSTRUMENTO DE VARIAÇÃO:** Diferenças ou mudanças nos instrumentos de mensuração, tais como entrevistadores ou observadores que explicam diferenças nas mensurações.

**INSTRUMENTOS:** São recursos utilizados dentro de cada técnica com a finalidade de coleta de dados.

**INTEIRAMENTE CASUALIZADO:** Delineamento utilizado em experimentos inteiramente casualizados.

**INTELIGÊNCIA ARTIFICIAL (AI):** Área generalizada da ciência da computação que lida com a criação de programas de computador que tentam imitar as propriedades de aprendizado do cérebro humano. O estudo sobre inteligência artificial tem se estendido em várias áreas mais especializadas, incluindo redes neurais e algoritmos genéticos.

**INTENSIDADE DE AMOSTRAGEM:** Razão entre o número de unidades da amostra e o número total de unidades na população, que também pode ser expressa pela razão entre a área amostrada e a área total da população. Grandeza adimensional.

**INTENSIDADE SELETIVA:** Diferencial de seleção expresso em unidades de desvio padrão fenotípico da população.

**INTERAÇÃO:** Ver modificação de efeito.



**INTERAÇÃO:** Ação de um fator experimental sobre outro.

**INTERAÇÃO:** Em inglês interaction.

**INTERAÇÃO:** Em análise da variância de dois critérios, o efeito de quando um dos fatores varia segundo diferentes categorias do outro fator.

**INTEGRAL:** Símbolo matemático dado por  $\int$ .

**INTEGRAL:** Técnica de determinação matemática que é o contrário de derivada dado por  $\int_{-\infty}^{+\infty} f(x) dx$ , o qual serve para determinar áreas sob curvas ou polígonos regulares e irregulares.

**INTEGRAIS PADRÕES:**

$y = f(x)$	$\int f(x) dx$
$a$	$ax + C$
$x^a$ ( $a \neq -1$ )	$\frac{x^{a+1}}{(a+1)} + C$
$\sqrt{x}$	$\frac{2x^{\frac{3}{2}}}{3} + C$
$\frac{1}{x}$	$\log_e  x  + C$
$\frac{1}{(a+x)}$	$\log_e  (a+x)  + C$
$e^{ax}$	$\frac{e^{ax}}{a} + C$
$y = f(x)$	$\int f(x) dx$
$\text{Sen } x$	$-\text{Cos } x + C$
$\text{Cos } x$	$\text{Sen } x + C$

$\tan x$	$-\log_e  \cos x  + C$
$\frac{1}{x^2 + a^2}$	$\frac{1}{a} \tan^{-1} \left( \frac{x}{a} \right) + C$
$\frac{1}{\sqrt{a^2 + x^2}}$	$\text{sen}^{-1} \left( \frac{x}{a} \right) + C$
$\log_e x$	$x \log_e  x  - x + C$

**INTERAÇÃO:** Em forma geral, quando um número de indivíduos ou itens está agrupado de acordo com vários fatores de classificação e estes fatores não são independentes se diz que existe interação entre eles. Em inglês Interaction.

**INTERAÇÃO:** Ocorre interação quando os efeitos de duas ou mais variáveis, em uma análise de regressão ou dois ou mais fatores em uma análise de variância não são independentes de cada outra.

**INTERAÇÃO DESORDINAL:** Forma de efeito de interação entre variáveis independentes que invalida a interpretação dos efeitos principais dos tratamentos. Uma interação desordinal é mostrada graficamente fazendo-se o gráfico das médias para cada grupo e tendo a interseção de retas. Nesse tipo de interação, as diferenças médias não apenas variam, dadas as combinações únicas de níveis de variável independente, mas a ordenação relativa de grupos também muda.

**INTERAÇÃO NEGATIVA:** Aquela em que um fator diminui o efeito de outro, quando agem simultaneamente.

**INTERAÇÃO ORDINAL:** Tipo aceitável de efeito de interação no qual as magnitudes de diferenças entre grupos variam, mas as posições relativas dos grupos permanecem constantes. É graficamente representada com valores médios e observando-se retas não-paralelas que não se interceptam.

**INTERAÇÃO POSITIVA:** Aquela em que um fator aumenta o efeito do outro, quando agem simultaneamente.

**INTERCAMBILIDADE DE ÍNDICES:** Termo cunhado por Paul Lazarsfeld referido à proposição lógica de que se alguma variável, todos os indicadores da variável devem ter aquela relação.

**INTERCEPTO:** É uma constante relativa ao valor de Y nas equações de regressão quando o escore de X é igual a zero, sendo simbolizada pela letra a. Representa, portanto, o valor da variável dependente quando o valor da variável preditiva é igual a zero.

**INTERCEPTO ( $b_0$ ):** Valor no eixo Y ou eixo da variável dependente onde a reta definida pela equação de regressão  $Y = b_0 + b_1 X_1$  cruza o eixo. É descrito pelo termo constante  $b_0$  na equação de regressão. Além de seu papel na previsão, o intercepto pode ter uma interpretação gerencial. Se a completa ausência da variável independente tem significado, então o intercepto representa essa quantia. Por exemplo, quando se estimam vendas a partir de gastos ocorridos com propaganda, o intercepto representa o nível de vendas

esperadas se a propaganda for eliminada. Contudo, em muitos casos, a constante tem apenas valor preditivo, porque não há situação na qual todas as variáveis independentes estejam ausentes. Um exemplo é prever a preferência sobre um produto com base em atitudes de consumidores. Todos os indivíduos têm algum nível de atitude, e assim o intercepto não tem uso gerencial, mas ainda auxilia na previsão.

**INTERCEPTO (DA RETA DE REGRESSÃO  $Y = a + b * X$ ):** É o valor de Y, quando  $X = 0$ . A maioria das vezes não faz sentido interpretar este valor.

**INTERCEPTO VERTICAL:** Ponto em que uma reta corta o eixo vertical.

**INTERCEPTO Y:** Ponto em que reta corta o eixo dos y.

**INTERCONEXÃO:** Bloco de construção básico de redes neurais que podem atuar como entrada, saída ou função de processamento/análise. Análogo, em operação, a um neurônio no cérebro.

**INTERCONEXÃO VIESADA:** Interconexão adicional a uma camada oculta, que tem um valor constante. Essa interconexão funciona de uma maneira parecida com o termo constante da regressão múltipla.

**INTERCORRELAÇÃO:** Termo usado para denotar a correlação de um número de variáveis entre elas, distinto da correlação entre elas e uma variável exógena. Em inglês Intercorrelation.

**INTERPRETAÇÃO:** Termo técnico usado em conexão com o modelo de elaboração. Representa o resultado de pesquisa em que se descobre que uma variável de controle é o fator mediador por meio de qual uma variável independente produz seu efeito sobre a variável dependente.

**INTERPOLAÇÃO:** Procedimento para achar valores interpolados. Pode existir em diversos tipos de análise: de séries, de tempo, espacial, entre outros. Em inglês Interpolation.

**INTERPOLAÇÃO:** É o processo matemático de ajuste de uma função que pode determinar o valor de uma variável de interesse em um ponto não amostrado. A interpolação pode ser feita em  $\hat{A}_1$ ,  $\hat{A}_2$  ou  $\hat{A}_3$ , usando as funções apropriadas para esse fim.

**INTERPOLAÇÃO EXATA:** É uma característica particular de algumas funções matemáticas. Observe-se que os polinômios, inclusive as superfícies de tendência não têm essa característica. Nesta classe de funções encontram-se as equações multiquádricas, a krigagem ordinária e o inverso da potência da distância. Este último método tem sua solução forçada fazendo com que o ponto interpolado seja igual ao ponto observado, pois  $W_1$  é indefinido para distância nula.

**INTERPOLAÇÃO HARMÔNICA:** É aquela em que se usam as recíprocas dos números de graus de liberdade para armar a regra de três para a obtenção do valor crítico ou tabelado em tabelas estatísticas de testes de significância e outras situações.

**INTERPOLAÇÃO LINEAR:** É aquela em que não se usam as recíprocas dos números de graus de liberdade, mas sim o próprio número de graus de liberdade para armar a regra de três para a obtenção do valor crítico ou tabelado em tabelas estatísticas de testes de significância e outras situações.

**INTERSEÇÃO OU PRODUTO LÓGICO:** Símbolo representado por  $\cap$ .

**INTERSEÇÃO:** A interseção de dois conjuntos é o conjunto que consiste de todos os elementos que estão em ambos os conjuntos. Antigamente era conhecida como operação chapéu.

**INTERSEÇÃO DE A E B:** O evento que contém todos os pontos amostrais que estão tanto em A como em B. A interseção é denotada por  $\cap$ .

**INTERSUBJETIVO:** Caracterizado por acordo entre observadores. Quando afirmamos que algo é objetivamente verdadeiro, usualmente queremos dizer que observadores diferentes concordam que é verdade.

**INTERVALAR:** Nível de mensuração de dados; caracteriza dados que podem ser dispostos em ordem, e para os quais as diferenças entre valores são significativos.

**INTERVALO:** O valor máximo de uma variável menos o valor mínimo para essa variável.

**INTERVALO:** Diferença entre o maior e o menor valor de uma lista.

**INTERVALO DE AMOSTRAGEM:** Distância padrão entre elementos selecionados uma população numa amostra sistemática.

**INTERVALO DE AMOSTRAGEM:** É o espaçamento adotado para coleta de amostras em trincheiras, canaletas, furos de sondagem, dentre outros. Este espaçamento foi definido em função da variabilidade natural do depósito ou corpo de minério que se deseja amostrar. Observe que este parâmetro pode provocar o erro geométrico no cálculo de reservas, quando o depósito não foi adequadamente amostrado ou subamostrado.

**INTERVALO DE AMOSTRAGEM:** Na amostragem sistemática, é o número constante K de elementos tal que, depois do inicial, todo subsequente K-ésimo elemento, após o último escolhido, irá também fazer parte da amostra.

**INTERVALO DE CLASSE:** Diferença entre limite superior real e o limite real de uma classe numa distribuição de frequência. Em inglês class interval.

**INTERVALO DE CLASSE:** Diferença entre dois limites inferiores (li) ou superiores (ls) consecutivos. É o quociente entre a amplitude total é o número de classes.

**INTERVALO DE CONFIANÇA PARA UM PARÂMETRO:** É o estimador por intervalo do parâmetro populacional a ser estimado, o qual consiste na construção de um intervalo de confiança delimitado por dois limites fiduciais ou de confiança, os quais geralmente depende para a sua determinação de um valor crítico ou tabelado sob a curva da distribuição de probabilidade do estimador em questão, do tamanho da amostra retirada da população em estudo, bem como dos valores das estimativas ou estatísticas amostrais e do erro padrão do estimador, e que sob o qual define-se uma probabilidade de confiança de que este intervalo contenha o valor desconhecido do parâmetro populacional.

**INTERVALO DE CONFIANÇA:** Intervalo numérico que, com certa probabilidade como por exemplo, 95 %, 99 %, 99,9 %, dentre outros, encerra um parâmetro desconhecido.

**INTERVALO DE CONFIANÇA:** Definido pelos limites de confiança, é o intervalo dentro do qual espera-se encontrar a média populacional (m), com um dado grau de probabilidade, aproximada pela média amostral ou experimental ( $\bar{X}$ ):  $IC = \bar{X} \pm \frac{tS}{\sqrt{n}}$ .

**INTERVALO DE CONFIANÇA:** Estimativa de um intervalo de um parâmetro, construído de tal forma que o intervalo tem uma probabilidade de confiança pré-determinada de abranger o parâmetro.

**INTERVALO DE CONFIANÇA:** Intervalo baseado em observações de uma amostra e construído de maneira que tenha uma probabilidade especificada de conter o verdadeiro valor desconhecido de um parâmetro é costume utilizar intervalos que tenham 95 % e chance de conter o verdadeiro valor.

**INTERVALO DE CONFIANÇA:** Uma estimativa probabilística da verdadeira média ou proporção da população baseada nos dados da amostra. Representa a margem de erro, a qual indica o nível de precisão obtido na amostragem.

**INTERVALO DE CONFIANÇA:** É um intervalo centrado na estimativa pontual, cuja probabilidade de conter o verdadeiro valor do parâmetro é igual ao nível de confiança.  $\alpha$ : É a probabilidade de erro na estimação por intervalo.

**INTERVALO DE CONFIANÇA:** A estimativa de um parâmetro populacional por meio de um intervalo de valores ao invés de um único valor. Em inglês confidence interval.

**INTERVALO DE CONFIANÇA:** É um intervalo centrado na estimativa pontual, cuja probabilidade de conter o verdadeiro valor do parâmetro é igual ao nível de confiança.  $1-\alpha$ , e  $\alpha$  é a probabilidade de erro na estimação por intervalo.

**INTERVALO DE CONFIANÇA:** Se for possível definir duas estatísticas  $t_1$  e  $t_2$  que são as funções de valores da amostra somente, de forma que, sendo  $\theta$  o parâmetro a estimar,  $\Pr(t_1 \leq \theta \leq t_2) = \alpha$ , onde  $\alpha$  é o coeficiente de confiança, o intervalo entre  $t_1$  e  $t_2$  é chamado intervalo de confiança. Em inglês Confidence interval.

**INTERVALO DE CONFIANÇA:** A estimativa de um parâmetro populacional por meio de um intervalo de valores ao invés de um único valor. Se for possível definir duas estatísticas  $t_1$  e  $t_2$  que são funções de valores da amostra somente, de forma que, sendo  $\theta$  o parâmetro a estimar,  $\Pr(t_1 \leq \theta \leq t_2) = \alpha$ , onde  $\alpha$  é o coeficiente de confiança, o intervalo entre  $t_1$  e  $t_2$  é chamado intervalo de confiança.

**INTERVALO DE CONFIANÇA:** Faixa de valores dentro da qual se estima que um parâmetro de população vá cair. Um survey, por exemplo, pode mostrar que 40 % de uma amostra são a favor do candidato A. Embora a melhor estimativa apoio entre todos os eleitores também seja de 40 %, não se deve esperar que ele seja exatamente este. Podemos, pois, computar um intervalo de confiança, por exemplo, de 35 % a 45 %, dentro do qual está provavelmente o verdadeiro percentual da população. É necessário especificar um nível de confiança juntamente com cada intervalo de confiança.

**INTERVALO DE CONFIANÇA:** Conjunto ou intervalo, de valores usados para estimar um parâmetro populacional com um nível especificado de confiança; chamado também estimativa intervalar.

**INTERVALO DE CONFIANÇA:** Área abrangida por  $n$  erros padrões em torno da estimativa pontual, para mais e/ou para menos, e definidos em termos probabilísticos pela Regra Empírica e pelo Teorema do Limite Central. Os intervalos de confiança mais usados são de 95% e 99%, correspondentes, no caso do modelo da distribuição normal, a  $\bar{x} \pm 1.96\sigma_{\bar{x}}$  e  $\bar{X} \pm 2.58\sigma_{\bar{x}}$ , respectivamente.

**INTERVALO DE CONFIANÇA A  $i$  % PARA A MÉDIA AMOSTRAL:** O intervalo de confiança (IC) é um intervalo onde se encontra, com alguma  $i$  % de certeza, a média da população. O IC a 95 % é o mais usado em apresentação de resultados e calcula-se como  $\bar{X} \pm 2EP$  (média da amostra mais ou menos duas vezes o erro padrão). Se de uma população, com média  $\mu_e$  desvio padrão  $\sigma$ , se retirarem várias amostras, todas do mesmo tamanho  $n$ , e para cada amostra se calcular o intervalo de confiança a 95 %

para a sua respectiva média, 95 % desses intervalos vão conter a média da população. Quanto mais curto é o intervalo de confiança (IC), maior é a precisão da média amostral. É incorreto dizer que a média da população tem a probabilidade de estar no intervalo de confiança (IC), uma vez que a média da população não é uma variável.

**INTERVALO DE CONFIANÇA ASSINTÓTICO:** Intervalo de confiança construído a partir de um resultado assintótico. Do inglês Asymptotic confidence interval.

**INTERVALO DE GERAÇÃO:** Idade média dos pais quando a descendência destinada a substituí-los nasce. Uma geração representa a taxa média de renovação de um rebanho.

**INTERVALO DE PREDIÇÃO:** Estimativa intervalar de confiança para um valor predito de  $y$ .

**INTERVALO DE PREDIÇÃO:** Intervalo construído de modo que haja uma probabilidade específica de que certa variável esteja contida nele.

**INTERVALO INTERQUARTIL:** Metade da diferença entre o primeiro e o terceiro quartil.

**INTERVALO INTERQUARTIL:** Diferença entre o primeiro e o terceiro quartis.

**INTERVALO INTERQUARTIL:** Valor do terceiro quartil menos o valor do primeiro quartil; metade dos elementos de uma lista está no intervalo interquartil.

**INTERVALO SEMI-INTERQUARTILE:** Metade da diferença entre o terceiro e o primeiro quartis de uma distribuição de frequências. É como o desvio-padrão e a amplitude, uma medida de dispersão. Em inglês semi-interquartile range.

**INTRODUÇÃO:** É o primeiro capítulo de um relatório de pesquisa, onde o pesquisador irá apresentar, em linhas gerais, o que o leitor encontrará no corpo do texto. Por isso, apesar do nome Introdução, é a última parte a ser escrita pelo autor.

**INTUICIONISMO:** É um ramo da lógica que afirma que a matemática tem prioridade sobre a lógica, os objetos da matemática são construídos e operados na mente pelo matemático e é impossível definir as propriedades dos objetos matemáticos simplesmente estabelecendo um certo número de axiomas. Em particular, os intuicionistas rejeitam a lei do meio excluído que permite a prova por contradição. Em inglês intuitionism.

**INVARIÂNCIA:** Este termo é usado em estatística para denotar a propriedade de não sofrer alterações sob algumas transformações. Em inglês Invariance.

**INVERSA COMUM DA MATRIZ A ( $A^{-1}$ ):** Uma inversa de uma matriz quadrada  $A$ , com determinante diferente de zero indicada por  $A^{-1}$  onde lê-se inversa de  $A$ , se existir, é uma matriz quadrada única tal que  $AA^{-1} = A^{-1}A = I$ . Em que  $I$  é uma matriz identidade cuja ordem é igual à de  $A$ . Por exemplo,

$$A = \begin{bmatrix} 2 & 4 \\ 6 & 8 \end{bmatrix} A^{-1} = \begin{bmatrix} -1 & \frac{1}{2} \\ \frac{6}{8} & -\frac{1}{4} \end{bmatrix} AA^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$$



**INVERSA GENERALIZADA:** A inversa generalizada de uma matriz  $A$   $m \times n$ , é uma matriz  $n \times m$   $A^+$  de forma que para qualquer  $y$  onde  $Ax = y$  consistente,  $x = A^+y$  é uma solução. Em inglês Generalized inverse.

**INVERSO DA POTÊNCIA DA DISTÂNCIA:** É o método mais comum para interpolação de dados em  $\hat{A}_1$ ,  $\hat{A}_2$  ou  $\hat{A}_3$ , pois depende simplesmente das distâncias entre os pontos de dados (amostras) e o ponto a ser interpolado, conforme a seguinte equação:  $T = \sum_{i=1}^n W_i \cdot T_i$ , onde:  $W_i$  é o ponderador associado ao

$i$ -ésimo ponto de dado;  $T_i$  é o teor do  $i$ -ésimo ponto de dado. O ponderador  $W_i$  é igual a:  $W_i = \frac{1}{d_i^p}$ , onde:

$d_i$  é a distância entre o  $i$ -ésimo ponto de dado e o ponto a ser interpolado;  $p$  é a potência. A distância  $d_i$  em  $\hat{A}_2$  entre o ponto de dado  $(X_i, Y_i)$  e o ponto a ser estimado  $(X, Y)$ , pode ser calculado como:

$$d_i = \left[ (X_i - X)^2 + (Y_i - Y)^2 \right]^{\frac{1}{2}}$$

A distância  $d_i$  em  $\hat{A}_3$  entre o ponto de dado  $(X_i, Y_i, Z_i)$  e o ponto a ser

estimado  $(X, Y, Z)$ , pode ser calculado como:

$$d_i = \left[ (X_i - X)^2 + (Y_i - Y)^2 + (Z_i - Z)^2 \right]^{\frac{1}{2}}$$

**INVERSO DO QUADRADO DA DISTÂNCIA:** Ou simplesmente IQD é o nome dado para o método do inverso da potência da distância usando potência igual a dois. O IQD é o melhor interpolador, pois dá uma interpretação intermediária entre os princípios da variação gradual e dos pontos mais próximos.

**INVESTIGAÇÃO HUMANÍSTICA:** Método de pesquisa no qual o pesquisador fica imerso no sistema ou grupo sob estudo.

**IN VITRO:** É Um processo biológico ou bioquímico que acontece fora de um organismo vivo.

**IN VIVO:** É Um processo biológico ou bioquímico que acontece dentro de um organismo vivo.

**IP:** Intervalo de previsão.

**IPSIS LITTERIS:** Significa pelas mesmas letras, literalmente. Utiliza-se para expressar que o texto foi transcrito com fidelidade, mesmo que possa parecer estranho ou esteja reconhecidamente escrita com erros de linguagem.

**IPSIS VERBIS:** Significa pelas mesmas palavras, textualmente. Utiliza-se da mesma forma que ipsi litteris ou sic.

**INQUÉRITO (SURVEY):** Termo usado com o sentido amplo de investigação não-experimental, na qual os indivíduos da amostra são interrogados ou examinados para determinar o nível, a presença e a ausência das características de interesse.

**INQUÉRITO EPIDEMIOLÓGICO:** O estudo epidemiológico das condições de morbidade por causas específicas, efetuado em amostra representativa ou no todo de uma população definida e localizada no tempo e no espaço. Estudo levado a efeito quando as informações são inexistentes ou, se existentes, são inadequadas em virtude de diagnóstico deficiente, mudança de comportamento epidemiológico de determinadas, dificuldade na avaliação de cobertura ou eficácia vacinais dentre outros.

**INQUÉRITO SOROLÓGICO EM EPIDEMIOLOGIA:** Estudo epidemiológico ou atividade baseada na identificação, com base em testes sorológicos, de mudanças nos níveis de anticorpos específicos em uma população. Esse método permite não só a identificação de casos clínicos, mas, também, dos estados de portador e as infecções latentes ou subclínicas.

**INTERAÇÃO (OU MODIFICAÇÃO DO EFEITO):** Interdependência entre os dois ou mais fatores, de que decorre alteração da magnitude de um dado efeito; a combinação do efeito de duas variáveis difere da soma de seus efeitos separadamente: por exemplo, infecção e desnutrição interagem para aumentar o risco de mortalidade.

**INTERVALO DE CONFIANÇA:** Estipula o erro de amostragem; extensão de valores de uma variável, definida por dois limites, de modo a incluir o verdadeiro valor do parâmetro, na população em geral, com 95 % de chance; informa-se dois valores, um menor e outro maior, calculados com os dados da amostra, que definem os limites, no interior dos quais se encontra o verdadeiro valor do parâmetro populacional: por exemplo, prevalência de desnutrição de  $10 \pm 2$  %, ou seja, entre 8 % e 12 %. Ver erro-padrão.

**INVERSAMENTE PROPORCIONAL:** Símbolo matemático dado por  $(\propto^{-1})$ .

**INVESTIGAÇÃO EPIDEMIOLÓGICA:** Estudo realizado, particularmente no caso de doenças transmissíveis, a partir de casos clinicamente declarados ou mesmo de portadores, com a finalidade de detectar as fontes de infecção e permitir a adoção das medidas profiláticas mais adequadas. Não é um estudo amostral, sendo utilizado na investigação de casos, de óbitos ou surtos.

**ISÓCLINA:** Curva de igual inclinação. Em inglês isochrone.

**ISOMORFISMO:** Equivalência lógica de duas teorias, no sentido de uma teoria ser obtida a partir da outra pela tradução ou interpretação de noções e símbolos básicos. Em inglês Isomorphism.

**ISS:** Software estatístico inteligente. Em inglês Intelligent Statistical Software.

**ITERAÇÕES (RUNNS TEST):** Destina-se a verificar se a amostra em questões foi obtida de maneira aleatória. A técnica deste teste é baseada na ordem ou sequência com que os escores originais forem amostrais.

**INTERVALOS ABERTOS:** São aqueles onde os limites da classe inferior e superior não pertencem a ela.

**INTERVALOS FECHADOS:** São aqueles onde os limites de classe superior e inferior pertencem à classe em questão.

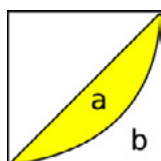
**INTERVALOS MISTOS:** São aqueles onde um dos limites pertence à classe, e o outro, não.

**ÍNDICE DE GINI (COEFICIENTE DE GINI):** É um cálculo usado para medir a desigualdade social, desenvolvido pelo estatístico italiano Corrado Gini em 1912. Esta estatística apresenta dados entre o número 0 e o número 1, onde zero corresponde a uma completa igualdade na renda, onde todos detêm a mesma renda per capita e o valor um que corresponde a uma completa desigualdade entre as rendas, onde um indivíduo ou uma pequena parcela de uma população, detêm toda a renda e os demais nada têm. Gini mede o coeficiente através de pontos percentuais que é igual ao coeficiente multiplicado por 100. Em uma linguagem mais simples, no resultado final, quanto mais um país se aproxima do número 1, mais desigual é a distribuição de renda e riqueza, e quanto mais próximo do número 0, mais igualitário será aquele país. Dados do PNUD (Plano das Nações Unidas para o Desenvolvimento) de 2010, pelo

índice de Gini, apontam o Brasil com o resultado de 0,56, sendo assim, o terceiro país mais desigual do mundo. O PNUD constatou, ainda no mesmo ano, que dos 15 países mais desiguais do mundo, segundo o índice de Gini do mundo, 10 se encontram na América Latina e no Caribe. O Coeficiente de Gini (G) é uma medida estatística de desigualdade, muito usada para indicar o grau de concentração de renda de uma região. Seu cálculo é baseado na Curva de Lorenz. No eixo X dispõem-se os percentuais acumulados da população, sempre em ordem crescente de renda, e no eixo Y os percentuais acumulados da renda. O cálculo do coeficiente de Gini é simples: divide-se a área de concentração pela área de perfeita desigualdade, ou seja, pela área do triângulo situado abaixo da linha de perfeita igualdade:  $G = \frac{\text{área de concentração}}{\text{área de perfeita desigualdade}}$ . Se não há concentração, o numerador é zero, e o coeficiente de Gini

resulta também em zero. Se a concentração é máxima, teremos o numerador igual ao denominador, e o coeficiente assume valor um, sendo então:  $0 \leq G \leq 1$ . Na representação gráfica do coeficiente de Gini o eixo horizontal representa o rendimento, e o eixo vertical, a quantidade de pessoas. A diagonal representa a igualdade perfeita de rendimento, e a área amarela é o coeficiente de Gini. A curva que delimita o coeficiente denomina-se curva de Lorenz. O coeficiente de Gini calcula-se como uma razão das áreas no diagrama da curva de Lorenz. Se a área entre a linha de perfeita igualdade e a curva de Lorenz é a, e a área abaixo da curva de Lorenz é b, então o coeficiente de Gini é igual a  $\frac{a}{(a+b)}$ . Esta razão se expressa como

percentagem ou como equivalente numérico dessa percentagem, que é sempre um número entre zero e um.



$I(\theta)$ : Informação de Fisher.

$\text{Int}[x]$ : Parte inteira de x.

# J

**JACKKNIFE:** Método proposto por Tukey (1958), que reduz o vício de estimação e fornece intervalos de confiança nos casos em que a teoria de distribuição ordinária não o permite. Este método é baseado na ideia que se  $t_n$  é uma estatística viciada na ordem de  $\frac{1}{n}$ ,  $t_{n-1}$  é a média das  $n$  estatísticas derivadas omitindo

um dos membros da amostra, então,  $n t_n - (n-1)t_{n-1}$  tem vício da ordem  $\frac{1}{n^2}$ . Em inglês Jackknife.

**JACOBIANA:** Uma matriz de derivadas parciais que é usada para determinar a distribuição de variáveis aleatórias transformadas.

**JACKKNIFE:** Método de reamostragem sem reposição.

**JACKKNIFING:** Um procedimento intensivo de computador que envolve parâmetros estimados e seus intervalos de confiança. Cada um dos  $n$  indivíduos do estudo é removido por vez e os parâmetros estimados da amostra de interesse incluindo os  $n-1$  indivíduos remanescentes. A distribuição dessas  $n$  estimativas é utilizada para obter uma estimativa geral e o intervalo de confiança para o parâmetro.

**JACKKNIFE (JACKKNIFE):** Uma ferramenta para vários propósitos. Na estatística, o jackknife é um processo de aleatorização ou Monte Carlo no qual uma observação é excluída do conjunto de dados, e a estatística é então recalculada. Isto é repetido tantas vezes quanto ha de observações. A estatística teste, calculada a partir dos dados originais, é comparada com a distribuição da estatística teste obtida pela repetida aplicação do jackknife, fornecendo um valor de  $P$  exato. Ver também bootstrap e função de influência.

**JACKKNIFE:** Procedimento para obter repetidas amostras baseado na omissão de uma observação quando se cria cada amostra. Viabiliza a estimação empírica de níveis de confiança de parâmetros ao invés de estimação paramétrica.

**JANELA DE PESQUISA (PARA CÁLCULO DE VARIOGRAMAS EXPERIMENTAIS):** É o domínio dentro do qual são pesquisados os pontos para cálculo de variogramas experimentais para dados dispersos ou não dispostos em malha regular. Esta janela é definida pela tolerância angular e pela tolerância na abertura.

**JOGO DE SOMA NULA:** É dito o jogo em que a soma algébrica dos lucros (perda = -lucro) de todos os que nele participam é sempre nula, não havendo produção nem destruição de bens. É da terminologia usada por Von Neumann e Morgenstern (1944).

**JUROS:** Percentagem de um capital dado em empréstimo, que é paga periodicamente como anualmente, mensalmente, dentre outros como recompensa pela utilização do dinheiro. Em inglês interest.

**JUSTIFICATIVA:** É a parte mais importante de um projeto de pesquisa já que é nesta parte que se formularão todas as intenções do autor da pesquisa. A justificativa num projeto de pesquisa, como o próprio nome indica, é o convencimento de que o trabalho de pesquisa é fundamental de ser efetivado. O tema



escolhido pelo pesquisador e a hipótese levantada são de suma importância, para a sociedade ou para alguns indivíduos, de ser comprovada. Deve-se tomar o cuidado, na elaboração da justificativa, de não se tentar justificar a hipótese levantada, ou seja: tentar responder ou concluir o que vai ser buscado no trabalho de pesquisa. A justificativa exalta a importância do tema a ser estudado, ou justifica a necessidade imperiosa de se levar a efeito tal empreendimento.

# K

**KOLMOGOROV-SMIRNOV:** Teste estatístico não paramétrico, de aderência ou de independência, para uma ou duas amostras, respectivamente. O teste de aderência é utilizado também para testar a normalidade dos escores amostrais.

**KEYWORD:** Palavra-chave, em inglês, utilizada para a indexação de uma obra acadêmico-científica.

**KOLMOGOROV-SMIRNOV:** Ver teste Kolmogorov-Smirnov.

**KURTOSE:** Perfil de alongamento que uma distribuição de dados apresenta. As distribuições em que os valores se agrupam basicamente no centro da curva são denominadas leptocúrticas. As distribuições mais achatadas, com os valores distribuídos de maneira mais uniforme e com caudas mais alongadas do que uma distribuição normal, são chamadas de platicúrticas. As distribuições intermediárias, nem tão agudas nem tão achatadas denominam-se mesocúrticas, como por exemplo, a distribuição normal é mesocúrtica. A kurtose pode ser expressa graficamente ou numericamente, como nos exemplos a seguir: valores negativos indicam uma distribuição com kurtose platicúrtica e valores positivos indicam uma tendência leptocúrtica. O valor zero representa a distribuição normal que é mesocúrtica.

**KRIGAGEM:** Técnica para estimar um efeito puramente espacial numa região meio de estimadores lineares cujos pesos são calculados de forma a minimizar a covariância espacial. Em inglês Kriging.

**KRIGAGEM:** É o nome genérico dado ao processo de estimativa baseado na Teoria das Variáveis Regionalizadas. O nome é homenagem a Danie Gerhardus Krige, engenheiro de minas sul africano que foi o pioneiro no uso de médias móveis para avaliação dos depósitos de ouro do Rand.

**KRIGAGEM DE BLOCO:** É uma feição original da krigagem, em que é possível avaliar o teor médio de um bloco e não somente de um ponto no centro do mesmo. A krigagem de bloco equivale a realização de uma série de krigagens pontuais nos centros dos sub-blocos e calcular a média dos teores dos sub-blocos, ou então ao cálculo do teor médio usando os ponderadores médios da krigagem de bloco, conforme o Teorema da Combinação das Estimativas de Krigagem.

**KRIGAGEM DISJUNTIVA:** Ao contrário dos demais métodos de krigagem, este não calcula os ponderadores que poderiam ser aplicados na equação da média ponderada, mas a estimativa é feita por meio da projeção do valor desconhecido  $Z(x_0)$  no espaço vetorial intermediário  $D_n$  gerado pela soma das  $n$  funções  $f(Z)$ :  $Z_{DK}^* = \sum_{\alpha=1}^n f_{\alpha}(Z_{\alpha})$ . A krigagem disjuntiva é por definição o processo de projeção da valor

desconhecido  $Z(x_0)$  em  $D_n$ .

**KRIGAGEM INDICADORA:** É uma adaptação da técnica da krigagem ordinária, agora não para estimativa de teores, mas sim para obtenção de uma distribuição de probabilidade acumulada numa dada localização, usando os pesos resultantes da krigagem ordinária. A grande diferença nesta estimativa é que a distribuição espacial das amostras é levada em consideração e, portanto, não é utilizado o mesmo peso para todas as amostras vizinhas. Assim, a krigagem indicadora tem por objetivo estimar a função

probabilidade acumulada  $\Phi^*(A; z)$ , que dá a exata proporção de teores  $Z(x)$  abaixo do teor de corte  $z$ , dentro de qualquer área  $A$ , conforme a seguinte expressão:  $\Phi^*(A; Z) = \sum_{j=1}^n \lambda_j \cdot i(X_j; Z)$ , com  $x_j \in A$ .

**KRIGAGEM DA MÉDIA:** Permite obter a média de um conjunto de pontos de dados, ponderando-os adequadamente segundo as relações espaciais reconhecidas no matriz das covariâncias. O estimador da média. é:  $m^* = \sum_{\alpha=1}^n \lambda_{\alpha} \cdot Z(X_{\alpha})$ .

**KRIGAGEM ORDINÁRIA:** É a forma mais simples de krigagem e deve responder satisfatoriamente à maioria dos problemas de estimativa.

**KRIGAGEM PONTUAL:** Refere-se à avaliação de um ponto por meio da krigagem ordinária, a partir das informações de pontos vizinhos, cujas relações espaciais são levadas em consideração usando as informações estruturais do variograma modelado.

**KRIGAGEM TRANSITIVA:** É a estimativa feita com base na informação estrutural da função de autocorrelação. De formulação bastante simples, é indicada para casos em que a variável regionalizada não é estacionária, contudo não é comum encontrar referência a sua utilização na literatura corrente em geoestatística aplicada.

**KRIGAGEM UNIVERSAL:** É a forma prevista originalmente para todos os casos em que a variável regionalizada não fosse estacionária. Este método é baseado inicialmente no ajuste de polinômios de baixo grau (1, 2 ou 3), que se encarregariam de remover a componente de deriva da variável regionalizada, e com os resíduos obter os variogramas que agora podem ser modelados, pois, em princípio, deveriam apresentar-se estacionários. A estimativa pela krigagem universal é feita por meio da equação da média ponderada, entretanto, os ponderadores são calculados simultaneamente com os coeficientes dos polinômios, o que exige maior tempo de computação.

**KRUSKALL-WALLIS:** Nome dado ao teste não paramétrico de Kruskall- Wallis. Em inglês Kruskall-Wallis.

**KURTOSE:** Um termo usado para descrever a extensão pela qual a curva de frequência unimodal é pontiaguda; ou seja, a extensão relativa do declive na vizinhança da moda. O termo foi introduzido por Pearson (1906) que propôs a razão de momentos  $\frac{\mu_4}{\mu_2^2}$  como uma medida de kurtose. É duvidoso, entretanto, se

qualquer razão única pode medir adequadamente a kurtose. Em inglês Kurtosis.



**LAMBDA:** Uma medida de associação para duas variáveis, das quais pelo menos uma é nominal em escala.

**LAMBDA:** Medida de associação apropriada a variáveis nominais.

**LAMBDA DE WILKS:** É um teste multivariado para verificar diferenças de médias para os casos de múltiplas variáveis dependentes intervalares e múltiplos grupos ou seja, mais de dois formados pelas variáveis independentes. O teste t, o T de Hotelling e o teste F são casos especiais deste teste. Em inglês Wilks's lambda.

**LAMBDA DE WILKS:** Uma das quatro principais estatísticas para testar a hipótese nula em análise de variância multivariada a chamada MANOVA. Também chamado de critério de máxima verossimilhança ou estatística U.

**LAMBDAGRAMA:** De uma série cronológica  $x_t$  ( $t = 1, 2, \dots, n$ ), cujo coeficiente de correlação serial de ordem  $J$  é  $r_j$ , é o diagrama que se obtém, marcando-se, para a abscissa  $n$  a variável, que é o valor do coeficiente  $\lambda_n = \frac{n-1}{n} \sum_{j=1}^n r_j$  o qual mede a divergência entre essa sequência de amostras de  $x$  e uma que fosse

constituída de elementos retirados ao acaso da respectiva população.

**LATÊNCIA:** Período na evolução clínica de uma doença parasitária, no qual os sintomas desaparecem apesar de estar o hospedeiro ainda infectado e de já ter sofrido o ataque primário, ou uma ou várias recaídas. Terminologia utilizada com frequência em relação à malária.

**LÁTICE:** O mesmo que reticulado.

**LAVAGEM DE DADOS:** Ver limpeza de dados.

**LEI:** Em teoria, uma generalização universal sobre um conjunto de fatos, por vezes também chamado de princípio. Um exemplo é a lei da gravidade.

**LEI CIENTÍFICA:** São hipóteses que foram confirmadas por um grande número de pesquisas. O termo lei comumente se refere a um relacionamento regular e predizível entre variáveis. Uma lei científica não deve se pensar como possuidora de uma existência independente em que algum cientista teve sorte em tropeçar. Uma lei científica não é parte da natureza; é apenas um meio de compreender a natureza.

**LEI DA RECIPROCIDADE QUADRÁTICA:** Fornece as condições sob as quais um número primo  $p$  é um resíduo quadrático módulo primo  $q$  em termos de  $q$  ser ou não um resíduo quadrático módulo  $p$ . Em inglês law of quadratic reciprocity.

**LEI DAS PROVAS REPETIDAS:** É o enunciado diferente da distribuição binomial a saber: a probabilidade  $P$  que um evento de probabilidade simples constantemente igual a  $p$  tem de verificar-se  $n-r$  vezes,



em  $n$  provas, é dada pelo valor do termo de ordem  $r+1$   $P = \frac{n!}{r!(n-r)!} p^r q^{n-r}$  do desenvolvimento do binômio  $(p+q)^n$ , onde  $q=1-p$  é a probabilidade contrária e  $r=0,1,2,\dots,n$ , dada por Bernoulli (1713).

**LEI DO EQUILÍBRIO DE HARDY-WEINBERG:** Na ausência de seleção, mutação e migração, uma grande população de acasalamento aleatório mantém constante as frequências gênicas e as frequências genotípicas, geração após geração. Diz-se que a população, nestas condições, está em equilíbrio. A propriedade de equilíbrio de uma população foi demonstrada primeiramente, simultânea e independentemente, em 1908, por Hardy e Weinberg. Em homenagem a ambos recebeu o nome de Lei de Hardy-Weinberg. Na demonstração dessa propriedade consideram-se três passos. i) os progenitores e os gametas que produzem; ii) a união dos gametas e genótipos dos zigotos produzidos; e, iii) os genótipos dos zigotos e as frequências gênicas da progênie que é a geração seguinte.

**LEI DOS GRANDES NÚMEROS:** Seja  $f(x)$  uma densidade com média  $\mu$  e variância finita  $\sigma^2$ , e seja  $m$  a média amostral de  $n$  valores selecionados de forma aleatória de  $f(x)$ . Sejam  $\varepsilon$  e  $\delta$  dois números que satisfazem  $\varepsilon > 0$  e  $0 < \delta < 1$ . Se  $n$  é um inteiro maior que  $\sigma^2/(\varepsilon^2\delta)$ , então  $P[-\varepsilon < m - \mu < \varepsilon] \geq 1 - \delta$ . Em inglês Law of large numbers.

**LEI DOS GRANDES NÚMEROS:** Lei que afirma que, se uma variável aleatória é observada muitas vezes, a média dessas observações tende para a média daquela variável.

**LEI DOS PEQUENOS NÚMEROS:** Termo sugerido por Von Bortkiewicz (1898) para descrever o comportamento de eventos raros que obedecem uma distribuição de Poisson. Em inglês Law of small numbers.

**LEI INTRÍNSECA:** É a lei de dispersão única no campo mineralizado, em que o campo geométrico, considerado geologicamente homogêneo, é até certo ponto intrínseco ou independente da posição nas características representando a variável regionalizada.

**LEIS DE DE MORGAN:** Em inglês De Morgan's laws.

**LEITOR DE PESSOAS:** Máquina que registra simultaneamente o material de leitura dos respondentes e as reações dos olhos.

**LEMA DE NEYMAN E PEARSON:** Seja  $x_1, \dots, x_n$  uma amostra aleatória de  $f(x; \theta)$ , onde  $\theta$  é um parâmetro que pode ter dois valores possíveis  $\theta_0$  ou  $\theta_1$ , e seja  $0 < \alpha < 1$  fixo. Seja  $k$  uma constante positiva e  $C$  um subconjunto do espaço amostral que satisfaz: (a)  $P_{\theta_0}[(x_1, \dots, x_n) \in C] = \alpha$ , (b)  $\lambda = L(\theta_0; x_1, \dots, x_n) / L(\theta_1; x_1, \dots, x_n) \leq k$  se  $(x_1, \dots, x_n) \in C$  e  $\lambda \geq k$  se  $(x_1, \dots, x_n) \in C^*$ . Então, o teste correspondente à região crítica  $C$  é o teste mais poderoso de tamanho  $\alpha$  para  $\theta = \theta_0$  contra  $\theta = \theta_1$ , sendo  $C$  e  $C^*$  complementares. Em inglês Neyman-Pearson lemma.

**LEI DOS GRANDE NÚMEROS (LAW OF LARGE NUMBERS):** Um teorema fundamental da probabilidade, a Lei dos Grandes Números diz que quanto maior a amostra, mais próxima a média aritmética se torna à expectativa probabilística de uma dada variável.

**LEI DOS GRANDE NÚMEROS:** Uma lei ou teorema que pode ser verificada por meio da desigualdade de Tchebycheff, senão veja,

$$P\{|X - E(X)| < \varepsilon\} \geq 1 - \frac{V(X)}{\varepsilon^2} \Rightarrow P\left\{\left|\frac{X}{n} - E\left(\frac{X}{n}\right)\right| < \varepsilon\right\} \geq 1 - \frac{V\left(\frac{X}{n}\right)}{\varepsilon^2}.$$

Dado que

$$E\left(\frac{X}{n}\right) = \frac{1}{n} E(X) = p \text{ e } V\left(\frac{X}{n}\right) = \frac{1}{n^2} V(X) = \frac{p(1-p)}{n}, \text{ tem-se que: } P\{|f - p| < \varepsilon\} \geq 1 - \frac{p(1-p)}{n\varepsilon^2}.$$

É chamada lei dos grandes números. Note que, para um dado valor de  $\varepsilon$ , se  $n \rightarrow \infty$ , então a probabilidade  $P\{|f - p| < \varepsilon\} \rightarrow 1$ . E, portanto, definido-se um valor de  $\varepsilon$  suficientemente pequeno, a frequência relativa  $f$  converge para a probabilidade  $p$  à medida que o tamanho da amostra  $n$  cresce.

**LEPTOCÚRTICA (LEPTOKURTIC):** Uma distribuição de probabilidades é dita leptocúrtica se ela tiver relativamente poucos valores no centro da distribuição e caudas pouco pesadas. A curtose de uma distribuição leptocúrtica é maior que 0. Comparar com platicúrtica.

**LEPTOCÚRTICA:** Tipo de kurtose que uma curva de distribuição de dados pode assumir, caracterizada por ser mais pontiaguda do que uma distribuição normal. Ver kurtose.

**LEMA:** É uma proposição auxiliar de um teorema. Em inglês lema.

**LETALIDADE:** Entende-se como maior ou menor poder que uma doença tem de provocar a morte das pessoas. Obtém-se a letalidade calculando-se a relação entre o número de óbitos resultantes de determinada causa e o número de pessoas que foram realmente acometidas pela doença, com o resultados expresso em percentual. A letalidade da escabiose é nula, e a da raiva é de 100%, havendo uma extensa gama de porções intermediárias entre esses extremos.

**LEVANTAMENTO (SURVEY), ESTUDOS:** Os estudos de levantamento são feitos para determinar o estado atual de um dado fenômeno. Embora muitas vezes empregado apenas como um meio para coletar dados em circunstâncias mal definidas, o levantamento pode ser muito útil para uma primeira exploração de fenômenos.

**LEVANTAMENTO EPIDEMIOLÓGICO:** Estudo realizado com dados existentes nos registros. Não é um estudo amostral e destina-se ao estabelecimento de coeficientes e de coberturas vacinais registradas, objetivando consolidação e análise dos dados existentes para avaliar o comportamento epidemiológico de doenças na população.

**LEVANTAMENTOS OU SURVEYS:** Estudos caracterizados pela coleta de dados, descrevendo-se os escores amostrais e calculando-se estatísticas que estimam os parâmetros da população investigada. Nos levantamentos pode-se, ainda, estudar a relação de causa e efeito de variáveis, mas sem o controle efetivo dos elementos causais.

**LABORATÓRIO:** Qualquer local de pesquisa onde se obtém coleta de dados nos quais determinados fatores ambientais, tais como temperatura, luminosidade, umidade relativa, etc. podem ser controlados pelo pesquisador. Contrapõe-se ao termo campo, porém pode ser também uma casa de vegetação, estufa ou ripado na pesquisa vegetal.

**LEI CIENTÍFICA:** Afirmação relativa a determinado aspecto da realidade, muito bem consolidadas por meio de pesquisas científicas e dados empíricos consistentes. Suposição confirmada por um elevado número de pesquisas. Generalização universalmente aceita.

**l<sub>i</sub>:** Limite superior da classe  $i$  e, simultaneamente, limite inferior da classe  $i + 1$ . Então,  $l_0$  é o limite inferior da primeira classe e  $l_c$  é o limite superior da última classe.

**LIMITES DE ADVERTÊNCIA:** Linhas horizontais adicionadas a um gráfico de controle em adição aos limites de controle que são usadas para tornar o gráfico mais sensível a causas atribuídas.

**LIMITES REAIS DE CLASSE:** Valores entre os quais situam todos os valores abrangidos por uma classe. Numa distribuição de valores cujas classes apresentam-se com os limites aparentes de 10 |— 19, 20 |— 29, 30 |— 39, os limites reais são 9,5 e 19,5; 19,5 e 29,5; 29,5 e 39,5.

**LIMITES DE CONTROLE:** Veja Gráfico de controle.

**LIMITES DE ESPECIFICAÇÃO:** Números que definem a região de medida para aceitar um produto. Geralmente, há um limite superior e um limite inferior, porém limites unilaterais podem também ser usados.

**LIMITES DE TOLERÂNCIA:** Um conjunto de limites entre os quais alguma proporção estabelecida dos valores de uma população tem de estar com nível especificado de confiança.

**LIMITES NATURAIS DE TOLERÂNCIA:** Um conjunto de limites simétricos que são três vezes os desvio-padrão da média do processo.

**LINHA CENTRAL:** Uma linha horizontal em um gráfico de controle, localizada no valor que estima a média da estatística plotada no gráfico. Veja Gráfico de controle.

**LINHA (OU CURVA) DE REGRESSÃO:** Um gráfico de um modelo de regressão, geralmente com a resposta  $y$  na ordenada e o regressor  $x$  na abcissa.

**LIGAÇÃO CANÔNICA:** Cada distribuição da família exponencial considerada em modelos lineares generalizados tem uma função de ligação especial para a qual existe uma estatística suficiente igual em dimensão a  $\beta$  do preditor linear  $\eta = \sum x_j \beta_j$ . Estas funções de ligação são chamadas de ligações canônicas. Em inglês Canonical link.

**LIGAÇÃO COMPLETA:** Algoritmo aglomerativo no qual a similaridade entre objetos é baseada na distância máxima entre objetos em dois agrupamentos através da distância entre os membros mais diferentes de cada agrupamento. Em cada estágio da aglomeração, os dois agrupamentos com a menor distância máxima ou seja, mais semelhantes são combinados.

**LIGAÇÃO INDIVIDUAL:** Procedimento de agrupamento hierárquico no qual a similaridade é definida como a distância mínima entre qualquer objeto em um agrupamento e qualquer objeto de outro. Isso simplesmente significa a distância entre os objetos mais próximos de dois agrupamentos. Esse procedimento tem o potencial de criar agrupamentos menos compactos ou mesmo em cadeia. É diferente do método de ligação completa, que usa a distância máxima entre objetos no agrupamento.

**LIGAÇÃO LOGARÍTMICA:** Seja  $\eta$  a função de ligação e  $\mu$  a média da parte aleatória, então a função de ligação logarítmica é  $\eta = \log(\mu)$ . Em inglês Log link.

**LIGAÇÃO LOGÍSTICA:** Seja  $\eta$  a função de ligação e  $\mu$  a média da parte aleatória, então a função de ligação logística é  $\eta = \log\left(\frac{\mu}{1-\mu}\right)$ . Esta função é comumente usada quando a parte aleatória de um

modelo linear generalizado segue uma distribuição de Bernoulli. Em inglês Logistic link.

**LIGAÇÃO MÉDIA:** Algoritmo usado em métodos aglomerativos que representa a similaridade como a distância média entre todos os objetos em um agrupamento e todos os objetos de outro agrupamento. Essa técnica tende a combinar agrupamentos com pequenas variâncias.

**LILLIEFORS:** Teste de aderência empregado para testar a normalidade amostral.

**LIMITE DE INFERIOR DE CONTROLE:** Fronteira usada em um gráfico de controle para separar pontos excepcionalmente baixos.

**LIMITE INFERIOR DE CRAMÉR-RAO:** Limite importante usado na inferência estatística, o qual mede o limite inferior para a variância de qualquer estimador não tendencioso de  $\tau(\theta)$ , e ele mostra ou permite dizer que a variância de qualquer estimador não tendencioso é igual ou superior a este limite. Em particular, se  $T$  é um estimador não tendencioso de  $\tau(\theta)$  cuja variância coincide com este limite,  $T$  é o melhor estimador não tendencioso possível para  $\tau(\theta)$ , conhecido como estimador UMVUE, pois nenhum outro estimador não tendencioso de  $\tau(\theta)$  poderá ter variância menor que a de  $T$ . Em inglês CRLB=Cramér-Rao lower bound.

**LIMITE DE SUPERIOR DE CONTROLE:** Fronteira usada em um gráfico de controle para separar pontos excepcionalmente altos.

**LIMITES DE CLASSE:** São os extremos de cada classe. O menor número é o limite inferior de classe (li) e o maior número, limite superior de classe (Li).

**LIMITES DE CLASSE:** As duas fronteiras de uma classe usada para agrupar dados.

**LIMITES DE CONFIANÇA:** São os limites que incluem o valor verdadeiro com um grau predeterminado de confiança é o nível de confiança.

**LIMITES DE CONTROLE:** Fronteira usada em um gráfico de controle para identificar pontos raros.

**LIMITES ESTATÍSTICOS DE TOLERÂNCIA:** Tipo de limites usados na construção de intervalos aplicados em controle estatístico de qualidade para avaliar a qualidade dos produtos manufaturados. Estes limites são determinados na consideração de que a qualidade é normalmente distribuída com média  $\mu$  conhecida e desvio padrão  $\sigma$  também conhecido. Estes limites podem ser formalmente conhecidos pela adição e pela subtração de  $\mu$  da quantidade  $\left(K_{\frac{\alpha}{2}}\right)(\sigma)$ . Infelizmente, na prática, os valores de  $\mu$  e  $\sigma$

raramente são conhecidos. Através da distribuição normal pode-se afirmar que 95 % dos produtos estarão dentro do intervalo  $\mu \pm 1,96\sigma$ . Esta afirmativa já não é válida para o intervalo  $\bar{X} \pm 1,96\sigma$  devido as quantidades  $\bar{X}$  e  $s$  serem variáveis aleatórias e os limites dependeram dos resultados de uma determinada amostra. Desta maneira torna evidente que a proporção de itens incluídos no intervalo  $\bar{X} \pm K_{\frac{\alpha}{2}}s$  não será

de  $1 - \alpha$ . É possível, entretanto, através da determinação de uma nova constante  $K$ , afirmar que  $100(\gamma)\%$  dos intervalos  $\bar{X} \pm Ks$ , estarão incluídos dentro do intervalo  $\mu \pm K_{\frac{\alpha}{2}}\sigma$ . Em outras palavras, que  $(1 - \alpha)\%$

dos valores de uma distribuição normal estarão incluídos no intervalo  $\bar{X} \pm Ks$  com uma probabilidade igual ou maior que  $100(\gamma)\%$ . Assim os limites estatísticos de tolerância para uma distribuição normal são determinados através das relações:  $L = \bar{X} - Ks$  e  $U = \bar{X} + Ks$ . Os valores de  $K$  para limites bilaterais e unilaterais de tolerância podem ser conhecidos em tabelas anexas em livros de estatística. Por exemplo:

Considerando o seguinte problema. As seguintes observações foram obtidas da medida do volume líquido, em litros, de vasilhames: 5,68; 5,65; 5,59; 5,64; 5,66; 5,61; 5,62; 5,64; 5,63; e 5,61. Encontre o intervalo de confiança de 95 % para a média do volume líquido da distribuição correspondente, considerando que os dados são normais. Solução: como o desvio padrão é desconhecido deve ser utilizada a distribuição de Student. (CHRISTMANN, 1978).

$$C' = \frac{s}{\sqrt{n}} = \frac{0,00267}{\sqrt{10}} = 0,00844. L_b = \bar{X} - C' t_{\frac{\alpha}{2}; n-1} = 5,633 - (0,00844)(t_{0,025;9});$$

$$L_b = 5,633 - (0,00844)(2,262) = 5,633 - 0,01909 = 5,6139;$$

$$U_b = \bar{X} + C' t_{\frac{\alpha}{2}; n-1} = 5,633 + 0,01909 = 5,65209.$$

**Donde o intervalo bilateral para o valor médio será:**  $(5,61391 \leq \mu \leq 5,652091)$ , com uma confiança de 95 %. Se for considerado que somente o limite inferior tem importância, o intervalo seria unilateral e determinado da seguinte maneira:  $L_u = \bar{X} - C' t_{\alpha; n-1} = 5,633 - (0,00844)(t_{0,05;9})$ ;  $L_u = 5,633 - (0,00844)(1,833) = 5,617531$ , e o intervalo unilateral com 95 % de confiança seria  $(5,61753 \leq \mu \leq \infty)$ . Agora considerando o problema apresentado anteriormente, a medida do volume líquido de vasilhames, encontre o intervalo que contenha 90 % dos vasilhames com uma probabilidade mínima de 95 %. Solução. Para este problema tem-se que:  $\alpha = 0,10$ ;  $\gamma = 0,95$ ;  $\bar{X} = 5,633$  e  $s = 0,0267$ , donde,

$$U = \bar{X} + K s = 5,633 + (2,839)(0,0267) = 5,70881;$$

$$L = \bar{X} - K s = 5,633 - (2,839)(0,0267) = 5,55721, \text{ ou seja, } 90 \% \text{ dos vasilhames apresentarão um volume líquido entre } 5,561 \text{ e } 5,711 \text{ com um confiança de } 95 \%$$

**LIMITES DE TOLERÂNCIA PARA DISTRIBUIÇÕES LIVRES (QUE NÃO SEJA DISTRIBUIÇÃO TEÓRICA DE PROBABILIDADE NORMAL OU GAUSSIANA):** Este método é aquele diferente do desenvolvido para a determinação dos limites de tolerância quando só são válidos se os dados provêm de uma distribuição normal. Sendo assim quando esta condição é violada, os limites devem ser obtidos de maneira diversa. Embora de uso industrial limitado, devido exigirem amostras muito grandes, a determinação dos limites de tolerância para distribuições não normais se baseia na menor e na maior observação da amostra. Para limites bilaterais de tolerância, o número de observações necessárias é conhecido através da relação

$$n = \left( \frac{2-\alpha}{\alpha} \right) \left( \frac{\chi_{1-\gamma;4}^2}{4} \right) + \frac{1}{2}. \text{ Para limites unilaterais, o tamanho da amostra pode ser conhecido através da}$$

$$\text{expressão } n = \frac{\log(1-\gamma)}{\log(1-\alpha)}. \text{ Considere a seguinte situação através do exemplo a seguir: (CHRISTMANN,}$$

1978). Na fabricação de válvulas eletrônicas, deseja-se saber, com probabilidade mínima de 99 % os limites bilaterais não normais que incluem 80 % das unidades. Solução. Utilizando a relação

$$n = \left( \frac{2-\alpha}{\alpha} \right) \left( \frac{\chi_{1-\gamma;4}^2}{4} \right) + \frac{1}{2} \text{ e considerando que } \alpha = 0,20 \text{ e } \gamma = 0,99, \text{ tem-se que: } n = \left( \frac{2-0,20}{0,20} \right) \left( \frac{\chi_{0,01;4}^2}{4} \right) + \frac{1}{2} \text{ e}$$

$$n = (9) \left( \frac{13,277}{4} \right) + \frac{1}{2} = 30 \text{ unidades, fazendo com que os limites de tolerância sejam iguais a e}$$

$$L = x_{\min} \text{ das } 30 \text{ observações } U = x_{\max} \text{ das } 30 \text{ observações.}$$

**LIMITES DE UM INTERVALO DE CONFIANÇA:** Dois números usados como fronteiras superior e inferior de um intervalo de confiança.

**LIMITES DE TOLERÂNCIA NÃO-PROBABILÍSTICOS:** São os determinados por processo que dispensa o conhecimento da distribuição da população de que provém a amostra sobre a qual se baseia aquela determinação.

**LIMITE DE UMA FUNÇÃO f(X) QUANDO X TENDE AO INFINITO:**  $\lim_{X \rightarrow \infty} f(x)$ .

**LIMITES INFERIORES DE CLASSE:** O menor número que pode efetivamente pertencer a diferentes classes em uma tabela de frequência.

**LIMITES REAIS DE CLASSE:** Valores entre os quais se situam todos os escores abrangidos por uma classe. Numa distribuição de escores cujas classes se apresentam com os limites aparentes de 10-19, 20-29, 30-39, os limites reais são 9,5 e 19,5, 19,5 e 29,5, 29,5 e 39,5. Em inglês actual class limits.

**LIMITES SUPERIORES DE CLASSE:** Os maiores números que podem pertencer a diferentes classes em uma tabela de frequência.

**LIMPEZA OU FILTRAGEM DE DADOS:** Remoção de erros aleatórios e sistemáticos de elementos de dados por meio de filtragem, fusão e tradução.

**LINEAR:** Modelo de variograma teórico usado em geoestatística dada pela seguinte equação:  $\gamma(h) = C_0 + p.h$ , onde:  $C_0$  é o efeito pepita;  $p$  é o coeficiente angular da reta.

**LINEARIDADE:** Usada para expressar o conceito de que o modelo possui as propriedades de aditividade e homogeneidade. Em termos gerais, os modelos lineares preveem valores que recaem em uma linha reta que tem uma mudança com unidade constante que é o coeficiente angular da variável dependente em relação a uma mudança com unidade constante da variável independente. No modelo população  $Y = b_0 + b_1 X_1 + E$ , o efeito de uma mudança de 1 em  $X_1$  deve acrescentar  $b_1$  (uma constante) unidades em  $Y$ .

**LINEARIDADE:** Termo usado para expressar o conceito de que o modelo possui as propriedades de aditividade e homogeneidade. De um modo geral, os modelos lineares preveem valores que estão sobre uma reta que tem uma taxa constante de variação (coeficiente angular) da variável dependente em relação a uma variação unitária constante na variável independente. No modelo populacional  $Y = b_0 + b_1 X_1 + \epsilon$ , o efeito de uma variação de 1 unidade em  $X_1$  será o acréscimo de  $b_1$  (uma constante) unidades em  $Y$ .

**LISREL:** Termo de abreviatura que significa Linear Structural Relations. É um programa, pacote ou software de computador utilizado para cálculos estatísticos, especialmente para a modelagem de equações estruturais.



**LINGUAGEM DE CONSULTA ESTRUTURADA (SQL):** Método para extração de informação a partir de uma base de dados de acordo com critérios especificados. Difere da mineração de dados no sentido de que são apenas fornecidos aqueles resultados que atendem as condições específicas solicitadas. SQL não pode executar qualquer processamento analítico, como o desenvolvimento de modelos de classificação ou a determinação de níveis de associação. Por exemplo, SQL pode extrair informação para todos os clientes entre as idades de 35 e 55 que compraram um produto durante o último mês com cartão de crédito, mas não pode agrupar clientes em segmentos sem critérios especificados.

**LINHA:** Conjunto de símbolos, especialmente números, que, numa tabela, se dispõem horizontalmente.

**LINHA:** Um dos tipos de bloco que ocorrem no quadrado latino.

**LINHA DE ACEITAÇÃO:** Lugar geométrico dos pontos que tem por ordenadas os números de aceitação de uma sequência de amostras na mesma ordem de sua extração e por abscissas os tamanhos dessas amostras.

**LINHA DE BASE:** Ponto no tempo ou conjunto de dados que servem como base para medir mudanças nas variáveis de interesse.

**LINHA DE REGRESSÃO:** Representação gráfica de uma equação de regressão. Linha traçada por meio de um conjunto de pontos que chega mais perto deles medida como a distância entre os pontos e a linha, fornecendo uma descrição gráfica da relação entre as variáveis.

**LINHA DE REJEIÇÃO:** Define-se como linha de aceitação, mutais mutandis. Ver linha de aceitação.

**LINHA DE CENTRO:** Linha usada em um gráfico de controle para representar um valor central das medidas características.

**LEITURAS:** Medidas primárias obtidas em amostras biológicas ou de animais, individualmente; também denominadas valores, dados originais.

**LEVANTAMENTO:** Examinamos um agregado de animais ou outras unidades em um estudo observacional para deduzir valores para estimativas de vários parâmetros da população.

**LEVANTAMENTO AMOSTRAL:** Um estudo observacional que utiliza dados amostrais para obter informações acerca da população da qual a amostra foi extraída.

**LEVANTAMENTO POPULACIONAL:** Um estudo observacional de toda a população; por exemplo, um censo.

**LEVERAGE:** O quanto os valores individuais das variáveis explicativas em uma análise de regressão diferem do valor médio das variáveis explicativas.

**LIGAÇÃO DE IDENTIDADE:** É a função de ligação utilizada em regressões simples e múltiplas; o valor médio da variável resposta nessa forma de modelo linear generalizado não é transformado quando está relacionado com uma função linear de co-variáveis.

**LIMITES DE CONCORDÂNCIA:** A amplitude de valores onde se encontra a maioria, normalmente 95 % das diferenças entre os pares de observações.

**LIMITES DE CONFIANÇA:** Os valores superiores e inferiores do intervalo de confiança.

**LINEAR:** Linha reta.

**LITERATURA CIENTÍFICA PUBLICADA:** Informação disponível ao estudante e ao profissional oriunda de trabalhos de outros e publicada em períodos científicos, livros-texto e revistas, bem como por via eletrônica por exemplo, na internet.

**LIVRO-CÓDIGO:** Documento usado em processamento e análise de dados que indica a localização de variáveis diferentes num arquivo de dados e o significado dos códigos usados para representar atributos diferentes destas variáveis.

**LOCAÇÃO DE PARCELAS:** Procedimento de instalação da parcela no campo.

**LOCALIZAÇÃO:** É uma característica das variáveis regionalizadas, cujos valores são dependentes de suas posições espaciais relativas dentro do campo geométrico ou depósito. Além disso, estes valores são dependentes do tamanho da amostra, forma e orientação. Ver suporte.

**LEVANTAMENTO:** Tipo de pesquisa descritiva com a finalidade de verificar o estado atual de um determinado fenômeno. Constitui-se basicamente na coleta de dados seguida de uma descrição dos mesmos, através das técnicas da chamada estatística descritiva ou análise exploratória de dados.

**LÓGICA:** É um termo usado na filosofia para distinguir o raciocínio correto do incorreto, sendo que neste caso raciocínio, indica a ideia do pensamento no qual se obtém conclusões a partir de premissas ou se realizam inferências.

**LOG-PROBABILIDADE ARITMÉTICA:** Ou gráfico de log-probabilidade aritmética, tem em sua ordenada a escala de probabilidade aritmética e em abscissa a escala logarítmica, de tal modo que se uma curva acumulativa for de uma distribuição lognormal ela será representada por uma reta, cuja inclinação será proporcional a variância amostral.

**LOGIT-REGRESSÃO LOGÍSTICA:** A Logit analisa o relacionamento entre variáveis dependentes ou respostas e variáveis independentes. A variável dependente é sempre categórica. O logaritmo das odds ratios da variável dependente é expresso como uma combinação linear dos parâmetros.

**LOGARITMO:** Se  $a^x = y$ , então  $x$  é o logaritmo de  $y$  na base  $a$ .

**LOGARITMO COMUM:** Logaritmo na base 10.

**LOGARITMO NATURAL:** Logaritmo de base  $e$ , onde  $e = 2,71828 \dots$

**LÓGICA:** Sistema de feitura de inferência. As regras por meio das quais podem-se derivar hipóteses ou conclusões de um conjunto de observações ou suposições iniciais. Embora alguns sistemas de lógica sejam rigorosamente explicitados na filosofia, todos nós operamos dentro de sistemas lógicos implícitos, por exemplo, racional.

**LOGÍSTICO (MODELO LOGÍSTICO):** Modelo matemático que utiliza logaritmo: por exemplo, para verificar a associação entre um ou mais fatores de risco e uma doença.

**LINHA DE BASE:** Valor médio de uma variável, obtido de um indivíduo, antes do mesmo ser submetido a uma determinada condição experimental.



**LOGIT:** Tipo de transformação em que se  $U$  for uma variável definida no intervalo  $[0; 1]$ , então a transformação que associa a cada valor  $u$ , neste intervalo, um valor  $v$  no intervalo  $(-\infty; +\infty)$  é denominada de logit ou logística e definida por:

$$V = \text{logit}(u) = \ln \left[ \frac{u}{(1-u)} \right]. \text{ A transformação inversa é executada por: } \text{logit}^{-1}(V) = \frac{e^V}{(1+e^V)}. \text{ Em inglês logit.}$$

**LOG-LINEAR:** Em inglês log-linear.

**LOG-LOG COMPLEMENTAR:** Para uma variável  $x$ , a função log-log complementar  $y$ , é dada por  $y = \log(-\log(1-x))$ . Em inglês Complementary log-log.

**LÓGICA:** É em essência um código de relacionamento que são inerentes entre aqueles dados internos e externos acessíveis aos nossos processos perceptuais diretos. Podemos até mesmo dizer que, embora a lógica resida na mente, suas raízes estão nos relacionamentos entre os próprios fatos externos.

**LOOP:** É um processo iterativo, no qual sua função irá executar uma sequência de comandos de linguagem de programação no computador até uma condição previamente estabelecida. É importante, nesses casos, que as iterações tenham uma condição finita.

**LONGITUDINAL, ESTUDO (TIME SERIES ANALYSIS):** Envolve observações contínuas ou intermitentes, feitas periodicamente com o fim de descrever mudanças que acontecem em dado ambiente. As observações podem ser controladas para que se possa testar hipóteses. Entretanto, neste tipo de estudo, as variáveis não estão sob controle do pesquisador. Estudos que envolvem a mensuração de indivíduos semelhantes ou diferentes duas ou mais vezes durante um período de tempo, comumente longo, como vários meses ou anos.

**LOTE:** Um grupo de itens tal como carregamento de matérias-primas ou peças compradas e também artigos de montagem final.

**LS:** Quadrados mínimos.

**$L(\theta)$ :** Função de verossimilhança.



**MAGNITUDE:** Na relação entre variáveis pode-se dizer que é a capacidade de se prevê uma baseada na outra.

**MAIOR OU IGUAL A:** Símbolo matemático dado por  $\geq$ .

**MAIOR QUE:** Símbolo matemático dado por  $>$ .

**MAIOR RAIZ CARACTERÍSTICA (GCR):** Estatística para testar a hipótese nula em MANOVA. Ela testa a primeira função discriminante das variáveis dependentes em relação à sua habilidade de distinguir diferenças de grupos.

**MAIS OU MENOS:** Símbolo matemático dado por  $\pm$ .

**MALHA DE SONDAGEM:** É o resultado do planejamento de uma campanha de sondagens e dá a localização dos furos de sondagem a serem realizados. A densidade de sondagem depende da fase e do objetivo da pesquisa, podendo ser baixa no início, a qual vai aumentando com o desenvolvimento da pesquisa.

**MALHA ÓTIMA:** É a malha de sondagem, definida por meio da análise geoestatística, cuja abertura situa-se dentro da amplitude do variograma ou zona de influência, podendo ser utilizado, por exemplo,  $0,67^*a$ .

**MALHA REGULAR:** É uma forma de disposição dos valores sobre os nós de um arranjo regular, que tem a vantagem de permitir o armazenamento otimizado sem suas coordenadas. Além disso, é a forma padrão de dados para mapas de curvas de nível e projeção em perspectiva de blocos diagrama.

**MANCOVA:** Análise de covariância multivariável.

**MANCOVA:** É a análise multivariada de covariância a qual pode ser usada em conjunção com a análise de variância multivariada (MANOVA) para remover após o experimento o efeito de quaisquer variáveis independentes métricas não controladas, conhecidas como covariáveis estatísticas sobre as variáveis dependentes. O procedimento é análogo ao envolvido na correlação parcial bivariada, na qual o efeito de uma terceira variável é removido da correlação. Termo em inglês: Multivariate analysis of covariance.

**MANOVA:** A análise de variância multivariada é utilizada para testar diferenças entre grupos ou vetores de médias utilizando várias variáveis métricas dependentes simultaneamente.

**MANCOVA:** A análise de covariância multivariada é uma extensão da Análise de Covariância onde existe mais de uma variável dependente e elas não podem ser combinadas.

**MANOVA:** Análise de variância multivariável.

**MANIPULAÇÃO AMBIENTAL:** O pesquisador provê aos respondentes várias situações, frequentemente inesperadas, que exigem alguma forma de resposta. Um bom exemplo deste tipo de manipulação é um estudo sobre a difusão de responsabilidade. Os autores hipotetizam que numa situação de emergência, quanto maior o número de pessoas envolvidas, tanto menos a probabilidade de que qualquer uma delas tente resolver a emergência.

**M DE BOX:** Teste estatístico para a igualdade das matrizes de covariância das variáveis independentes nos grupos da variável dependente. Se a significância estatística é maior que o nível crítico como por exemplo, 0,01, então a igualdade das matrizes de covariância encontra sustentação. Se o teste mostra significância estatística, os grupos são considerados diferentes e a suposição é violada.

**MÉDIA:** Em sua acepção mais ampla, qualquer promédio de uma distribuição de frequência. Em sentido estrito, média aritmética.

**MÉDIA ARITMÉTICA:** Soma de todos os valores dividida pelo número de valores.

**MEDIANA:** Medida de tendência central que divide uma distribuição de frequências em duas partes de igual valor.

**MEDIDAS DE TENDÊNCIA CENTRAL:** Valores que, em estatística, caracterizam os valores médios.

**MODA:** Valor mais frequente de uma distribuição. Numa representação gráfica, a moda é o escore correspondente ao ponto mais alto da curva. Em outras palavras, é o valor que está na moda.

**MAP:** Máximo a posteriori.

**MAPEAMENTO LINEAR:** Este procedimento tem a finalidade de, a partir de um conjunto de variáveis quantitativas, obter um mapa de distância entre as observações, representadas na forma de um gráfico de duas dimensões ou dois eixos, com objetivo de facilitar a interpretação dos resultados, formar grupos dentre outros. Neste procedimento são calculados altovalores que, após um processo de otimização, são representados nos eixos horizontal e vertical. Para a execução da análise de agrupamento, é necessária a construção de uma matriz de similaridade ou de dissimilaridade.

**MAPEAMENTO NÃO-LINEAR:** Este procedimento tem a finalidade de, a partir de um conjunto de variáveis quantitativas, obter um mapa de distância entre as observações, representadas na forma de um gráfico de duas dimensões ou dois eixos, com objetivo de facilitar a interpretação dos resultados, formar grupos dentre outros. Este é um processo iterativo, em que, a cada iteração, procura-se novo ângulo de visão para representar os valores de todas variáveis em dois eixos. Este é um processo iterativo; portanto, o usuário deverá informar o número de iterações a serem processadas pelo procedimento.

**MAPA DE NÍVEIS:** Gráfico que mostra as curvas em um plano (x,y) correspondente aos valores da constante k em uma equação definida como  $f(x,y)=k$ . É similar aos contornos de mesma altitude em um mapa geográfico. Em inglês Level map.

**MAPA ESPACIAL:** Ver mapa perceptual.

**MAPA PERCEPTUAL:** Representação visual de percepções que um respondente tem sobre objetos em duas ou mais dimensões. Geralmente esse mapa tem níveis opostos de dimensões nos extremos dos eixos X e Y, tais como de doce a azedo nos extremos do eixo X e de caro a barato nos extremos do eixo Y. Cada objeto então tem uma posição espacial no mapa perceptual que reflete a similaridade ou preferência relativa a outros objetos no que se refere às dimensões do mapa perceptual.

**MAPA TERRITORIAL:** Representação gráfica dos escores de corte em duas dimensões. Quando é combinado com os gráficos de casos individuais, a dispersão de cada grupo pode ser vista e as classificações ruins de casos individuais podem ser diretamente identificadas a partir do mapa.

**MAPEAMENTO GENÉTICO:** Determinação das posições relativas dos genes em uma molécula de DNA, cromossomo ou plasmídeo e da distância, em unidades de ligação ou unidades físicas, entre eles.

**MARCO DE AMOSTRAGEM:** Ver cadastro de amostragem.

**MARGEM DE ERRO:** Diferença máxima provável com probabilidade  $1-\alpha$  entre a estatística amostral observada e o verdadeiro valor do parâmetro populacional.

**MARGINAIIS:** Nome que é dado as distribuições de probabilidades de variáveis unidimensionais em distribuição de probabilidade bidimensionais. Em inglês marginals.

**MARKETING:** O processo de planejar e executar a concepção, apreçamento, promoção e distribuição de ideias, bens e serviços para criar intercâmbios que satisfaçam aos objetivos individuais e organizacionais.

**MARTINGALE:** Uma sequencia de variáveis aleatórias  $X_1, X_2, \dots, X_n$  (com expectâncias finitas) é um martingale se a expectância condicional de  $X_{n+1}$  dado os valores de  $X_1, X_2, \dots, X_n$  é  $E(X_{n+1}/X_1=x_1, X_2=x_2, \dots, X_n=x_n) = x_n$ . Um exemplo é quando  $X_n$  é a quantidade ganha por um jogador após a n-ésima sequencias de jogos honestos. Em inglês martingale.

**MAS:** Seleção auxiliada por marcadores.

**MARGEM DE ERRO:** Diferença máxima que um pesquisador aceita assumir entre o valor real de um parâmetro e a sua estimativa.

**MANIPULAÇÃO:** Meio usado pelo pesquisador para interferir no comportamento de uma variável resposta ou também chamada de independente.

**MATRIZ:** Tipo de organização de valores em que os mesmos apresentam uma disposição ou estrutura em forma de uma tabela com linhas e colunas dispostas horizontalmente e verticalmente respectivamente.

**MATRIZ DE CORRELAÇÃO:** Estrutura de dados em forma de tabela que mostra todas as correlações possíveis entre as variáveis numa pesquisa, notadamente nas pesquisas descritivas correlacionais. Em cada célula da matriz, normalmente temos duas informações: o coeficiente de correlação designado genericamente por  $r$  e o nível de significância da correlação ( $p$ ).

**MANOVA:** Ver análise de variância multivariada.

**MATRIZ (MATRIX):** Um objeto matemático frequentemente usado como variável constituído por muitas observações em linhas e muitas variáveis nas colunas.

**MATRIZ DE CLASSIFICAÇÃO (CLASSIFICATION MATRIX):** Um resultado de uma análise discriminante. Uma tabela onde as linhas representam os grupos observados a que cada observação pertence e as colunas representam a predição do grupo a que cada observação pertence. Os valores nesta tabela representam o número de observações colocadas em cada par observado ou predito. Em uma classificação perfeita uma análise discriminante sem nenhum erro, somente os elementos da diagonal da matriz de classificação terão valores diferentes de zero.

**MATRIZ DE COLUNAS ORTONORMAIS (COLUMN-ORTHONORMAL MATRIX):** Uma matriz  $m \times n$  cujas colunas foram normalizadas de maneira que a distância euclidiana de cada coluna = 1.

**MATRIZ DE GRÁFICO DE DISPERSÃO (SCATTERPLOT MATRIX):** Um gráfico feito com vários gráficos de dispersão é usado para ilustrar dados multivariados. Cada gráfico é de dispersão simples e eles são agrupados por linhas com variáveis Y em comum e por colunas com variáveis X em comum. O arranjo desses gráficos em uma matriz ilustra todas as relações bivariadas possíveis entre todas as variáveis medidas.

**MATRIZ JACOBIANA:** Dado uma função vetorial de várias variáveis  $F: \mathfrak{R}^m \rightarrow \mathfrak{R}^n$  com  $F(X) = (f_1(X), \dots, f_n(X))$ , a representação matricial da derivada, quando existe, é denominada de matriz jacobiana e é definido como sendo:

$$JF(x_1, \dots, x_m) = \begin{bmatrix} f_1 \\ \vdots \\ f_n \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \dots & \frac{\partial f_1}{\partial x_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \dots & \frac{\partial f_n}{\partial x_m} \end{bmatrix}$$

Quando  $m = n$ , a matriz jacobiana é uma matriz quadrada e o seu determinante

$$\frac{\partial(f_1, \dots, f_n)}{\partial(x_1, \dots, x_m)} = \det \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \dots & \frac{\partial f_1}{\partial x_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \dots & \frac{\partial f_n}{\partial x_m} \end{bmatrix}$$

é denominado de função jacobiana. Os livros de cálculo costumam designar a função jacobinana simplesmente como jacobiana.

**MATRIZ HESSIANA:** Dada uma função real de várias variáveis,  $F: \mathfrak{R}^n \rightarrow \mathfrak{R}$ , a matriz jacobiana, também chamada de derivada, do gradiente que é uma função vetorial é denominada de matriz hessiana de  $f$ . Assim,

$$Hessf(x_1, \dots, x_n) = J \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \dots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \dots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{bmatrix}. \text{ A matriz hessiana sempre é uma matriz qua-}$$

drada. O determinante da matriz hessiana é denominado de função hessiana que não pode ser confundido com a matriz hessiana. Nos livros de cálculo a função hessiana costuma ser referenciada simplesmente de hessiana.

**MEDIANA:** Tipo de medida de tendência central que representa o valor do meio de uma distribuição de dados, quando dispomos os valores em ordem crescente ou decrescente. Por exemplo: considere-se o seguinte conjunto de valores: 5; 7; 15; 10; 22. Colocando-os em ordem crescente, ficariam desta forma: 5; 7; 10; 15; 22. A mediana ( $\tilde{x}$  ou  $x_{\text{til}}$ ) representa o  $\frac{(n+1)}{2}$ -ésimo valor central ordenado, isto é, 10 (terceiro

valor ordenado). Quando o número de valores da distribuição for ímpar, a mediana será exatamente o valor do meio da lista. Por outro lado, quando o número de valores for par, a mediana representará a média dos dois valores do meio. Exemplo: conjunto de valores: 5; 7; 10; 12; 15; 22. Neste caso, a mediana será a média entre 10 e 12, isto é, 11.

**MEDIDA:** Ver mensuração.

**MEDIDA NÃO OBTRUSIVA:** Medida do comportamento realizada sem o conhecimento do sujeito. Esse tipo de procedimento é comum em pesquisas na área de marketing e comportamento do consumidor, por exemplo. Sua aplicação é, em muitos casos, eticamente contestável.

**MEDIDAS DE DISPERSÃO:** Valores que procuram mensurar a dispersão ou variabilidade de um conjunto de dados. Ver dispersão.

**MEDIDAS DE TENDÊNCIA CENTRAL:** São valores típicos ou representativos em uma distribuição de dados. As medidas de tendência central mais utilizadas são a média, a mediana e a moda.

**MENSURAÇÃO:** Designação de números ou símbolos a objetos ou eventos, de acordo com regras específicas. Normalmente, a mensuração refere-se à quantificação das diferentes características de uma variável. Tomemos como exemplo a variável cor dos olhos. Poderíamos assumir que esta variável deva ser mensurada da seguinte forma: {azul, verde, castanho claro, castanho escuro}. Outro exemplo: a variável temperatura pode ser mensurada em graus Celsius com duas casas decimais, como por exemplo, 27,52° ou 15,75° dentre outros, ou então numa escala mais subjetiva de três posições: quente, morno, frio.

**MESOCÚRTICA:** Tipo de kurtose que uma curva de distribuição de dados pode assumir, caracterizada por não ser tão pontiaguda quanto a distribuição leptocúrtica nem tão alongada e achatada quanto a distribuição platicúrtica. O tipo mais conhecido de distribuição mesocúrtica é a distribuição normal. Ver kurtose para um exemplo gráfico.

**META-ANÁLISE:** Síntese estatística dos resultados de vários estudos dedicados a investigar uma mesma questão. Normalmente, os resultados de uma meta-análise são expostos graficamente por meio de um blobograma. Ver blobograma para exemplo.

**META-ANÁLISE:** Análise por meio de quaisquer critérios científicos dos resultados de diversos estudos dedicados a investigar uma mesma questão.

**MODA:** Tipo de medida de tendência central que representa o valor mais frequente em uma distribuição de dados. Por exemplo: considere o seguinte conjunto de valores: 3; 7; 15; 7; 9; 24; 14. A moda normalmente identificada como M ou  $M_o$  dessa distribuição é 7, pois é o valor que mais aparece na distribuição.

**MODELAGEM CAUSAL:** Técnica matemática utilizada para inferir a causalidade entre variáveis, porém sem a utilização de uma pesquisa experimental. Muitos cientistas sociais rejeitam a denominação modelagem causal por entenderem que a análise da causalidade somente pode ser realizada por meio de estudos experimentais, desta forma, a denominação mais utilizada atualmente é a de modelagem de equações estruturais.

**MODELAGEM DE EQUAÇÕES ESTRUTURAIS:** Técnica utilizada que tem por objetivo verificar a causalidade entre conjuntos de variáveis encadeadas ou estruturas de covariância. Por exemplo: numa pesquisa descritiva, busca-se analisar as relações entre diversos componentes que levam ao grau de comprometimento organizacional dos funcionários de uma empresa. Pode-se modelar a seguinte sequência de variáveis: nível salarial influencia o grau de motivação no trabalho, que influencia o grau de comprometimento organizacional, ou seja, X influencia Y e influencia W. O modelo apresentado pode ser testado



por meio de modelagem de equações estruturais e essas relações de causa e efeito podem ser hipotetizadas teoricamente, sem a realização de uma pesquisa experimental clássica.

**MODELO:** Representação simplificada da realidade construída para ajudar em nossa compreensão dessa realidade.

**MÉTODO DE SOBREVIDA DE KAPLAN-MEIER:** Tipo de método usado na análise de sobrevivência onde os intervalos de tempo não são fixos, mas determinados pelo aparecimento de uma falha, por exemplo, o óbito. Nessa situação, o número de óbitos em cada intervalo deve ser um. Esse é um método não paramétrico, ou seja, que independe da distribuição de probabilidade, e para calcular ou estimar os estimadores, primeiramente, deve-se ordenar os tempos de sobrevivência em ordem crescente representados por  $(t_1 \leq t_2 \leq t_3 \dots \leq t_n)$ . Os sobreviventes ao tempo  $t$  ( $l_t$ ) são ajustados pela censura, ou seja, os pacientes censurados entram no cálculo da função de probabilidade de sobrevivência acumulada até o momento de serem considerados como perda. Isto propicia o uso mais eficiente das informações disponíveis.

**MÉTODO DE ESTIMAÇÃO DE MÍNIMOS QUADRADOS INDIRETOS:** Mínimos quadrados indiretos (MMQI) é um método de calcular valores consistentes dos parâmetros estruturais de equações exatamente identificadas, em um sistema de equações simultâneas. O MMQI envolve o uso inicial do método dos mínimos quadrados (MMQ) para estimar as equações na forma reduzida do sistema, após o que utiliza-se os parâmetros estimados na forma reduzida, calculando-se as estimativas únicas e consistentes dos parâmetros estruturais. Uma desvantagem de se usar o MMQI é que ele não dá o erro padrão dos parâmetros estruturais calculados, e é mais complicado calculá-los. Outra desvantagem do MMQI é que ele não pode ser usado para calcular estimativas únicas e consistentes dos parâmetros estruturais a partir dos coeficientes das formas reduzidas para as equações superidentificadas de um modelo de equações simultâneas.

**MÉTODO DE ESTIMAÇÃO DOS MÍNIMOS QUADRADOS DE DOIS ESTÁGIOS:** Mínimos quadrados de dois estágios (MQ2E) é um método de estimação de parâmetros estruturais consistentes em equações exatamente identificadas ou superidentificadas de um sistema de equações simultâneas. Para equações exatamente identificadas, o MQ2E apresenta o mesmo resultado que o método dos mínimos quadrados indiretos (MMQI). A estimação por MQ2E envolve a aplicação do método dos mínimos quadrados ordinários (MMQ) em dois estágios. No primeiro estágio, cada variável endógena é regredida sobre todas as variáveis pré-determinadas do sistema. Estas são agora as equações nas formas reduzidas. No segundo estágio, os valores esperados ao invés dos valores observados das variáveis endógenas são utilizados para estimar as equações estruturais do modelo. Os valores esperados das variáveis endógenas são obtidos pela substituição dos valores observados das variáveis exógenas nas equações das formas reduzidas. Os valores esperados das variáveis endógenas são não correlacionados com os termos erros, conduzindo as estimativas consistentes dos parâmetros estruturais pelo MQ2E. Uma vantagem do MQ2E sobre o MMQI é que o MQ2E pode ser utilizado para obter estimativas consistentes dos parâmetros estruturais de equações superidentificadas, bem como para equações exatamente identificadas, em um sistema de equações simultâneas. Outra importante vantagem é que o MQ2E, mas não o MMQI dá o erro padrão das estimativas dos parâmetros estruturais, diretamente. Como a maioria dos modelos identificados são de fato superidentificados, o MQ2E é muito útil. De fato, o MQ2E é o mais simples e um dos melhores métodos de estimação de equações simultâneas.

**MODELOS DE DEFASAGENS DISTRIBUÍDAS:** É o tipo de modelo estatístico onde é frequente o caso em que o valor corrente da variável dependente é uma função, ou depende da soma ponderada do

valor presente,  $t$ , e dos valores passados da variável independente e o termo erro, com pesos geralmente diferentes atribuídos nos vários períodos de tempo:  $Y_t = a + b_0X_t + b_1X_{t-1} + b_2X_{t-2} + \dots + u_t$ .

A estimação do modelo de defasagens distribuídas apresenta duas dificuldades. Primeira, os dados de uma observação ou período de tempo são perdidos para cada valor defasado de  $X$ . Segunda, é provável que os  $X$ s estejam relacionados um com o outro, tal que pode ser difícil ou impossível isolar o efeito de cada  $X$  sobre  $Y$ .

**MODELOS DE DEFASAGENS DISTRIBUÍDAS:** Tipos de modelos estatísticos utilizados na análise de regressão e econometria. Como exemplo de aplicação considere que em muitas vezes o efeito de uma variável política pode ser distribuído para um período de série de tempo isto é, a variável dependente pode responder lentamente a uma mudança de política, exigindo uma série de variáveis explicativas defasadas para calcular o ajustamento completo através do tempo. Um modelo de defasagens distribuídas é aquele em que o valor corrente da variável dependente,  $Y_t$ , depende da soma ponderada dos valores presente e passados das variáveis independentes isto é,  $X_t, X_{t-1}, X_{t-2}$ , etc. e o termo erro, com pesos geralmente diferentes para vários períodos de tempo geralmente declinando sucessivamente para o primeiro período de tempo. Veja os modelos.

$$Y_t = a + b_0X_t + b_1X_{t-1} + b_2X_{t-2} + \dots + u_t$$

$$Y_t = a + b_0X_t + b_1X_{t-1} + b_2X_{t-2} + \dots + b_kX_{t-2k} + u_t$$

Observe que nas equações,  $a$  é constante, enquanto  $b_0$  é o coeficiente de  $X_t$ .

Na estimação de um modelo de defasagens distribuídas, a inclusão de cada termo defasado usa mais de um grau de liberdade. Quando o número de termos defasados independentes,  $k$ , pequeno, o modelo pode ser estimado pelo método dos mínimos quadrados (MMQ). Contudo, o  $k$  elevado em relação ao tamanho da série de tempo, um número inadequado de graus de liberdade pode ser abandonado na estimativa do modelo, ou para a significância dos parâmetros estimados. E ainda, é provável que as variáveis explicativas defasadas, em um modelo de defasagens distribuídas, sejam fortemente correlacionadas, com o que pode tornar-se difícil separar adequadamente seus efeitos independentes sobre a variável dependente.

**MODELOS DE EQUAÇÕES SIMULTÂNEAS:** Quando a variável dependente em uma equação é também uma variável explicativa em alguma outra equação, temos um sistema ou modelo de equações simultâneas. As variáveis dependentes em um sistema de equações simultâneas são chamadas variáveis endógenas. As variáveis determinadas por fatores externos ao modelo são chamadas variáveis exógenas. Há uma equação comportamental ou estrutural para cada variável endógena no sistema. Usar o método de mínimos quadrados simples para estimar equações estruturais, resulta em estimativas viesadas e inconsistentes dos parâmetros. Este é o chamado viés de equações simultâneas. Para obter estimativas consistentes dos parâmetros, deve-se obter primeiro as formas reduzidas das equações do modelo. Estas expressam cada variável endógena do sistema apenas como função da variável exógena do modelo. Por exemplo. As duas equações seguintes representam um modelo macroeconômico simples:

$$M_t = a_0 + a_1Y_t + u_{1t}$$

$$M_t = b_0 + b_1M_t + b_2I_t + u_{2t}$$

Onde  $M_t$  é oferta de modelo no período de tempo t, Y é renda e I é investimento. Como M depende de Y na primeira equação e Y depende de M e I na segunda equação, M e Y são determinadas conjuntamente, o que define um modelo de equações simultâneas. Me E são variáveis endógenas, enquanto I é exógena ou determinada fora do modelo. Uma mudança em afeta na primeira equação. Esta, por sua vez, afeta na segunda equação. Como resultado, e são correlacionadas, deixando as estimativas das equações de M (e Y), através do método de mínimos quadrados simples, viesadas e inconsistentes.

**MULTICOLINEARIDADE:** Termo usado para se referir ao caso em que duas ou mais variáveis explicativas ou independentes do modelo de regressão são perfeitamente ou altamente correlacionadas ou colineares ou seja se uma ou mais das variáveis puderem ser expressas como uma combinação linear da(s) outra(s) variável(is). Por exemplo, há perfeita multicolinearidade entre e se = 2 ou . Se duas ou mais variáveis explicativas são perfeitamente correlacionados linearmente, seria impossível calcular as estimativas de mínimos quadrados simples dos parâmetros, porque o sistema de equações normais poderia conter duas ou mais equações que não são independentes. Multicolinearidade alta, mas não perfeita, refere-se ao caso em que duas ou mais variáveis independentes no modelo de regressão são altamente correlacionados. Isto pode tornar difícil ou impossível isolar o efeito que cada uma das variáveis explicativas altamente colineares tem sobre a variável dependente. Contudo, os coeficientes estimados pelo método dos mínimos quadrados (MMQ) são ainda não viesados se o modelo está especificado corretamente. Além disso, se o alvo principal é a predição, a multicolinearidade não é um problema, se a mesma multicolinearidade padrão persistir durante o período projetado. O caso clássico e multicolinearidade ocorre quando nenhuma das variáveis explicativas na regressão é estatisticamente significantes e alguma pode mesmo ter o sinal errado, ainda que  $R^2$  possa ser alto por exemplo com valores entre 0,7 e 1,0. Em casos menos evidentes detectar a multicolinearidade pode ser mais difícil. Os coeficientes de correlação simples ou parcial entre variáveis explicativas são algumas medidas usadas para detectar a multicolinearidade. Contudo, pode ainda ocorrer multicolinearidade elevada, mesmo com coeficientes de correlação simples ou parcial relativamente baixos isto é, menos do que 0,5. O problema da multicolinearidade pode algumas vezes ser corrigido usando algumas técnicas tais como: i) aumentando o tamanho da amostra, ii) utilizando uma informação a priori por exemplo, pode-se saber, através de um estudo prévio, que  $b_2 = 0,25b_1$ , iii) transformando a relação funcional, ou iv) excluindo uma das variáveis altamente colinear mas isto pode conduzir a um viés ou erro de especificação, se a teoria informar que a variável abandonada deveria estar incluída no modelo.

**MATRIZ:** Disposição retangular de números que representam, por exemplo, os coeficientes de um sistema de equações. Uma matriz é representada por A (m, n), onde m = número de linhas e n = número de colunas. Um elemento genérico da matriz é representado por  $a_{ij}$ . Em inglês matrix.

**MATRIZ:** Uma matriz é um quadrado retangular de números dado por.

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

As m n – uplas horizontais

$(a_{11}, a_{12}, \dots, a_{1n}), (a_{21}, a_{22}, \dots, a_{2n}), \dots, (a_{m1}, a_{m2}, \dots, a_{mn})$  são camadas de linhas de a, e as n m – uplas verticais, suas colunas.

$$\begin{pmatrix} a_{11} \\ a_{21} \\ \cdots \\ a_{m1} \end{pmatrix}, \begin{pmatrix} a_{12} \\ a_{22} \\ \cdots \\ a_{m2} \end{pmatrix}, \dots, \begin{pmatrix} a_{1n} \\ a_{2n} \\ \cdots \\ a_{mn} \end{pmatrix}$$

Note que o elemento  $a_{ij}$ , chamado a ij-ésima entrada, aparece na interseção da i-ésima linha com a j-ésima coluna. Representamos frequentemente tal matriz por  $A = (a_{ij})$ . Uma matriz com m linhas e n colunas é chamada matriz m por n e escrevemos mat m X n; se m = n, então é chamada matriz quadrado ou matriz n-quadrada. Notamos também que a matriz com uma única linha pode ser olhada como um vetor, e vice-versa.

**MATRIZ:** Conjunto retangular eventualmente quadrado de números, em que se distinguem p linhas e q colunas, com um total de pq números.

**MATRIZ:** Representação tabular de um conjunto de dados caracterizada uma dimensionalidade medida pelo número de fileiras e colunas. Se ambas são iguais, a matriz será chamada de quadrada, e se o triângulo superior é igual em valores ao triângulo inferior, a matriz será chamada de simétrica. Se for empregado métodos algébricos a matriz será simbolizada por uma letra capital.

**MATRIZ COLUNA:** É toda matriz do tipo mx1 ( $m \in R^+$ ).

$$B = \begin{bmatrix} a_1 \\ \vdots \\ a_{m1} \end{bmatrix}. \text{ Por exemplo: } A = \begin{bmatrix} 1 \\ 2 \\ 6 \end{bmatrix}$$

**MATRIZ DE CLASSIFICAÇÃO:** Matriz que avalia a habilidade preditiva da(s) função(ões) discriminante(s) ou da regressão logística. É também chamada de matriz confusão, designação ou de previsão. Criada pela tabulação cruzada dos membros do grupo real com os do grupo previsto, essa matriz consiste em números na diagonal, que representam classificações corretas, e números fora da diagonal, que representam classificações incorretas.

**MATRIZ DE CORRELAÇÃO:** É uma matriz quadrada, simétrica, cuja diagonal é formada pela unidade, pois se trata da correlação da variável com ela mesma, e em cada interseção linha (i) coluna (j) a correlação das variáveis Xi e Xj.

**MATRIZ DE CORRELAÇÃO:** Formato tabular para apresentação de correlações bivariadas dentre diversas variáveis. A lista de variáveis é apresentada horizontalmente no topo da tabela e para baixo do lado. Qualquer célula da tabela contém a correlação entre as duas variáveis cuja interseção define aquela célula. Em inglês correlation matrix.

**MATRIZ DE CORRELAÇÃO ANTI-IMAGEM:** Matriz das correlações parciais entre variáveis após análise fatorial, que representa o grau em que os fatores explicam um ao outro nos resultados. A diagonal contém as medidas de adequação da amostra para cada variável, e os demais valores são correlações parciais entre variáveis.

**MATRIZ DE CORRELAÇÃO:** É uma matriz quadrada, simétrica, cuja diagonal é formada pela unidade, pois trata-se da correlação da variável com ela mesma, e em cada interseção linha (i) coluna (j) a correlação das variáveis  $X_i$  e  $X_j$ .

**MATRIZ DE CORRELAÇÃO:** Tabela que mostra as intercorrelações entre todas as variáveis.

**MATRIZ DE CORRELAÇÃO GENÉTICA EPISTÁTICA (DO TIPO ADITIVO x ADITIVO) ENTRE OS INDIVÍDUOS EM AVALIAÇÃO:** Denotada pelo símbolo  $A\#A$ .

**MATRIZ DE CORRELAÇÃO GENÉTICA EPISTÁTICA (DO TIPO ADITIVO x DOMINANTE) ENTRE OS INDIVÍDUOS EM AVALIAÇÃO:** Denotada pelo símbolo  $A\#D$ .

**MATRIZ DE CORRELAÇÃO GENÉTICA EPISTÁTICA (DO TIPO DOMINANTE x DOMINANTE) ENTRE OS INDIVÍDUOS EM AVALIAÇÃO:** Denotada pelo símbolo  $D\#D$ .

**MATRIZ DE CORRELAÇÃO GENÉTICA EPISTÁTICA (DO TIPO ADITIVO x ADITIVO x ADITIVO) ENTRE OS INDIVÍDUOS EM AVALIAÇÃO:** Denotada pelo símbolo  $A\#A\#A$ .

**MATRIZ DE COVARIÂNCIA:** É uma matriz quadrada, simétrica, cuja diagonal contém a variância da variável e em cada interseção linha (i) coluna (j) a covariância das variáveis  $X_i$  e  $X_j$ .

**MATRIZ DE VARIÂNCIA E COVARIÂNCIA (VARIANCE-COVARIANCE MATRIX):** Uma matriz quadrada  $C$  ( $n$  linhas  $\times$   $n$  colunas, onde  $n$  é o número de medidas ou observações individuais para cada observação multivariada) em que os elementos da diagonal  $c_{i,j}$  são as variâncias de cada variável e os elementos fora da diagonal  $c_{i,j}$  são as covariâncias entre a variável  $i$  e a  $j$ . A matriz de variância-covariância é uma matriz quadrada que contém as variâncias e covariâncias associadas a diversas variáveis. Os elementos diagonais da matriz contém as variâncias das variáveis, e os elementos fora da diagonal contém as covariâncias entre todos os pares possíveis de variáveis. A partir da matriz  $X$  de dados de ordem ' $n \times p$ ' pode-se fazer uma estimativa da matriz de covariância  $\Sigma$  da população  $\pi$  que é representada por  $S$ . A matriz  $S$  é simétrica e de ordem ' $p \times p$ '.

$$S = \begin{bmatrix} \hat{\text{Var}}(x_1) & \hat{\text{Cov}}(x_1x_2) & \hat{\text{Cov}}(x_1x_3) & \cdots & \hat{\text{Cov}}(x_1x_p) \\ \hat{\text{Cov}}(x_2x_1) & \hat{\text{Var}}(x_2) & \hat{\text{Cov}}(x_2x_3) & \cdots & \hat{\text{Cov}}(x_2x_p) \\ \hat{\text{Cov}}(x_3x_1) & \hat{\text{Cov}}(x_3x_2) & \hat{\text{Var}}(x_3) & \cdots & \hat{\text{Cov}}(x_3x_p) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{\text{Cov}}(x_px_1) & \hat{\text{Cov}}(x_px_2) & \hat{\text{Cov}}(x_px_3) & \cdots & \hat{\text{Var}}(x_p) \end{bmatrix}$$

**MATRIZ DIAGONAL (DIAGONAL MATRIX):** Uma matriz quadrada em que todos os valores, exceto os da diagonal principal, são iguais a zero.

**MATRIZ IDENTIDADE (IDENTITY MATRIX):** Uma matriz diagonal em que todos os termos da diagonal são = 1 e todos os outros = 0.

**MATRIZ INVERSA (INVERSE MATRIX):** Uma matriz  $A^{-1}$  tal que  $AA^{-1} = I$ , a matriz identidade.

**MATRIZ QUADRADA (SQUARED MATRIX):** Uma matriz com o mesmo número de linhas e de colunas.

**MATRIZ TRIANGULAR (TRIANGULAR MATRIX):** Uma matriz na qual todos os valores em um dos lados da diagonal são = 0.

**MATRIZ TRIANGULAR INFERIOR (LOWER TRIANGULAR MATRIX):** Uma matriz na qual todos os elementos acima da diagonal são = 0.

**MATRIZ TRIANGULAR SUPERIOR (UPPER TRIANGULAR MATRIX):** Uma matriz na qual todos os elementos abaixo da diagonal = 0.

**MÁXIMA VEROSSIMILHANÇA (MAXIMUM LIKELIHOOD):** O valor mais provável da distribuição de verossimilhanças é calculado pegando a derivada da verossimilhança e determinando quando a derivada é igual a 0. Para muitos métodos estatísticos paramétricos ou frequentistas, a estatística teste assintótica é o valor de máxima verossimilhança daquela estatística.

**MATRIZ DE DADOS:** É uma espécie de planilha eletrônica onde os dados, da maior parte dos testes estatísticos, devem ser introduzidos. Sinônimo: Grid geral.

**MATRIZ DE DECOMPOSIÇÃO DA VARIÂNCIA DOS COEFICIENTES DE REGRESSÃO:** Método para determinar a contribuição relativa de cada autovalor para cada coeficiente estimado. Se dois ou mais coeficientes são altamente associados com um dado autovalor que é o índice de condição, um nível inaceitável de multicolinearidade é apontado.

**MATRIZ DE ESTRUTURA FATORIAL:** Uma matriz fatorial obtida em uma rotação oblíqua que representa as correlações simples entre variáveis e fatores, incorporando a variância única e as correlações entre fatores. A maioria dos pesquisadores prefere usar a matriz de padrão fatorial no momento da interpretação de uma solução oblíqua.

**MATRIZ DE FATORES:** A matriz de coeficientes ( $a_{ij}$ ) que aparece na relação entre variáveis e fatores numa análise fatorial. Em inglês Factor matrix.

**MATRIZ DE IMPORTÂNCIA-DESEMPENHO:** Abordagem bidimensional para auxiliar o pesquisador na rotulação das dimensões. O primeiro eixo é das percepções da importância como por exemplo, como se mede em uma escala que vai de extremamente importante até nada importante) julgada pelo respondente. O outro eixo é de desempenho como, por exemplo, como se mede em uma escala que vai de altamente provável para desempenhar até altamente improvável para desempenhar para cada marca, ou produto ou serviço, sobre vários atributos. Cada objeto é representado por seus valores de importância e desempenho.

**MATRIZ DE PADRÃO FATORIAL:** Uma de duas matrizes fatoriais em uma rotação oblíqua que é mais comparável com a matriz fatorial em uma rotação ortogonal.

**MATRIZ DE PRECISÃO:** Inversa da matriz que contém as variâncias e covariâncias de um vetor de variáveis aleatórias. Em inglês Precision matrix.

**MATRIZ DE PROJEÇÃO:** Matriz que contém valores para cada observação na diagonal, conhecidos como valores de projeção, os quais representam o impacto da variável dependente observada sobre seu valor previsto. Se todos os casos têm igual influência, cada um tem um valor de  $p/n$ , onde  $p$  é o número de variáveis independentes + 1 e  $n$  é o número de casos. Se um caso não tem influência, seu valor é  $-\frac{1}{n}$ , e o total domínio por um único caso resultaria em um valor de  $-\frac{(n-1)}{n}$ . Valores que excedem  $\frac{2p}{n}$



para amostras maiores, ou  $\frac{3p}{n}$  para amostras menores ( $n \leq 30$ ), são candidatas à classificação como observações influentes.

**MATRIZ DE VARIÂNCIA-COVARIÂNCIA:** Para um vetor de variáveis aleatórias, a matriz de variância-covariância é aquela que contém as variâncias das variáveis na diagonal principal e as covariâncias entre essas variáveis fora da diagonal principal. Por exemplo, se tivermos três variáveis aleatórias envolvidas num estudo então a matriz pode ser escrita conforme a matriz M abaixo. Em inglês Variance-covariance matrix.

$$M = \begin{bmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \text{Cov}(X_1, X_3) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) & \text{Cov}(X_2, X_3) \\ \text{Cov}(X_3, X_1) & \text{Cov}(X_3, X_2) & \text{Var}(X_3) \end{bmatrix}$$

**MATRIZ DEFINIDA NEGATIVA:** Uma matriz A é definida negativa se  $b'Ab < 0$ , para todo vetor b não nulo. Em inglês Negative definite matrix.

**MATRIZ DEFINIDA POSITIVA:** Uma matriz A é definida positiva se  $b'Ab > 0$ , para todo vetor b não nulo. Em inglês Positive definite matrix.

**MATRIZ DIAGONAL:** É toda matriz quadrada em que os elementos não pertencentes à diagonal principal são todos nulos, e os da diagonal forem distintos de zero. Por exemplo:

$$K = \begin{bmatrix} 69 & 0 \\ 0 & -37 \end{bmatrix} \quad L = \begin{bmatrix} 95 & 0 & 0 \\ 0 & 43 & 0 \\ 0 & 0 & 26 \end{bmatrix}$$

**MATRIZ ESTOCÁSTICA:** Uma matriz quadrada  $P = (p_{ij})$  é chamada matriz estocástica, se cada uma de suas linhas é um vetor de probabilidade, isto é, se cada entrada de P é não negativa e a soma das entradas em cada linha é igual a um (1). Considere as seguintes matrizes.

$$A = \begin{pmatrix} \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{3}{4} & \frac{1}{2} & -\frac{1}{4} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}, \quad B = \begin{pmatrix} \frac{1}{4} & \frac{3}{4} \\ \frac{1}{3} & \frac{1}{3} \end{pmatrix} \quad e \quad C = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{2} & \frac{1}{6} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} & 0 \end{pmatrix}$$

A matriz A não é uma matriz estocástica, pois a entrada na segunda linha e terceira coluna é negativa, já a matriz B não é uma matriz estocástica, pois a soma das entradas na segunda linha não é igual a 1, e a matriz C é uma matriz estocástica, pois cada linha é um vetor de probabilidade. Se A e B são matrizes estocásticas, então o produto AB é uma matriz estocástica. Consequentemente, todas as potências de A<sup>n</sup> são matrizes estocásticas.

**MATRIZES ESTOCÁSTICAS REGULARES:** Uma matriz estocástica P é considerada regular se todas as entradas de alguma potência P<sup>m</sup> são positivas. Ver matriz estocástica.

**MATRIZ FATORIAL:** Tabela das cargas fatoriais de todas as variáveis sobre cada fator.

**MATRIZ IDENTIDADE:** É toda matriz especial diagonal em que os elementos da diagonal principal são iguais a 1, e denotada por  $I_m$ . Para indicar uma matriz identidade de ordem n, utilizamos a notação:

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

**MATRIZ IDENTIDADE:** Uma matriz quadrada é dita identidade se todos os elementos da diagonal principal forem iguais a um e os demais iguais a zero. Em inglês identity matrix.

**MATRIZ LINHA:** É toda matriz do tipo  $1 \times n$  ( $n \in \mathbb{R}^*$ ).

$$A = [a_1 \quad a_2 \quad a_3 \quad \dots \quad a_{1n}]. \quad \text{Por exemplo: } A = [3 \quad -8 \quad 5]_{1 \times 3} \quad e \quad B = [5 \quad 7 \quad 3 \quad 2]_{1 \times 4}.$$

**MATRIZ NÃO SINGULAR:** Matriz que possui um determinante não nulo. Em inglês Non-singular matrix.

**MATRIZ NULA:** É toda matriz do tipo  $m \times n$  cujos elementos são todos nulos. Para indicar uma matriz nula utiliza-se a seguinte notação:

$$O_{2 \times 2} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \quad e \quad O_{3 \times 2} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

**MATRIZ ORTOGONAL:** Uma matriz quadrada A de ordem p é ortogonal se  $A'A = AA' = I_p$ , sendo  $I_p$  a matriz identidade de ordem p. Em inglês Orthogonal matrix.

**MATRIZ QUADRADA:** É toda matriz cujo número de linhas é igual ao número de colunas. Assim, chamamos matriz quadrada de ordem n toda matriz do tipo  $n \times n$ . Por exemplo:

$$A = \begin{bmatrix} 12 & 22 \\ 60 & 35 \end{bmatrix} \quad e \quad B = \begin{bmatrix} 2 & 3 & -8 \\ 4 & 5 & -2 \\ 0 & 15 & 8 \end{bmatrix}.$$

**Toda matriz quadrada possui duas diagonais:** a diagonal principal, composta por elementos  $a_{ij}$  tais que  $i = j$ , isto é:  $a_{11}, a_{22}, a_{33}, \dots, a_{nn}$  e a diagonal secundária, em que os elementos  $a_{ij}$  são tais que,  $i + j = n + 1$ .

**MATRIZ QUADRADA:** Uma matriz é dita quadrada se o número de linhas é igual ao número de colunas. Em inglês square matrix.

**MATRIZ SIMÉTRICA:** Seja uma matriz A com valores  $a_{ij}$ . A é uma matriz simétrica se  $a_{ij} = a_{ji}$ , para todo i e j. Em inglês Symmetric matrix.

**MATRIZ SINGULAR:** Matriz que possui um determinante igual a zero. Em inglês Singular matrix.

**MATRIZ TRIANGULAR SUPERIOR:** Matriz que contiver apenas elementos nulos abaixo da diagonal.

**MATRIZ TRIANGULAR INFERIOR:** Matriz que contiver apenas elementos nulos acima da diagonal.

**MATURACÃO:** Mudanças nos indivíduos ocorridas durante um experimento ao qual não estão relacionadas, mas que podem afetar as respostas dos indivíduos ao fator experimental.



**MATURAÇÃO (AMEAÇA À VALIDADE INTERNA):** Esta ameaça à validade interna de uma pesquisa ocorre quando há mudanças de natureza física e ou psicológica nos sujeitos, geralmente, no período de tempo transcorrido entre preteste e posteste. Tais mudanças podem afetar os desempenhos dos sujeitos na variável dependente.

**MÁXIMO:** É o maior valor encontrado numa determinada variável.

**MÁXIMO (Max):** É o maior valor de um conjunto de dados. Em inglês maximum.

**MASCARADO:** Consultar teste cego.

**MÉDIA:** i) Do inglês average, uma medida de posição de um conjunto de observações que indica uma tendência central; por exemplo, média, mediana e moda; ii) Do inglês mean, abreviação de média aritmética uma medida de posição; é a soma das observações dividida pelo número de observações do conjunto.

**MÉDIA ARITMÉTICA:** Consultar média.

**MÉDIA GEOMÉTRICA:** Uma medida de posição. É o antilog da média aritmética dos valores de uma variável transformados em log.

**MEDIANA:** Uma medida de posição. É o valor central do conjunto de observações que foram dispostas em ordem de postos.

**MEDICINA VETERINÁRIA BASEADA EM EVIDÊNCIA (MVBE):** É o uso cauteloso, explícito e criterioso da melhor evidência atual, a fim de tomar decisões acerca do tratamento de animais, individualmente.

**MEDIDA DE CONCORDÂNCIA:** Indica quão bem os pares de observações concordam um com o outro; pode ser calculada como duas vezes o desvio-padrão (DP) das diferenças entre pares de medidas.

**MEDIDA SUMÁRIA:** Um valor que reduz um conjunto de medidas a um valor único que representa uma importante característica do conjunto de dados; por exemplo, a resposta máxima em um conjunto de medidas ao longo do tempo, para um indivíduo. O uso de uma medida sumária simplifica consideravelmente a análise de medidas de dados repetidos por reduzir uma série de valores de cada indivíduo em um único valor.

**MEDIDAS DE DISPERSÃO:** Indicam quão dispersas estão as observações, em qualquer direção, em relação à sua média.

**MEDIDAS DE POSIÇÃO:** Alguma forma de média que indica a tendência central do conjunto de dados.

**MEDIDAS NUMÉRICAS:** Uma característica que assume um valor quantitativo.

**METANÁLISE (PANORAMA GERAL):** Uma abordagem sistemática para combinar a informação de vários estudos independentes de uma dada condição, de modo que se obtenha um sumário estatístico dos resultados numéricos.

**MÉTODO:** Sequência ordenada de mecanismos que norteiam a concretização de um objetivo.

**MÉTODO CAPTURE-TAG-RECAPTURE:** Um método de estimar o tamanho de uma população de animais selvagens.

**MÉTODO DOS QUADRADOS MÍNIMOS:** Uma técnica matemática empregada para obter a melhor linha de ajuste em uma série de pontos. Depende da minimização da soma de resíduos ao quadrado.

**METODOLOGIA CIENTÍFICA:** Conjunto que envolvem várias técnicas e métodos científicos usadas pelos pesquisadores em geral.

**MÉTODO MANTEL-HAENSZEL:** É o procedimento correto para combinar os resultados contidos em tabelas de contingência que se relacionam a diferentes estratos da população.

**MÉTODO CIENTÍFICO:** Conjunto de ações validadas e aceitos por um conjunto de pesquisadores científicos que irá assegurar a qualidade e a confiabilidade do conhecimento obtido.

**MÉDIA (AVERAGE-MEAN):** Em inglês, existem duas palavras que podem ser traduzidas como média, average e mean. Portanto, a média pode ter 2 significados: i) no contexto de average, indica a média aritmética de uma variável. Com frequência é usada de maneira incorreta para indicar o valor mais comum de uma variável. Comparar com moda, mediana e com o segundo significado de média. ii) no contexto de mean, é o valor mais provável de uma variável aleatória ou de um conjunto de observações. A média deve ser classificada em: aritmética, geométrica ou harmônica. Se não for classificada, assume-se que é a média aritmética. Comparar com moda e mediana.

**MÉDIA AJUSTADA (ADJUSTED MEAN):** O valor esperado de um grupo tratamento caso todos os grupos tenham a mesma covariável média.

**MÉDIA ARITMÉTICA (ARITHMETIC MEAN):** É a média de um conjunto de observações de uma dada variável. É uma medida de tendência central calculada como a soma de todas as observações dividida pelo número de observações, e é simbolizada por uma barra horizontal, ou um acento circunflexo sobre a letra m para denominar o nome deste estimador, por exemplo,  $\bar{Y}$  ou  $\hat{m}$ .

**MÉDIA GEOMÉTRICA (GEOMETRIC MEAN):** É a raiz de ordem n ou a n-ésima raiz do produto de n observações. Uma média útil da expectativa para variáveis log-normal. A média geométrica é sempre menor que ou igual à média aritmética.

**MÉDIA HARMÔNICA (HARMONIC MEAN):** É o inverso da média aritmética dos inversos das observações. A média harmônica é uma estatística útil para calcular o tamanho efetivo da população ou o tamanho de uma população com acasalamentos completamente aleatórios. A média harmônica sempre é menor que ou igual à média geométrica.

**MODA:** A observação que ocorre mais frequentemente em um conjunto de observações.

**MODELO:** Descrição algébrica da relação entre duas ou mais variáveis.

**MODELO DETERMINÍSTICO:** Modelo que estabelece relação racional a qual pode ser lógica ou matemática precisa entre as variáveis envolvidas, devido ao grau de conhecimento relativamente alto do sistema sob estudo.

**MODELO PROBABILÍSTICO:** Modelo que procura explicar ou prever um resultado, em geral usando probabilidades, sem se preocupar com as especificações de todos os agentes causais ou determinantes desse resultado. Também denominado de não determinístico, de estatístico, de estocástico, ou, ainda, de aleatório.

**MODELO DE DEFINIÇÃO DE PROBABILIDADE:** Consultar probabilidade a priori.

**MODELO DE EFEITOS ALEATÓRIOS:** Um modelo de regressão para uma estrutura de dados hierárquicos em que o efeito aleatório para um modelo de dois níveis é a fonte de erro atribuída às unidades de nível 2. Também conhecido como modelo multinível, modelo de momentos sequenciais transverso, modelo hierárquico ou modelo misto.

**MODELO DE INTERCEPTAÇÃO ALEATÓRIA:** Uma forma particular de modelo de efeitos aleatórios em que, para uma estrutura de dois níveis ou seja, unidades de nível 1 dentro dos conglomerados, as equações de regressões lineares para os conglomerados têm a mesma inclinação, mas com intercepções que variam aleatoriamente ao redor da intercepção média.

**MODELO DE REGRESSÃO DE RISCO PROPORCIONAL DE COX:** Um procedimento avançado de regressão para testes de sobrevivência, utilizado quando se pretende investigar o efeito de diversas variáveis na sobrevivência, bem como quando desejamos considerar os dados censurados.

**MODELO DE REGRESSÃO DE RISCOS PROPORCIONAIS:** Consultar modelo de regressão de risco proporcional de Cox.

**MODELO DE REGRESSÃO MULTIVARIÁVEL:** Um modelo matemático que compreende uma variável dependente e duas ou mais variáveis explicativas.

**MODELO DE REGRESSÃO UNIVARIÁVEL:** Um modelo matemático que inclui uma variável dependente e uma variável explicativa.

**MODELO DE SENSIBILIDADE:** Indica o quanto as estimativas dos parâmetros em um modelo de regressão são influenciadas pelos dados de um ou mais indivíduos.

**MODELO HIERÁRQUICO:** Consultar modelo de efeitos aleatórios.

**MODELO LINEAR GENERALIZADO (MLG):** Uma forma geral para um modelo de regressão em que o valor médio da variável resposta, que tem uma distribuição de probabilidade conhecida está relacionado, via função de ligação, com uma combinação linear das variáveis explicativas.

**MODELO MISTO:** Consultar modelo de efeitos aleatórios.

**MODELO MULTINÍVEL (MME):** Consultar modelo de efeitos aleatórios.

**MODELO SATURADO:** Envolve um número de variáveis que é igual a, pelo menos, o número de indivíduos.

**MULTICOLINEARIDADE:** Consultar colinearidade.

**MULTICOLINEARIDADE:** Este termo foi registrado por Frisch (1934), e significava originalmente a existência de uma perfeita ou exata relação linear entre algumas ou todas as variáveis explicativas de um modelo de regressão. Para a regressão de  $k$  variáveis envolvendo as variáveis explicativas  $X_1, X_2, \dots, X_k$  (em que  $X_1 = 1$  para todas as observações, permitindo o termo do intercepto), dizemos que existe uma relação linear exata se for satisfeita a seguinte condição:  $\lambda_1 X_1 + \lambda_2 X_2 + \dots + \lambda_k X_k = 0$ , em que  $\lambda_1, \lambda_2, \dots, \lambda_k$  são constantes de modo que nem todos sejam simultaneamente iguais a zero. Atualmente, porém, o termo multicolinearidade é usado em um sentido mais amplo, para incluir o caso da perfeita multicolinearidade, como mostrado em  $\lambda_1 X_1 + \lambda_2 X_2 + \dots + \lambda_k X_k = 0$ , bem como o caso em que as variáveis  $X$  são intercorrelacionadas, mas não tão perfeitamente, como segue:  $\lambda_1 X_1 + \lambda_2 X_2 + \dots + \lambda_k X_k + \nu_i = 0$ ,

em que  $\nu_i$  é um termo de erro estocástico. Vale lembrar que se houver somente duas variáveis explicativas, a intercorrelação pode ser medida pelo coeficiente de correlação linear simples ou de ordem zero. Mas se houver mais de duas variáveis  $X$ , a intercorrelação pode ser medida pelos coeficientes de correlação parcial ou pelo coeficiente de correlação múltipla  $R$  de uma variável  $X$  com todas as outras variáveis  $x$  tomadas conjuntamente.

**MVBE:** Consultar medicina veterinária baseada em evidências (MVBE).

**MEDIANA (MEDIAN):** O valor de um conjunto de observações que está exatamente no centro de todos os valores: metade dos valores são menores e metade são maiores. O cento do quinto decil, ou o 50-ésimo percentil. Comparar com média e moda.

**METADADOS (METADATA):** Dados sobre dados. A prosa descritiva que descreve todas as variáveis e características de um conjunto de dados.

**MÉTODO DE BONFERRONI (BONFERRONI METHOD):** Um ajuste do valor crítico,  $\alpha$ , ao qual a hipótese nula será rejeitada. Ele é usado quando são feitas observações múltiplas, e é calculado como o  $\alpha$ , geralmente igual a 0,05, dividido pelo número de comparações. Comparar com o método de Dunn-Sidak.

**MÉTODO DE DUNN-SIDAK (DUNN-SIDAK METHOD):** Um ajuste do valor crítico,  $\alpha$ , no qual a hipótese nula será rejeitada. É usado quando múltiplas comparações são feitas e é calculado como  $\alpha_{ajustado}$  é igual a  $1 - (1 - \alpha_{nominal})^{\frac{1}{k}}$ , onde  $\alpha_{nominal}$  é o valor crítico do  $\alpha$ , geralmente 0,05 e  $k$  é o número de comparações. Comparar com o método de Bonferroni.

**MÉTODO HIPOTÉTICO DEDUTIVO (HYPOTHETICO DEDUCTIVE METHOD):** O método científico creditado a Karl Popper. Como na indução, o método hipotético dedutivo se inicia com uma observação individual. A observação pode ser predita por muitas hipóteses, cada uma das quais fornece predições adicionais. As predições são testadas uma a uma e as hipóteses alternativas vão sendo eliminadas até que reste apenas uma. Os praticantes do método hipotético-dedutivo nunca confirmam uma hipótese; eles apenas as rejeitam, ou falham em rejeitá-las. Quando os livros-texto se referem ao método científico, geralmente estão se referindo ao método hipotético dedutivo. Comparar com indução, dedução e inferência bayesiana.

**MÉDIA:** Valor igual à soma de uma lista de valores dividida pelo número de valores; a média de uma variável aleatória é o valor médio que resultaria se observássemos a variável muitas vezes o mesmo que esperança, sendo simbolizada pela letra grega  $\mu$  (mu).

**MÉDIA:** Em sua acepção mais ampla, qualquer promédio de uma distribuição de freqüências. Em sentido estrito, média aritmética. Em inglês mean.

**MÉDIA:** A soma dos valores para todas as observações de uma variável dividida pelo número de observações.

**MÉDIA:** A média é uma medida de tendência central. A média de um conjunto é calculada como o quociente da soma de todos os valores pelo número de valores somados. A fórmula da média para um conjunto de valores  $x_1, x_2, x_3, \dots, x_n$ , é então,

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{\sum_{i=1}^n X_i}{n}$$

Por convenção, usa-se a letra grega  $\mu$  (miu) para a média da população e  $\bar{X}$  para a média da amostra. Dada a sua natureza, a média apenas se deve usar em situações de distribuição simétrica pois é uma medida bastante sensível a valores extremos. Neste caso pode-se optar pela mediana já que é uma medida mais estável.

**MÉDIA:** É o parâmetro  $m$  ou  $\mu$  de uma distribuição normal.

**MÉDIA:** Valor intermediário, calculado como sendo a soma de todos os valores observados dividida pelo número total de observações.

**MÉDIA:** Medida computada somando os valores de diversas observações e dividindo pelo número de observações.

**MÉDIA:** Termo ambíguo, normalmente sugerindo típico ou normal. Exemplos específicos de médias são a média aritmética, a moda e a mediana.

**MÉDIA:** Qualquer uma das várias medidas destinadas a revelar a tendência central de uma coleção de dados. Em inglês average.

**MÉDIA (mean):** Soma de um conjunto de valores dividida pelo número respectivo.

**MÉDIA:** Geralmente utiliza-se média como abreviação de média aritmética que é a soma dos valores observados, dividida pelo número de elementos considerados na soma. Em inglês Mean.

**MÉDIA A PRIORI:** Valor esperado do parâmetro usando a distribuição a priori. Em inglês Prior mean.

**MÉDIA AMOSTRAL:** É a média calculada para uma amostra particular. Em inglês Sample mean.

**MÉDIA AMOSTRAL:** É a estatística definida por:  $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$ .

**MÉDIA AMOSTRAL ( $\bar{X}$ ):** O símbolo entre parêntese se lê X barra ou X traço, e é uma variável aleatória, função dos valores da amostra, é definida como a soma de todos os valores da amostra dividido pelo número de observações da amostra. Serve para estimar a média populacional.

**MÉDIA (ARITMÉTICA):** O quociente da soma dos valores pelo número de mensurações; informa onde está localizado o valor central do grupo.

**MÉDIA APARADA:** Se  $0 < P < 1$ , uma média aparada a 100P % é obtida eliminando-se 100P % das menores observações e 100P % das maiores observações e calculando-se a média aritmética dos restantes dos valores. Por exemplo, se tivermos 10 observações ordenadas  $x_{(1)} < x_{(2)} < \dots < x_{(10)}$ , a média aparada, podada ou truncada a 10% é determinada pela fórmula:

$$\bar{X}(0,10) = \frac{x_{(2)} + x_{(3)} + \dots + x_{(9)}}{8}. \text{ Se } P = 0,25, \text{ então, } \bar{X}(0,25) \text{ é denominada meia-média.}$$

**MÉDIA ARITMÉTICA:** Dos N dados de uma amostra é a soma desses dados dividida por N. Geralmente se indica por  $\hat{m}$ , notação que indica tratar-se de uma estimativa da média verdadeira  $m$ , desconhecida.

**MÉDIA ARITMÉTICA:** Soma de todos os escores dividida pelo número de escores.

$$M = \frac{\sum_{i=1}^N X}{N}$$

Em inglês arithmetic mean.

**MÉDIA ARITMÉTICA:** É calculada como:  $\bar{X} = \sum_{i=1}^n P(X_i) X_i$ . Porém, assumindo-se que as probabi-

lidades de ocorrência das n amostras são iguais entre si, tem-se:  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ , que é a equação usualmente empregada na estatística.

**MÉDIA ARITMÉTICA:** Medida de tendência central de largo emprego em estatística, representada pela razão do somatório dos valores pelo número de escores.

**MÉDIA (MÉDIA ARITMÉTICA):** O centro matemático dos dados, levando em conta não só a localização dos dados acima ou abaixo do centro, mas também sua distância relativa ao centro.

**MÉDIA ARITMÉTICA ( $\bar{X}$ ):** É igual ao quociente entre a soma dos valores do conjunto e o número total dos valores.

**MÉDIA ARITMÉTICA:** Medida de tendência central com características de que uma relação empírica de esperança matemática está condicionada às exigências dos teoremas limites que justificam a interpretação de frequência relativa. Essa esperança ou média depende do conjunto de valores da série, uma vez que todos eles entram em seu cálculo com igual peso, sendo uma condição para que seja valor representativo. Isto significa que está univocamente determinada em toda a série; pode ser tratada algebricamente e é afetada por valores extremos muitos grandes ou muitos pequenos. As principais propriedades da média são: a soma algébrica dos desvios em relação à média é zero; a soma dos quadrados dos desvios em relação a qualquer número  $X_0$  é o mínimo quando a  $X_0$  = média aritmética. Tais características se ajustam (teste) aos dados de pesquisa? A média tem como aplicação de ser muito utilizada para sintetizar dados de uma série. A estimativa é única. Deve ser aplicada para dados que procedem de uma distribuição normal e/ou estatisticamente aceita como tal, existem testes estatísticos para este propósito.

**MÉDIA ARITMÉTICA ( $\bar{X}$ ):** Soma de um conjunto de valores dividida pelo respectivo número; referido em geral como média.

**MÉDIA ARITMÉTICA SIMPLES:** É o quociente da soma de todos os dados não classificados pelo número desses dados. Com n = número de observações da amostra.

**MÉDIA ARITMÉTICA PONDERADA:** É o quociente entre o somatório do produto de cada dado classificado pela sua frequência absoluta e o número desses dados. Com n = número de observações da amostra.

**MÉDIA DE UMA VARIÁVEL ALEATÓRIA:** Também chamada de valor esperado, valor obtido ao longo prazo da experimentação. Em inglês Mean of a random variable.

**MÉDIA DOS QUADRADOS:** Outro nome da variância, que é definida como sendo a soma dos quadrados dividida pelo número apropriado de graus de liberdade.

**MÉDIA DOS QUADRADOS DENTRO DOS GRUPOS:** O mesmo que Variância dentro dos grupos.



**MÉDIA DOS QUADRADOS ENTRE OS GRUPOS:** O mesmo que Variância entre os grupos.

**MÉDIA EXTREMA:** Extremo de um conjunto de médias aritméticas.

**MÉDIA GEOMÉTRICA:** É a raiz n-ésima do produto de todos eles.

**MÉDIA GEOMÉTRICA:** Medida de locação que consiste em calcular a raiz n-ésima do produto das quantidades amostrais. Em inglês Geometric mean.

**MÉDIA GEOMÉTRICA:** Média de tendência central representada pela raiz n do produto dos escores e indicada, sobretudo, quando os dados estão dispostos em progressão geométrica.

**MEDIA GEOMÉTRICA:** Medida de tendência central com características de que sua principal vantagem é a de que o produto da razão dos desvios dela é igual à unidade. Tem como limitações o fato de que todos os números da série devem ser positivos e não nulos. É aplicada para médias de razões e para médias de proporções de variação, em números índices. Indispensável para calcular valores intermediários de séries com tendência exponencial.

**MÉDIA GEOMÉTRICA:** É definida como sendo a raiz enésima do produtório de um conjunto de n observações:

$$\bar{X}_G = \sqrt[n]{\prod_{i=1}^n X_i} = \sqrt[n]{X_1 \cdot X_2 \cdot \dots \cdot X_n} . \text{ Contudo, a equação apresentada nem sempre é eficiente em termos}$$

computacionais, pois frequentemente o resultado da produtório pode exceder a capacidade máxima de armazenamento do computador, sendo comum encontrar:

$$\bar{X}_G = e^{\bar{\mu}}, \text{ onde: } \bar{\mu} \text{ é a média dos logaritmos naturais das observações;}$$

As duas equações são perfeitamente equivalentes, como se demonstra. Tomando-se o logaritmo natural da primeira expressão, tem-se:  $\ln(\bar{X}_G) = \frac{1}{n}[\ln X_1 + \ln X_2 + \dots + \ln X_n]$

$$\text{Como: } \bar{\mu} = \frac{1}{n}[\ln X_1 + \ln X_2 + \dots + \ln X_n]. \text{ Então, } \ln(\bar{X}_G) = \bar{\mu} \Rightarrow \bar{X}_G = e^{\bar{\mu}}.$$

**MÉDIA GLOBAL:** Média de todos os elementos em um teste de análise de variância.

**MÉDIA HARMÔNICA:** A média harmônica H de um conjunto de dados é o recíproco da média aritmética de seus recíprocos. Pode ser escrita no caso de dados discretos para n quantidades  $x_1, \dots, x_n$  como como:  $H^{-1} = n^{-1} \sum_{i=1}^n (x_i)^{-1}$ . Em inglês Harmonic mean..

**MÉDIA HARMÔNICA:** Medida de tendência central para grandezas inversamente proporcionais como, por exemplo, tempo e velocidade.

**MÉDIA HARMÔNICA PONDERADA:** De uma coleção de n números não-nulos  $x_i$ , aos quais se acham associados pesos respectivos  $p_i$ , é o quociente da soma dos pesos pela soma dos quocientes dos pesos pelos valores que eles afetam.

$$\bar{X}_{HP} = \frac{\sum_{i=1}^n p_i}{\sum_{i=1}^n \left( \frac{p_i}{x_i} \right)}$$

**MÉDIA MÓVEL:** Se uma série de tempo é dada por  $x_1, \dots, x_n$  e se existem pesos dados por  $w_0, w_1, \dots, w_k$ , de forma que a soma desse pesos é 1 (um), a série de valores  $u_t = \sum_{j=0}^k w_j x_{t+j}$ ,  $t=1, 2, \dots, n-k$ , é chamada de série de médias móveis. Em inglês Moving average.

**MÉDIA MÓVEL:** É qualquer processo de filtragem de uma seqüência de dados, seja ela uni ou bidimensional, com o objetivo de promover a suavização da seqüência de dados. Para uma seqüência unidimensional pode-se ter como exemplos:

$$X_{m_i} = \frac{X_{i-1} + 2 \cdot X_i + X_{i+1}}{3} \text{ ou } X_{m_i} = \frac{0,5 \cdot X_{i-2} + X_{i-1} + 2X_i + X_{i+1} + 0,5 \cdot X_{i+2}}{5}$$

$$\text{E para uma seqüência bidimensional: } X_{m_{i,j}} = \frac{X_{i,j-1} + 2 \cdot X_{i,j} + X_{i,j+1} + X_{i-1,j} + X_{i+1,j}}{6}$$

Observe-se que diferentes efeitos serão produzidos conforme os ponderadores escolhidos, bem como com relação ao número de termos, quanto maior a janela, maior será a suavização produzida.

**MÉDIA MÓVEL:** Valor médio das observações em uma série temporal, que estão mais próximas de um valor particular.

**MÉDIA PONDERADA:** Média de uma coleção de valores aos quais foram atribuídos diferentes graus de importância.

**MÉDIA POPULACIONAL (M):** É o valor que representa um conjunto de valores da população. Definida como a soma de todos os valores da população dividido pelo número de observações. Por exemplo: renda per capita de um país, esperança de vida, renda familiar média, pontuação média na escala de atitudes em relação à Estatística, dentre outros. Também simbolizada pela letra grega  $\mu$ , que é  $\mu$ .

**MÉDIA QUADRÁTICA:** É a raiz quadrada da média aritmética dos quadrados.

**MINITAB:** Programa, pacote estatístico ou software de computador utilizado para programação, tabulação de dados, cálculos matemáticos e estatísticos.

**MÉTODOS DE MONTE CARLO (MONTE CARLO METHODS):** Métodos estatísticos que recaem sobre a aleatorização ou o remanejamento dos dados, com frequência usando métodos de bootstrap ou de jackknife. O resultado de uma análise de Monte Carlo é um valor exato de probabilidade para os dados, dada a hipótese nula estatística.

**MÉTRICA (METRIC):** Um tipo de medida de distância ou dissimilaridade que satisfaz quatro propriedades: i) a distância mínima é igual a 0 e a entre objetos idênticos é igual a 0; ii) a distância entre dois objetos não idênticos é positiva; iii) as distâncias são simétricas: a do objeto a para o b é igual à distância do objeto b para o objeto a; e iv) as medidas das distâncias satisfazem a desigualdade dos triângulos. Comparar com semimétrica e não métrica.

**MÍNIMOS QUADRADOS APARADOS (LEAST-TRIMMED SQUARES):** Um método de regressão para minimizar o efeito de dados discrepantes. Calcula a inclinação da regressão após remover alguma proporção especificada pelo analista das maiores somas dos quadrados dos resíduos.

**MODA (MODE):** O valor mais comum em um conjunto de observações. Coloquialmente, a média ou média da população em geral é confundida com a moda. Comparar com média e mediana.



**MODELO (MODEL):** Uma função matemática que relaciona um conjunto de variáveis a outro por meio de um conjunto de parâmetros. Todos os modelos são errados, mas alguns são úteis.

**MEDIANA (Md):** É o valor da variável, para dados não classificados, que ocupa a posição central da distribuição.

**MEDIANA:** O valor da variável que representa o ponto médio dos dados. Metade deles têm valores abaixo da mediana e metade, acima desta medida.

**MEDIANA:** Valor tal que metade dos números de uma lista está acima dele e metade abaixo dele.

**MEDIANA:** Observação que se situa no meio quando os dados são dispostos em ordem, do valor mais baixo para o valor mais alto.

**MEDIANA:** O valor central quando as mensurações são alinhadas do menor ao maior valor; ponto ou posição que divide a distribuição em duas metades.

**MEDIANA:** Medida de tendência central com características de que seu valor não depende de valores particulares da série ou amostra, mas da densidade ou ordenação dos mesmos; esta característica resulta de grande interesse por seu claro e expressivo significado. A estatística não é afetada por valores extremos. Não utiliza muita informação do conjunto. Entretanto, a mediana não possui propriedades estatísticas interessantes que a média possui; é de difícil manejo algébrico; é difícil de calcular quando o número de observações é muito grande. Tem sua aplicação útil para pequenos conjuntos de observações que contém um ou mais valores extremos. É um valor típico da série que, em geral, se torna expressivo. Tem, com respeito à média, a vantagem de ser dado pelo número e ordem dos termos da série.

**MEDIANA (Md):** Medida de tendência central que separa os escores em dois grupos: 50 % inferiores e 50 % superiores à mediana.

**MEDIANA (Md):** É o valor situado de tal forma no conjunto que o separa em dois subconjuntos de mesmo número de elementos.

**MEDIANA:** A mediana é o valor da variável que divide a frequência total em duas metades iguais. Em inglês Median.

**MEDIANA (Md):** É o valor da variável, para dados não classificados, que ocupa a posição central da distribuição. Com  $n = n^\circ$  de observações da amostra.

**MEDIANA:** Medida representando o valor do caso no meio de um conjunto ordenado de observações. Se as idades de cinco homens são por exemplo 16, 17, 20, 54 e 88, então a mediana seria 20. A média seria 39.

**MEDIANA:** A mediana é uma medida de tendência central. É o valor que divide a distribuição em duas partes iguais; a metade dos valores estão abaixo da mediana e a outra metade acima desta. Seja  $x_1, x_2, x_3, x_4, x_5$  um conjunto de 5 valores ordenados crescentemente. Neste caso a mediana é  $x_3$ . No caso do número de valores do conjunto ser par, por exemplo,  $x_1, x_2, x_3, x_4$ , a mediana calcula-se como a média dos dois valores centrais:

$\frac{X_2 + X_3}{2}$ . A mediana apenas tem em conta a posição de um valor em relação a ela, se está acima ou

abaixo, independentemente da grandeza deste. Assim, apesar de ser menos informativa é mais estável do que a média.

**MEDIANA:** A observação sob a qual e enquadram 50 % das observações, ou seja 50% das observações são menores que a mediana e 50 % são menores que ela..

**MEDIANA:** Medida de tendência central por meio de cujo valor a distribuição de frequência é dividida em duas partes de igual frequência. Cinquenta por cento dos casos dos casos se situam acima e cinquenta por cento dos casos se situam abaixo da mediana. Em inglês median.

**MEDIANA (ESTATÍSTICA):** Em uma amostra, disposta por ordem crescente dos seus elementos, é o número do meio. No caso dessa amostra ter um número par de elementos a mediana será a média dos 2 centrais. Exemplo: a mediana da amostra: 1, 3, 4, 6, 7, 11, 23 é 6.

**MEDIANA AMOSTRAL:** É a mediana calculada para uma amostra particular. Em inglês Sample median.

**MEDIANA PONDERADA:** De um conjunto de valores aos quais se acham associados pesos é o valor da variável tal que a soma dos pesos dos valores que lhes são inferiores é igual à dos que lhe são superiores. O conceito acha-se registrado em Yule e Kendall (1940).

**MEDIÇÃO:** Processo fundamental de pesquisa, no qual unidades de análise são identificadas com atributos específicos nas variáveis estudadas. Por exemplo, o processo de decidir que alguns eleitores são liberais; outros conservadores dentre outros.

**MEDIDA (PROPRIEDADES MATEMÁTICAS):** As propriedades matemáticas de uma medida são: i) é expressa por um número positivo e real; ii) é completamente aditiva, isto é, a soma das medidas é a medida da soma: partes medidas independentemente ou como um todo têm que ter a mesma medida; iii) possui uma ordem garantida, isto é, um conjunto menor tem uma medida ainda menor sobre ela; iv) possui um zero absoluto.

**MEDIÇÃO DIRETA:** Técnica de coleta de dados que envolve a contagem, medição ou teste direto dos dados.

**MEDIDA COMPOSTA:** Ver escala múltipla.

**MEDIDA DA VARIÂNCIA EXTRAÍDA:** Quantia de variância compartilhada ou comum entre os indicadores ou variáveis manifestas para um construto. Valores maiores representam um maior grau de representação compartilhada dos indicadores com o construto.

**MEDIDAS DE ACHATAMENTO:** Medida de formas da distribuição de frequência com características de que é dada por um coeficiente  $a_4$  ou quarto momento que não depende da localização ou da variância da distribuição, podendo, por conseguinte, ser usado para fins de comparação. A distribuição normal, em forma de sino, e principalmente usada como referência para medida de achatamento, sendo seu valor referencial  $a_4=3$ . Tem como aplicação o fato de poderem ser especialmente úteis no ajustamento de curvas. E também para caracterizar séries de frequência, no tratamento preliminar de dados.

**MEDIDA DE ADEQUAÇÃO DA AMOSTRA (MSA):** Medida calculada tanto para toda a matriz de correlação quanto para cada variável individual e que permite avaliar o quão adequada é a aplicação da análise fatorial. Valores acima de 0,50 para a matriz toda ou para uma variável individual indicam tal adequação.

**MEDIDA DE AJUSTE ABSOLUTO:** Medida de ajuste geral para ambos os modelos estrutural e de mensuração coletivamente. Esse tipo de medida não faz qualquer comparação com um modelo nulo especificado que é a medida de ajuste incremental ou ajuste para o número de parâmetros no modelo estimado neste caso a medida de ajuste parcimonioso.

**MEDIDA DE AJUSTE INCREMENTAL:** Medida de qualidade de ajuste que compara o modelo corrente com um modelo nulo especificado para determinar o grau de melhora sobre o modelo nulo. Complementa os outros dois tipos de medidas de qualidade de ajuste, as medidas de ajuste absoluto e de ajuste parcimonioso.

**MEDIDA DE AJUSTE PARCIMONIOSO:** Medida de qualidade de ajuste geral representando o grau de ajuste do modelo por coeficiente estimado. Essa medida tenta corrigir qualquer superajuste e avalia a parcimônia do modelo comparada com a qualidade do ajuste.

**MEDIDA DE ASSOCIAÇÃO:** Um valor que fornece uma quantificação da força do relacionamento e direção entre duas ou mais variáveis. Em inglês Measure of association.

**MEDIDA DE ASSOCIAÇÃO:** Uma medida da força e direção da relação entre as duas variáveis.

**MEDIDA DE DISPERSÃO:** Uma das diversas medidas destinadas a refletir a quantidade de variação em um conjunto de valores.

**MEDIDA DE DEPENDÊNCIA ESPACIAL:** É a relação existente em um conjunto de valores, quando estes são comparados entre si segundo distâncias crescentes até uma distância máxima, numa dada direção, tal como medida pelo variograma. Contudo, a explicação matemática do variograma pode ser obtida como se expõe a seguir: Supondo um conjunto de pares de valores separados por uma distância  $h$ , pode-se fazer a comparação dos mesmos, lançando-os em diagramas de dispersão e medindo o quanto a nuvem de pontos distancia-se da reta 1:1. Obviamente a dispersão da nuvem de pontos aumenta com o aumento da distância  $h$ , pois os valores de amostras distantes tenderão a ser diferentes e, portanto, os pontos nos diagramas de dispersão estarão distantes da reta 1:1. Para o desenvolvimento da relação de dependência entre valores  $(x,y)$ , separados por uma distância  $h$ . A distância  $d_i$  entre o  $i$ -ésimo ponto  $(x_i, y_i)$  e a reta ideal é:  $d_i = |X_i - Y_i| \cdot \cos 45^\circ$ . Elevando-se ao quadrado, tem-se:  $d_i^2 = \frac{1}{2}(X_i - Y_i)^2$ . Havendo  $n$  pares de pontos, pode-se calcular o momento de inércia em torno da reta  $45^\circ$ , como:

$$\gamma_{XY} = \frac{1}{n} \sum_{i=1}^n d_i^2 = \frac{1}{n} \sum_{i=1}^n \frac{1}{2} (X_i - Y_i)^2 = \frac{1}{2n} \sum_{i=1}^n (X_i - Y_i)^2$$

Quanto maior a dispersão, maior o momento de inércia e menor a correlação. Se não houver dispersão, com todos os pares de pontos sobre a reta  $45^\circ$ , o momento de inércia é zero e o coeficiente de correlação é igual a 1 que é a máxima correlação. A última expressão não é nada mais que o próprio variograma e, sobretudo, a origem do semivariograma.

**MEDIDA DE DESAJUSTE:** Proporção da variância das disparidades dos dados otimamente escalonados não explicada pelo modelo MDS (escalonamento multidimensional). Esse tipo de medida varia de acordo

com o tipo de programa e de dados em análise. A medida de desajuste ajuda a determinar o número adequado de dimensões a serem incluídas no modelo.

**MEDIDAS DE DISPERSÃO:** É um conjunto de medidas como por exemplo, amplitude, variância e desvio padrão utilizadas no estudo da variabilidade de uma determinada distribuição, permitindo obter uma informação mais completa acerca da forma da mesma.

**MEDIDA DE INTERVALO:** Nível de medição que descreve uma variável cujos atributos são ordenados, com a distância igual entre atributos adjacentes. Exemplo: a escala fahrenheit de temperatura, já que a distancia entre  $17^\circ$  e  $18^\circ$  é a mesma que entre  $89^\circ$  e  $90^\circ$ . Ver medida nominal, medida ordinal e medida de razão.

**MEDIDA DE LOCAÇÃO:** Uma qualidade que pretende localizar uma distribuição ou um conjunto de valores amostrais por meio de quantidades que são, de alguma forma, centrais ou típicos. Em inglês Measure of location.

**MEDIDA DE RAZÃO:** Nível de medição descrevendo uma variável cujos atributos possuem todas as qualidades das medidas nominais, ordinais e intervalo, além de se basearem num genuíno ponto zero. Idade seria um exemplo.

**MEDIDA DE VARIABILIDADE:** Medidas de uma distribuição que indicam o quanto estão dispersos os dados numa distribuição.

**MEDIDAS DERIVADAS:** Procedimento para obter percepções de dados de similaridades por parte dos respondentes. Similaridades derivadas geralmente são baseadas em uma série de escores dados aos estímulos pelos respondentes, que são então combinados de alguma maneira. A escala diferencial semântica freqüentemente é usada para deduzir tais escores.

**MEDIDA MULTIVARIADA:** Uso de duas ou mais variáveis como indicadores de uma única medida composta. Por exemplo, um teste de personalidade pode oferecer as respostas a diversas questões individuais que são os indicadores, as quais são então combinadas para formar um escore único ou seja, uma escala múltipla, que representa o tipo de personalidade.

**MEDIDA NOMINAL:** Nível de medição descrevendo uma variável cujos atributos são apenas diferentes, em contraste com medidas ordinais, intervalo ou de razão. Sexo é um exemplo de medida nominal.

**MEDIDA ORDINAL:** Nível de medida que descreve uma variável cujos atributos podem ser ordenados ao longo de algum dimensão. Um exemplo seria status socioeconômico composto dos atributos alto, médio e baixo. Ver medida nominal, medida de intervalo e medida de razão.

**MEDIDAS DE DENSIDADE DE INCIDÊNCIA:** Medidas de frequência de eventos adversos em saúde que são utilizadas, por exemplo, para determinar o melhor momento para a administração de uma nova vacina e a duração da imunidade produzida.

**MEDIDAS DE DISPERSÃO:** São medidas que indicam o grau de variação ou dispersão dos valores de uma variável. As medidas de dispersão mais utilizadas são: desvio padrão, variância, âmbito máximo e mínimo

**MEDIDAS DE LOCALIZAÇÃO:** É um conjunto de medidas como por exemplo, média, mediana, moda e quartis, que representam de uma forma global um conjunto de dados.

**MEDIDAS DE POSIÇÃO:** São as estatísticas que representam uma série de dados orientando-nos quanto à posição da distribuição em relação ao eixo horizontal do gráfico da curva de frequência.

**MEDIDAS DE SUMÁRIO:** Uma das formas de resumir a informação é usar determinadas medidas ditas de sumário. Estas medidas ainda podem ser divididas em medidas de tendência central, que indicam o centro da distribuição e medidas de dispersão, que indicam o grau de variabilidade dos dados. As medidas de sumário mais usadas são: medidas de tendência central tais como média, mediana e moda e as medidas de dispersão o desvio padrão, o âmbito, a variância, o valor máximo e o valor mínimo.

**MEDIDA DE TENDÊNCIA CENTRAL:** Estatísticas que fornecem um número resumido para caracterizar o que é típico ou médio para determinados dados. Média, moda e mediana são as três medidas de tendência central.

**MEDIDAS DE ASSIMETRIA:** Medida de formas da distribuição de frequência com características de que é dada por um coeficiente  $a_3$  ou terceiro momento utilizado para escolha nas comparações. A média, a mediana e a moda serão iguais quando a distribuição de frequência for simétrica ou normal. Tem como aplicação o fato de poderem ser especialmente úteis no ajustamento de curvas. E também para caracterizar séries de frequência, no tratamento preliminar de dados.

**MEDIDAS DE TENDÊNCIA CENTRAL:** Valores que, em estatística, caracterizam os valores médios. Em inglês measures of central tendency.

**MEDIDAS DE TENDÊNCIA CENTRAL:** As medidas de tendência central englobam-se dentro das medidas de posição e indicam o centro da distribuição de uma determinada variável. As principais medidas de tendência central são: média, mediana e a moda.

**MEDIDAS DE TENDÊNCIA CENTRAL OU PROMÉDIAS:** O mesmo que medidas de localização.

**MEDIDAS DE TENDÊNCIA CENTRAL OU PROMÉDIAS:** Verifica-se uma tendência dos dados observados a se agruparem em torno dos valores centrais. As mais utilizadas são: média aritmética, moda e mediana. Outros promédios menos usados são as médias: geométrica, harmônica, quadrática, cúbica e biquadrática.

**MEDIDAS REPETIDAS:** Uso de duas ou mais variáveis respostas de um único indivíduo em uma análise de variância univariada (ANOVA) ou em uma análise de variância multivariada (MANOVA). O propósito de um delineamento de medidas repetidas é controlar as diferenças de nível individual que possam afetar a variância interna no grupo. As medidas repetidas são uma forma de falta de independência do respondente.

**MEDIDAS SOCIOMÉTRICAS:** Destinadas a obter informação sobre escolhas sociais e interações entre pessoas, a fim de satisfazer necessidades da pesquisa de um grupo. Basicamente, elas avaliam atrações e ou repulsas em um dado grupo.

**MICRONUMEROSIDADE:** Conforme Goldberger (1991), ocorre micronumerosidade exata a contrapartida da multicolinearidade exata quando  $n$ , o tamanho da amostra, é zero, caso em que se torna impossível qualquer espécie de estimativa. Quase micronumerosidade, tal como a quase multicolinearidade, ocorrerá quando o número de observações mal superar o número de parâmetros a serem estimados num estudo de regressão.

**MEDIDOR DE PESSOAS:** Um sistema de classificação computadorizada por microondas que transmite informações demográficas da noite para o dia para medir audiências nacionais da televisão.

**MEIOS IRMÃOS:** Animais que possuem o mesmo pai ou a mesma mãe.

**MELHOR ESTIMADOR:** A estimação de parâmetros de uma população a partir da informação proveniente de uma amostra levanta a questão de se existe o melhor estimador. A resposta depende principalmente dos critérios que classificam a bondade de um estimador. Se existe um critério que diferencia um estimador de outro como o melhor, e se existe esse estimador, então ele é chamado de melhor estimador. Em inglês Best estimator.

**MELHOR ESTIMADOR LINEAR NÃO-VICIADO:** Um estimador é dito estimador linear se ele é uma combinação linear das observações amostrais. É dito o melhor estimador linear não-viciado se de todas as combinações lineares da amostra ele apresentar a menor variância. Em inglês best linear unbiased estimator.

**MELHOR ESTIMATIVA:** Estimativa obtida com o modelo estatístico que produz o menor somatório dos erros elevados ao quadrado.

**MENOR OU IGUAL A:** Símbolo matemático dado por  $\leq$ .

**MENOR QUE:** Símbolo matemático dado por  $<$ .

**MENSURAÇÃO:** Designação de numerais a objetos ou eventos, segundo regras específicas. Refere-se simplesmente à designação de número a observações particulares, de tal forma a refletir as variações entre as observações.

**MENSURAÇÃO:** O processo de designar números ou rótulos a pessoas ou objetos de acordo com regras específicas para representar quantidades ou qualidades de atributos.

**MENSURAÇÃO COMPLETA:** Os postulados subjacentes ao sistema de número real podem ser agrupados em três classes:

Postulados relativos a equivalência

i) Ou  $A = B$  ou  $A \neq B$  (equivalência non-equivalência); ii) Se  $A = B$ , então  $B = A$  (simetria); iii) Se  $A = B$  e  $B = C$ , então  $A = C$  (transitividade)

Postulados relativos à ordinalidade

i) Se  $A \neq B$ , então  $A > B$  ou  $A < B$  (conexão); ii) Se  $A > B$ , então  $B > A$  (assimetria); iii) Se  $A > B$ , e  $B > C$ , então  $A > C$  (transitividade)

Postulados relativos à aditividade

i)  $A + B = B + A$  (lei comutativa); ii) Se  $A = P$ ,  $B = Q$ , então  $A + B = P + Q$  (axioma de iguais); iii)  $A + (B + C) = (A + B) + C$  (lei associativa).

**MENSURAÇÃO IRREATIVA:** Uma mensuração é irreativa quando em nada altera aquilo que está sendo medido. Um exemplo é contar o número de vezes que os alunos levantam suas mãos, sorriem ou participam em discussões de classe, como índices de interesse numa dada aula ou palestra.



**MODELO DE REGRESSÃO LINEAR (LINEAR REGRESSION MODEL):** Um modelo estatístico no qual a relação entre o preditor e a variável resposta pode ser ilustrada por uma linha reta. Um modelo é dito linear porque seus parâmetros são constantes e afetam a variável preditora apenas aditiva ou multiplicativamente. Comparar com modelo de regressão não linear.

**MODELO DE REGRESSÃO NÃO LINEAR (NON-LINEAR REGRESSION MODEL):** Um modelo estatístico no qual a relação entre as variáveis preditora e resposta não é uma linha reta e que as variáveis preditoras afetam a variável resposta de forma não aditiva ou não multiplicativa. Comparar com modelo de regressão linear.

**MODELO DE SÉRIE TEMPORAL (TIME SERIES MODEL):** Um modelo estatístico em que o tempo é incluído de maneira explícita, como uma variável preditora. Muitos modelos de séries temporais também são modelos autorregressivos.

**MODELO FATORIAL (FACTOR MODEL):** O resultado de uma análise fatorial após os fatores comuns de suas cargas correspondentes terem sido padronizadas e da retenção apenas dos fatores comuns úteis ou significativos.

**MODELO LOG-LINEAR (LOG-LINEAR MODEL):** Um modelo estatístico em que o logaritmo dos parâmetros é linear. Por exemplo, o modelo  $Y = aX^b$  é um modelo log-linear, pois o logaritmo transforma os resultados em um modelo linear:  $\log(Y) = \log(a) + b \times \log(X)$ .

**MODELO MISTO (MIXED MODEL):** Um modelo de análise de variância univariada (ANOVA) que inclui tanto efeitos aleatórios quanto fixos.

**MODELOS HIERÁRQUICOS (HIERARCHICAL MODELS):** Um grupo de modelos estatísticos em que os parâmetros dos modelos mais simples são um subconjunto de parâmetros de modelos mais complexos.

**MOMENTO CENTRAL (CENTRAL MOMENT):** É a média aritmética das diferenças entre cada observação de uma variável e a média aritmética daquela variável, elevada a r-ésima potência. O termo de momento central precisa sempre ser classificado: Por exemplo, primeiro momento central, segundo momento central, etc. O primeiro momento central é igual a 0 e o segundo momento central é a variância.

**MENU:** Lista de opções que é ativada com um clique do mouse. É encontrado na parte superior da tela.

**MERCADO DE TESTE:** Testar um novo produto ou algum elemento do mix de marketing usando projetos experimentais ou quase experimentais.

**METANÁLISE:** Conjunto de procedimentos destinados a examinar, de modo simultâneo, vários estudos ou diversas pesquisas, sobre um mesmo tópico. Largamente utilizado na esfera das ciências médicas.

**METANÁLISE:** Processo de utilizar métodos estatísticos para combinar resultados de diferentes estudos.

**METANÁLISE:** Cada vez mais utilizada em medicina, na tentativa de obter uma análise metodológica de dados agregados, quantitativos ou qualitativos, referente às pesquisas publicadas sobre um determinado assunto.

**METADADOS:** Descrição completa de um elemento de dados, definindo não somente sua localização de armazenagem e tipo de dados, mas também quaisquer atributos, dimensões ou associações com outros elementos de dados. Em termos mais simples, dados sobre dados, que são guardados em um repositório centralizado para comum acesso a todos os usuários da base de dados.

**MÉTODO:** A palavra método deriva do grego e quer dizer caminho. Método então, no nosso caso, é a ordenação de um conjunto de etapas a serem cumpridas no estudo de uma ciência, na busca de uma verdade ou para se chegar a um determinado conhecimento.

**MÉTODO BAYESIANO:** Abordagem estatística em que o pesquisador parte de uma distribuição a priori e utiliza os resultados de observações para calcular uma distribuição atualizada.

**MÉTODO CENTRÓIDE:** Algoritmo aglomerativo no qual a similaridade entre agrupamentos é medida como a distância entre centróides de agrupamentos. Quando dois agrupamentos são combinados, um novo centróide é computado. Logo, os centróides de agrupamentos migram, ou se movem, conforme os agrupamentos são combinados.

**MÉTODO CIENTÍFICO:** É o método que se refere ao conjunto de processos utilizado para se conhecer um fenômeno ou uma realidade, desenvolver certos comportamentos ou produzir um objeto. O método deve permitir a fiel reprodução do sistema de conhecimento, além de ser compartilhável, verificável e transmissível. Tem como principal característica a produção de conhecimento aplicado para explicação, previsão e controle de fenômenos. É importante lembrar que não há ciência sem a utilização de método científico, mas que o inverso não é necessariamente verdadeiro.

**MÉTODO CIENTÍFICO:** É essencialmente um rodízio de fatos a hipóteses, a leis, a teorias, e de volta a fatos como bases para o teste e refinamento de hipóteses mais adequadas.

**MÉTODO CIENTÍFICO (INTERPRETAÇÕES):** É errôneo pensar que existe um só método científico. Há muitos e variados métodos. Métodos científicos mudam com a descoberta de novos instrumentos de observação, novas técnicas de experimentação, e com novos instrumentos, tal como o cálculo. Métodos científicos diferem de uma ciência para a outra. Todavia, há alguns princípios que são comuns a todas as ciências.

**MÉTODO CITY-BLOCK:** Método de calcular distâncias com base na soma das diferenças absolutas das coordenadas para os objetos. Esse método assume que as variáveis não são correlacionadas e que as escalas das unidades são compatíveis.

**MÉTODO CLÁSSICO DE TESTES DE HIPÓTESES:** Método de teste de uma hipótese baseado na comparação estatística de teste com valores críticos.

**MÉTODO COMPOSICIONAL:** Abordagem alternativa aos métodos de composicionais mais tradicionais de mapeamento perceptual que determina avaliações gerais de similaridade ou preferência a partir de avaliações de atributos separados por respondente. Essas avaliações de atributos separados são combinadas para uma avaliação geral. Os exemplos mais comuns de métodos composicionais são as técnicas de análise fatorial e análise discriminante.

**MÉTODO DA BISSEÇÃO:** O método da biseção determina uma raiz  $x$  de uma função  $f(x)$  num intervalo  $[X_a, X_b] \in \mathcal{R}$  onde  $f(x_a) * f(x_b) < 0$ . A ideia é diminuir o intervalo por meio de repetidas divisões ao meio do intervalo  $[X_a, X_b]$ , de tal forma que o valor de  $X_a$  tenda ao valor de  $X_b$ , ou seja, que a raiz  $X \approx X_a \approx X_b$  e que a função  $f(x)$  seja aproximadamente nula dentro de uma certa tolerância.

**MÉTODO DA BISSEÇÃO:** É um método de iteração numérica bastante simples, e útil no dia a dia, cujo algoritmo pode ser exemplificado assim: identificar dois pontos tais que:  $f(x_1)$  e  $f(x_r)$  tenham



sinais opostos. Por exemplo, sejam as funções,  $f(x) = x^4 - 9x^3 - 2x^2 + 120 - 130$ ;  $f(x_1) = -20, 0$  e  $f(x_r) = 46, 0$ . Tomar um ponto  $(\bar{x})$  dentro do intervalo  $x_1$  e  $x_r$ , em seguida avaliar a função  $f(\bar{x})$ .

Considere três casos:

No caso 1 temos que i)  $f(\bar{x}) = 0 \rightarrow$ , é a própria raiz;  $f(\bar{x}) < 0, \rightarrow \bar{x} < x_0 < x_r$  e  $f(\bar{x}) > 0, \rightarrow x_l < x_0 < \bar{x}$ .

No Caso 2 temos que:  $f(\bar{x}) < 0; \rightarrow \bar{x} < x_0 < x_r$ , e assim temos que  $x_l \leftarrow \bar{x}$  e  $x_r = x_r$ .

No Caso 3 temos o seguinte:  $f(\bar{x}) > 0; x_l < x_0 < \bar{x}; x_l = x_l$  e  $x_r \leftarrow \bar{x}$ , um novo valor é dado por  $\bar{x} = \frac{1}{2}(x_r + x_l)$ .

O processo continua, por exemplo, até  $|\bar{x} - x_0| \leq 0,0001$ . As vantagens é que o processo é muito simples, e as desvantagens são as seguintes: i) é necessário determinar os valores iniciais de  $x_l$  e  $x_r$ ; ii) converge muito devagar, ou seja, é lento; iii) tem problemas de precisão e iv) é necessário limitar o número de iterações.

**MÉTODO DA MÁXIMA VEROSSIMILHANÇA:** É aquele método onde se obtém os estimadores por maximização da função de verossimilhança da amostra:

$$L = \prod_{i=1}^n f(X_i, \theta), \text{ ou, de uma forma equivalente, de } \log L.$$

**MÉTODO DA MÁXIMA VEROSSIMILHANÇA RESTRITA (REML):** O método da máxima verossimilhança restrita (REML) foi desenvolvido por Patterson e Thompson (1971), que apresentaram uma correção ao método da máxima verossimilhança (ML), eliminando o seu vício. No método da máxima verossimilhança restrita (REML), somente a porção da verossimilhança que é invariante aos efeitos fixos é maximizada. Assim, o método da máxima verossimilhança restrita (REML) mantém as demais propriedades do método de máxima verossimilhança (ML), é não viciado e permite também a imposição de restrições de não negatividade. Dessa forma, o método da máxima verossimilhança restrita (REML) é o procedimento ideal, por exemplo, na estimação de componentes de variância com dados desbalanceados. O procedimento da máxima verossimilhança restrita (REML) é também denominado de máxima verossimilhança marginal e, na Europa, por máxima verossimilhança residual. Os métodos de máxima verossimilhança (ML) e da máxima verossimilhança restrita (REML) requirem solução iterativa, pela não linearidade das equações, fato que dificulta a derivação de estimadores explícitos. Os métodos da máxima verossimilhança restrita (REML) e de máxima verossimilhança (ML) exigem normalidade para que os estimadores tenham propriedades desejáveis. Entretanto, tais estimadores podem ser robustos aos desvios da normalidade, gerando estimativas razoáveis mesmo quando a forma da distribuição não é especificada.

**MÁSCARA V:** Uma figura geométrica usada para detectar causas atribuídas em um gráfico de controle de soma cumulativa. Com valores apropriados para os parâmetros, conclusões idênticas podem ser feitas a partir de uma máscara V e um CUSUM (gráfico de controle de soma acumulada em inglês Cumulative Sum Control Charts) tabular.

**MATRIZ CHÁPEU:** Em regressão múltipla, a matriz  $H = X(X'X)^{-1}X'y = Hy$ . Essa é a matriz de projeção que mapeia o vetor de valores observados das respostas em um vetor de valores ajustados por  $\hat{y} = X(X'X)^{-1}X'y = Hy$ .

**MATRIZ DE CORRELAÇÃO:** Uma matriz quadrada que contém as correlações entre o conjunto de variáveis aleatórias,  $X_1, X_2, \dots, X_k$ . Os elementos da diagonal principal da matriz são unitários e os elementos fora da diagonal,  $r_{ij}$ , são as correlações entre  $X_i$  e  $X_j$ .

**MATRIZ DE COVARIÂNCIA:** Uma matriz quadrada que contém as variâncias e as covariâncias entre o conjunto de variáveis aleatórias  $X_1, X_2, \dots, X_k$ . Os elementos da diagonal principal da matriz são as variâncias das variáveis aleatórias e os elementos fora da diagonal são as covariâncias entre  $X_i$  e  $X_j$ . Também chamada de matriz variância-covariância. Quando as variáveis aleatórias são padronizadas para se ter variâncias unitárias, a matriz de covariância se torna a matriz de correlação.

**MATRIZ DE PLANEJAMENTO:** Uma matriz que fornece os testes que devem ser conduzidos em um experimento.

**MÉDIA:** A média geralmente se refere tanto ao valor esperado de uma variável como à média aleatória de um conjunto de dados. Veja média aritmética.

**MÉDIA ARITMÉTICA:** A média aritmética de um conjunto de números  $x_1, x_2, \dots, x_n$ , é a sua soma dividida pelo número de observações,  $\left(\frac{1}{n}\right) \sum_{i=1}^n x_i$ . A média aritmética é geralmente denotada por  $\bar{x}$  e é frequentemente chamada de média.

**MÉDIA CONDICIONAL:** A média da distribuição de probabilidades condicionais de uma variável aleatória.

**MÉDIA DA AMOSTRA:** A média aritmética das observações em uma amostra. Se as observações são  $x_1, x_2, \dots, x_n$ , então a média da amostra é  $\left(\frac{1}{n}\right) \sum_{i=1}^n x_i$ . A média da amostra é geralmente denotada por  $\bar{X}$ .

**MÉDIA GEOMÉTRICA:** A média geométrica de um conjunto de n valores positivos de dados é a n-ésima raiz do produto dos valores dos dados; ou seja,

$$g = \left( \prod_{i=1}^n x_i \right)^{\frac{1}{n}}.$$

**MÉDIA HARMÔNICA:** A média harmônica de um conjunto de valores de dados é a recíproca da média aritmética dos recíprocos dos valores de dados;

$$\text{isto é, } \bar{h} = \left[ \left( \frac{1}{n} \right) \sum_{i=1}^n \frac{1}{x_i} \right]^{-1}.$$

**MÉDIA QUADRÁTICA:** Em geral, uma média quadrática é determinada dividindo a soma dos quadrados pelo número de graus de liberdade associado com a soma dos quadrados.

**MÉDIA QUADRÁTICA DO ERRO:** A soma dos quadrados do erro dividida pelo seu número de graus de liberdade.

**MÉDIA AMOSTRAL:** Medida resumo, de posição ou tendência central cujo símbolo  $\bar{X}$ , se lê X barra ou X traço, é uma variável aleatória, função dos valores da amostra, é definida como a soma de todos os valores da amostra dividido pelo número de observações da amostra. Serve para estimar a média populacional.

**MEDIANA:** A mediana de um conjunto de dados é aquele valor que divide os dados em duas metades iguais. Quando o número de observações é par, isto é  $2n$ , é costume definir a mediana como a média entre o  $n$ -ésimo e o  $(n+1)$ -ésimo valores, devidamente ordenados. A mediana pode também ser definida para uma variável aleatória. Por exemplo, no caso de uma variável aleatória contínua  $X$ , a mediana  $M$  pode ser definida como

$$\int_{-\infty}^M f(x)dx = \int_M^{\infty} f(x)dx = \frac{1}{2}.$$

**MÉTODO DA MÍNIMA DIFERENÇA SIGNIFICATIVA DE FISHER (LSD):** Uma série de testes de hipóteses par a par das médias dos tratamentos em um experimento, de modo a determinar que pares de médias diferem.

**MÉTODO DA SOMA EXTRA DOS QUADRADOS:** Um método usado na análise de regressão para conduzir um teste de hipóteses para a contribuição adicional de uma ou mais variáveis a um modelo.

**MÉTODO DO GRADIENTE ASCENDENTE:** Uma técnica que permite um experimentalista se mover eficientemente em direção a um conjunto de condições experimentais ótimas, seguindo a direção do gradiente. O método do gradiente ascendente é geralmente empregado em conjunção com o ajuste por superfície de resposta de primeira ordem e com a decisão de que a região corrente de operação não é apropriada.

**MÉTODO(S) ESTATÍSTICO(S) NÃO-PARAMÉTRICO(S):** Veja método(s) livre(s) de distribuição.

**MÉTODO(S) LIVRE(S) DE DISTRIBUIÇÃO:** Qualquer método de inferência tais como aplicação de testes de hipóteses ou construção de intervalo de confiança que não depende da forma da distribuição básica considerada das observações. Algumas vezes chamados de método(s) não-paramétrico(s).

**MÍNIMOS QUADRADOS (MÉTODO DOS):** Um método de estimação de parâmetros em que os parâmetros de um sistema são estimados pela minimização da soma dos quadrados das diferenças entre os valores observados e os valores previstos ou ajustados do sistema.

**MODA:** A moda de uma amostra é aquele valor observado que ocorre mais frequentemente. Em uma distribuição de probabilidades  $f(x)$  com primeira derivada contínua, a moda é um valor de  $x$  para o qual  $\frac{df(x)}{dx} = 0$  e  $\frac{d^2f(x)}{dx^2} < 0$ . Há talvez mais do que uma moda de uma amostra ou de uma distribuição.

**MODELO DE EFEITOS ALEATÓRIOS:** Em um contexto de análise de variância, isso se refere a um modelo que envolve somente fatores aleatórios.

**MODELO DE PRIMEIRA ORDEM:** Um modelo que contém somente os termos de primeira ordem. Por exemplo, o modelo de primeira ordem de superfície de resposta em duas variáveis é  $y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \varepsilon$ . Um modelo de primeira ordem é chamado também de modelo de efeitos principais.

**MODELO DE SEGUNDA ORDEM:** Um modelo que contém os termos de segunda ordem. Por exemplo, o modelo de superfície de resposta de segunda ordem em duas variáveis é  $y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_{12}x_1x_2 + \beta_{11}x_1^2 + \beta_{22}x_2^2 + \varepsilon$ . Os termos de segunda ordem nesse modelo são  $\beta_{12}x_1x_2$ ,  $\beta_{11}x_1^2$ ,  $\beta_{22}x_2^2$ .

**MODELO DE UM FATOR (ONE-WAY):** No contexto de uma análise de variância, ele envolve uma variável única ou fator único com a diferentes variáveis.

**MODELO EMPÍRICO:** Um modelo para relacionar uma resposta a um ou mais regressores ou fatores, que é desenvolvido a partir de dados obtidos do sistema.

**MODELO INTRINSECAMENTE LINEAR:** Em uma análise de regressão, uma função não linear que pode ser expressa como uma função linear, depois de uma transformação linear apropriada, é chamada de linear intrinsecamente.

**MODELO LINEAR:** Um modelo em que as observações são expressas como uma função linear dos parâmetros desconhecidos. Por exemplo,  $y = \beta_0 + \beta_1x + \varepsilon$  e  $y = \beta_0 + \beta_1x + \beta_2x^2 + \varepsilon$  são modelos lineares.

**MODELO MECANICISTA:** Um modelo desenvolvido de conhecimento teórico ou experiência, em contraste com um modelo desenvolvido a partir de dados. Veja modelo empírico.

**MODELO MISTURADO:** Em um contexto de análise de variância, um modelo misturado contém ambos os fatores: fixos e aleatórios.

**MOMENTO (OU MOMENTO DE POPULAÇÃO):** O valor esperado de uma função de uma variável aleatória tal que  $E(X - c)^r$ , para  $r$  e  $c$  constantes. Quando  $c = 0$ , diz-se que o momento está em torno da origem. Veja função geradora de momento.

**MOMENTO DA AMOSTRA:** A grandeza  $\left(\frac{1}{n}\right) \sum_{i=1}^n X_i^k$  é chamada de  $k$ -ésimo momento da amostra.

**MUDANÇA DE ESCALA (CHANGE OF SCALE):** A transformação de uma variável aleatória normal realizada multiplicando-a por uma constante. Ver também operação de deslocamento.

**MULTICOLINEARIDADE (MULTICOLLINEARITY):** As correlações indesejadas entre duas ou mais variáveis preditoras. Ver também colinearidade.

**MULTINORMAL (MULTINORMAL):** Abreviação para a distribuição teórica de probabilidade normal multivariada.

**MULTICOLINEARIDADE:** Uma condição ocorrendo em regressão múltipla, em que alguns dos preditores ou variáveis regressoras são aproximadamente linearmente dependentes. Essa condição pode levar à instabilidade nas estimativas dos parâmetros do modelo de regressão.

**MÉTODO DA REFERÊNCIA PARALELA:** Procedimento de agrupamento não-hierárquico que seleciona as sementes de agrupamentos simultaneamente no início. Objetos dentro das distâncias de referência são designados para a semente mais próxima. Distâncias de referência podem ser ajustadas para incluir menos ou mais objetos nos agrupamentos. Esse método é o oposto do método da referência sequencial.

**MÉTODO DA REFERÊNCIA SEQUENCIAL:** Procedimento de agrupamento não-hierárquico que começa pela seleção de uma semente de agrupamento. Todos os objetos dentro de uma distância pré-especificada são então incluídos no agrupamento. Sementes de agrupamentos subsequentes são selecionadas até que todos os objetos estejam reunidos em um agrupamento.

**MÉTODO DAS ISOLINHAS:** É um método clássico para avaliação de reservas, baseado no princípio da variação gradual. Observe que este método deriva do método dos triângulos, pois a rede de triângulos é utilizada para realizar a interpolação linear das linhas de iso-espessuras e isoteores. As espessuras e os teores são integrados dentro do bloco de cubagem, obtendo-se ao final a espessura e o teor médio, os quais aplicados à área do bloco de cubagem resulta na reserva do mesmo.

**MÉTODO DAS MÉDIAS MÓVEIS:** Trata-se de um método empregado em análise de séries temporais que objetiva suavizar as variações das séries por um processo de sucessivas médias. Os cálculos necessários para obtenção das médias são fáceis, e o método é extremamente flexível, pois a tendência não é forçada a adaptar-se a qualquer função matemática. Senão vejamos. Dado um conjunto de números  $Y_1, Y_2, Y_3, Y_4, \dots, Y_n$ , define-se a média aritmética móvel de ordem  $K$ , à sequência de médias dadas por:

$$\frac{Y_1 + Y_2 + Y_3 + Y_4 + \dots + Y_K}{K}, \frac{Y_2 + Y_3 + Y_4 + \dots + Y_{K+1}}{K}, \frac{Y_3 + Y_4 + \dots + Y_{K+2}}{K}, \dots, \frac{Y_{n+1-K} + Y_{n+2-K} + \dots + Y_n}{K}$$

Quanto maior a ordem  $K$ , maior o efeito de suavização das variações. A escolha da ordem dependerá da amplitude dos movimentos sazonais ou cíclicos da série. Ciclos periódicos podem ser eliminados fazendo-se  $K$  igual ao período do ciclo ou um múltiplo dele. Assim, se uma série apresenta um movimento estacional de três meses, deve-se usar uma média móvel de ordem 3 para isolar a tendência. A construção do gráfico facilitará a determinação da ordem a ser empregada.

**MÉTODO DE AMOSTRAGEM:** Refere-se ao tipo de unidade amostral utilizada num levantamento. Por exemplo: método de parcelas, método de quadrantes, método de Bitterlich, dentre outros.

**MÉTODO DE BOOTSTRAP:** Método de reamostragem que utiliza a função de distribuição empírica. Em inglês Bootstrap method.

**MÉTODO DE COMPARAÇÃO PAREADA (OU AOS PARES):** Método de apresentar um par de estímulos a um respondente para avaliação, em que o respondente seleciona um estímulo como preferido.

**MÉTODO DE CONCORDÂNCIA (BUSCA DE CAUSAS):** Se duas circunstâncias que levam a um dado evento têm tido em todos os casos um fator comum, esse fator pode ser a causa buscada, principalmente se ele for o único fator comum.

**MÉTODO DE COX:** O mesmo que método dos azares proporcionais.

**MÉTODO DE DIFERENÇA (BUSCA DE CAUSAS):** Se dois conjuntos de circunstâncias diferem em apenas um fator e se o conjunto que contém o fator leva ao evento e o outro não, esse fator pode ser considerado a causa do evento.

**MÉTODO DE FELLEGI:** Método proposto por Fellegi (1963) o qual é um método de amostragem sem reposição com probabilidade proporcional ao tamanho, que permite a rotação da amostra simultaneamente com os cálculos exatos de probabilidades conjuntas de seleção de conjuntos de unidades. Em inglês Fellegi's method.

**MÉTODO DE GAUSS E SEIDEL:** Método clássico para a solução iterativa de um conjunto de equações lineares, particularmente para aquelas que surgem como soluções de mínimos quadrados. Em inglês Gauss-Seidel method.

**MÉTODO DE INTERPOLAÇÃO:** Consiste na escolha e ajuste de uma função matemática aos pontos de observação, com o objetivo de interpolar os valores da variável de interesse em pontos não amostrados. As funções de interpolação podem ser classificadas em globais e locais. As funções globais dependem de todo o conjunto de pontos de observação, de modo que a retirada ou inclusão de um único ponto de dado altera a função, sendo necessário novo ajuste, enquanto as funções locais são definidas para pequenas porções da área em estudo, de modo que o seu ajuste depende única e exclusivamente do subconjunto de pontos de dados encontrado na vizinhança do ponto a ser interpolado. Como exemplo de funções globais tem-se

os polinômios bivariados ou superfícies de tendência e as equações multiquádricas e de funções locais, tem-se os métodos da média ponderada como o inverso da potência da distância e a krigagem ordinária, entre outros. Os métodos locais tem predominado no cálculo de reservas, pois o teor de um ponto está mais intimamente relacionado a um ponto próximo daquele que em relação a qualquer outro ponto distante.

**MÉTODO DE KAPLAN-MEYER PARA ANÁLISE DA TÁBUA DE VIDA:** Abordagem mais comumente utilizada em medicina para análises de sobrevivência. O método de Kaplan-Meyer difere do método atuarial no sentido de que a ocorrência de cada óbito define o fim do período de observação e o começo do próximo. Portanto, por esse método, é variável a duração dos períodos de observação pelos quais se determina a sobrevivência. Ver também Análise da tábua de vida.

**MÉTODO DE MÁXIMA VEROSSIMILHANÇA:** Método de estimação de parâmetros de uma população por meio de valor ou valores que maximiza ou maximizam a função de verossimilhança de uma amostra. Em inglês Maximum likelihood method.

**MÉTODO DE MÁXIMA VEROSSIMILHANÇA (ML):** É um método de estimação que baseia-se na obtenção do ponto de máximo de uma função de verossimilhança que é a função densidade de probabilidade conjunta dos pontos amostrais. E este máximo é obtido por derivação da função de verossimilhança ( $L$ ) em relação ao parâmetro de interesse. O método de máxima verossimilhança (ML) foi desenvolvido por Fisher (1922), mas somente após um período de 45 anos, Hartley e Rao (1967) apresentaram a especificação matricial de um modelo misto, fato que permitiu a derivação de equações de máxima verossimilhança (ML) para várias classes de modelos. É importante destacar que em situações de dados desbalanceados, os estimadores de máxima verossimilhança (ML) apresentam propriedades desejáveis de um bom estimador as quais são: suficiência, consistência, eficiência e invariância a translação. Outra vantagem do método de máxima verossimilhança é a geração de estimativas não negativas dos componentes de variância, por exemplo. Por outro lado, os estimadores de máxima verossimilhança (ML) são viciados em decorrência da perda de graus de liberdade devida à estimação dos efeitos fixos.

**MÉTODO DE MÍNIMOS QUADRADOS:** Uma técnica de estimação pela qual as quantidades a estimar são determinadas minimizando-se uma certa forma quadrática das observações e aquelas quantidades. Em inglês Least squares method.

**MÉTODO DE MOMENTOS:** Método de estimação de parâmetros baseado na construção de sistemas de equações em que são igualados os momentos centrais com seus respectivos momentos amostrais. Em inglês Method of moments.

**MÉTODO DE MONTE CARLO EM CADEIAS DE MARKOV (MCMC):** A ideia central por trás dos métodos de MCMC é criar uma cadeia de Markov fácil de simular e tem distribuição de equilíbrio dada pela distribuição de interesse. Em inglês Monte Carlo Markov Chain methodology.

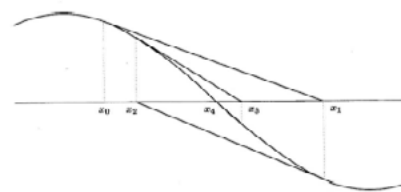
**MÉTODO DE NEWTON-RAPHSON:** Existe um método numérico para se achar as raízes de uma função denominado de método da bisseção, o qual é um método que apresenta uma vantagem, pois ele sempre converge para a raiz, desde que exista uma no intervalo inicial dado, no entanto tem duas desvantagens, ele é um método lento, e se a função não muda de sinal, a raiz não é encontrada. Existe um outro método conhecido como de Newton – Raphson, o qual cobre as desvantagens do método da bisseção, isto é, é mais rápido e encontra raízes que tocam o eixo  $X$ , mas também apresenta duas desvantagens: i) nem sempre converge e ii) nem precisa do cálculo da derivada da função, o que nem sempre é uma tarefa fácil.



Para desenvolver este método, pode-se utilizar a expansão de uma função em série de Taylor em torno do ponto  $x_0$ . Ela é escrita como:  $f(x) = f(x_0) + (x - x_0)f'(x_0) + \frac{(x - x_0)^2}{2!} f''(x_0) + \dots$ . Mantendo apenas os dois primeiros termos da série temos que  $f(x) \approx f(x_0) + (x - x_0)f'(x_0)$ . Esta é a equação de

uma reta que passa pelo ponto  $f(x_0)$  com inclinação  $f'(x_0)$ , isto é, ela é tangente à curva no ponto  $x_0$ . Suponha que a função  $f(x)$  seja bem aproximada por uma reta, o ponto que essa reta cruza o eixo  $x$ , está próximo ao ponto que a função cruza o eixo  $x$ . Este ponto  $x$  para o qual a função cruza o zero será:  $0 = f(x_0) + (x - x_0)f'(x_0); -f(x_0) = (x - x_0)f'(x_0)$  e  $x = x_0 - \frac{f(x_0)}{f'(x_0)}$ . Esse ponto  $x$  é

então usado no lugar de  $x_0$  como um novo valor inicial melhorando assim a aproximação. Essa idéia de se usar um valor para gerar um outro melhor é chamada de iteração. A figura a seguir ilustra a idéia do método.



**Figura:** Método de Newton-Raphson em ação. O processo vai Começando em  $x_0$ , as sucessivas iterações se aproximam do zero da função. A posição de  $x_4$  e do zero real da função são indistinguíveis nesta escala.

Mudando ligeiramente a notação, pode-se calcular o  $(i + 1)$ ésimo valor usando o  $i$ -ésimo valor por meio

da expressão  $x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}$ . vamos considerar a função  $f(x) = \cos(x) - x$ . A derivada da função é

$$f'(x) = -\text{sen}(x) - 1 \text{ e vamos dar o chute inicial } x_0 = \frac{\pi}{4} = 0,785398163.$$

Usando a expressão  $x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}$  temos:

$$x_1 = \frac{\pi}{4} - \frac{\cos\left(\frac{\pi}{4}\right) - \frac{\pi}{4}}{-\text{sen}\left(\frac{\pi}{4}\right) - 1} = 0,739536134$$

e usamos esse resultado para obter  $x_2$ ,  $x_2 = x_1 - \frac{\cos(x_1) - x_1}{-\text{sen}(x_1) - 1} = 0,739085178$ . E continuamos o processo até

que a precisão desejada seja alcançada, o que tipicamente é da ordem de  $10^{-8}$ . A precisão pode ser a diferença entre  $x_{i+1} - x_i$  ou o que chamamos de erro relativo  $\frac{x_{i+1} - x_i}{x_i}$ . Vale salientar que no método da

bisseção nós damos o limite inferior e superior da região que deve conter a raiz. No método de Newton-Raphson damos um valor inicial e dependendo deste valor nem sempre o método converge, pois podemos ter o caso em que a reta tangente à função no ponto inicial não representa bem a função naquele ponto. Esse método é conhecido também como método das tangentes.

**MÉTODO DE PRODAN:** Método que utiliza pontos como unidades amostrais, distribuídos na população de acordo com determinado delineamento amostral. Em cada ponto mede-se um certo número de árvores que são as mais próximas do ponto  $n$ , e considera-se que estejam numa parcela circular com raio definido pela distância da  $n$ -ésima árvore até o ponto.

**MÉTODO DE QUADRANTES:** Método de amostragem em que a unidade amostral é um ponto, a partir do qual são identificados os quadrantes, para então medir os atributos de interesse na árvore mais próxima do ponto em cada um dos quadrantes. A densidade de árvores na floresta é estimada com base na distância média das árvores em relação aos pontos amostrais ou na área média por árvore, assumindo que o padrão espacial das árvores é completamente aleatório.

**MÉTODO DE TRIANGULARIZAÇÃO DE GAUSS:** É o método computacional para resolução direta e exata de sistemas de equações lineares. Este método caracteriza-se pela eliminação de elementos abaixo da diagonal principal obtida pela multiplicação dos elementos da linha da diagonal principal pelo multiplicador definido como o valor do elemento que se quer eliminar dividido pelo valor do elemento sobre a diagonal principal, com o sinal trocado, somados com os elementos da linha que contém o elemento que se deseja eliminar. Muitas vezes a causa da propagação de erros na resolução de sistemas lineares está nos multiplicadores, principalmente quando estes forem calculados sobre elementos da diagonal cujos valores em módulo são muito pequenos. Este problema pode ser evitado aplicando-se a condensação pivotal.

**MÉTODO DE WARD:** Procedimento de agrupamento hierárquico no qual a similaridade usada para juntar agrupamentos é calculada como a soma de quadrados entre os dois agrupamentos somados sobre todas as variáveis. Esse método tende a resultar em agrupamentos de tamanhos aproximadamente iguais devido a sua minimização de variação interna.

**MÉTODO DE WOOLHOUSE:** Processo ou mecanismo de perequação mecânica que, à ordenada  $y_1$  do ponto observado  $A_1 (X_1, Y_1)$ , substitui a média aritmética em  $x_1$ , das ordenadas de cinco parábolas do segundo grau, cada qual passando pelo ponto observado, de abscissas  $X_{i-7}, X_{i-2}, X_{i+3}, \dots, X_{i-5}, X_i, X_{i+5}, \dots, X_{i-3}, X_{i+2}, X_{i+7}$ , dado por Woolhouse (1870).

**MÉTODO DECOMPOSICIONAL:** Método de mapeamento perceptual associado a técnicas MDS (escalonamento multidimensional), no qual o respondente fornece apenas uma avaliação geral de similaridade ou preferência entre objetos. Esse conjunto de avaliações gerais é então decomposto, em um conjunto de dimensões que melhor representam as diferenças de objetos.

**MÉTODO DIVISIVO:** Procedimento de agrupamento que começa com todos os objetos em um único agrupamento, que é então dividido em cada passo em dois agrupamentos que contêm os objetos mais distintos. Esse método é o oposto do método aglomerativo.

**MÉTODO DOS AZARES PROPORCIONAIS:** Método usado para testar diferenças entre curvas de sobrevivência de Kaplan-Meyer enquanto se faz o controle para outras variáveis. Esse método também é usado para determinar quais variáveis estão associadas com uma sobrevivência maior.

**MÉTODO EPIDEMIOLÓGICO:** Um tipo de metodologia variante do método científico associada à epidemiologia.



**MÉTODO ESTATÍSTICO EXATO:** Usado em estimação por intervalos para afirmar que a distribuição de probabilidades envolvida é completamente conhecida e os níveis de probabilidade são valores exatamente calculados. Em inglês Exact statistical method.

**MÉTODO ESTATÍSTICO:** Diante da impossibilidade de manter as causas constantes nas ciências sociais, admitem todas essas causas presentes variando-as, registrando essas variações e procurando determinar, no resultado final, que influências cabem a cada uma delas.

**MÉTODO EXPERIMENTAL:** É o método que consiste em manter constante todas as causas, menos uma, que sofre variação para se observar seus efeitos, caso existam.

**MÉTODO INDUTIVO-DEDUTIVO:** Consiste de um começo indutivo, indo das observações às hipóteses e, depois, das hipóteses às suas implicações, testando a validade delas em termos de compatibilidade com o conhecimento já existente. O pesquisador usa o método indutivo-dedutivo quando: i) coleta dados; ii) formula hipóteses para explicá-los; iii) busca consequência lógicas de suas hipóteses; e iv) obtém mais dados para testar a veracidade delas.

**MÉTODO INVARIANTE:** Princípio de estimação ou teste de hipótese que requer um estimador ou hipótese que permaneça invariante quando os dados sofrem alguma transformação. Em inglês Invariance method.

**MÉTODO DE PERFIL COMPLETO:** Método de apresentação de estímulos aos respondentes para avaliação que consiste em uma descrição completa dos estímulos com todos os atributos. Por exemplo, considere que um doce tenha sido descrito por três fatores com dois níveis cada: preço 15 ou 25 centavos, sabor cítrico ou caramelo e cor branco ou vermelho. Um estímulo de perfil completo seria definido por um nível de cada fator. Um exemplo de estímulo de perfil completo seria um doce vermelho de caramelo que custasse 15 centavos.

**MÉTODO DE TROCA:** Método para apresentar estímulos a respondentes no qual os atributos são descritos dois por vez e os respondentes ordenam todas as combinações dos níveis em termos de preferência.

**MÉTODO DOS MÍNIMOS QUADRADOS:** É o método geralmente empregado no ajuste de funções matemáticas lineares nos coeficientes, ou então, funções que após transformações tornam-se lineares nos coeficientes. Este método permite o ajuste de qualquer função linear com ou sem transformação nos coeficientes, de tal modo que a somatória dos resíduos ao quadrado é o mínimo possível.

**MÉTODO DOS MÍNIMOS QUADRADOS ORDINÁRIOS (MQO):** É o método de estimação que utiliza como modelo uma curva reta no caso de duas variáveis, plano no caso de três variáveis e hiperplano no caso de mais variáveis cuja soma dos quadrados de distâncias entre os dados e a curva seja a menor possível.

**MÉTODO DOS MOMENTOS:** Consiste na obtenção de estimadores a partir dos momentos amostrais, como por exemplo, estimar a variância de uma população normal a partir da variância de uma sua amostra. Em inglês method of moments.

**MÉTODO DOS PERFIS LINEARES:** É um método clássico para avaliação de reservas, baseado no princípio dos pontos mais próximos. O bloco de cubagem é definido sobre a seção de amostragem, estendendo-se até as meias distâncias para as seções adjacentes. A reserva é obtida multiplicando-se a área da seção central pela soma das meias distâncias para as seções adjacentes e pelo teor médio da seção.

**MÉTODO DOS MOMENTOS:** É aquele onde os estimadores obtêm-se por substituição dos momentos da amostra nas expressões que representam os momentos na população, por exemplo  $E[X] = \mu = \bar{X}$ .

**MÉTODO DOS MÍNIMOS QUADRADOS:** São usualmente utilizados no âmbito da regressão linear, o qual consiste em estimar os parâmetros do modelo minimizando a soma de quadrados dos desvios, afastamentos, erros ou resíduos  $SQD = \sum_{i=1}^n e_i^2$ .

**MÉTODO DA MÁXIMA VEROSSIMILHANÇA:** É o método de estimação mais importante. Geralmente, os estimadores de máxima verossimilhança gozam das propriedades desejáveis num bom estimador: são os mais eficientes e consistentes. Embora, usualmente, não sejam centrados costumam ser assintoticamente não enviesados. Uma forma de obter uma estimativa pontual de um parâmetro da população, por exemplo,  $\mu$ ,  $\sigma$ ,  $\sigma^2$  e  $p$  é retirar uma amostra aleatória representativa dessa população e calcular o valor da estatística correspondente, por exemplo,  $\bar{x}$ ,  $s$ ,  $s^2$ ,  $\bar{p}$ .

**MÉTODO DOS MOMENTOS:** Este método de estimação é um dos mais simples e mais antigos para obter estimadores de um ou mais parâmetros de uma distribuição. A ideia base é utilizar os momentos da amostra para estimar os correspondentes momentos da população, e, a partir daí, estimar os parâmetros de interesse. Seja  $X_1, X_2, \dots, X_n$  uma amostra aleatória de uma dada população com função densidade de probabilidade abreviada como f. d. p.,  $f(x; \theta_1, \theta_2, \dots, \theta_k)$  que depende de  $k$  parâmetros. Admitindo que existam os momentos ordinários,  $\mu_r$  da população  $x$ , estes são funções dos  $k$  parâmetros,

$$\mu'_r = E[X^r] = \begin{cases} \sum_{i=1}^N x^r f(x; \theta_1, \theta_2, \dots, \theta_k) & \text{para distribuições discretas} \\ \int_{-\infty}^{+\infty} x^r f(x; \theta_1, \theta_2, \dots, \theta_k) dx & \text{para distribuições contínuas} \end{cases}$$

Os correspondentes momentos amostrais são dados por  $m'_r = \frac{1}{n} \sum_{i=1}^n x_i^r$

O método dos momentos consiste em considerar que os estimadores dos momentos ordinários são dados pelos momentos ordinários amostrais, ou seja,  $\hat{\mu}'_r = m'_r, r = 1, \dots, k$ .

**MÉTODO DA MÁXIMA VEROSSIMILHANÇA:** O método da máxima verossimilhança foi introduzido por Fisher (1922). Sua introdução, em muitos aspectos, determinou o começo da teoria estatística moderna. Para apresentar o seu conceito, considere que  $X_1, X_2, \dots, X_n$  seja uma amostra aleatória de uma população com densidade  $f(x)$ , determinada pelos parâmetros  $\theta_i, i = 1, 2, \dots, k$ . Inicialmente, é considerada a situação específica de apenas um parâmetro  $\theta$  ( $k=1$ ), por facilitar o entendimento do conceito a ser apresentado. Para uma amostra particular  $X_1, X_2, \dots, X_n$ , o estimador de máxima verossimilhança ( $\hat{\theta}$ ) do parâmetro  $\theta$  é aquele que maximiza a densidade conjunta de  $X_1, X_2, \dots, X_n$ . Em razão do fato de os valores amostrais  $X_1, X_2, \dots, X_n$  serem independentes é possível definir a densidade conjunta ou função de verossimilhança ( $L$ ) pelo produto das densidades de cada  $X_i$  ( $i = 1, 2, \dots, n$ ). Assim, a função de verossimilhança,  $L$ , é definida pela seguinte equação:

$$L = f(x_1)f(x_2)f(x_3)\dots f(x_n) = \prod_{i=1}^n f(x_i).$$

O estimador de máxima verossimilhança é aquele que maximiza o valor de L na equação anterior. Para obter o estimador de máxima verossimilhança ( $\hat{\theta}$ ), basta tomar a primeira derivada de L em relação ao parâmetro  $\theta$ , igualar a zero e resolver para  $\theta$ . A solução é o estimador de máxima verossimilhança. Nem sempre uma solução explícita existe e métodos numéricos são utilizados para se obterem as estimativas. Quando se tem mais de um parâmetro, tomam-se as derivadas parciais de L com respeito a cada um deles. Igualam-se cada derivada a zero e resolve-se o sistema formado obtendo-se os estimadores de máxima verossimilhança dos parâmetros.

Algumas propriedades matemáticas da função L garantem a possibilidade de usar a função  $S = \ln(L)$  em seu lugar, uma vez que apresentam o máximo para o mesmo valor de  $\theta$ . Isso é feito para tornar mais fácil a obtenção do máximo, uma vez que o produto se transforma em somatório. Essa função é denominada de função suporte. Na estimação paramétrica, este método só pode ser aplicado se a distribuição da população for conhecida.

A Definição desse método pode ser explicada da seguinte maneira:

Seja  $X_1, X_2, \dots, X_n$ , uma amostra aleatória de uma dada população com função densidade de probabilidade, abreviada por f.d.p.,  $f(x; \theta)$ . Então a f.d.p. conjunta das variáveis que constituem a amostra é dada por:

$$f = (x_1, x_2, \dots, x_n; \theta) = f(x_1; \theta) f(x_2; \theta) \dots f(x_n; \theta) \prod_{i=1}^n f(x_i; \theta).$$

Considerando uma amostra aleatória em concreto, designa-se por função de verossimilhança a função do parâmetro  $\theta$  e da amostra tal que:

$$L(x_1, x_2, \dots, x_n; \theta) = L(x; \theta) = \prod_{i=1}^n f(x_i; \theta).$$

O Método da Máxima Verossimilhança pode ainda ser descrito da seguinte forma: consiste em encontrar o estimador  $\hat{\theta}$  que maximiza o valor a função de verossimilhança para uma determinada amostra, ou seja, o valor de  $\hat{\theta}$  que torna aquela amostra concreta mais provável, isto é, mais verossímil. Para se obter o estimador de máxima verossimilhança devem-se fazer as seguintes operações:

- i) Determinar a função de verossimilhança  $L(x; \theta)$ ;
- ii) Se necessário, aplicar a transformação logarítmica à função de verossimilhança  $\ln L(x; \theta)$ . Geralmente, esta transformação torna o problema da maximização mais simples.

iii) Determinar os pontos onde a 1ª derivada da função ( $L(x; \theta)$  ou  $\ln L(x; \theta)$ ) em relação a  $\theta$  se anula (condição de primeira ordem):

$$\frac{\partial L(x; \theta)}{\partial \theta} = 0 \text{ ou } \frac{\partial \ln L(x; \theta)}{\partial \theta} = 0$$

iv) Verificar se a 2ª derivada da função em relação a  $\theta$  é negativa (condição de segunda ordem):

$$\frac{\partial^2 L(x; \theta)}{\partial \theta^2} < 0 \text{ ou } \frac{\partial^2 \ln L(x; \theta)}{\partial \theta^2} < 0$$

O método da máxima verossimilhança pode ser utilizado para estimar mais do que um parâmetro em simultâneo.

**MÉTODO DOS QUADRADOS MÍNIMOS:** É um outro importante método de estimação que aparece na teoria estatística é o dos quadrados mínimos. Grande parte da teoria em que se fundamenta a estatística experimental e baseada nesse método. Para apresentar o conceito e ilustrar o método, considerar o

modelo linear a seguir para cada observação da amostra aleatória  $X_1, X_2, \dots, X_n$ . A observação amostral é modelada como resultante da soma de dois componentes básicos: um componente fixo que é a constante  $\mu$  e outro de natureza aleatória  $e_i$ .  $X_i = \mu + e_i$  em geral, o componente  $e_i$  é suposto normal com média 0 e variância constante  $\sigma^2$  para todo valor de  $i, i = 1, 2, \dots, n$ . Pela observação do modelo linear apresentado, verifica-se que  $\mu$  e  $\sigma^2$  são os parâmetros desconhecidos desse modelo. O parâmetro  $\sigma^2$  não é, entretanto, mencionado explicitamente nesse modelo. O método de estimação de quadrados mínimos baseia-se na minimização da soma de quadrados da variável aleatória denominada erro ou resíduo,  $e_i$ . Assim, isolando-se  $e_i$  no modelo anterior resulta em:  $e_i = X_i - \mu$ . Tomando-se a soma dos quadrados (SQ) de  $e_i$  para as  $n$  observações amostrais tem-se:

$$SQ = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (X_i - \mu)^2$$

Para se obter o estimador de quadrados mínimos e necessário minimizar o valor dessa soma de quadrados (SQ). Para se obter esse mínimo, é necessário derivar em relação a cada parâmetro, igualar as derivadas a zero e resolver o sistema de equações formado. Nesse modelo, em que se esta exemplificando o método, a soma de quadrados SQ só depende de um parâmetro que é a média  $\mu$ , então,

$\frac{dSQ}{d\mu} = -2 \sum_{i=1}^n (X_i - \mu)$ . Igualando a zero a derivada primeira de SQ em relação ao parâmetro  $\mu$  tem-se:

$$\begin{aligned} -2 \sum_{i=1}^n (X_i - \hat{\mu}) &= 0 \\ -\sum_{i=1}^n X_i + \sum_{i=1}^n \hat{\mu} &= 0 \\ n\hat{\mu} = \sum_{i=1}^n X_i \quad \therefore \hat{\mu} &= \frac{\sum_{i=1}^n X_i}{n} = \bar{X} \end{aligned}$$

A segunda derivada de SQ em relação a  $\mu$  é:  $\frac{d^2SQ}{d\mu^2} = 2n > 0$ .

Esse resultado indica que  $\mu = \bar{X}$  representa um ponto de mínimo da função SQ. Consequentemente, o estimador de quadrados mínimos da média  $\mu = \bar{X}$ . O estimador de momentos da variância  $\sigma^2$ , não viesado, é:  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \hat{X}_1)^2$ . Em que,  $\hat{X}_1 = \hat{\mu} = \bar{X}$  é o preditor de quadrados mínimos de  $X_i$  logo,

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

**MÉTODO DOS PERFIS PADRÃO:** É um método clássico para avaliação de reservas, baseado no princípio da variação gradual. O bloco de cubagem é definido entre duas seções adjacentes de amostragem. O teor médio do bloco é obtido pela média ponderada dos teores das seções pelas respectivas áreas. A reserva é o produto da área média pela distância entre as seções adjacentes e pelo teor médio do bloco.

**MÉTODO DOS POLÍGONOS:** É um método clássico para avaliação de reservas, baseado no princípio dos pontos mais próximos. Chama-se método dos polígonos porque a aplicação do princípio dos pontos mais próximos para teor e espessura de um furo de sondagem, resulta em um polígono em torno do mesmo, pois é o espaço comum em torno da zona de influência do furo em relação a outros furos. Aplicando-se

ao conjunto de furos de sondagem, obtém-se ao final da operação um conjunto de polígonos. A reserva pelo método dos polígonos é obtida pela somatória das reservas parciais dos furos, as quais são obtidas aplicando-se à área definida por cada polígono, a espessura do furo e o teor médio do furo.

**MÉTODO DOS TRIÂNGULOS:** É um método clássico para avaliação de reservas baseado no princípio das variações graduais ou lei da variação linear. A aplicação deste princípio, ou seja, que a variação entre dois pontos contíguos de amostragem é linear e contínua, em um conjunto de pontos distribuídos em área resulta numa malha de triângulos e, dão o nome de método dos triângulos. A reserva pelo método dos triângulos é obtida pela somatória das reservas parciais dos triângulos, as quais são obtidas aplicando-se à área do triângulo, a espessura média e o teor médio dos furos situados nos vértices de cada triângulo.

**MÉTODO LIVRE DE DISTRIBUIÇÃO:** Um método que não depende da forma da distribuição considerada. No caso de teste de hipótese, a expressão deve ser entendida no sentido de que o teste é independente da distribuição da hipótese nula. Em inglês *Distribution free method*.

**MÉTODO PASSO ATRÁS (BACKWARD):** Critério utilizado na análise de regressão para a seleção de variáveis regressoras para serem incorporadas ao modelo estudado, o qual é um procedimento que caracteriza-se por incorporar, inicialmente, todas as variáveis auxiliares em um modelo de regressão linear múltipla e percorre etapas, nas quais uma variável por vez pode vir a ser eliminada. Se em uma dada etapa não houver eliminação de alguma variável, o processo é então interrompido e as variáveis restantes definem o modelo final.

**MÉTODO PASSO A FRENTE (FORWARD):** Critério utilizado na análise de regressão para a seleção de variáveis regressoras para serem incorporadas ao modelo estudado, o qual é um procedimento que caracteriza-se por considerar, inicialmente, um modelo de regressão linear simples usando como variável auxiliar a variável de maior coeficiente de correlação amostral observado com a variável resposta, Y. Etapas se sucedem, quando uma variável por vez pode vir a ser incorporada. Se em uma etapa não houver uma inclusão, o processo é interrompido e as variáveis selecionadas até esta etapa definem o modelo final.

**MÉTODO PASSO A PASSO (STEPWISE):** Critério utilizado na análise de regressão para a seleção de variáveis regressoras para serem incorporadas ao modelo estudado, o qual é um procedimento que é uma generalização do procedimento passo a frente, quando após cada etapa de incorporação de uma variável, temos uma etapa em que uma das variáveis já selecionada pode ser descartada. As etapas de eliminação e incorporação de variáveis são efetuadas conforme descrito nos procedimentos todas as regressões possíveis, método passo atrás e método passo a frente. O procedimento chega ao final quando nenhuma variável é incluída ou descartada.

**MÉTODO SUBJETIVO:** Um método de atribuir probabilidades com base no julgamento.

**MÉTODO TRADICIONAL DE TESTES DE HIPÓTESES:** Método de teste de uma hipótese baseado na comparação estatística de teste com um valor crítico.

**MÉTODO NÃO-PARAMÉTRICO:** Método estatístico que não faz hipótese sobre as formas específicas das distribuições, não se preocupando, assim, em estimar valores desconhecidos de parâmetros.

**MÉTODOS:** Métodos são técnicas suficientemente gerais para se tornarem comuns a todas as ciências ou uma significativa parte delas.

**MÉTODOS AGLOMERATIVOS:** Procedimento hierárquico que começa com cada objeto ou observação em um grupo separado. Em cada passo que se segue, os dois agrupamentos de objetos que são mais parecidos são combinados para construir um novo agrupamento. O processo é repetido até que todos os objetos sejam finalmente combinados em um único agrupamento.

**MÉTODOS CLÁSSICOS:** São aqueles desenvolvidos e utilizados desde os primórdios da mineração e caracterizam-se por simplificações no cálculo de teores médios, os quais são aplicados em grandes volumes do depósito, à custa de grandes erros de estimativa, como o chamado efeitos de regressão.

**MÉTODOS DE ATRIBUIÇÃO:** Processo de estimação de dados perdidos de uma observação baseado em valores válidos das outras variáveis. O objetivo é empregar relações conhecidas que possam ser identificadas nos valores válidos da amostra para auxiliar na representação ou mesmo na estimação das substituições dos valores perdidos.

**MÉTODOS DE DECISÃO MÚLTIPLA:** Termo geral para descrever técnicas estatísticas para tratar problemas em que os possíveis resultados ou decisões são mais de dois. Em inglês *Multiple decision methods*.

**MÉTODOS DE INTEGRAÇÃO NUMÉRICA:** Coleção de métodos usados para resolver problemas de integração em inferência, principalmente quando a dimensão do espaço de parâmetros é moderada. Estes métodos são usados quando a solução analítica é inviável. Em inglês *Numerical integration methods*.

**MÉTODOS DE MONTE CARLO:** Termo usado para denotar a solução de problemas matemáticos em um contexto matemático por meio de experimentos de amostragem. Em inglês *Monte Carlo methods*.

**MÉTODOS DE SIMULAÇÃO:** Coleção de técnicas que consistem na geração aleatória de amostras de distribuições de interesse. Em inglês *Simulation methods*.

**MÉTODOS ESTATÍSTICOS:** Referem-se ao conjunto de técnicas utilizadas para a coleta, organização, análise e interpretação de dados, visando determinar os parâmetros da população amostrada.

**MÉTODOS E DEFINIÇÕES:** Definições são sempre melhoráveis. À luz de novos dados modificamos velhas definições. Frequentemente a definição é selecionada por que o cientista social tem o método disponível para estudar o problema, contanto que ele o defina de forma que o método se faça aplicável. Mais tarde, ele busca problemas úteis para aplicação do seu método.

**MÉTODOS NÃO PROBABILÍSTICOS:** São amostragens em que há uma escolha deliberada dos elementos da amostra. Não é possível generalizar os resultados das pesquisas para a população, pois as amostras não-probabilísticas não garantem a representatividade da população.

**MÉTODOS PROBABILÍSTICOS:** Exige que cada elemento da população possua determinada probabilidade de ser selecionado. Normalmente possuem a mesma probabilidade. Assim, se N for o tamanho da população, a probabilidade de cada elemento será  $\frac{1}{N}$ . Trata-se do método que garante cientificamente a aplicação das técnicas estatísticas de inferências.

**METODOLOGIA:** Methodo significa caminho; logia significa estudo. É o estudo dos caminhos a serem seguidos para se fazer ciência.



**METODOLOGIA DE LE E ZIDEK:** Metodologia bayesiana proposta em 1992 para a estimação de valores não observados em uma região por meio de um modelo de regressão Normal. Em inglês Le and Zidek methodology.

**MI:** Progenies de meios-irmãos.

**MINERAÇÃO DE DADOS:** Obtenção de informação acionável, válida e previamente desconhecida de grandes bancos de dados e a aplicação dela em modelos de negócios.

**MÍNIMO:** É o menor valor encontrado numa determinada variável, abreviadamente min..

**MÍNIMOS QUADRADOS:** É um método em regressão que consiste em adotar como estimativa dos parâmetros os valores que minimizem a soma dos quadrados dos desvios.

**MÍNIMOS QUADRADOS:** Procedimento de estimação usado em regressão simples e múltipla no qual os coeficientes de regressão são estimados de modo a minimizar a soma total dos quadrados dos resíduos.

**MISTURA DE DISTRIBUIÇÕES:** Um processo de obter uma média ponderada de um grupo de funções de distribuição que é uma nova função de distribuição. Em inglês Mixture of distributions.

**MIX DE MARKETING:** A singular mescla de produto/serviço, apreçamento, promoção, ofertas e distribuição destinadas a atender às necessidades de um grupo específico de consumidores.

**ML:** Máxima verossimilhança. Do inglês Maximum Likelihood o método da máxima verossimilhança consiste em determinar estimadores que possuam as características fundamentais de um estimador desejável, ou seja: linearidade, não enviesamento e variância mínima.

**MODA:** Medida representando o valor ou atributo mais frequentemente observado. Se a amostra tem 1.500 evangélicos, 288 católicos e 37 judeus, evangélicos é a categoria modal.

**MODA:** O valor mais comumente observado em uma distribuição, isto, é, o valor com o número mais alto de observações.

**MODA:** Medida de tendência central com características de que seu valor não é necessariamente único e às vezes nem existe. É o máximo da função de frequência. É de difícil manejo algébrico. Não possui características estatísticas desejáveis como a média. Tem como aplicação o fato de poder ser usada tanto para dados numéricos como dados não-numéricos. É um valor central expressivo com significado claro, porém, usar com precaução.

**MODA:** É o valor mais freqüente de um conjunto de dados.

**MODA:** A categoria mais freqüente de uma escala. Na série 1, 2, 2, 3, 3, 4 e 5, a moda é o valor 2, pois ela aparece três vezes.

**MODA:** A categoria ou valor dos dados que se caracteriza como possuidor da maior frequência de resposta.

**MODA (Mo):** É o valor que ocorre com maior frequência em uma série de valores.

**MODA (M):** Observação que ocorre com maior frequência numa amostra.

**MODA:** O valor que ocorre com maior frequência.

**MODA:** Valor que ocorre com maior frequência em uma lista de números.

**MODA:** A observação mais freqüente, ou a mais comum, ou ainda a que mais se repete em um conjunto de dados. Em inglês Mode.

**MODA:** Valor mais freqüente de uma distribuição. Numa representação gráfica, a moda é o escore correspondente ao ponto mais alto da curva. Em inglês mode.

**MODA:** A moda é uma medida de tendência central. É o valor com maior frequência. É normalmente usada para variáveis categóricas uma vez que as frequências de valores de uma variável contínua não são muito informativas.

**MODELAGEM DE EQUAÇÕES ESTRUTURAIS (SEM):** Técnica multivariada que combina aspectos de regressão múltipla examinando relações de dependência e análise fatorial representando conceitos não medidos, que são os fatores com múltiplas variáveis para estimar uma série de relações de dependência interrelacionadas simultaneamente.

**MODELO:** Um modelo é uma expressão formalizada de uma teoria ou da situação de causa que se assume ter gerado um conjunto de dados observados. Em análise estatística o modelo é representado geralmente por símbolos, isto é, em forma matemática, porém também existem modelos representados em grafos. Em inglês Model.

**MODELO:** O termo modelo pode significar uma representação matemática ou quantitativa de algum problema. Um modelo é um relacionamento entre variáveis específicas e parâmetros do sistema. Um modelo é uma secção de um sistema que representa uma parte específica de realidade, um objeto de interesse, ou um assunto de pesquisa da vida real.

**MODELO:** Forma simbólica de um princípio físico expressado por uma equação ou por uma fórmula.

**MODELO:** Conjunto especificado de relações de dependência que podem ser testadas empiricamente, ou seja, uma operacionalização de uma teoria. O propósito de um modelo é fornecer concisamente uma representação abrangente das relações a serem examinadas. O modelo pode ser formalizado em um diagrama de caminhos ou em um conjunto de equações estruturais.

**MODELO ADAPTATIVO:** Técnica para simplificar a análise conjunta combinando os modelos conjuntos auto-explicados e tradicionais. O exemplo mais comum é a análise conjunta adaptativa da Sawtooth Software.

**MODELO ADITIVO:** Modelo baseado na regra de composição aditiva, que considera que indivíduos apenas adicionam as utilidades parciais para calcular um escore geral ou valor total que indica utilidade ou preferência. É também conhecido como um modelo de efeitos principais. É o modelo conjunto mais simples em termos do número de avaliações e procedimento de estimação necessários.

**MODELO ADITIVO:** Modelo em que os fatores que influenciam a variável dependente possuem um efeito aditivo. Do inglês Additive model.

**MODELO ADITIVO GENERALIZADO:** Modelo alternativo aos modelos lineares generalizados, em que, para covariáveis  $x$ , o preditor linear é dado por  $\eta = \alpha + \sum_j f_j(x_j)$  no qual  $f_j(x_j)$  são funções suaves estimadas a partir dos dados. Em inglês Generalized additive model.



**MODELO ARIMA:** É o nome dado a um modelo muito utilizado na modelagem e previsões em séries temporais e que pode representar o comportamento dinâmico que muitos conjuntos de dados apresentam. Arima significa modelo autoregressivo integrado de média móvel. Em inglês AutoRegressive Integrated Moving Average.

**MODELO AUTOEXPLICADO:** Técnica composicional para executar análise conjunta na qual o respondente fornece as estimativas das utilidades parciais diretamente, sem fazer escolhas.

**MODELO AUTOREGRESSIVO:** Um modelo econométrico que utiliza a própria variável de interesse defasada no tempo para explicar os valores atuais. Em inglês Autoregressive model.

**MODELO AUTOREGRESSIVO:** É o modelo envolvendo um ou mais valores defasados da variável dependente entre suas variáveis explicativas. Cujas equações são dadas por:  $Y_t = \alpha + \beta_0 X_t + \gamma Y_{t-1} + e_t$ .

**MODELO BILINEAR:** Se um modelo linear com duas covariáveis pode ser arrumado na forma  $(\beta_0 + \beta_2 x_2) + (\beta_1 + \beta_2 x_2)x_1$ , ele mostra uma relação linear em  $x_1$  para a qual tanto intercepto como coeficiente angular são funções lineares de  $x_2$ , este modelo é conhecido por modelo bilinear. Em inglês Bilinear model.

**MODELO COMPOSICIONAL:** Classe de modelos multivariados que baseia a relação de dependência em observações do respondente referentes às variáveis dependentes e independentes. Esses modelos calculam ou compõem a variável dependente a partir dos valores fornecidos pelo respondente para as variáveis independentes. Os principais entre esses métodos são a análise de regressão e a análise discriminante. Esses modelos estão em contraste direto com os modelos decomposicionais.

**MODELO FIXO OU DO TIPO I:** É aquele cujos parâmetros nele existentes, são todos de efeito fixo; excessão feita ao erro.

**MODELO ALEATÓRIO OU DO TIPO II:** É aquele o que contém apenas efeitos aleatórios, excessão da média.

**MODELO MISTO OU DO TIPO III:** É aquele onde aparecem no modelo efeitos fixos e efeitos aleatórios.

**MODELO DE ARRHENIUS:** Em uma ampla variedade de reações químicas onde existe, por exemplo, corrosão de metais, queda de rendimento de lubrificantes, ou difusão de materiais semicondutores, o comportamento da taxa em que acontecem estes eventos segue a equação seguinte: taxa do processo =  $A \exp\left\{-\frac{Q}{R}\right\}$ , onde A é uma constante, Q é a energia de reação, R é a constante de gases, e T é a temperatura

absoluta. Do inglês Arrhenius model.

**MODELO DE BOX E JENKINS:** Modelo de previsão de séries temporais não estacionárias, com ou sem efeito sazonal. Este modelo depende de autoregressões de m-éssima ordem da série, com resíduos de média móvel. Em inglês Box and Jenkins model.

**MODELO DE COX:** O mesmo que Modelo dos azares proporcionais.

**MODELO DE CRESCIMENTO LINEAR (MCL):** Modelo polinomial de primeira ordem que inclui um parâmetro adicional para descrever o crescimento de um processo. Em inglês Linear growth model (LGM).

**MODELO DE DEFASAGEM DISTRIBUÍDA:** É o modelo usado na análise de regressão envolvendo dados de séries temporais, que inclui não somente os valores correntes, mas também os valores defasados ou passados das variáveis explicativas (X's). A equação é dada por:  $Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + e_t$ .

**MODELO DINÂMICO:** É o modelo envolvendo dados de séries temporais que retrata o caminho temporal da variável dependente em relação ao(s) seu(s) valor(es) passado(s).

**MODELO DE ELABORAÇÃO:** Método lógico para entender relações causais entre variáveis. A relação entre duas variáveis é examinada entre subconjuntos criados por uma variável de controle.

**MODELO DE ESCALA:** X tem um modelo de escala se uma função g e a quantidade  $\sigma$  existem de forma que a distribuição de X dado  $\sigma$  satisfaz  $f(x|\theta) = \sigma^{-1} g\left(\frac{x}{\sigma}\right)$ . Neste caso,  $\sigma$  é chamado de parâmetro de escala. Em inglês Scale model.

**MODELO DE LOCAÇÃO:** X tem um modelo de locação se uma função g e a quantidade  $\theta$  existem de forma que a distribuição de X, dado  $\theta$  satisfaz  $f(x|\theta) = g(x-\theta)$ . Neste caso,  $\theta$  é chamado de parâmetro de locação. Em inglês Location model.

**MODELO DE LOCAÇÃO-ESCALA:** X tem um modelo de locação e escala se uma função g e as quantidades  $\theta$  e  $\sigma$  existem de forma que a distribuição de probabilidade de X dado  $(\theta, \sigma)$  satisfaz  $f(x|\theta, \sigma) = \sigma^{-1} g((x-\theta)/\sigma)$ . Neste caso,  $\theta$  é chamado de parâmetro de locação e  $\sigma$  o parâmetro de escala. Em inglês Location-scale model.

**MODELO LOGIT:** É um modelo obtido da seguinte forma: Considere a seguinte equação.  $P_i [Y_i = 1 | X_i] = \frac{1}{1 + e^{-(\beta_1 + \beta_2 X_i)}}$ , que pode ser escrito como segue:  $P_i = \frac{1}{1 + e^{-Z_i}}$ , em que  $Z_i = \beta_1 + \beta_2 X_i$ ,

a equação  $P_i = \frac{1}{1 + e^{-Z_i}}$  é conhecida como função distribuição logística acumulada. Se escrevermos

$1 - P_i = \frac{1}{1 + e^{Z_i}}$ , podemos também escrever  $\frac{P_i}{1 - P_i} = \frac{1 + e^{Z_i}}{1 + e^{-Z_i}} = e^{Z_i}$ . Onde  $\frac{P_i}{1 - P_i}$  representa a razão

de probabilidade. Se aplicarmos o log natural em  $\frac{P_i}{1 - P_i} = \frac{1 + e^{Z_i}}{1 + e^{-Z_i}} = e^{Z_i}$  obtemos um resultado bastante

interessante, dado por:  $L_i = \ln\left(\frac{P_i}{1 - P_i}\right) = Z_i = \beta_1 + \beta_2 X_i$ , ou seja, L, o log da razão de probabilidade é não

somente linear em X, mas também nos parâmetros, e assim L é chamada de logit, daí o nome modelo logit para o modelo do tipo da equação

$$L_i = \ln\left(\frac{P_i}{1 - P_i}\right) = Z_i = \beta_1 + \beta_2 X_i$$

**MODELO DE MENSURAÇÃO:** Submodelo em SEM que significa modelagem de equações estruturais, em inglês, structural equation modeling, que especifica os indicadores para cada construto, e avalia a confiabilidade de cada construto para estimar as relações causais. O modelo de mensuração é semelhante em forma à análise fatorial; a principal diferença reside no grau de controle fornecido ao pesquisador. Em análise fatorial, o pesquisador pode especificar apenas o número de fatores, mas todas as variáveis têm cargas ou seja, elas atuam como indicadores para cada fator. No modelo de mensuração, o pesquisador

especifica quais variáveis são indicadoras de cada construto, com as demais variáveis sem cargas exceto aquelas em seu construto especificado.

**MODELO DE PREDIÇÃO:** Modelo completo, com coeficientes, para ser usado na predição de um determinado desfecho, dada a presença de diversas variáveis independentes.

**MODELO DE PROBABILIDADE LINEAR [MPL]:** É um modelo simples dado pela equação  $Y_i = \beta_1 + \beta_2 X_i + e_i$ , que expressa o  $Y_i$ , dicotômico como uma função linear de uma ou mais variáveis explicativas  $X_i$ , já que  $E[Y_i|X_i]$ , a esperança matemática condicional de  $Y_i$  dado  $X_i$ , pode ser interpretada como a probabilidade condicional de que o evento aleatório ocorrerá dado  $X_i$ , ou seja,  $P_r[Y_i = 1|X_i]$ .

**MODELO PROBIT:** É um modelo de estimativa que surge da função de probabilidade acumulada normal ou função de repartição ou de distribuição normal, cuja equação dada a seguir expressa a probabilidade de ocorrer um evento, senão vejamos.

$$P_i = P_r(Y = 1) = P_r(I_i^* \leq I_i) = F(I_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{I_i} e^{-t^2/2} dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\beta_1 + \beta_2 X_{ii}} e^{-t^2/2} dt,$$

em que  $t$  é uma variável normal padronizada, isto é com média zero e variância igual a 1, cuja notação é dada por  $t \sim N(0,1)$ .

**MODELO TOBIT:** É uma extensão do modelo probit. Um modelo também conhecido como modelo de regressão censurada ou modelos de variável dependente limitada, por causa da restrição imposta sobre os valores assumidos pelo regressando que são as variáveis independentes  $X$ . Matematicamente, pode-se expressar o modelo tobit como mostrado na seguinte equação:  $Y_i = \beta_1 + \beta_2 X_{2i} + e_{2i}$  se  $LD > 0$  e  $Y_i = 0$ , caso contrário, em que  $LD$  é o lado direito, outras variáveis  $X$  podem ser facilmente acrescentadas no modelo.

**MODELO DE QUASE VEROSSIMILHANÇA:** Os modelos de quase-verossimilhança (MQV's) pertencentes à classe dos modelos lineares generalizados (MLG's), pois estes modelos apresentam uma componente sistemática com estrutura linear das variáveis independentes e função de ligação que relaciona a média ( $\mu_i$ ) da variável dependente à estrutura linear das variáveis independentes ( $x_i$ ),  $i = (1, \dots, p)$ . Porém para estes modelos, não há a necessidade de assumir a princípio alguma distribuição de probabilidade para a variável dependente. Por conseguinte, a esperança matemática e a variância da variável aleatória não são conhecidas a priori.

**MODELO DE REGRESSÃO MULTIVARIADA:** Modelo de regressão com mais de uma variável dependente. Em inglês Multivariate regression model.

**MODELO DE SIMULAÇÃO:** Modelo de um sistema dinâmico complicado para achar uma solução analítica explícita, porém pode ser simulado para diversas condições em forma iterativa usando valores iniciais numéricos adequados. Em inglês Simulation model.

**MODELO DOS AZARES PROPORCIONAIS:** Modificação da regressão logística múltipla para permitir a modelagem multivariável dos dados de uma análise de tábua de vida.

**MODELO DECOMPOSICIONAL:** Classe de modelos multivariados que decompõem a preferência do respondente. Essa classe de modelos apresenta ao respondente um conjunto predefinido de variáveis independentes, geralmente na forma de um produto ou serviço hipotético ou real, e então pede uma avaliação ou preferência geral do produto ou serviço. Uma vez dada, a preferência é decomposta relacionando-se

os atributos conhecidos do produto que se tomam as variáveis independentes com a avaliação da variável dependente. O principal entre esses modelos é a análise conjunta e algumas formas de escalonamento multidimensional.

**MODELO DETERMINÍSTICO:** Um modelo matemático no qual os parâmetros e variáveis não estão sujeitos a flutuações randômicas ou aleatórias, tal que o sistema inteiramente definido pelas condições iniciais escolhidas em qualquer tempo. Contrasta com um modelo aleatório.

**MODELO DINÂMICO:** Um modelo é considerado dinâmico se possuir uma ou ambas propriedades: a primeira é que ao menos uma variável ocorre nas equações estruturais com valores considerados em diferentes pontos de tempo, dentre outros; e a segunda é que ao menos uma equação contém uma função do tempo. Em inglês Dynamic model.

**MODELO ESPAÇO-TEMPORAL:** Modelo estatístico que considera, simultaneamente, as características temporais e espaciais de um fenômeno. Em inglês Space-time model.

**MODELO ESTRUTURADO POR FAIXA ETÁRIA:** Modelo matemático que leva em consideração a divisão da população do hospedeiro em diferentes faixas etárias. Tais modelos são usados quando se consideram fatores tais como infecção idade-dependente, taxas de morbidade ou mortalidade por faixa de idades, ou esquemas de vacinação específico para determinadas faixas etárias.

**MODELO ESTRUTURAL:** Conjunto de uma ou mais relações de dependência conectando os construtos hipotetizados do modelo. O modelo estrutural é mais útil para representar as inter-relações de variáveis entre relações de dependência.

**MODELO EXATAMENTE IDENTIFICADO:** Modelo estrutural com zero graus de liberdade que atende exatamente a condição de ordem. Isso corresponde a um modelo perfeitamente ajustado, mas não tem generalização.

**MODELO HÍBRIDO:** Ver modelo adaptativo.

**MODELO KOHONEN:** Modelo de rede neural que é delineado para problemas de agrupamento e opera em um modo de aprendizado não-supervisionado.

**MODELO LINEAR:** Um modelo cujas equações que unem as variáveis estão em uma forma linear. Em inglês Linear model.

**MODELO LINEAR GERAL:** Modelo geral que apresenta a relação linear da primeira ordem entre múltiplas variáveis independentes e uma variável dependente que é o efeito. A análise de variância (ANOVA), a análise de covariância (ANCOVA), a regressão linear múltipla e outras técnicas multivariáveis são variações desse modelo básico.

**MODELO LINEAR GENERALIZADO:** Um modelo linear generalizado é uma extensão dos modelos lineares que satisfaz: i) as componentes aleatórias,  $Y$ , têm distribuição independente que pertence à família exponencial com valor esperado igual a  $\mu$  e variância constante; ii) as covariáveis produzem um preditor linear dado por  $\eta = \sum_j x_j \beta_j$ ; iii) a ligação entre as componentes aleatória e sistemática é dada por  $g(\mu) = \eta$ . Em inglês tem a sigla GLM ou seja Generalized linear model.

**MODELOS LINEARES MISTOS GENERALIZADOS HIERÁRQUICOS (HGLMM):** São modelos em que uma distribuição de probabilidade e uma função de ligação podem ser especificados para cada fator aleatório, por exemplo, a situação em os dados seguem distribuição de Poisson e a função de ligação especificada para os resíduos é a logarítmica, e assim nesse caso uma suposição mais apropriada para os demais fatores aleatórios é uma distribuição gama com função de ligação logarítmica. Como os fatores aleatórios nem sempre são os de classificação hierárquica, uma denominação alternativa é modelos lineares mistos generalizados estratificados, HGLMM's.

**MODELO LOGITO:** Modelo linear generalizado com ligação logística. Em inglês Logit model.

**MODELO LOG-LINEAR:** Modelo usado em técnica de análise multivariada de dados categóricos onde é comum se ter tabelas de contingência complexas, com grande número de variáveis. Em inglês log-linear model.

**MODELO MATEMÁTICO:** Suporte formal para ideias sobre a interação entre os componentes da interação parasita-hospedeiro. A construção de um modelo requer três tipos de informação: i) uma clara compreensão da interação entre o agente infeccioso e o seu hospedeiro, ii) o modo e a taxa de transmissão entre os indivíduos, e iii) características da população hospedeira, tais como demografia e comportamento. Os modelos matemáticos podem ajudar na exploração do comportamento do sistema sob várias condições para determinar os fatores dominantes que dão origem aos padrões e fenômenos observados. Eles também ajudam na interpretação de coleções de dados e nas estimativas de parâmetros, e fornecem ferramentas para identificar possíveis abordagens para o controle e para avaliar o impacto potencial de medidas intervencionistas.

**MODELO NULO:** Ponto de referência ou padrão de comparação usado em índices de ajuste incremental. O modelo nulo é considerado, por hipótese, como o mais simples que pode ser teoricamente justificado. O exemplo mais corriqueiro é o modelo de um só construto relacionado com todos os indicadores e sem erro de mensuração.

**MODELO NORMIT:** O mesmo que modelo probit.

**MODELO PARA RESPOSTA BINÁRIA:** Quando os dados assumem dois valores possíveis, que geralmente representam a ausência ou presença de uma característica, devem ser tomados cuidados especiais na hora de definir o modelo. Um exemplo de modelo para este tipo de dado é o modelo de regressão logística. Em inglês Model for binary response.

**MODELO PROBITO:** Em modelos lineares generalizados, o modelo probito considera a proporção de sobreviventes,  $\pi_x$ , para a dose  $x$  como função de  $x$ , que é geralmente medida na escala logarítmica. Em inglês Probit model.

**MODELO SUBIDENTIFICADO:** Modelo estrutural com um número negativo de graus de liberdade. Isso indica uma tentativa de estimar mais parâmetros do que o possível com a matriz de dados.

**MODELO SUPERIDENTIFICADO:** Modelo estrutural com um número positivo de graus de liberdade, indicando que algum nível de generalização pode ser possível. O objetivo é atingir o máximo ajuste de modelo com o maior número de graus de liberdade.

**MODELO TEMPORAL DISCRETO:** Um modelo no qual o sistema salta de um estado ao seguinte em intervalos fixos de tempo. Estes modelos discretos são simples de entender mas frequentemente

difíceis analisar; contrasta com os modelos temporais contínuos. Os parâmetros em tais modelos referem à quantidade de mudança em cada etapa finita; por isso eles são às vezes chamados de taxas finitas. Via de regra, uma equação diferencial é o que se origina de uma equação diferença quando os intervalos de tempo se tornam infinitamente pequenos.

**MODELO TRIDIMENSIONAL DE BLOCOS:** É o equivalente da malha regular em três dimensões. Os blocos, cujas dimensões são definidas em função das informações de pesquisa e do objetivo do cálculo de reservas, são dispostos num arranjo regular tridimensional, que deve representar aproximadamente o depósito em estudo. Observe que nem todos os blocos poderão conter informações e por isso não serão avaliados, além de outros que poderão estar fora dos limites do depósito. A reserva será computada efetivamente a partir dos blocos avaliados dentro dos limites físicos do depósito.

**MODELOS ANINHADOS:** Modelos que têm os mesmos construtos, mas diferem em termos do número ou tipos de relações causais representadas. A forma mais comum de modelo aninhado acontece quando uma única relação é acrescentada ou eliminada de um outro modelo. Logo, o modelo com menos relações estimadas está aninhado dentro do modelo mais geral.

**MODELOS EQUIVALENTES:** Modelos comparáveis que têm o mesmo número de graus de liberdade que são modelos aninhados, mas diferem em um ou mais caminhos. O número de modelos equivalentes expande muito rapidamente conforme a complexidade do modelo aumenta.

**MODELOS MULTIVARIÁVEIS:** Modelos estatísticos que têm uma variável dependente o qual é o desfecho, porém incluem mais de uma variável independente.

**MODELOS DE QUASE-VEROSSIMILHANÇA:** É uma generalização dos modelos lineares generalizados (MLG) no sentido de assumirem uma função de variância para a variável resposta bem como uma relação funcional entre a média e o vetor paramétrico  $b$ , no entanto não requer mais o conhecimento da distribuição da resposta. A distribuição da variável resposta ficará determinada quando a função de variância escolhida coincidir com a função de variância de alguma distribuição da família exponencial.

**MODELOS ESTOCÁSTICOS:** O oposto de modelo determinístico. Modelos matemáticos que levam em consideração a presença de alguma randomicidade em um mais de seus parâmetros ou variáveis. As previsões do modelo, contudo, não dão um resultado específico, mas uma distribuição de probabilidades das possíveis estimativas. Podemos discernir estocasticidade demográfica, que se origina do caráter discreto de indivíduos e de eventos individuais tais como nascimentos, e estocasticidade ambiental, que se origina de interações mais ou menos imprevisíveis com o mundo externo.

**MODELOS LINEARES GENERALIZADOS (MLG):** Os modelos lineares generalizados constituem uma extensão dos modelos lineares de regressão múltipla e foram apresentados pela primeira vez por Nelder e Wedderburn (1972). A passagem dos modelos lineares de regressão múltipla, na sua versão clássica, para os modelos lineares generalizados permitiu alargar as hipóteses admitidas. A variável resposta do modelo passa a poder provir de um universo que siga uma lei de distribuição da família exponencial, deixando de ter obrigatoriamente uma distribuição normal. Além disso, a passagem dos modelos lineares de regressão clássicos para os modelos lineares generalizados contém outras generalizações. Se nos primeiros a relação entre o valor médio da variável resposta e a combinação linear das variáveis explicativas é a função identidade, nos segundos aquela relação pode ser estabelecida por qualquer função monótona e diferenciável. Um modelo linear generalizado, segundo McCullagh e Nelder (1989), pode ser caracterizado em três



pontos. Considerem-se  $n$  variáveis aleatórias  $Y_i$  ( $i = 1, \dots, n$ ) independentes de média  $\mu_i$ , respectivamente, e função de probabilidade ou função densidade de probabilidade pertencente à família exponencial, isto é,

$$f(y_i | \theta_i, \phi) = \exp \left\{ \frac{y_i \theta_i - b(\theta_i)}{a_i(\phi)} + c(y_i, \phi) \right\},$$

em que  $a_i(\cdot)$ ,  $b(\cdot)$  e  $c(\cdot)$  são funções específicas para cada

distribuição. Se  $\phi$  for conhecido tem-se uma distribuição da família exponencial com parâmetro canônico  $\theta$ . Caso  $\phi$  seja desconhecido a distribuição pode ou não pertencer a família exponencial, e a então que considerar apenas os casos em que a distribuição pertence àquela família. A função  $a_i(\phi)$  assume, na maior parte dos casos, a forma  $a_i(\phi) = \frac{\phi}{\omega_i}$ , onde se assume que  $\omega_i$  é conhecido, enquanto que  $\phi$ , deno-

minado parâmetro de dispersão ou de escala, pode sê-lo ou não. Note-se que, num modelo linear generalizado a distribuição de todas as variáveis aleatória  $Y_i$  seguem a mesma forma como: a da Normal, Poisson, Gama, ou outra da família exponencial, e o parâmetro de escala  $\phi$ , é constante, isto é, não varia com o índice  $i$  das variáveis aleatórias  $Y_i$ . Suponha-se agora a existência de  $p$  variáveis explicativas  $X_j$  ( $j = 1, \dots, p$ ). A  $n$  observações destas  $p$  variáveis constituem a matriz  $X$ . Isto é, cada elemento  $x_{ij}$  da matriz  $X$  designa o valor da  $j$ -ésima variável explicativa para a observação  $i$  ( $i = 1, \dots, n$ ). A partir dessa matriz define-se um preditor linear  $\eta_i$ ,  $i = 1, \dots, n$ , da forma  $\eta_i = \sum_{j=1}^p x_{ij} \beta_j$  constituindo os  $\beta_j$ ,  $j = 1, \dots, p$ , um vetor de parâ-

metro desconhecidos, a ser estimados a partir dos dados. Na forma matricial temos:  $\vec{\eta} = X\vec{\beta}$ . Dessa forma fica especificada a componente sistemática do modelo; ii) As duas componentes anteriores relacionam-se por meio de uma função de ligação  $g$ , que se admite existir, ser monótona e diferenciável, e que transforma  $\mu_i$  em  $\eta_i$ , ou seja,  $\eta_i = g_i(\mu_i)$ ,  $i = 1, 2, \dots, n$ . De fato, teoricamente pode definir-se uma função de ligação para cada observação  $i$ , mas na prática apenas o caso em que  $\eta_i = g_i(\mu_i)$ ,  $i = 1, 2, \dots, n$ , tem aplicabilidade. Dessa forma ficam definidas as componentes de um modelo linear generalizado.

**MODELOS LINEARES DE PROBABILIDADES:** São modelos frequentemente chamados de análise logit, os quais são uma combinação de regressão múltipla com análise discriminante múltipla. Essa técnica é similar à análise de regressão múltipla no sentido em que uma ou mais variável (is) independente(s) é ou são empregada(s) para prever uma única variável dependente. O que diferencia um modelo linear de probabilidade de regressão múltipla é que a variável dependente é não métrica, como no caso da análise discriminante. A escala não métrica da variável dependente requer diferenças no método de estimação e nas suposições sobre o tipo de distribuição inerente, ainda que em muitos outros aspectos seja bastante semelhante à regressão múltipla. Assim, uma vez que a variável dependente seja corretamente especificada e a técnica de estimação apropriada seja empregada, os fatores básicos considerados em regressão múltipla também serão usados aqui. Os modelos lineares de probabilidade diferem da análise discriminante principalmente no sentido em que eles acomodam todos os tipos de variáveis independentes métricas e não métricas e não requerem a suposição de normalidade multivariada. No entanto, em muitos casos, particularmente com mais de dois níveis de variável dependente, a análise discriminante é a técnica mais apropriada.

**MODELOS LINEARES HIERÁRQUICOS:** Em uma abordagem bayesiana, um modelo que define uma relação linear para a média das observações, cujos parâmetros apresentam uma estrutura de hierarquia

com os hiperparâmetros que podem, por sua vez, ter outra relação hierárquica com hiperparâmetros de mais alto nível, e assim por diante. Em inglês Hierarchical linear models.

**MODELOS LOG-LINEARES:** Forma de análise de dados que usa cálculos logarítmicos para simplificar a análise de complexas tabulações cruzadas multivariadas.

**MODELOS TEÓRICOS DE VARIOGRAMAS:** São os modelos que permitem descrever a variabilidade espacial de um fenômeno em função da distância.

**MODERADOR DE GRUPO DE FOCO:** A pessoa contratada pelo cliente para liderar um grupo de foco. Essa pessoa talvez precise ter uma formação em psicologia ou sociologia ou, pelo menos, em marketing.

**MODIFICAÇÃO DE EFEITO:** Ocorre quando a força, ou mesmo a direção da influência de um fator causal sobre um desfecho é alterada por uma terceira variável, a modificadora de efeito.

**MÓDULO:** Dizemos que os números  $m$  e  $n$  são iguais módulo  $p$ , se  $p$  divide exatamente  $m-n$ . Em inglês modulo.

**MÓDULO OU VALOR ABSOLUTO DE X:** Símbolo matemático representado por  $|X|$ .

**MOLDURA DE AMOSTRAGEM:** Lista, ou quase lista, de unidades que compõem uma população da qual a amostra é selecionada. Para que a amostra seja representativa da população é essencial que a moldura da amostragem inclua todos ou quase todos os membros da população.

**MOMENTO:** Em geral, um momento é o valor médio da potência de uma variável aleatória,  $\int x^r dF(x)$ , ou, quando se fala em momento em torno de um valor central,  $m$ , é dada pela expressão  $\int (x-m)^r dF(x)$ . Em inglês Moment.

**MOMENTOS NEGATIVOS:** Os momentos negativos de uma distribuição de frequências são os momentos dos recíprocos das potências de uma variável. Em inglês Negative moments.

**MORBIDADE:** Variável característica das comunidades de seres vivos refere-se ao conjunto dos indivíduos que adquiram doenças num dado intervalo de tempo. Denota-se morbidade ao comportamento das doenças e dos agravos à saúde em uma população exposta.

**MORBIDADE, ESTATÍSTICAS DE:** Referente ao número de casos de doenças ou de agravos à saúde, de maneira geral.

**MORTALIDADE:** Sujeitos participantes do experimento que abandonam ou são afastados dele, não podendo, destarte, continuar no experimento até a sua conclusão.

**MORTALIDADE:** Perda de unidades de teste ou de indivíduos durante o curso de um experimento. O problema é que aqueles que desistem podem estar sistematicamente diferentes dos que permanecem.

**MORTALIDADE:** Variável característica das comunidades de seres vivos, refere-se ao conjunto dos indivíduos que morreram num dado intervalo de tempo.

**MORTALIDADE:** Refere-se às árvores que morrem na amostra no período entre duas medições de uma parcela permanente. A mortalidade pode ser definida em termos absolutos com base no número de árvores por unidade de área, ou ainda com base na área basal, volume, biomassa, dentre outros. Em termos relativos, a mortalidade é normalmente definida na forma de uma taxa instantânea assumindo-se que essa taxa é constante ao longo do tempo que é chamado modelo exponencial de mortalidade.



**MORTALIDADE, ESTATÍSTICAS DE:** Referente ao número de óbitos.

**MORTALIDADE INFANTIL:** Termo para designar todos os óbitos de crianças menores de 1 ano, ocorridos em determinada área, em dado período de tempo. Ver coeficiente de mortalidade infantil.

**MORTALIDADE INFANTIL PROPORCIONAL:** Proporção de óbitos de crianças menores de 1 ano no conjunto de todos os óbitos. Seu índice é calculado dividindo-se o número de óbitos menores de 1 ano pelo total de óbitos e multiplicando-se por 100.

**MQE:** Média de quadrados dos resíduos ou erros de ajuste.

**QQR:** Média de quadrados dos desvios explicados pela regressão.

$m'_r$  : Valor do momento amostral simples ou ordinário de ordem r.

$m_r$  : Valor do momento amostral centrado de ordem r.

**MUITO MAIOR:** Símbolo matemático dado por >>.

**MUITO MENOR:** Símbolo matemático dado por <<.

**MULTICOLINEARIDADE:** Grau em que uma variável pode ser explicada pelas outras variáveis na análise.

**MULTICOLINEARIDADE:** Problema que surge na análise da regressão quando duas ou mais variáveis independentes estão estreitamente correlacionadas.

**MULTICOLINEARIDADE:** Em análise de regressão, uma situação em que existe uma relação linear que conecta as variáveis independentes. Em inglês Multicollinearity.

**MULTICOLINARIEDADE:** Extensão em que uma variável pode ser explicada pelas outras variáveis na análise. À medida que a multicolineariedade aumenta, fica mais complicada a interpretação da variável estatística, uma vez que se torna mais difícil verificar o efeito de qualquer variável, devido suas inter-relações.

**MULTICOLINEARIDADE:** Grau em que uma variável pode ser explicada pelas outras variáveis na análise. Quando a multicolinearidade aumenta, ela complica a interpretação da variável estatística por ser mais difícil de determinar o efeito de qualquer variável individual, devido às inter-relações entre variáveis.

**MULTICOLINEARIDADE:** Este é um pressuposto importante na análise regressão múltipla, pois se uma variável independente for uma combinação linear de outras variáveis independentes, coloca em risco toda a análise. Não adianta tentar modelar uma variável em função de várias variáveis correlacionadas, essas não incrementam o poder explicativo do modelo. Neste caso, use o modelo de regressão stepwise, backward ou forward.

**MULTICOLINEARIDADE:** Termo cunhado por Frisch (1934), que significa a existência de uma perfeita ou exata relação linear entre algumas ou todas as variáveis explicativas X de um modelo de regressão, e também serve para representar os casos em que as variáveis independentes X são intercorrelacionadas. Rigorosamente falando, multicolinearidade se refere à existência de mais de uma relação linear exata, enquanto colinearidade se refere à existência de uma única relação linear. Mas tal distinção raramente se sustenta na prática, de modo que multicolinearidade se refere a ambos os casos.

**MULTICOLINEARIDADE:** Ver colinearidade.

**MULTIMODAL:** Que tem mais de duas modas.

**MULTIPLICADOR DE LAGRANGE:** É a técnica aplicada para obtenção do mínimo da função variância de estimativa, sujeito a condição de não enviesamento, conforme a colocação do problema a seguir: Minimizar a função:  $\bar{C}(V, V) - 2\sum_{\alpha} \lambda_{\alpha} \bar{C}(V, V_{\alpha}) + \sum_{\alpha} \sum_{\beta} \lambda_{\alpha} \cdot \lambda_{\beta} \bar{C}(V_{\alpha}, V_{\beta})$ .

**Sujeita à condição de não enviesamento:**  $\sum_{\alpha} \lambda_{\alpha} = 1$ . Para resolver esse problema de otimização, forma-se o Lagrangiano:

$$L(\lambda_1, \lambda_2, \dots, \lambda_n, \mu) = \bar{C}(V, V) - 2\sum_{\alpha} \lambda_{\alpha} \bar{C}(V, V_{\alpha}) + \sum_{\alpha} \sum_{\beta} \lambda_{\alpha} \cdot \lambda_{\beta} \bar{C}(V_{\alpha}, V_{\beta}) - 2\mu \left( \sum_{\alpha} \lambda_{\alpha} - 1 \right),$$

onde:  $L(l_1, l_2, \dots, l_n, m)$  é o lagrangiano; m é o multiplicador de Lagrange. Fazendo cada uma das n derivadas parciais  $\frac{dL}{dl_i}$  iguais a zero tem-se n equações:  $\sum_{\beta=1}^n \lambda_{\beta} \bar{C}(V_{\alpha}, V_{\beta}) - \mu = \bar{C}(V_{\alpha}, V)$  para o valor de  $\alpha = 1, \dots, n$ . Fazendo  $\frac{dL}{dm}$

igual a zero, tem-se:  $\sum_{\beta=1}^n \lambda_{\beta} = 1$ , ou seja, as (n+1) equações do sistema de equações de krigagem.

**MUTUAMENTE EXCLUDENTE:** Qualidade desejada num conjunto de respostas que acompanha uma questão fechada, que apenas uma resposta se aplique a cada respondente.

**MATERIAL EXPERIMENTAL PARA USO EM ENSAIOS AGRÍCOLAS:** Todo pesquisador da área agrícola deve atentar para os seguintes itens a seguir no planejamento dos experimentos a serem conduzidos por ele, os quais são: i) escolha da cultura e da época de cultivo adequada; ii) escolha do local do experimento que deve ser típico das condições de campo na região investigada e uniforme em sua topografia, fertilidade do solo, profundidade, estrutura e qualidade do subsolo. ii.1) o histórico do local: verificar a sequência de cultivos, aplicação de adubos, tratamentos, culturas etc. nos períodos anteriores; ii.2) condições fitossanitárias: implementar um controle rigoroso particularmente em relação a doenças e organismos localizados no solo, como podridão da raiz, nematóides e a murcha do Guandu provocada por Fusarium que aparece entre dois e três anos. Uma área recém desmatada não pode ser usada para experimentação de campo sem um cultivo prévio de uniformização de pelo menos um ano. Uma regra geral é escolher o local durante o ano anterior ao experimento. No preparo da área experimental o pesquisador deve ter atenção no preparo do solo que deve ser efetuado segundo as práticas comuns de cada cultura, mas deve ser feito com bastante cuidado. As operações devem ser preferencialmente mecanizadas. Efetuar a nivelção do terreno, pois reduz o perigo de empoçamento de água da chuva e melhora a superfície de semeadura assegurando, portanto, um stand ou número de plantas uniforme. O plantio de um experimento deverá ser preferencialmente feito, em um dia no caso de delineamento inteiramente casualizado ou 1 repetição no caso do delineamento em blocos casualizados, evitar plantar partes de uma repetição em diferentes dias. Todas as operações que antecedem a semeadura devem ser efetuadas transversalmente às parcelas, ou seja, paralelamente ao eixo prolongado dos blocos. Isso assegurará um mínimo de variação entre as parcelas dentro do bloco.

**MEDIDAS ESTATÍSTICAS OU ÍNDICES COMUNS DE DISTÂNCIAS OU DISSIMILARIDADE UNIVARIADAS E MULTIVARIADAS ENTRE DOIS GRUPOS OU AMOSTRAS DE TRATAMENTOS EXPERIMENTAIS:** Os ecólogos, cientistas, pesquisadores ou investigadores da área ambiental e estatísticos desenvolveram muitas medidas para quantificar a distância entre duas amostras ou populações multivariadas. Essas medidas dividem-se em duas categorias, que são as distâncias métricas e distâncias semimétricas. As distâncias métricas têm quatro propriedades as quais são: i) a distância mínima é igual a zero ou seja,  $d = 0$ , se dois objetos ou amostras  $x_1$  e  $x_2$ , são idênticos, então a distância d entre eles

também é igual a 0:  $x_1 = x_2 \Rightarrow d(x_1, x_2) = 0$ ; ii) a medida de distância  $d$  é sempre positiva se dois objetos ( $x_1$  e  $x_2$ ) não são idênticos:  $x_1 \neq x_2 \Rightarrow d(x_1, x_2) > 0$ ; iii) a medida de distância  $d$  é simétrica:  $d(x_1, x_2) = d(x_2, x_1)$  e iv) a medida de distância  $d$  satisfaz a desigualdade triangular: para três objetos  $x_1, x_2$  e  $x_3$ ,  $d(x_1, x_2) + d(x_2, x_3) \geq d(x_1, x_3)$ . As distâncias euclidiana, de Manhattan, de Chord, de Mahalanobis, Chi quadrado e de Jaccard são exemplos de distâncias métricas. As distâncias semimétricas satisfazem apenas as três primeiras dessas propriedades, mas podem violar a desigualdade triangular. As medidas de Bray-Curtis e de Sørensen são semimétricas. Um terceiro tipo de medida de distância, não usada por ecólogos, e a não métrica, que viola a segunda propriedade das distâncias métricas e pode assumir valores negativos. As distâncias euclidiana, de Manhattan ou City Block, de Corda, e de Bray-Curtis são usadas principalmente para dados numéricos contínuos. As de Jaccard e Sørensen são usadas para medir o afastamento entre duas amostras que são descritas por dados de presença e ausência. Nas distâncias de Jaccard e Sørensen,  $a$  é o número de objetos, como por exemplo de espécie que ocorrem apenas em  $y_i$ ,  $b$  é o número de objetos que ocorrem apenas em  $y_j$  e  $c$  é o número de objetos que ocorrem em ambos, em  $y_i$  e em  $y_j$ . A distância de Mahalanobis se aplica apenas para grupos de amostras  $y_i$  e  $y_j$ , em que cada um contém  $m_i$  e  $m_j$  amostras respectivamente. Na equação da distância de Mahalanobis,  $d$  é o vetor de diferenças entre as médias das  $m$  amostras em cada grupo,  $V$  é a matriz conjunta de variância e covariância amostral dentro de grupos, calculada como demonstrado, e  $C_i$  é a matriz de variância e covariância amostral dada por  $E = \sum_{k=1}^g \sum_{l=1}^g (y_k - \bar{Y}_k)(y_l - \bar{Y}_l)^T$  para  $y_i$ . Para medir distâncias

entre amostras ou grupos de tratamentos experimentais, não apenas entre indivíduos. Pode-se aplicar a fórmula da distância euclidiana para medir essas distâncias entre médias de qualquer número arbitrário de grupos  $g$ , dada por:

$$d_{i,j} = \sqrt{\sum_{k=1}^g (\bar{Y}_{i,k} - \bar{Y}_{j,k})^2}. \text{ Onde, } \bar{Y}_{i,k} \text{ é a média da variável } i \text{ no grupo } k. \text{ O } Y_i, \text{ na equação } d_{i,j} = \sqrt{\sum_{k=1}^g (\bar{Y}_{i,k} - \bar{Y}_{j,k})^2}$$

, pode ser univariado ou multivariado. Mais uma vez, para que nenhuma variável domine o cálculo de distância, em geral se padroniza os dados através da equação  $Z = \frac{(\bar{Y}_i - \bar{Y}_j)}{s}$  antes de calcular as médias e as

distâncias entre grupos. Muitas vezes se pergunta por que se usa o índice de Jaccard como uma medida de distância ou dissimilaridade, isto se deve porque o índice de Jaccard (1901) foi desenvolvido para descrever quão similares duas comunidades são em termos de espécies compartilhadas  $s_{i,j}$ , ou seja  $s_{i,j} = \frac{c}{a+b+c}$ , onde  $a$  é o número de espécies que ocorrem apenas na comunidade  $i$ ,  $b$  é o número de

espécies que ocorre apenas na comunidade  $j$ , e  $c$  é o número de espécies que elas têm em comum. Como as medidas de similaridade assumem seu valor máximo quando os objetos são mais similares, e as medidas de dissimilaridade ou distância assumem seu valor máximo quando os objetos são mais diferentes, qualquer medida de similaridade pode ser transformada em uma medida de dissimilaridade ou distância. Se uma medida de similaridade  $s$  varia de 0 a 1 como, por exemplo, o coeficiente de Jaccard, ela pode ser transformada em uma medida de distância  $d$  usando uma das três equações a seguir:  $d = 1 - s$ ,  $d = \sqrt{1 - s}$ , ou  $d = \sqrt{1 - s^2}$ . Deste modo, na tabela anterior, a distância de Jaccard é igual a 1 - o coeficiente de similaridade de Jaccard. A transformação reversa, como por exemplo,  $s = 1 - d$  pode ser usada para transformar medidas de distância em medidas de similaridade. Se a medida de distância não é delimitada, isto, é, varia de 0 a  $\infty$ , ela deve ser normalizada para variar entre 0 e 1, ou seja,

$$d_{norm} = \frac{d}{d_{max}} \text{ ou } d_{norm} = \frac{d - d_{min}}{d_{max} - d_{min}}.$$

Sendo assim na tabela a seguir estão descritas algumas medidas comuns de distância ou dissimilaridade usadas por ecólogos e pesquisadores do meio ambiente.

Nome da medida de distância ou dissimilaridade	Fórmula	Propriedade
<b>Euclidiana</b>	$d_{i,j} = \sqrt{\sum_{k=1}^g (\bar{Y}_{i,k} - \bar{Y}_{j,k})^2}$	<b>Métrica</b>
<b>Manhattan (ou City Block)</b>	$d_{i,j} = \sum_{k=1}^g  y_{i,k} - y_{j,k} $	<b>Métrica</b>
<b>Corda</b>	$d_{i,j} = \sqrt{2x \left( 1 - \frac{\sum_{k=1}^n y_{i,k} y_{j,k}}{\sqrt{\sum_{k=1}^n y_{i,k}^2 \sum_{k=1}^n y_{j,k}^2}} \right)}$	<b>Métrica</b>
<b>Mahalanobis</b>	$d_{y_i, y_j} = d_{i,j} V^{-1} d_{i,j}^T$ $V = \frac{1}{m_i + m_j - 2} [(m_i - 1)C_i + (m_j - 1)C_j]$	<b>Métrica</b>
<b>Chi-quadrado</b>	$d_{i,j} = \sum_{i=1}^m \sum_{j=1}^n y_j x \sqrt{\sum_{k=1}^n \left( \frac{1}{\sum_{k=1}^n y_k} x \left( \frac{y_k}{\sum_{k=1}^n y_k} - \frac{y_k}{\sum_{k=1}^n y_k} \right)^2 \right)}$	<b>Métrica</b>
<b>Bray-Curtis</b>	$d_{i,j} = \frac{\sum_{k=1}^n  y_k - y_j }{\sum_{k=1}^n (y_k + y_j)}$	<b>Semimétrica</b>
<b>Jaccard</b>	$d_{i,j} = \frac{a+b}{a+b+c}$	<b>Métrica</b>

<b>Sørensen</b>	$d_{i,j} = \frac{a+b}{a+b+2c}$	<b>Semimétrica</b>
-----------------	--------------------------------	--------------------

**MÉDIA ARITMÉTICA:** A média aritmética, ou simplesmente média de um conjunto de n observações,  $x_1, x_2, \dots, x_n$  é definida como:

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

. O somatório ( $\Sigma$ ) corresponde à soma de todos os valores obtidos.

**MÉTODOS USADOS NA DETERMINAÇÃO DO NÚMERO DE CLASSES (K) PARA CONSTRUÇÃO DE TABELAS DE DISTRIBUIÇÕES OS SÉRIES ESTATÍSTICAS DE FREQUÊNCIA:** A primeira pergunta que deve ser respondida quando se vai iniciar a construção de uma distribuição de frequências é a de qual deve ser o número de classes (K) ideal que são formadas para sumariar os dados de sua pesquisa ou levantamento amostral. Existem vários critérios que podem ser utilizados para orientar o pesquisador a ter uma idéia do melhor número de classes, porém tais critérios servirão apenas como indicação e nunca como regra fixa, pois caberá sempre ao pesquisador estabelecer o melhor número tendo sempre em mente que a escolha dependerá antes da natureza dos dados e da unidade de medida em que eles forem expressos do que de regras muitas vezes arbitrárias e pouco flexíveis. De qualquer forma apresentaremos alguns dos principais métodos utilizados para se determinar o melhor número de classes (K). Na literatura especializada os autores recomendam utilizar de 5 a 20 classes, outros recomendam escolher um número entre 4 e 20, dependendo do número de observações.

i) Critério dos valores mínimo e máximo baseado no número de observações;

**Tabela:** Números mínimo e máximo de classes baseado no número de observações.

Número de dados observados (n)	Número de classes (K)	
	Mínimo	Máximo
Até 50	5	10
51 a 100	8	16
101 a 200	10	20
201 a 300	12	24
301 a 500	15	30
Mais de 500	20	40

ii) Método do Modo Prático

ii.1.  $K = 5$  para  $n \leq 25$ ;

ii.2.  $K = \sqrt{n}$  para  $n > 25$

iii) Fórmula de Yule

$$K = 2,5\sqrt[4]{n}$$

iv) **Fórmula de Sturges ou Regra do Logaritmo:** o número de classes de uma série estatística é dada pela equação:  $K = 1 + 3,3 \cdot \text{Log } n$ . Essa fórmula apresenta dois inconvenientes que são o de propor um número demasiado de classes para um número pequeno de observações e relativamente poucas classes, quando o total de observações for grande. Pode-se construir uma tabela do número de classes em função do número de dados, como pode ser observado abaixo.

**Tabela:** Número de classes obtidas através do método de Sturges.

Número de observações (n)	Número de classes (K)
1	1
2	2
3 a 5	3
6 a 11	4
12 a 23	5
24 a 46	6
47 a 93	7
94 a 187	8
188 a 376	9
377 a 756	10
757 a 1519	11
1520 a 3053	12

v) **Tabela de Kelley**

Segundo Toledo e Ovalle (1987), Truman L. Kelley, em the grouping data for graphic portrayal, sugere os seguintes números de classes (K), com base no número total de observações (n), para efeito de representação gráfica.

**Tabela:** Número de classes obtidas através do método de Kelley em função do número de observações.

n	5	10	25	50	100	200	500	1000
K	2	4	6	8	10	12	15	18

vi) **Determinação gráfica**

Conforme Costa Neto (2002), a determinação do número de classes pode ser feita utilizando um gráfico ou diagrama onde o número de leituras os de dados a serem agrupados n fica no eixo das abscissas e o número de classes k fica no eixo das ordenadas então uma curva é traçada e nela identifica-se o melhor número de classes k a ser usado na construção da tabela ou série de frequência, isto é o número de classes K estará sempre em função do número de observações n coletadas na pesquisa.

vii) **Regra da potência de 2:**

Uma outra regra que pode ser utilizada para se ter uma idéia aproximada do número de classes K é a chamada regra da potência de 2, dada por: Menor valor inteiro de K, tal que  $2^K \geq n$ , onde K é o número de classes e n é o número de dados ou observações. Pode-se construir uma tabela do número de classes em função do número de dados, como pode ser observado abaixo.

**Tabela:** Número de classes obtidos através do método da potência de 2.

Número de observações (n)	Número de classes (K)
1 a 2	1
3 a 4	2
5 a 8	3
8 a 16	4
17 a 32	5
33 a 64	6
65 a 128	7
129 a 256	8
257 a 512	9
513 a 1024	10
1025 a 2048	11
2049 a 4096	12

**viii) Primeiro critério empírico**

Um critério empírico usado para a determinação do melhor número de classes (K) é o de considerar um número de classes entre 5 e 20, em função do conhecimento do pesquisador sobre os dados de sua investigação.

**ix) Segundo critério empírico**

Esse outro critério empírico baseia-se numa função do tamanho amostral. Esse critério está apresentado na seguinte tabela.

**Tabela:** Número de classes obtidos em função do tamanho da amostra.

Tamanho da amostra (n)	Número de classes (K)
Até 100	$\sqrt{n}$ (inteiro mais próximo)
Acima de 100	$5 \text{Log}_{10} n$ (inteiro mais próximo)

**x) Critério proposto por Scott (1979)**

Esse critério é indicado para dados provenientes de uma amostragem de uma distribuição de probabilidade denominada distribuição normal de probabilidade, a qual é vista com mais detalhes quando estudamos o capítulo das distribuições teóricas ou especiais de probabilidades. Embora esse critério tenha sido proposto e desenvolvido para essa condição, ele tem um bom desempenho em situações de distribuições não simétricas ou naquelas distribuições que tenham um maior ou menor grau de achatamento ou curtose que a distribuição normal. O critério de Scott é dado pela seguinte expressão:

$$K = 1 + \frac{A\sqrt[3]{n}}{3,49.S}$$

Em que A é amplitude, n o tamanho da amostra e S o desvio padrão. As estatísticas A e S são definidas nas equações dadas a seguir:

$$A = X_{(n)} - X_{(1)}, \text{ e } S = \sqrt{\frac{1}{n-1} \left[ \sum_{i=1}^n X_i^2 - \frac{\left( \sum_{i=1}^n X_i \right)^2}{n} \right]}$$

A representação  $X_{(i)}$ , cujo índice se encontra dentro dos parênteses, refere-se ao valor da variável X cuja ordem na amostra ordenada de forma crescente é igual a i. Esse tipo de representação aparece na equação acima da amplitude total. Essas estatísticas são chamadas de estatísticas de ordem. Assim,  $X_{(1)}$  é a estatística do menor valor e  $X_{(n)}$  é a do maior valor. O valor de K deve ser tomado como o inteiro mais próximo ao valor encontrado usando a expressão de K dada anteriormente.

**xi) Critério proposto por Doane (1976)**

Esse critério é dado pela seguinte equação:

$$K = 1 + \text{Log}_2 \left[ 1 + \frac{|\sqrt{b_1}|}{S_{(\sqrt{b_1})}} \right] + \text{Log}_2(n), \text{ em que, } S_{(\sqrt{b_1})} \text{ é o erro padrão do coeficiente de assimetria } (\sqrt{b_1}), \text{ os}$$

quais são dados pelas seguintes equações:

$$\sqrt{b_1} = \frac{\sum_{i=1}^n (X_i - \bar{X})^3}{\left[ \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} \right]^{3/2}}, \text{ e, } S_{(\sqrt{b_1})} = \sqrt{\frac{6(n-2)}{(n+1)(n+3)}}$$

xii)  $K = \sqrt{n} + 1$ , onde n é o tamanho da amostra. Esse critério é proposto por Francisco (1995). O autor argumenta que geralmente o número de classes varia de 8 a 12 classes. Entretanto, existe uma fórmula empírica para se determinar o melhor número de classes K que é dada pela equação acima.

xiii) Critério de Oliveira ou a Regra do Quadrado. O número de classes é igual a raiz do número de observações, ou seja,  $K = \sqrt{n}$ . Deve-se utilizar o número quadrado perfeito mais próximo.

**MATERIAL EXPERIMENTAL PARA USO EM ENSAIOS AGRÍCOLAS:** Todo pesquisador da área agrícola deve atentar para os seguintes itens a seguir no planejamento dos experimentos a serem conduzidos por ele, os quais são: i) escolha da cultura e da época de cultivo adequada; ii) escolha do local do experimento que deve ser típico das condições de campo na região investigada e uniforme em sua topografia, fertilidade do solo, profundidade, estrutura e qualidade do subsolo. ii.1) o histórico do local: verificar a sequência de cultivos, aplicação de adubos, tratamentos, culturas etc. nos períodos anteriores; ii.2) condições fitossanitárias – implementar um controle rigoroso particularmente em relação a doenças e organismos localizados no solo, como podridão da raiz, nematóides e a murcha do Guandu provocada por Fusarium que aparece entre dois e três anos. Uma área recém desmatada não pode ser usada para experimentação de campo sem um cultivo prévio de uniformização de pelo menos um ano. Uma regra geral é escolher o local durante o ano anterior ao experimento. No preparo da área experimental o pesquisador deve ter atenção no preparo do solo que deve ser efetuado segundo as práticas comuns de cada cultura, mas deve



ser feito com bastante cuidado. As operações devem ser preferencialmente mecanizadas. Efetuar a nivelção do terreno, pois reduz o perigo de empoçamento de água da chuva e melhora a superfície de semeadura assegurando, portanto, um stand ou número de plantas uniforme. O plantio de um experimento deverá ser preferencialmente feito, em um dia no caso de delineamento inteiramente casualizado ou 1 repetição no caso do delineamento em blocos casualizados, evitar plantar partes de uma repetição em diferentes dias. Todas as operações que antecedem a semeadura devem ser efetuadas transversalmente às parcelas, ou seja, paralelamente ao eixo prolongado dos blocos. Isso assegurará um mínimo de variação entre as parcelas dentro do bloco.

$m_X(t)$  : Função geradora de momentos calculada em  $t$ .

$\mu'_r$  : Momento simples de ordem  $r$ .

$\mu_r$  : Momento centrado de ordem  $r$ .

$\mu_X$  : Valor esperado da variável aleatória  $X$ .



**n:** símbolo utilizado para representar o tamanho de um conjunto de dados ou de uma amostra.

**n:** Número total de observações.

**n:** Em estatística é o número de graus de liberdade (GL). Em inglês degrees of freedom.

**n:** Símbolo usado para tamanho da amostra.

**N:** Símbolo utilizado para representar o tamanho de um conjunto de dados, geralmente uma de uma população ou universo estatístico.

**N:** É o símbolo usado para tamanho da população ou universo estatístico.

**IN:** Conjunto dos números naturais.

**NADO-MORTO:** O mesmo que nascido morto, é a forma usada em Portugal. Conhecido também como natimorto.

**NÃO-ALEATÓRIO:** Qualquer método que não esteja em conformidade com a definição estatística de acaso; termo usado pelos estatísticos para enfatizar a natureza de um processo fortuito ou sistemático. Veja também não-casual.

**NÃO-CASUAL:** O mesmo que não-aleatório.

**NÃO CONTÉM:** Símbolo matemático dado por  $\not\subset$ .

**NÃO-LINEAR:** Que não segue as propriedades de proporcionalidade. Ver linear.

**NÃO ERGÓDICO:** Nome dado aos procedimentos nos quais a média, a esperança matemática ou o valor esperado, da distribuição de probabilidade de uma variável aleatória não pode ser estimado a partir de resultados obtidos anteriormente.

**NÃO PERTENCE A:** Símbolo matemático dado por  $\notin$ .

**NÃO RECURSIVA:** Relação em um diagrama de caminhos indicando uma relação mútua ou recíproca entre dois construtos. Cada construto no par tem uma relação causal com o outro construto. É descrita por uma seta de dois sentidos entre ambos os construtos.

**NÃO-SUPERVISIONADO:** Usa-se em análise de cluster quando o analista não impõe qualquer estrutura à classificação, deixando simplesmente ela emergir dos dados. Somente depois é que investigamos a classificação para verificar se ela fecha com outros critérios de agrupamento como por exemplo, sexo ou espécie.

**NÃO-VIESADO:** Isento de viés ou vício. Também se diz não-viciado e imparcial.

**NASCIDO VIVO:** Produto da concepção que, após ser completamente removido do corpo da mãe, apresenta qualquer sinal de vida.

**NATALIDADE, ESTATÍSTICAS DE:** Referente ao número de nascimentos.

**NATIMORTO:** O mesmo que óbito fetal tardio.

**NATIMORTO:** O mesmo que nado-morto.

**NATIMORTO:** É o nascimento de um feto que pesa mais de 500 g e que não tem evidências de vida depois de nascer. Para fins de cálculos padronizados dos coeficientes de mortalidade perinatal para comparação internacional, serão incluídos somente aqueles fetos mortos que pesam 1.000 g ou mais ao nascer.

**NASCIDO MORTO:** Ver natimorto.

**NATUREZA SEQUENCIAL DO TEOREMA BAYES:** O teorema de Bayes pode ser aplicado de forma seqüencial conforme a informação coletada dos dados. O resultado final será o mesmo que o obtido com todos os dados de uma única vez. Do inglês Bayes sequential nature.

**NEGATIVAMENTE CORRELACIONADAS:** Refere-se a duas grandezas relacionadas de modo que uma tende a aumentar quando a outra tende a diminuir.

**NELÓIDE:** Família de sólidos geométricos de base circular cujo diâmetro da secção transversal aumenta segundo o inverso de uma parábola à medida que se desloca do topo para a base do sólido. O fator de forma absoluto do nelóide é sempre menor que um terço, no caso do nelóide ordinário o fator é exatamente um quarto.

**NEUTRALIDADE ÉTICA:** Refere-se às conseqüências das descobertas científicas. A busca de conhecimento novo deve prosseguir independentemente da sensibilidade do assunto. É claro que em certas áreas há conflito entre interpretações científicas, religiosas, políticas e ideológicas.

**NEWMAN-KEULS:** Procedimento de comparação entre médias de vários tratamentos, no sentido de verificar a significância estatística das diferenças entre essas medidas de tendência central. A comparação é efetuada após a análise da variância e somente se o valor de F for significativo.

**NEWTON, FÓRMULA DE:** Fórmula de cubagem cuja aproximação do volume de uma tora resulta do produto do comprimento da tora por uma média ponderada das áreas transversais em três pontos da tora: nas extremidades e no meio. A média ponderada é construída dando-se peso 1 para as áreas transversais das extremidades e peso 4 para área transversal do meio da tora. A fórmula produz resultados exatos para os seguintes sólidos: cilindro, parabolóide ordinário como parabolóide quadrático, cone e nelóide ordinário. Essa fórmula também é atribuída a Cavalieri. (Bonaventura Francesco Cavaliere, 1598-1647).

**N:** Símbolo representativo de número de observações. Pode representar o número de elementos usados em uma pesquisa, o número de indivíduos numa população ou mesmo o número de variáveis utilizadas num estudo. A letra n geralmente indica o número de indivíduos numa amostra, ou o tamanho desta.

**NÍVEL:** Valor específico que descreve um fator. Cada fator deve ser representado por dois ou mais níveis, mas o número de níveis normalmente jamais excede quatro ou cinco. Se o fator é métrico, ele deve ser reduzido a um pequeno número de níveis. Por exemplo, os inúmeros valores possíveis de tamanho e preço podem ser representados por um pequeno número de níveis: tamanho (15, 18 ou 20 quilogramas); ou preço (R\$ 3,20, R\$ 4,35 ou R\$ 6,09). Se a variável é não-métrica, os valores originais

podem ser usados como nesses exemplos: cor (amarelo ou verde); marca (A, B ou C); ou aditivo amaciante de fábrica (com ou sem).

**NÍVEL ALFA:** É a probabilidade máxima de cometer um erro falso-positivo que o investigador está disposto a aceitar.

**NÍVEL DE AMOSTRAGEM:** Expressão usada em controle de qualidade para indicar a fração dos artigos, que se estão produzindo continuamente, que deve ser inspecionada. Exemplo: de cada 80 peças examina-se sete unidades.

**NÍVEL DE CONFIANÇA:** É a probabilidade de que um intervalo de confiança contenha o valor do parâmetro que ele se propõe a estimar.

**NÍVEL:** São as diferentes manifestações de um fator, por exemplo; as doses de adubação empregadas, os espaçamentos utilizados, as cultivares ou linhagens que se testam, diferentes temperaturas de cocção, marcas de pneus, etc.

**NÍVEL DE CONFIANÇA:** Está associado a uma afirmação baseada em dados disponíveis e mede a probabilidade de que a afirmativa seja verdadeira.

**NÍVEL DE CONFIANÇA (ESTATÍSTICA):** Uma probabilidade associada com o teste de uma hipótese, através do uso de técnicas estatísticas que determina se deve-se ou não rejeitar a hipótese. Comumente, os níveis de confiança usados em pesquisas físicas, sociais e biológicas, por exemplo, é de 0,95 e 0,99.

**NÍVEL DE CONFIANÇA:** É a probabilidade de que o intervalo de confiança contenha o verdadeiro valor do parâmetro.

**NÍVEL DE SIGNIFICÂNCIA (ALFA):** É definida como a probabilidade de cometer o erro de tipo I, ou seja, rejeitar a hipótese nula ( $H_0$ ), quando ela é verdadeira. Por exemplo:  $H_0$ : As atitudes em relação à Matemática não interferem na formação das atitudes em relação à Estatística ( $r_{xy} = 0$ ).

**NÍVEL DE CONFIANÇA:** A probabilidade de que determinado intervalo de confiança inclua o real valor da população.

**NÍVEL DE CONFIANÇA:** Veja grau de confiança.

**NÍVEL DE CONFIANÇA:** Grau de confiança associado à precisão das medições extraídas dos dados da amostra.

**NÍVEL DE CONFIANÇA:** Grau de confiança associado a um intervalo de confiança a probabilidade de o intervalo de confiança conter o verdadeiro valor do parâmetro.

**NÍVEL DE CONFIANÇA:** A probabilidade estimada de que um parâmetro de população caia dentro de um determinado intervalo de confiança. Podemos ter, assim, 95 % de confiança de que entre 35 % e 45 % dos eleitores são a favor do candidato.

**NÍVEL DE CONFIANÇA (1- $\alpha$ ):** É a probabilidade de que um intervalo de confiança contenha o valor do parâmetro que ele se propõe a estimar. Em inglês confidence level.

**NÍVEL DE INCIDÊNCIA:** Expressão de ordem geral se refere à incidência realmente observada da doença. Nos estudos epidemiológicos, são especificados, de uma forma genética, dois níveis de incidência: o nível endêmico e o nível epidêmico.

**NÍVEL DE LINGUAGEM:** Uma diretriz para a elaboração de perguntas que instrui o pesquisador a estar consciente da população a ser pesquisada quando escolher as palavras, as expressões e o jargão a ser usado nas perguntas.

**NÍVEL DE MENSURAÇÃO:** As variáveis diferem em quão bem elas podem ser medidas, isto é, em quanta informação seu nível de mensuração pode prover. Há obviamente algum erro em cada medida, o que determina o montante de informação que se pode obter, mas basicamente o fator que determina a quantidade de informação que uma variável pode prover é o seu tipo de nível de mensuração. Sob este prisma as variáveis são classificadas como nominais ordinais e intervalares.

**NÍVEL DE PREÇOS:** Nível médio de todos os preços no país; pode ser medido pelo deflator PIB, pelo índice de preços ao consumidor, ou pelo índice de preços ao produtor.

**NÍVEL DE SIGNIFICÂNCIA ( $\alpha$ ):** É a probabilidade de se cometer erro do tipo I no teste de hipóteses, isto é, a probabilidade de se rejeitar a hipótese nula quando ela é verdadeira. Em inglês alpha level.

**NÍVEL DE SIGNIFICÂNCIA:** De um processo de teste de hipótese, probabilidade de cometer um erro tipo I.

**NÍVEL DE SIGNIFICÂNCIA:** Um termo associado com testes de hipóteses. Também designado alfa, ou nível  $\alpha$ , o nível de significância é um critério usado para se tomar decisões sobre uma hipótese. Em ciências sociais e biológicas por exemplo, os níveis de significância mais comuns são 0,05 e 0,01, apesar de hoje em dia com o advento da computação se utilizar como probabilidade de risco do erro tipo I, o valor p, ou p value.

**NÍVEL DE SIGNIFICÂNCIA ( $\alpha$ ):** Valor probabilísticas do limite de rejeição da hipóteses de nulidade, do erro alfa ou do tipo I. Em geral adota-se o valor alfa de 0,05 ou 0,01, admitindo-se, ao rejeitar a hipótese de nulidade, a probabilidade de ocorrência de 1 erro em 20 equivale a 5 % e 1 em 100 corresponde a 1 %.

**NÍVEL DE SIGNIFICÂNCIA:** Em teste de hipótese, é a probabilidade de rejeitar uma hipótese verdadeira. Também chamado de probabilidade de erro tipo I. Em inglês Level of significance.

**NÍVEL DE SIGNIFICÂNCIA ( $\alpha$ ):** No contexto de teste de significação estatística, é o grau de probabilidade de que uma relação empírica, observada, possa ser atribuída a erro de amostragem. Uma relação é significativa no nível 0,05 se sua probabilidade de ser apenas função de erro de amostragem não for maior que 5 em 100.

**NÍVEL DE SIGNIFICÂNCIA ( $\alpha$ ):** É definida como a probabilidade de cometer o erro de tipo I, ou seja, rejeitar a hipótese nula ( $H_0$ ), quando ela é verdadeira. Por exemplo:  $H_0$ : As atitudes em relação à Matemática não interferem na formação das atitudes em relação à Estatística ( $r_{xy} = 0$ ) e  $H_1$ : As atitudes em relação à Matemática interferem na formação das atitudes em relação à Estatística. ( $r_{xy} \neq 0$ ).

**NÍVEL DE SIGNIFICÂNCIA (PROBABILIDADE DE ERRO TIPO I):** Ver  $\alpha$ .

**NÍVEL DE SIGNIFICÂNCIA:** É denotado por  $\alpha$  e indica a probabilidade de cometer um erro tipo I. Na maioria dos softwares da área de métodos quantitativos e áreas afins, a significância estatística é expressa pelo nível descritivo, valor p ou p value. Os níveis de significância mais utilizados são 5 %, 0,1 %, 1 % e 10 %.

**NÍVEL DE SIGNIFICÂNCIA OU ERRO TIPO I:** É a probabilidade de risco ou erro fixada a priori pelo pesquisador, que ele assume de rejeitar uma hipótese de nulidade ( $H_0$ ) quando na realidade ela for verdadeira. Geralmente para experimentos de campo adota-se nível de  $\alpha = 5\%$   $\alpha = 1\%$  ou  $\alpha = 10\%$  em ensaios de laboratório usa-se geralmente  $\alpha = 1\%$  ou  $\alpha = 0,1\%$ .

**NÍVEL DE CONFIANÇA:** É a probabilidade de não ocorrer risco, ou seja, do pesquisador aceitar uma hipótese de nulidade quando na realidade esta for verdadeira é e dado por  $1-\alpha$ .

**NÍVEL SOCIOECONÔMICO:** Medida utilizada para se determinar o perfil social de uma pessoa, que pode incluir diversos índices demográficos, tais como salário, grau de instrução, profissão, evolução patrimonial, etc.

**NÍVEL ENDÊMICO DE INCIDÊNCIA:** É uma qualidade de ordem genérica, atribuída às medidas de incidência cujos valores se situem abaixo do limite superior da faixa endêmica, qualquer que seja o patamar desta.

**NÍVEIS DE UM FATOR:** As condições usadas para um fator em um experimento.

**NÍVEL DE CONFIANÇA:** Um outro termo para coeficiente de confiança.

**NÍVEL DE SIGNIFICÂNCIA:** Se a estatística de teste Z para uma hipótese e a distribuição de Z, quando a hipótese é verdadeira, são conhecidas, então podemos encontrar as probabilidades  $P(Z \leq z_l)$  e  $P(Z \geq z_u)$ . A rejeição da hipótese é geralmente expressa em termos do valor observado de Z caindo fora do intervalo de  $z_l$  a  $z_u$ . As probabilidades  $P(Z \leq z_l)$  e  $P(Z \geq z_u)$  são geralmente escolhidas tendo valores pequenos com 0,01; 0,025; 0,05 e 0,1, e são chamadas de níveis de significância. Os níveis reais escolhidos são, de algum modo, arbitrários e frequentemente expressos em porcentagens, tal como nível de 5 % de significância.

**NÍVEL EPIDÊMICO DE INCIDÊNCIA:** Designação de incidência que apresenta medidas correspondentes, ocorrendo na região de valores epidêmicos, isto é, acima do limite superior endêmico ou limiar epidêmico, no gráfico do diagrama de controle.

**NÍVEL DE CONFIANÇA:** Probabilidade associada a determinado teste de hipótese. Todas as afirmações feitas em pesquisa devem ser consideradas a partir de determinado nível de confiança, ou seja, com certo grau de certeza que, nas ciências sociais, costuma ser de 95 %, sendo que em alguns campos específicos podem ser utilizados valores mais altos, como por exemplo 99 % em certas pesquisas das ciências físicas, biológicas e sociais. Por exemplo, numa pesquisa pré-eleitoral, poderia-se dizer que determinado candidato possui 70 % das intenções de voto. Mas esta afirmação deve ser qualificada com certo nível de confiança, por exemplo, de 95 %. Pode-se afirmar que a margem de erro é de 2 % para cima ou para baixo que faz parte do intervalo de confiança. Isto significa que pode-se afirmar com 95 % de certeza que, se as eleições fossem hoje, o candidato teria entre 68 % e 72 % dos votos.

**NÍVEL DE MENSURAÇÃO:** Quando fazemos uma pesquisa, muitas vezes necessitamos recorrer à mensuração que é o uso de medidas numéricas das variáveis para representar a realidade. Assim, por exemplo, podemos medir o peso, sendo a variável de pesquisa, das pessoas em quilogramas que é a



unidade de medida, ou a altura, outra variável de pesquisa, dos indivíduos em centímetros que é outra unidade de medida.

**NÍVEL DE SIGNIFICÂNCIA:** Também conhecido como alfa, nível  $\alpha$  ou valor  $p$ , refere-se à probabilidade de erro em determinada afirmação científica ou teste de hipóteses. Valor complementar ao nível de confiança. Quanto menor seu valor, maior o nível de confiança de determinada afirmação. Os valores situados entre 0,05 e 0,01 ou 5 % a 1 % respectivamente, representam as margens de erro tradicionalmente aceitas nas ciências sociais e naturais, respectivamente.

**NÚMEROS ALEATÓRIOS:** Conjunto de números, normalmente gerados por computador, na qual não existe relação conhecida entre os seus elementos. Uma tabela de números aleatórios serve para selecionar, por exemplo, os sujeitos em uma pesquisa que se utilize de procedimentos de amostragem probabilística.

**NÃO-TENDENCIOSA:** Livre de tendência ou de erro sistemático.

**NÍVEIS:** Uma categoria particular de uma variável ou fator qualitativo; por exemplo, dose de um medicamento, tratamentos diferentes.

**NÍVEL DE SIGNIFICÂNCIA:** Um valor de corte para o valor de  $p$ , de modo que a hipótese nula seja rejeitada caso o valor de  $p$  de um teste de hipótese seja inferior a esse valor de corte; o valor de corte é definido antes de o teste ser realizado, sendo frequentemente definido como 0,05. Logo, caso  $p < 0,05$ , dizemos que o teste é significativo ao nível de 5 %. O nível de significância é a chance máxima de cometer um erro Tipo I.

**NOMOGRAMA DE ALTMAN:** Uma representação gráfica mostrando a relação entre o tamanho, poder e diferença padronizada da amostra; pode ser utilizado para determinar qualquer um deles, com o conhecimento dos outros dois. A diferença padronizada é um reflexo da variabilidade das observações e da magnitude mínima do efeito de tratamento considerado importante.

**NOMOGRAMA DE FAGAN:** Um diagrama utilizado em testes diagnósticos para transformar a probabilidade pré-teste em probabilidade pós-teste, pela verossimilhança.

**NÚMERO DE ANIMAIS NECESSÁRIOS PARA TRATAR (NNT):** O número de animais que o clínico precisa para tratar com um novo medicamento, ao invés de um tratamento controle, para evitar a ocorrência de uma reação indesejável. Uma maneira simples de mostrar o benefício de um novo medicamento.

**NNH:** Número necessário para causar (harm) um evento desfavorável num determinado período de tempo (t).

**NNT:** Número necessário para tratar e prevenir um evento desfavorável num determinado período de tempo (t).

**NOCEBO:** Reação colateral do placebo.

**NORMA:** É o mesmo que moda. Em inglês norma.

**NORMAS:** Conjunto de informações que permitem uma interpretação adequada dos escores de um teste, pela apresentação dos resultados obtidos nesse teste por um específico grupo de pessoas, que serve como referência. As normas são frequentemente apresentadas em tabelas que fornecem os escores típicos de uma série de grupos homogêneos, como os estudantes de certo nível ou de determinada idade. Em inglês norms.

**NORMAS DE CONHECIMENTO CIENTÍFICO:** O conhecimento científico poder ser definido de forma útil como conhecimento coletado e organizado de acordo com quatro normas: i) validade empírica; ii) clareza lógica; iii) consistência lógica de proposições e iv) generalidade de princípios.

**NORMAS:** Conjunto de informações que permitem uma interpretação adequada dos valores de um teste pela apresentação dos resultados obtidos nesse teste por um específico grupo de pessoas, que serve como referência. As normas são frequentemente apresentadas em tabelas que fornecem os escores típicos de uma série de grupos homogêneos, como os estudantes de certo nível ou determinada idade.

**NÃO MÉTRICA:** Um tipo de medida de distância ou dissimilaridade que satisfaz duas propriedades; i) A distância mínima é igual a 0 e a entre objetos idênticos é igual a 0; ii) as distâncias são simétricas, ou seja, a distância do objeto a para o b é a mesma do objeto b para o a. Comparar com métrica e semimétrica. Em inglês Non-metric.

**NÍVEL:** O valor individual de um dado tratamento ou fator experimental em uma análise de variância ou em um delineamento tabular. Em inglês Level.

**NMDS:** Ver escalonamento multidimensional não métrico.

**NOMINAL:** Nível de mensuração de dados; caracteriza dados que consistem somente em nomes, rótulos ou categorias.

**NORMAL:** Uma variável quantitativa segue uma distribuição normal, se sua distribuição de frequências tem o formato similar ao de um sino, ou seja a maioria dos valores se concentram em torno da média e, a medida que se afasta do centro as observações são cada vez mais raras. Essa distribuição é simétrica. Muitas variáveis têm essa distribuição, tais como altura das pessoas adultas do sexo masculino, coeficiente de inteligência, dentre outros. Para examinar visualmente, você pode fazer o histograma com a curva ajustada, o SPSS faz isso. O teste que checa a normalidade de uma variável é o teste de Lilliefors, que se encontra no comando EXPLORE do SPSS. A hipótese nula é que a variável segue uma distribuição normal, por isso você espera aceitar  $H_0$ , e espera que o p-valor seja maior do que o nível de significância especificado pelo pesquisador.

**NORMAL:** Diversos significativos: o que segue uma norma; características biológica habitual ou prevalente, em geral, cobre 95 % da população; aquilo que é desejável biológica ou culturalmente; o que está associado a baixo risco à saúde, no futuro; variação dos resultados de um exame diagnóstico, além do qual a doença específica está presente ou a terapia traz mais benefícios do que malefícios.

**NORMAL:** Uma variável quantitativa contínua que é descrita pelo modelo de uma distribuição teórica, segue uma distribuição normal, se sua distribuição de frequências tem o formato similar ao de um sino, campânula, chapéu de Napoleão, ou seja, a maioria dos valores se concentram em torno da média e, a medida que se afasta do centro as observações são cada vez mais raras. Essa distribuição é simétrica e mesocúrtica. Muitas variáveis têm essa distribuição, tais como altura das pessoas adultas do sexo masculino, coeficiente de inteligência, peso, comprimento, taxas de colesterol, dentre outros. Para examinar visualmente, você pode fazer o histograma com a curva ajustada, o SPSS faz isso. O teste que checa a normalidade de uma variável é o teste de Lilliefors, que se encontra no comando EXPLORE do SPSS. A hipótese nula é que a variável segue uma distribuição normal, por isso você espera aceitar  $H_0$ , e espera que o p-valor seja maior do que o nível de significância especificado por você. O pressuposto de normalidade é chave para toda a estatística paramétrica, por essa razão você sempre deve checar a validade do

mesmo. Contudo, quando sua amostra for suficientemente grande ( $n > 30$ ), dependendo do formato da distribuição, o Teorema Central do Limite garante a convergência da média amostral para a normalidade (MORETIN; BUSSAB, 2003). O pressuposto de independência é chave para a maioria das estatísticas. Isto significa que o resultado de uma observação não interfere no resultado de outra observação. Por exemplo, a nota de um aluno não interfere na nota de um outro aluno. Já esse pressuposto é quebrado para amostras relacionadas como, por exemplo, passar uma prova antes de uma intervenção e a mesma prova (ou outra) depois da intervenção; essas duas medidas são correlacionadas, uma vez que os sujeitos são os mesmos. Neste caso, deve-se utilizar o teste para dados emparelhadas, ou ANOVA com medidas repetidas. Logo, cuidado com os estudos longitudinais, em que se acompanha os mesmos sujeitos em vários momentos. Homocedasticidade ou igualdade de variâncias: Este pressuposto exige que o nível de dispersão da variável dentro dos grupos seja similar. O programa estatístico SPSS automaticamente testa esta hipótese:  $H_0: s_1 = s_2 = s_3 = s_4$ , por meio de um teste de Levene.

**NORMAL:** Que segue ou se comporta de acordo com o modelo normal, curva de Gauss. Em inglês normal.

**NORMAL:** Uma variável quantitativa segue uma distribuição normal, se sua distribuição de frequências tem o formato similar ao de um sino, ou seja, a maioria dos valores se concentram em torno da média e, a medida que se afasta do centro as observações são cada vez mais raras. Essa distribuição é simétrica. Muitas variáveis têm essa distribuição, tais como altura das pessoas adultas do sexo masculino, coeficiente de inteligência, dentre outros. Para examinar visualmente, o investigador pode fazer o histograma com a curva ajustada, o SPSS faz isso. O teste que checa a normalidade de uma variável é o teste de Lilliefors, que se encontra no comando explore do SPSS. A hipótese nula é que a variável segue uma distribuição normal, por isso o pesquisador espera aceitar  $H_0$  e espera que o p-valor seja maior do que o nível de significância especificado por você. Em inglês normal.

**NORMALIDADE:** Grau em que a distribuição dos dados da amostra corresponde a uma distribuição normal.

**NORMALIDADE:** Termo usado em estatística que quer dizer que há normalidade ou que os dados são normalmente distribuídos e assim significa que eles seguem uma distribuição normal, isto é, valores concentrados simetricamente em torno da média e quanto maior a distância da média, menor a frequência das observações.

**NORMALIDADE ASSINTÓTICA** : Uma distribuição dependente de um parâmetro  $n$ , geralmente um tamanho de amostra, é definida como assintoticamente normal se, quando  $n$  tende ao infinito e a distribuição tende para a forma Normal. Em inglês Asymptotic Normality.

**NORMALIZAÇÃO:** É o processo de transformação de valores originais, pelo qual o valor normalizado é igual ao valor original subtraído do valor mínimo e dividido pelo valor máximo menos o valor mínimo:

$$X_n = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

**NOTAÇÃO CIENTÍFICA:** Notação que serve para expressar números muito grandes ou muito pequenos. O segredo é multiplicar um número pequeno por uma potência de 10. A forma de uma Notação científica é:  $m \cdot 10^e$ , onde  $m$  significa mantissa e  $e$  significa ordem de grandeza. A mantissa sempre será um valor em módulo entre 1 e 10. Para transformar um número como por exemplo um número grande qualquer

em notação científica, devemos deslocar a vírgula para a esquerda até o primeiro algarismo da seguinte forma: 200 000 000 000 para 2,00 000 000 000 note que a vírgula avançou 11 casas para a esquerda, então em notação científica este número fica:  $2 \cdot 10^{11}$ . Para transformar valores muito pequenos, é só mover a vírgula para a direita, e a cada casa avançada, diminuir 1 da ordem de grandeza: Por exemplo 0,0000000586 faz-se a transformação movendo-se a vírgula para direita então temos 5,86 (avanço de 8 casas) e assim  $5,86 \cdot 10^{-8}$ . No caso de -12.000.000.000.000 temos  $-1,2 \cdot 10^{13}$ .

**N.S.:** Não-significativo, do ponto de vista estatística. Ver p, valor p ou p value.

**NÚMERO ALEATÓRIO:** Número gerado num computador por meio de um algoritmo recursivo. Na realidade o correto seria dizer número pseudo aleatório uma vez que as sequências geradas são reproduzíveis. Em inglês **random number**.

**NÚMERO CASUAL OU ALEATÓRIO:** Número gerado por um processo aleatório definido.

**NÚMERO DE BERNOULLI:** São números que foram definidos por Jacob jacques Bernoulli em conexão com a avaliação da soma  $S_{ik}$ . A sequência  $B_0, B_1, B_2, \dots$  pode ser gerada utilizando a fórmula:

$$\frac{x}{(e^x - 1)} = \frac{S(B_{nxx})}{n!}, \text{ embora várias notações diferentes sejam utilizadas. Os primeiros números são: } B_0 = 1, B_1 = -\frac{1}{2}, B_2 = \frac{1}{6}, B_4 = -\frac{1}{30}, B_6 = \frac{1}{42}, \dots$$

Eles ocorrem em diversas áreas da matemática, incluindo a expansão da série de  $\tan(x)$  e no último teorema de Fermat. Em inglês Bernoulli number.

**NÚMERO DE CLASSES:** É o número de linhas de uma tabela de frequências ou distribuição de frequências. Deve ser pré-estabelecido entre 4 a 12 ou determinada por uma fórmula matemática. Fórmula de Sturges:  $n(k) = 1 + 3,3 \log n$ .

**NÚMERO DE INFORMAÇÃO DE KULLBACK-LIEBLER:** O número de informação  $I_x(1:2)$  denota a informação média por observação para discriminar entre duas hipóteses  $H_1$  e  $H_2$  quando  $H_1$  é verdadeira:  $I_x(1:2) = -\int f_1(x) \log(f_1(x)/f_2(x)) \psi(x)$  onde  $\psi(x)$  é uma medida comum. Em inglês Kullback-Liebler information number.

**NÚMERO DE GRAUS DE LIBERDADE:** De uma soma de quadrados de desvios ou resíduos é o número pelo qual ela deve ser dividida para que o quociente que é chamado quadrado médio tenha, quando válida a hipótese de nulidade, esperança matemática igual à variância ( $\sigma^2$ ).

**NÚMERO ÍNDICE:** Um número índice é uma quantidade que mostra, pela sua variação, as mudanças no tempo ou espaço de uma magnitude que não é suscetível de medição direta ou observação direta na prática. Em inglês Index number.

**NÚMERO ÍNDICE:** É uma medida estatística idealizada para mostrar as variações de uma variável, ou de um grupo de variáveis, correlacionadas ao tempo, à localização geográfica, ou a outras características como rendimento, profissão, dentre outras. Uma coleção de números índices de diversos anos, localidades, dentre outros, é frequentemente denominada série de índices. No Brasil, os índices mais conhecidos são o Índice Geral de Preços, calculado pela Fundação Getúlio Vargas do Rio de Janeiro, e o Índice Nacional de Preços ao Consumidor (INPC).

**NÚMERO ÍNDICE:** São medidas estatísticas frequentemente usadas por administradores, economistas e engenheiros, para comparar grupos de variáveis relacionadas entre si e obter um quadro simples e resumido das mudanças significativas em áreas relacionadas como preços de matérias primas, preços de produtos acabados, volume físico de produção, dentre outros.

**NÚMERO MAIS PROVÁVEL (NMP):** De acordo com a teoria estatística, é o número que, com maior probabilidade que qualquer outro, fornece a estimativa do número de bactérias em uma amostra. Expresso com densidade de organismos por 100 ml.

**NÚMEROS EQUIPROVÁVEIS:** São números, dispostos em tábua, usados para a seleção acidental de amostras; admite-se que toda sequencia desses números, tais como aparecem tabulados, é acidental, isto é, a probabilidade de escolha de um é a mesma que a de qualquer outro. Também chamados de números aleatórios de Tippett ou de números de Tippett.

**NÚMEROS ÍNDICES:** É uma técnica ou metodologia estatística idealizada para comparar, quantitativamente, as variações de um fenômeno complexo no tempo ou em outras situações diversas. Os números índices não se constituem em medida alguma, mas são indicadores de comportamento ou de tendência de uma ou mais variáveis componentes de um fenômeno.

**NÚMERO NECESSÁRIO A TRATAR (NNT):** O mesmo que número necessário para beneficiar (NNB).

**NÚMERO NECESSÁRIO PARA BENEFICIAR (NNB):** Número de pacientes que precisam ser tratados, em média, para que um deles experimente os efeitos benéficos de um determinado tipo de tratamento, O NNB está baseado na redução do risco absoluto (RRA). Se, por exemplo, o risco de beneficiar é igual a 0,10 com o tratamento e igual a 0,04 sem o tratamento, o RRA = 0,06. O NNB é calculado como sendo  $\frac{1}{RRA}$ . Nesse exemplo ele seria  $\frac{1}{0,06} = 1,67$ . Comparar com número necessário para prejudicar.

**NÚMERO PRIMO:** Um número é dito primo se ele tiver dois e somente dois divisores os quais são o número 1 e o próprio número. O número 1 por ter apenas um divisor e não é considerado primo. Em inglês prime number.

**NUVEM DE PONTOS:** O mesmo que diagrama de dispersão.

**$n_i$  :** Frequência absoluta da classe i.

**$n!$  :** fatorial de n.

**$N(\mu, \sigma^2)$  :** Distribuição normal ou gaussiana.

**$N_i$  :** Frequência acumulada absoluta da classe i.

**$\binom{n}{r}$  :** Combinação de n objetos r a r.



**OBJETIVIDADE (MENSURAÇÃO):** Um instrumento de mensuração é considerado objetivo quando examinadores ou usuários dele, igualmente capazes, chegam aos mesmos resultados quando o usam.

**OBJETO:** Pessoa, produto ou serviço, empresa ou qualquer outra entidade que possa ser avaliada quanto a uma determinada quantidade de atributos.

**OBJETO:** Qualquer estímulo, incluindo entidades tangíveis, produto ou objeto físico, ações ou serviço, percepções sensoriais como cheiro, sabor, impressões visuais, ou mesmo pensamentos, ideias, slogans, que podem ser comparados e avaliados pelo respondente.

**OBJETIVOS:** A definição dos objetivos determina o que o pesquisador quer atingir com a realização do trabalho de pesquisa. Objetivo é sinônimo de meta, fim. Os objetivos podem ser separados em objetivos gerais e objetivos específicos.

**OBJETIVOS DE INFORMAÇÃO OU LEVANTAMENTO:** A informação para a tomada de decisões que se busca pelo questionário.

**OBJETIVO DE PESQUISA DE MARKETING:** A informação específica necessária para resolver um problema de pesquisa de marketing.

**OBSERVAÇÃO:** O conjunto de medidas obtidas para um único elemento.

**OBSERVAÇÃO:** Uma técnica de coleta de informações que envolve o estudo direto do comportamento, quando este ocorre, pela observação dos objetos do estudo sem intervir sobre eles.

**OBSERVAÇÃO:** Observar como o próprio nome diz é olhar, ouvir, prestar atenção a manifestações ambientais, dentre outros. A observação permite colher dados sobre o domínio cognitivo, afetivo, psicomotor, mais especialmente sobre estes dois últimos. Todo educador deve utilizar as observações em situações, tais como: i) desenvolvimento de um experimento; ii) desempenho na utilização de instrumentos, quaisquer que sejam: musicais, científicos, dentre outros; iii) habilidades básicas de ler, ouvir, dentre outros; iv) habilidades físicas: correr, nadar, jogar futebol; v) participação em trabalhos em grupo, dentre outros. É indispensável, para que se efetue uma observação condizente, que se construam instrumentos de registros de fatos a fim de evitar que opiniões subjetivas do observador sejam tomadas como fatos. Entre os instrumentos para o registro da observação destacam-se: o registro anedótico, a lista de checagem e as escalas de classificação.

**OBSERVAÇÃO:** Forma de obtenção de dados que requer de um observador que utilize um tipo de técnica ou método de observação, tais como, simulada, direta, casual, sistemática, participante, etc.

**OBSERVAÇÃO NATURALÍSTICA:** É aquela em que o investigador não interfere nos resultados do fenômeno. Contrapõe-se à observação participante. Refere-se ao tipo de observação do comportamento que ocorre em locais naturais, isto é, em ambientes não controlados. Contrapõe-se à observação planejada.



**OBSERVAÇÃO ABERTA:** O processo de monitorar pessoas que não sabem que estão sendo monitoradas.

**OBSERVAÇÃO ATÍPICA:** Uma observação que é substancialmente diferente das outras, isto é, tem um valor extremo. O mais importante é a sua representatividade da população.

**OBSERVAÇÃO CASUAL:** Tipo de observação que envolve o monitoramento de fenômenos relevantes, tais como, a opinião de eleitores numa pesquisa de opinião pública, o comportamento de consumidores num Shopping Center, numa pesquisa de mercado referente ao uso de determinado produto, etc., desde que não especifique antecipadamente os resultados esperados. É o contrário ao conceito de observação sistemática.

**ORDENAÇÃO (ORDINATION):** Métodos de redução de dados multivariados através do arranjo das observações ao longo de um número de eixos menor que o de variáveis.

**ORDENADA (ORDINATE):** A vertical, ou eixo y, de um gráfico. O oposto da abscissa.

**OPERAÇÃO DE DESLOCAMENTO (SHIFT OPERATION):** A transformação de uma variável aleatória normal através da adição de uma constante. Ver também mudança de escala.

**ORTOGONAL (ORTHOGONAL):** Em uma análise de variância (ANOVA) multifatorial, é a propriedade de que todas as combinações dos tratamentos sejam representadas. Em um delineamento de regressão múltipla, é a propriedade de que todos os valores de uma variável preditora sejam encontrados em combinação com cada valor de outra variável preditora.

**ORTOGONAL 2 (ORTHOGONAL):** Em álgebra de matrizes, dois vetores cujo produto é igual a zero (0) são ditos ortogonais.

**ORTONORMAL (ORTHONORMAL):** Uma transformação aplicada a uma matriz de forma que as distâncias euclidianas das linhas, colunas ou ambas sejam igual a 1.

**OBSERVAÇÃO ATÍPICA:** Em termos estritos, é uma observação que tem uma diferença substancial entre seus valores reais e os previstos da variável dependente, resultando num grande resíduo, ou entre seus valores das variáveis independentes e os de outras observações. O objetivo de denotar as observações atípicas é identificar as observações que são representações inadequadas da população da qual a amostra é extraída, de modo que elas possam ser desconectadas ou mesmo eliminadas da análise como não-representativas.

**OBSERVAÇÃO ATÍPICA:** Em termos estritos, uma observação que tem uma diferença substancial entre o valor real para a variável dependente e o valor previsto. Casos que são substancialmente diferentes, em relação às variáveis dependentes ou às variáveis independentes, também são chamados de atípicos. O objetivo é identificar as observações que são representações inadequadas da população da qual a amostra é obtida, de forma que elas podem ser ignoradas ou mesmos eliminadas da análise como não representativas.

**OBSERVAÇÃO CENSURADA:** Observação na análise de sobrevivência que não cumpriu o seguimento estabelecido, desconhecendo-se o motivo, como abandono, dentre outros.

**OBSERVAÇÃO CIENTÍFICA (CRITÉRIOS):** A observação, como qualquer outro método científico de coleta de dados, requer cuidados especiais a fim de que as informações tenham validade e fidedignidade. Os seguintes critérios são fundamentais: i) a observação científica é sistemática em vez de oportunista

ou casual; ii) a observação científica deve ser objetiva e livre de tendências; iii) sempre que possível, a observação científica deve ser quantitativa; iv) a observação deve satisfazer os critérios de fidedignidade, usabilidade, e especialmente de validade.

**OBSERVAÇÃO DE ALTA-INFERÊNCIA:** Embora os sistemas de observação estejam em um contínuum, frequentemente são agrupados em duas categorias: alta-inferência e baixa-inferência. Um sistema de alta inferência é aquele que requer do observador inferências substanciais sobre quais competências estão sendo demonstradas e sobre a qualidade de tais demonstrações. O comportamento per se não é registrado. Comumente, um observador trabalha a partir de um grupo de indicadores e descritores das competências e, então, faz um julgamento sobre quão bem os indicadores estão sendo desempenhados.

**OBSERVAÇÃO DE BAIXA-INFERÊNCIA:** Observação de baixa-inferência foca comportamentos específicos que são definidos com tal precisão que observadores concordarão quanto à sua ocorrência ou não. O processo consiste em registrar comportamentos observados: nem mais, nem menos. O registro é objetivo. Pode ser feito por um auxiliar ou por computador. Não há necessidade de inferências.

**OBSERVAÇÃO DE INTERAÇÃO:** Observar interação social e tomar sistematicamente nota de comportamentos ocorrentes entre duas ou mais pessoas que estão em geral juntas fisicamente e são psicologicamente associadas.

**OBSERVAÇÃO DESESTRUTURADA:** Estudo no qual o observador simplesmente anota o comportamento observado.

**OBSERVAÇÃO DISCREPANTE:** É qualquer observação que se destaca de todo o conjunto de observações, por ter seu valor muito afastado dos outros valores observados, podendo ser muito grande ou muito pequeno e consequentemente fornecendo um valor residual fora dos padrões apresentados pela maioria das observações.

**OBSERVAÇÃO ESTRUTURADA:** Estudo no qual o observador preenche um formulário tipo questionário ou conta o número de vezes em que uma atividade ocorre.

**OBSERVAÇÃO INDIRETA:** Fazer um registro de comportamento anterior.

**OBSERVAÇÃO INFLUENTE:** Observação com uma influência desproporcional sobre um ou mais aspectos das estimativas de regressão. Essa influência pode ter como base i) diferenças substanciais dos outros casos no conjunto variáveis independentes, ii) valores observados extremos altos ou baixos para variáveis critério, ou iii) uma combinação desses efeitos. As observações influentes podem ser boas, reforçando o padrão dos demais dados, ou ruins, quando um único caso ou um pequeno conjunto de casos afeta excessivamente que são as tendências, as estimativas de regressão.

**OBSERVAÇÃO PARTICIPANTE:** Mecanismo de observação no qual o cientista faz parte do conjunto que ele próprio está estudando, onde utiliza-se desta condição para obter informações referente a esse conjunto.

**OBSERVAÇÃO INFLUENTE:** Uma observação que exerce uma influência desproporcional sobre um ou mais aspectos das estimativas de regressão. Essa influência pode ser baseada em valores extremos das variáveis independentes ou da variável dependente, ou de ambas. As observações influentes podem ser boas, reforçando o padrão dos demais dados, ou serem ruins, quando um único caso ou um pequeno conjunto de casos afeta excessivamente as estimativas. Não é necessário que a observação seja atípica,

apesar de que muitas vezes observações atípicas também podem ser classificadas como influentes. Também é conhecida como ponto influente.

**OBSERVAÇÃO INFLUENTE:** Qualquer observação que tem uma influência desproporcional sobre os parâmetros estimados.

**OBSERVAÇÃO PLANEJADA:** Mecanismo de observação do comportamento que ocorre em ambientes controlados, tais como em laboratório, casa de vegetação, ripado, estufa, etc. É o contrário da observação naturalística.

**OBSERVADOR:** Sujeito ou elemento que obtém valores de uma variável numa determinada pesquisa científica.

**OBSERVAÇÃO OUTLIER:** Em um conjunto de dados é aquela observação que parece ser inconsistente com o conjunto de dados remanescentes. Os outliers podem indicar algumas características importantes sobre um modelo, como modelo incompatível com os dados e omissão de variáveis importantes. É importante lembrar que a rejeição automática de outliers não é um procedimento correto e as regras propostas para rejeição de outliers devem incluir a reanálise sem essas observações., que, dependendo das circunstâncias, podem ser portadoras de informações vitais dos indivíduos de uma população. E ainda uma observação é chamada de outlier de regressão por exemplo, ou outlier na direção Y, se ela se afasta do padrão linear definido pelas outras ou pela maioria das outras observações, este tipo de outlier em geral dá origem à grandes resíduos. Também conhecido como dado aberrante ou atípico.

**OBSERVAÇÃO PARTICIPANTE:** Técnica de pesquisa na qual pesquisadores participam diretamente dos eventos estudados, seja revelando, seja escondendo sua identidade de pesquisadores.

**OBSERVAÇÃO SISTEMÁTICA:** Processo observacional no qual o investigador estabelece antecipadamente os elementos que compõem o fenômeno que irá estudar.

**OBSERVAÇÃO PARTICIPANTE:** O termo descreve o processo geral de pesquisa em que um observador, aceito como legítimo membro de um grupo, usa essa posição privilegiada para obter informações sobre o grupo.

**OBSERVAÇÃO POR ESPELHO-ESPIÃO:** A prática de observar um grupo por trás de um espelho-espião sem que os observadores sejam notados.

**OCORRÊNCIA:** Ver frequência.

**OCTANTE:** É o procedimento de pesquisa e localização de pontos vizinhos próximos pela subdivisão da região do ponto a ser interpolado em oito setores, nos quais são escolhidos um determinado número de amostras mais próximas. Este procedimento garante uma boa amostragem espacial.

**OBJETIVOS DO APRENDIZADO:** Indicam o que você deve estar apto a fazer após tomar conhecimento dos conceitos para testar o seu entendimento; se conseguir realizar as tarefas contidas nos objetivos de aprendizado, significa que você entendeu os conceitos.

**OBSERVAÇÕES PAREADAS:** Cada observação em um grupo é pareada ou individualmente igualada com uma observação do outro grupo.

**ORDEM DE POSTOS:** Uma disposição sistematizada dos valores dos dados em ordem crescente ou decrescente.

**ODDS RATIO:** Teste estatístico simbolizado por OR para determinar a vantagem ou desvantagem de um evento em relação ao outro, no estudo de casos nos quais o número de controles é fixado previamente pelo investigador.

**ODDS RATIO:** Chance de se observar casos expostos ao fator de risco sobre a chance de se observar controles expostos ao fator de risco. Se a exposição ao fator de risco for a mesma para casos e controles o odds ratio vale 1. Também é chamado de razão de chances. Em inglês odds ratio.

**ODDS-RATIO (OR):** O mesmo que odds relativa. Outros sinônimos: chance relativa, relação de chances, razão de chances, razão de probabilidade, razão cruzada ou razão dos produtos cruzados. A última denominação advém da maneira como são feitos os cálculos, nos estudos de caso-controle: o produto das células ad, de uma tabela 2x2, é dividido pelo produto bc. OR é utilizado como medida de associação em estudos de caso-controle; é estimativa do risco relativo; OR constitui também uma medida própria de risco. As outras denominações advém do fato de o OR corresponder à chance de exposição em casos dividida pela chance de exposição em controles.

**ODDS-RATIO:** Indicador para estimativa do risco a partir dos valores levantados em estudos de caso-controle. É uma medida do efeito do fator em estudo no desenvolvimento do desfecho clínico, representando a razão entre as chances em favor da exposição nos casos como doentes e as chances em favor da exposição nos não-casos como sadios. Traduções propostas, mas insuficientes: razão de chances, razão de produtos cruzados, estimativa de risco relativo.

**ODDS-RATIO AJUSTADO:** Odds-ratio em que uma ou mais variáveis confundidoras foram neutralizadas.

**ODDS-RELATIVA (OR):** O mesmo que odds ratio.

**O ERRO DE TIPO I:** É a probabilidade  $\alpha$  de rejeitar erradamente a hipótese nula  $H_0$ , quando na realidade ela for verdadeira.

**O ERRO DE TIPO II:** É a probabilidade  $\beta$  de aceitar erradamente a hipótese  $H_0$  quando na verdade ela for falsa, o que é uma função dos parâmetros dados em  $H_0$  assim como os da distribuição verdadeira, e, portanto do valor verdadeiro  $\theta_1$  dos parâmetros da hipótese alternativa. Chama-se função potência do teste a  $1 - \beta(\theta_1)$ .

**ORDENADA:** Em uma representação gráfica no sistema de eixos cartesianos, refere-se ao eixo dos Y's ou eixo vertical.

**OGIVA:** Nome genérico da ogiva de Galton. Em inglês Ogive.

**OGIVA:** Representação gráfica de uma tabela de frequências acumuladas.

**OGIVA:** Gráfico de uma distribuição acumulada.

**OGIVA CRESCENTE:** É a que representa uma distribuição de frequência acumulada a partir dos valores mais baixos, no sentido dos valores mais altos da ordem de classificação. Opõe-se a ogiva decrescente.

**OGIVA DECRESCENTE:** É a que representa uma distribuição de frequência acumulada a partir da frequência que corresponde ao valor mais alto da ordem de classificação. Opõe-se a ogiva crescente.

**OGIVA DE GALTON:** Sob certas condições, isto é, quando a distribuição de frequências é unimodal, a curva de distribuição tem o formato de S ou seja, em forma de ogiva, termo emprestado por Galton (1875) da arquitetura e usado particularmente para a curva de distribuição da distribuição Normal. Em inglês Galton ogive.

**OBSERVAÇÃO INFLUENTE:** Uma observação em uma análise de regressão que tem um grande efeito nos parâmetros estimados no modelo. A influência é medida pela mudança nos parâmetros quando a observação influente é incluída e excluída da análise.

**ORDEM ALEATÓRIA:** Uma sequência ou ordem para um conjunto de objetos que seja executada de tal maneira que cada ordenação possível seja igualmente provável. Em planejamento de experimentos, as corridas do experimento são tipicamente arranjadas e executadas em uma ordem aleatória.

**ORTOGONAL:** Neste termo existem vários significados relacionados, incluindo: o sentido matemático de perpendicular; a definição de que duas variáveis são ortogonais se elas são estatisticamente independentes; ou em planejamento de experimentos em que um planejamento é ortogonal se se admitem estimativas estatisticamente independentes dos efeitos.

**OS 14 PONTOS DE DEMING:** Uma filosofia de gerência promovida por Deming (2003) que enfatiza a importância de mudança e da qualidade.

**OUTLIER(S):** Uma ou mais observações em uma amostra que esteja(m) tão longe do corpo principal dos dados que dão origem à questão de elas serem de uma outra população.

**O INTERVALO:** Medida de dispersão com características de que informa apenas os valores extremos, utilizando poucas informações, sendo, portanto pouco confiável. Mede a extensão da variação total durante um período. É fácil de calcular e tem sua aplicação como a primeira informação da amplitude de variação.

**OLS:** Quadrados mínimos ordinários.

**OPERACIONALISMO MÚLTIPLO:** Basicamente este princípio envolve o reconhecimento de que nenhuma observação oferece suficiente informação para definir um conceito teórico.

**OPERACIONALIZAÇÃO:** Um passo além da conceituação. Operacionalização é o processo de desenvolver definições operacionais.

**OPERADOR DE DERIVAÇÃO PARCIAL DE FUNÇÕES DE DUAS OU MAIS VARIÁVEIS (DEL):** Simbolizado por  $\partial$ .

**OPERADOR DE INTEGRAÇÃO (INTEGRAL DE):** Simbolizado por  $\int$ .

**OPERADOR TRAÇO MATRICIAL:** Simbolizado por tr.

**OPORTUNIDADE DE UM SISTEMA DE VIGILÂNCIA EPIDEMIOLÓGICA:** Medida quantitativa de avaliação de um sistema de vigilância epidemiológica. Avaliação da oportunidade de informação basear-se-á na disponibilidade dos dados. Pode ser definida como a quantidade de tempo decorrido desde a ocorrência de um agravo, e sua notificação ou seu conhecimento pelos serviços de saúde responsáveis pelas medidas de prevenção e controle. Aplica-se mais às afecções agudas, onde muitas vezes são necessárias medidas de controle imediatas.

**OPUS CITATUM OU OP.CIT.:** Significa obra citada.

**ORDEM ACIDENTAL:** De um conjunto é a que foi obtida pro um processo tal que todas as possíveis ordens dele decorrentes são equiprováveis.

**ORDEM CÍCLICA:** É a disposição de termos quaisquer a,b,c, ..., z em que existe uma relação de precedência da forma: a precede b, que precede c, ..., que precede z, que precede a, não havendo, assim, nem primeiro, nem ultimo termo. Opõe-se a ordem hierárquica.

**ORDEM DE UMA CADEIA DE MARKOV:** Número de tempos discretos do passado que determinam a probabilidade de ocorrência de um estado numa cadeia de Markov. Em inglês Order of a Markov chain.

**ORDEM HIERÁRQUICA:** É a disposição a,b,c, ..., z de termos quaisquer em que existe uma relação de precedência da forma: a precede b, que precede c, ..., que precede z, havendo, assim, um primeiro e um último termo. Opõe-se a ordem cíclica.

**ORDEM TEMPORAL DE OCORRÊNCIA APROPRIADA:** Para ser considerada uma causa provável de uma variável dependente, uma mudança na variável independente precisa ocorrer antes de uma mudança observada na variável dependente.

**ORDENAÇÃO:** Classe de técnicas multivariadas aplicadas a dados ecológicos. Via de regra, são métodos geométricos que buscam reduzir dados multivariados a poucas dimensões.

**ORDINAL:** Nível de mensuração de dados; caracteriza dados que podem ser dispostos em ordem, mas as diferenças entre valores de dados ou não podem ser determinadas ou não têm significado.

**ORGANIZAÇÃO DO QUESTIONÁRIO:** O exame dos questionários preenchidos e devolvidos para verificar sua exatidão, legibilidade e consistência, sendo frequentemente chamado de processo de limpeza.

**ORGANIZAÇÃO DOS DADOS:** Consiste em resumir os dados por meio de sua contagem e agrupamento.

**ORIENTAÇÃO DE METAS:** Enfoque na realização de metas empresariais com limite estabelecido na orientação para o consumidor.

**ORIENTAÇÃO DE SISTEMAS:** Criação de sistemas para monitorar o ambiente externo e fornecer o mix de marketing ao mercado alvo.

**ORIENTAÇÃO PARA CONSUMIDORES:** Identificar e focar pessoas ou firmas mais propensas a comprar um produto ou proporcionar um bem ou serviço que atenda às suas necessidades de forma mais eficiente.

**ORIGEM:** Denominação do ponto de coordenadas nulas, isto é, (0, 0). Em inglês origen.

**ORTOGONAL:** Dois vetores u e v são ortogonais se o produto escalar entre eles for zero, isto é,  $u.v = 0$ . Podem ser dito ainda que os vetores são perpendiculares. Em inglês orthogonal.

**ORTOGONAL:** Independência matemática ou sem correlação de eixos fatoriais, um em relação ao outro, ou seja, em ângulos retos ou de 90 graus.

**ORTOGONAL:** Independência estatística ou ausência de associação. As variáveis estatísticas ortogonais explicam a variância única, sem qualquer explicação de variância compartilhada entre elas. Contrastes



ortogonais são comparações planejadas estatisticamente independentes que representam comparações únicas de médias de grupos.

**ORTOGONAL:** Restrição matemática que exige que as estimativas das utilidades parciais sejam independentes uma da outra. Na análise conjunta, a ortogonalidade se refere à habilidade de medir o efeito de mudança de cada nível de atributo e de sua separação dos efeitos de mudança de outros níveis de atributo e de erro experimental.

**ORTOGONAL:** Restrição matemática que especifica que as funções canônicas são independentes uma da outra. Em outras palavras, as funções canônicas são obtidas de modo que cada uma forma um ângulo reto com todas as outras funções, quando representadas em um espaço multivariado, garantindo assim independência estatística entre as funções canônicas.

**ORTONORMAL:** Dois vetores unitários são ortonormais se forem ortogonais. Em inglês orthonormal.

**OSCILAÇÃO:** Diferença entre valores máximo e mínimo de uma função. Em inglês oscillation.

**OTIMIZAÇÃO:** É o processo de determinação do ponto ótimo de uma função por meio da maximização ou minimização da mesma, dependendo do problema colocado.

**OTIMIZAÇÃO:** Otimizar é chegar a melhor solução possível. Todo problema que requer decisão tem um espaço solucional. Otimizar é encontrar aquele ponto no espaço solucional que produzirá o melhor valor para aquele que toma a decisão.

**OU:** Símbolo matemático dado por  $\vee$ .

**OUTLIER:** Valor ou dado discrepante, atípico ou aberrante, obtidos em uma determinada distribuição de frequências ou de probabilidades que pode prejudicar os resultados da análise dos dados, mascarando os valores de medidas descritivas ou resumo como as de tendência central, variabilidade, assimetria e curtose, dentre outras.

**OUTLIER:** Valor, em uma observação, que é significativamente diferente das outras observações; valor discrepante.

**OUTLIERS (VALORES EXTREMOS):** Valores extremamente raros, no sentido de que estão muito afastados da maioria dos dados.

**OUTLIERS:** São valores que superam em uma vez e meia a amplitude interquartílica que é o quartil 3 menos o quartil 1. Valores extremos quando superam três vezes essa amplitude.

**OUTLIERS:** São valores que superam em uma vez e meia a amplitude interquartílica a qual é quartil 3 menos quartil 1. Valores extremos quando superam três vezes essa amplitude. Também conhecidos como dados aberrantes ou dados atípicos.

**O PRESSUPOSTO DE NORMALIDADE:** É chave para toda a estatística paramétrica, por essa razão você sempre deve checar a validade do mesmo. Contudo, quando sua amostra for suficientemente grande ( $n > 30$ ), dependendo do formato da distribuição, o Teorema Central do Limite garante a convergência da média amostral para a normalidade.

**O PRESSUPOSTO DE INDEPENDÊNCIA:** É chave para a maioria das estatísticas. Isto significa que o resultado de uma observação não interfere no resultado de outra observação. Por exemplo, a nota de um aluno não interfere na nota de um outro aluno. Já esse pressuposto é quebrado para amostras relacionadas como, por exemplo, passar uma prova antes de uma intervenção e a mesma prova ou outra depois da intervenção; essas duas medidas são correlacionadas, uma vez que os sujeitos são os mesmos. Neste caso, deve-se utilizar o teste para dados emparelhadas, ou análise de variância (ANOVA) com medidas repetidas. Logo, cuidado com os estudos longitudinais, em que se acompanha os mesmos sujeitos em vários momentos.





# P

**P (DE PROBABILIDADE):** Estimativa estatística de que um resultado é devido ao acaso; um valor encontrado  $p=0,05$  e as vezes  $p=0,01$  é estabelecido arbitrariamente, como ponto de corte para tomada de decisões na inferência estatística.

**PARÂMETRO:** Valor típico ou característica numérica representativo de uma distribuição de frequência ou de probabilidade de uma população, geralmente fixo ou constante e desconhecido, simbolizado por letras do alfabeto grego, como por exemplo a média populacional  $\mu$ .

**PACIENTE:** No contexto de pesquisa, o termo refere-se, sempre, ao paciente que participa, ou foi convidado para participar, da pesquisa.

**PACOTES DE SOFTWARE DE COMPUTADOR:** Programas estatísticos ou ambientes computacionais pré-preparados para uso em computadores, que facilitam a análise dos dados de pesquisas. Como exemplo pode ser citado o SAS, SANEST, R, SAEG, MINITAB, SPSS, S-PLUS, dentre outros.

**PADRÃO DE FATORES:** Numa matriz de fatores, é possível assumir a priori que alguns elementos são iguais a zero; por exemplo, se o  $j$ -ésimo fator não aparece na  $i$ -ésima variável,  $a_{ij}=0$ . O padrão de coeficientes diferentes de zero é chamado de padrão de fatores. Em inglês Factor pattern.

**PADRÃO DE SEQUÊNCIA:** Necessidade de pular certas perguntas se o respondente der determinada resposta à pergunta anterior.

**PADRÃO ESPACIAL:** Refere-se a distribuição ou dispersão espacial dos indivíduos de uma comunidade ou população no espaço, geralmente um espaço bidimensional ou plano. Reconhece-se três formas gerais de padrão espacial. Padrão completamente aleatório: a localização dos indivíduos não tem nenhum padrão identificável sendo completamente aleatória. Padrão agrupado ou agregado: os indivíduos se agregam formando grupos ou regiões com maior densidade de indivíduos. Padrão regular: os indivíduos se localizam seguindo alguma forma de regularidade o que resulta em zonas de exclusão de indivíduos. Existe uma gradação contínua do extremo do padrão mais agregado ao extremo do padrão mais regular, sendo que regularidade e agregação têm graus variados. Somente o padrão completamente aleatório pode ser concebido como um ponto nessa gradação contínua.

**PADRÃO-OURO (GOLD STANDARD):** O mesmo que critério-padrão.

**PADRÃO PONTUAL:** Conjunto de dados que consiste numa série de locais de uma região geográfica em estudo, na qual são observadas algumas características de interesse. Em inglês Point pattern.

**PADRÕES DE CONSUMIDORES:** Desenhos que registram os passos de um comprador dentro de uma loja.

**PADRÕES ESPACIAIS:** Padrão de comportamento de uma variável no espaço. Em inglês Spatial patterns.

**PADRONIZAÇÃO:** Processo no qual os dados originais são transformados em novas variáveis com uma média de 0 e um desvio-padrão de 1. Quando os dados são transformados desse modo, o termo  $b_0$  ou

intercepto da equação de regressão assume um valor 0. Quando se usam dados padronizados, os coeficientes de regressão são conhecidos como coeficientes beta, os quais permitem ao pesquisador comparar diretamente o efeito relativo de cada variável independente sobre a variável dependente.

**PADRONIZAÇÃO:** É a transformação de valores originais, feita pela subtração da sua média e divisão pelo desvio padrão, fazendo com que a distribuição padronizada tenha média igual a zero e variância igual a 1, como uma distribuição teórica normal:

$$Z = \frac{X - \bar{X}}{S}$$

A variável aleatória resultante dessa transformação denomina-se variável aleatória padronizada. Veja que esta transformação permite fazer a comparação entre conjuntos de valores com variações muito diferentes, por exemplo, de elementos traços em rochas.

**PADRONIZAÇÃO:** Conjunto de operações capazes de garantir a um teste precisão, validade, normas padronizadas, instruções para sua aplicação, avaliação e interpretação. Em inglês standardization.

**PADRONIZAÇÃO:** É a uniformização de dados brutos usualmente apresentados em diferentes escalas para uma única escala. Isto inclui conversão para escores normalizados, ou seja, normalmente distribuídos e com média em 0 e desvio padrão de 1 ou reescalando em um intervalo para dar 0 como valor mínimo e 1 como valor máximo.

**PADRONIZAÇÃO CENTRADA EM LINHA:** Ver padronização dentro de casos.

**PADRONIZAÇÃO DIRETA:** É atribuída a mesma distribuição etária a duas populações a serem comparadas. Então, essa distribuição é aplicada aos coeficientes de mortalidade específicos de cada grupo etário para determinar o número de óbitos que ocorreria em cada uma das duas populações se elas fossem idênticas na distribuição por idade.

**PADRONIZAÇÃO DIRETA:** Método preferido de padronização, se os coeficientes específicos provém de grandes populações e os dados necessários estão disponíveis. O método direto dos coeficientes padronizados de mortalidade, por exemplo, aplicam na distribuição etária de uma população, a população padrão, os coeficientes de mortalidade específicos por idade de populações diferentes a serem comparadas. Isto remove o viés que ocorreria se uma população de velhos fosse comparadas com uma de jovens.

**PADRONIZAÇÃO INDIRETA:** Método de padronização utilizado quando as populações a serem comparadas são pequenas de forma que os coeficientes de mortalidade específicos por idade são instáveis, ou os coeficientes de mortalidade específicos por idade de uma ou mais populações não estão disponíveis, mas os dados referentes à distribuição por idade e ao coeficiente geral de mortalidade estão. Aqui, os coeficientes de mortalidade padronizados da população-padrão são aplicados aos grupos etários correspondentes, nas diferentes populações a serem estudadas. O resultado é um coeficiente de mortalidade esperado geral padronizado para cada população em estudo. Esses valores esperados são aqueles que seriam esperados, se os coeficientes de mortalidade-padrão fossem verdadeiros para a população em estudo. Após, calcula-se a razão de mortalidade padronizada.

**PADRONIZAÇÃO INDIRETA:** Usada quando os coeficientes de mortalidade específicos por idade não estão disponíveis na população em estudo ou se a população em estudo é pequena e, portanto, produziria

coeficientes de mortalidade que seriam estatisticamente instáveis. Os coeficientes da população padrão são aplicados à distribuição etária do grupo em estudo, que é conhecida.

**PADRONIZAÇÃO INTERNA:** Método de padronização no qual as respostas de um respondente não são comparadas com a amostra geral mas, em vez disso, são comparadas com suas próprias respostas. Também é conhecida por usar as respostas médias dos respondentes para padronizar suas próprias respostas.

**P:** Abreviação de probabilidade ou nível alfa. Em metodologia, normalmente as afirmações estatísticas são seguidas de expressões do tipo:  $p < 0.05$ , o que significa, neste exemplo, que a probabilidade de que a afirmação esteja incorreta é menor do que 5 %. Conhecido também como valor p, p-value ou nível descritivo do teste de hipótese.

**PACOTE ESTATÍSTICO:** Ver planilha eletrônica de dados.

**PAINEL:** É um tipo de estudo longitudinal no qual o pesquisador avalia o comportamento das variáveis durante determinado período de tempo, utilizando sempre das mesmas pessoas.

**PERCENTIL:** Posição numa escala de escores abaixo da qual se encontra determinada porcentagem de escore. Por exemplo, se um estudante obtém uma pontuação numa prova que o coloca no percentil 82%, isso significa que ele saiu-se melhor do que 82 % dos alunos com os quais está sendo comparado. Quando se divide determinada distribuição de escores em quatro grupos de igual tamanho que são os quartis, os três pontos utilizados para realizar essa divisão são os percentis 25, 50 e 75, também conhecidos como primeiro, segundo e terceiro quartis, respectivamente.

**PESQUISA QUALITATIVA:** i) Modalidade de pesquisa na qual os dados são coletados por meio de interações sociais, por exemplo, estudos etnográficos e pesquisas participantes, e analisados subjetivamente pelo pesquisador; ii) Enquanto a pesquisa quantitativa investiga fatos, a pesquisa qualitativa preocupa-se com fenômenos (MARTINS; BICUDO, 1989), sendo que um fato é tudo o que pode ser objetivamente observado e definido por consenso social, enquanto um fenômeno remete-nos à interpretação de um fato feita por um observador. Ou seja, o fenômeno é a interpretação subjetiva do fato.

**PESQUISA QUANTITATIVA:** i) Modalidade de pesquisa na qual variáveis predeterminadas são medidas e expressas numericamente. Os resultados também são analisados com o uso preponderante de métodos quantitativos, por exemplo, estatística; ii) Modalidade de pesquisa que investiga fatos.

**PERÍODO DE RETORNO T:** O período de retorno de uma variável X define-se como o número de anos que deve, em média, decorrer para que o valor dessa variável ocorra, ou seja, superado. Designando a probabilidade de a variável aleatória X assumir um valor menor ou igual a x, por probabilidade de não excedência, simbolizada por F(x) com,  $F(x) = P(X \leq x)$  e designando a probabilidade de a variável aleatória ou estocástica X assumir um valor superior a x, por probabilidade de excedência G(x), com resultado dado por  $G(x) = P(X > x) = 1 - P(X \leq x) = 1 - F(x)$ , pode exprimir-se o período de retorno por,

$$T = \frac{1}{G(x)} = \frac{1}{1 - F(x)}$$

**PLACEBO:** Nas pesquisas experimentais, constitui-se no tratamento fictício dado ao grupo controle, para efeito de comparação com os resultados obtidos pelo grupo experimental, que é submetido ao tratamento verdadeiro. O placebo pode ser, por exemplo, uma injeção de água esterilizada ou uma pílula inócua de farinha, de forma que o indivíduo acredita estar sendo tratado com ingredientes ativos reais.

**PLANILHA ELETRÔNICA:** Tabela que organiza os dados em linhas e colunas. Programa de computador ou software que disponibiliza uma tabela para a entrada de dados na qual as linhas representam os sujeitos e as colunas representam as variáveis do estudo. Esses softwares apresentam normalmente uma série de funções e operações para os cálculos estatísticos necessários para a análise dos dados que foram tabulados. Por exemplo: BioEstat, BMDP, Excel, Lisrel, Minitab, SAS, SPSS, S-plus, Statistica dentre outros. Ver também célula e tabulação.

**PLATICÚRTICA:** Tipo de kurtose que uma curva de distribuição de dados pode assumir, caracterizada por ser mais alongada e achatada do que uma distribuição normal. Ver kurtose para um exemplo gráfico.

**PODER DO TESTE:** Capacidade de um teste estatístico em detectar relacionamentos ou diferenças entre variáveis, por exemplo, determinado estudo pode ser poderoso o suficiente para detectar uma diferença de 30 %, mas não ser capaz de detectar uma de 20 %; Probabilidade de se rejeitar uma hipótese nula falsa. O poder do teste pode ser calculado subtraindo-se a probabilidade de um Erro Tipo II de 1.

**POPULAÇÃO:** Grupo de pessoas, objetos ou eventos que possui um conjunto de características comuns que o definem. Totalidade de pessoas, objetos ou eventos que se deseja estudar e realizar sobre a qual realizar-se-ão generalizações. Para tanto, normalmente, lança-se mão de uma amostra dessa população, que será submetida efetivamente aos procedimentos de pesquisa e a partir da qual se efetuarão as generalizações para toda a população.

**PROJETO DE PESQUISA:** Tipo de projeto associado a um tipo de investigação científica previamente planejado pelo cientista constituído das seguintes etapas: formulação do problema; delinear o experimento; condução do ensaio e coleta dos dados; tabulação e análise de dados e discussão dos resultados bem como conclusão.

**PESQUISA:** Termo usado quando se investigam coisas novas, por exemplo, a criação de novas cultivares.

**PROBABILIDADE:** Chance de ocorrência de um evento. Formalmente, o termo probabilidade, pode ser definido como a frequência relativa de um evento. Por exemplo, a probabilidade de tirarmos um ás de um baralho com as cartas misturadas é de  $\frac{4}{52}$ , ou aproximadamente 7,7 %, pois há apenas quatro ases num

baralho completo de 52 cartas ao total. Outro exemplo clássico: a probabilidade de obtermos cara ao jogarmos uma moeda para o alto é de  $\frac{1}{2}$  (ou 50 %), dado que há apenas duas possibilidades: cara ou coroa;

Área da matemática devotada ao estudo das probabilidades.

**PARADIGMA (PARADIGM):** Uma estrutura de pesquisa sob a qual existe uma concordância geral. Kuhn (1969) usou o termo paradigma ao descrever ciência ordinária, que é a atividade de cientistas engajados em ajustar as observações aos paradigmas. Quando muitas observações não se ajustam, é hora de um novo paradigma. Kuhn chamou as mudanças de paradigmas de revoluções científicas.

**PARAMÉTRICO (PARAMETRIC):** O pressuposto de que quantidades estatísticas podem ser estimadas usando distribuições de probabilidades com constantes definidas e fixas. O requisito de que vários testes estatísticos conformem uma distribuição de probabilidades conhecida e definível.

**PARÂMETRO (PARAMETER):** Uma constante em distribuições e equações estatísticas que precisam ser estimadas a partir dos dados.

**PARÂMETROS PARCIAIS DO MODELO DA REGRESSÃO (PARTIAL REGRESSION PARAMETERS MODELS):** Os parâmetros em um modelo de regressão múltipla cujos valores refletem a contribuição de todos os outros parâmetros no modelo.

**PROBLEMA:** Questão a ser investigada numa pesquisa, colocada na forma interrogativa. O problema é uma especificação do tema de pesquisa, devendo ser circunscrito e bem definido. A definição do problema constitui-se numa etapa crucial para o desenvolvimento da pesquisa, e, dependendo de sua correta formulação e análise, decidirá que tipo e delineamento de pesquisa deverão ser adotados. Possui estreita relação com o tema e as hipóteses de pesquisa. Alguns exemplos auxiliarão a clarear a questão. Exemplo 1: Tema: Estresse em estudantes de pós graduação. Problema: Estudantes universitários de matemática do período matutino possuem menor nível de estresse do que seus congêneres do período noturno? Hipótese: Os estudantes do período noturno, por estarem, em geral, submetidos a maior carga de atividade diárias, têm maior incidência de estresse do que seus congêneres do período matutino. Exemplo 2: Tema: Epidemiologia: Papiloma vírus humano (HPV). Problema: Existem associações entre fatores epidemiológicos e a infecção genital pelo papiloma vírus humano (HPV)? Hipótese: Variáveis ligadas ao comportamento sexual e ao nível socioeconômico devem estar associadas à incidência da infecção pelo HPV.

**PROCEDIMENTO EXPERIMENTAL:** Num experimento, diz-se da aplicação da variável independente sobre o(s) grupo(s) experimental(is). Conjunto de procedimentos aos quais os sujeitos do(s) grupo(s) experimental(is) são submetidos pelos pesquisadores e que não afetarão os sujeitos do grupo controle.

**PROBIT:** A Função Probit é a inversa da Distribuição Normal Acumulada ou a Função Quantil associada com a Distribuição Normal Padrão. Ela é utilizada na modelagem de variáveis de resposta binárias.

**PROPORÇÃO:** Valor que varia entre 0 e 1 e que pode ser calculado dividindo-se o número de objetos que possuem determinada característica pelo número total de objetos de um conjunto. Por exemplo: de um total de 745 alunos do curso de engenharia, apenas 92 são mulheres, ou seja, a proporção é de 0,12. Para calcularmos a porcentagem a partir da proporção, basta multiplicar esse valor por 100 ( $0,12 \times 100 = 12\%$ ). Ver também distribuição de frequência e probabilidade e tipo de nível de mensuração de uma variável, também conhecido como nível racional.

**PSICOMETRIA:** Corpo de conhecimentos acerca da medida ou mensuração nas ciências do comportamento, como por exemplo, psicologia, sociologia, antropologia dentre outras. Concepção segundo a qual é viável a utilização de símbolos matemáticos ou números no estudo científico dos fenômenos naturais e sociais.

**PARCIMÔNIA:** Critério adotado na ciência segundo o qual o cientista deve sempre optar pela simplicidade e economia. Por exemplo, num modelo de regressão deve-se escolher aquele mais simples ou menos complexo possível com menor número de parâmetros, desde que com um elevado valor do coeficiente de determinação e atendendo os pressupostos básicos da análise de regressão em estudo.

**PAINÉIS PELO CORREIO:** Participantes pré-contactados e selecionados aos quais são enviados questionários periodicamente.

**PALAVRAS E FRASES EMOCIONAIS:** Palavras e frases que promovem reações emocionais em vez de respostas racionais e objetivas.



**PALAVRAS E FRASES NÃO-ESPECÍFICAS:** Redação confusa que um entrevistado pode interpretar de mais de uma maneira.

**PANMÍTICA:** População grande cujos componentes reproduzem-se por acasamentos ao acaso.

**PARÁBOLA:** É uma curva de segundo grau ou cônica. Ela pode ser definida como o lugar geométrico dos pontos que são equidistantes de uma reta que é a diretriz e de um ponto fixo, ou seja, o foco. Outra forma é defini-la por meio de sistema cartesiano como sendo o conjunto de pontos do plano que satisfazem a equação  $y = x^2$ . Em inglês parabola.

**PARABOLÓIDE:** Família de sólidos geométricos de base circular cujo diâmetro da secção transversal aumenta segundo uma parábola à medida que se desloca do topo para a base do sólido. O fator de forma absoluto é sempre maior que um terço, no caso do parabolóide quadrático ou ordinário o fator é exatamente um meio.

**PARADIGMA:** Modelo ou quadro de referência para observação e entendimento. Nas ciências sócias, três paradigmas importantes são funcionalismos, interacionismo e teoria do conflito.

**PARADOXO DE LINDLEY:** A incerteza a priori tem um papel crucial na comparação bayesiana de hipóteses nulas. No limite, quando a variância da distribuição a priori tende ao infinito, o fator de Bayes também aumenta de forma infinita. Este é o resultado de usar prioris não informativas e é conhecido como paradoxo de Lindley. Em inglês Lindley's paradox.

**PARADOXO DE SIMPSON:** Refere-se à reversão da direção de uma comparação ou de uma associação quando se combinam dados de vários grupos para formar um grupo único. As variáveis ocultas do paradoxo de Simpson são categóricas. Isto é, elas separam por exmplo as pessoas em grupos, como no caso de pacientes de cirurgia que são classificados em boas condições ou em más condições. O paradoxo de Simpson é apenas uma forma extrema do fato de que asociações observadas podem ser enganosas quando há variáveis ocultas.

**PARÂMETRO:** Característica de uma população.

**PARÂMETRO:** Medida derivada da população ou de todo o universo e não de uma amostra.

**PARÂMETRO:** Valor usualmente desconhecido que caracteriza uma população. Por exemplo, a média populacional, a variância e o desvio-padrão populacional são parâmetros.

**PARÂMETRO:** Valor relativo a uma população, geralmente simbolizado por uma letra do alfabeto grego. Como por exemplo a média da população indicada pela grega  $\mu$  ou  $\mu$ . Em inglês parameter.

**PARÂMETRO:** Esta palavra é usada no seu significado matemático comum de uma quantidade desconhecida que pode variar numa amplitude de valores. Em estatística, é mais encontrada em expressões que definem distribuições de probabilidade que representam parâmetros populacionais ou em modelos que descrevem situações estocásticas como parâmetros de regressão, por exemplo. O domínio de variação permissível de parâmetros define a classe de população ou modelo sob consideração. Em inglês Parameter.

**PARÂMETRO:** Quantidade ou medida característica da população. Por exemplo,  $\mu$  e  $\sigma^2$  são os símbolos usados para os parâmetros populacionais de média ( $\mu$ ) e variância ( $\sigma^2$ ). Estes normalmente são estimados

a partir dos dados da amostra em que a média aritmética da amostra é utilizada como uma medida da média populacional e a variância da amostra é empregada para estimar a variância da população.

**PARÂMETRO:** É uma medida usada para descrever, de forma resumida, uma característica da população. Por exemplo, a média populacional ( $m$ ), a proporção populacional ( $p$ ), a variância populacional ( $s$ ), o coeficiente de correlação ( $r$ ), dentre outros. Os parâmetros, via de regra, são valores desconhecidos e deseja-se estimar, ou testar, a partir dos dados de uma amostra.

**PARÂMETRO:** Números, ou conjunto de números, usados para sumariar a descrição de uma população estudada. Os números podem ser médias, percentagens, razões ou outras expressões numéricas.

**PARÂMETRO DA POPULAÇÃO:** O valor que define a verdadeira característica de uma população total.

**PARÂMETRO DE ESCALA:** Um parâmetro de uma distribuição de probabilidade que está relacionado funcionalmente com a escala da variável, por exemplo, o desvio padrão em uma distribuição normal. Em inglês Scale parameter.

**PARÂMETRO DE LOCAL:** Um parâmetro que localiza a distribuição de frequência no sentido de definir o centro ou valor típico como uma média ou moda. Em inglês Location parameter.

**PARÂMETRO DE POPULAÇÃO:** Um valor numérico usado como medida de sumário para uma população como por exemplo, a média de população, a variância de população e o desvio-padrão.

**PARÂMETROS:** São valores singulares que existem na população e que servem para caracterizá-la. Para definirmos um parâmetro devemos examinar toda a população.

**PARÂMETROS:** Característica determinada para uma população.

**PARÂMETROS:** Valores numéricos das populações representando constantes de cada variável do universo investigado.

**PARÂMETROS DE DISTÚRPIO:** Nas teorias tanto de estimação quanto de testes de significância surge o problema de encontrar uma distribuição de amostragem que seja independente de certos parâmetros desconhecidos da população. Embora estes parâmetros sejam essenciais para a especificação, eles possuem um distúrbio na formação de afirmações exatas sobre outros parâmetros. O caso típico de parâmetro de distúrbio aparece na definição de intervalos de confiança para a média, que dependem da variância da população; a dificuldade neste caso é contornada usando a distribuição de Student que não depende da variância. Em inglês Nuisance parameters.

**PARCELA (UNIDADE AMOSTRAL):** Parcelas podem ser de área fixa ou variável. As parcelas de área fixa mantêm a mesma área e forma em todo levantamento. As parcelas de área variável são baseadas em pontos de amostragem, onde a inclusão de indivíduos não é definida por uma área pré-estabelecida, mas por algum atributo como por exemplo proximidade do centro da parcela, como no método de Prodan. Todos os eventos de interesse encontrados em seu interior são amostrados. O conjunto de parcelas forma a amostra que será analisada.

**PARCELA:** Porção de material experimental a que se aplica um tratamento num experimento. Também se chama unidade experimental.

**PARCELA:** É a unidade que vai receber o tratamento e fornecer os dados que deverão refletir o efeito do tratamento.

**PARCELA DE ÁREA FIXA:** Parcela cuja área e forma são definidas antes do trabalho de campo, sendo as mesmas para todas as unidades amostrais num dado levantamento florestal. Podem ter várias formas, sendo as mais comuns: **retangular**, **em faixa** a qual é retangular com uma das dimensões bem maior que a outra, **circular** e **quadrada**.

**PARCELA DE ÁREA VARIÁVEL:** Amostragem por enumeração angular.

**PARCELA DE RAIOS VARIÁVEL:** Amostragem por enumeração angular.

**PARCERIA ESTRATÉGICA:** Duas ou mais empresas de pesquisa de marketing com habilidades e recursos singulares formam uma aliança para oferecer um novo serviço aos clientes, proporcionar suporte estratégico para cada empresa ou, de alguma maneira, criar benefícios mútuos.

**PARCIMÔNIA:** Grau em que um modelo conquista qualidade de ajuste para cada coeficiente estimado. A meta não é minimizar o número de coeficientes ou maximizar o ajuste, mas maximizar a quantidade de ajuste por coeficiente estimado e evitar superajuste do modelo com coeficientes adicionais que atingem apenas pequenos ganhos no ajuste.

**PAREAMENTO:** O mesmo que emparelhamento.

**PARTICIPANTE:** O mesmo que sujeito, isto é, pode ser um paciente ou apenas voluntário que participa de um estudo.

**PARTIÇÃO DO ESPAÇO AMOSTRA:** Sejam  $A_1, A_2, \dots, A_n$  eventos de um espaço amostra  $S$ . Diremos que a coleção  $\{A_i\}$  forma uma partição do espaço amostra  $S$ , se as seguintes condições forem verificadas: (i)  $A_i \cap A_j = \emptyset$ , para todo  $i \neq j$ . (ii)  $A_1 \cup A_2 \cup \dots \cup A_n = S$  e (iii)  $P(A_i) > 0$ , para todo  $i$ . Em inglês sample space partition.

**PASSEIO ALEATÓRIO:** Seja  $\{\varepsilon_t\}$  um processo discreto puramente aleatório com média  $\mu$  e variância  $\sigma_\varepsilon^2$ . Um processo  $\{X_t\}$  é chamado de passeio aleatório se  $X_t = X_{t-1} + \varepsilon_t$ . Fazendo-se substituições sucessivas obtém-se que

$$X_t = X_{t-2} + \varepsilon_{t-1} + \varepsilon_t$$

$$X_t = X_{t-3} + \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t$$

**M**

$$X_t = X_0 + \sum_{j=1}^t \varepsilon_j$$

E iniciando o processo em  $X_0 = 0$  não é difícil verificar que  $E(X_t) = t\mu$  e  $Var(X_t) = t\sigma_\varepsilon^2$ . Como a média e a variância dependem de  $t$  este processo é não estacionário. No entanto é interessante notar que a primeira diferença de um passeio aleatório é estacionária já que  $\nabla X_t = X_t - X_{t-1} = \varepsilon_t$ . Os exemplos mais conhecidos de séries temporais que se comportam como um passeio aleatório são os preços de ações em dias sucessivos.

**PATAMAR:** É o valor de variância no qual o variograma estabiliza-se, indica a passagem do campo estruturado para o campo aleatório.

**PERÍODO DE ACOMPANHAMENTO:** Metodologia adotada em pesquisas na área de saúde ou epidemiologia, principalmente psicologia e medicina, e que diz respeito ao período de tempo no qual os indivíduos serão observados após o final da pesquisa.

**PEDIDO DE PESQUISA:** Documento que descreve um projeto potencial de pesquisa, seus benefícios para a organização e os custos estimados. Em muitas organizações, um projeto não pode começar até que o pedido de pesquisa tenha sido aprovado formalmente.

**PEPITA PURO:** É o modelo de variograma de um fenômeno aleatório, descrito pela seguinte equação:  $\gamma(h) = C$ . Observe-se que neste modelo, qualquer que seja a distância  $h$ , a variabilidade é constante e, portanto, não há como reconhecer a configuração espacial dos dados, pois se este modelo for utilizado qualquer amostra contribuirá com o mesmo peso ( $1/n$ ,  $i=1, n$ ) independente de sua distância ao ponto a ser interpolado.

**PERCENTIL:** O ponto numa escala de escores abaixo do qual está uma dada porcentagem dos escores ou casos. Por exemplo, um indivíduo que obteve um escore correspondente ao 68º percentil, pode gabar-se de estar se saindo melhor que 68 % das pessoas com quem ela está sendo comparada. Os percentis de 25, 50 e 75 dividem uma distribuição de escores em quatro quartos, e são chamados o primeiro, segundo e terceiro quartis, respectivamente.

**PERCENTIL:** A série de valores que dividem a frequência total em cem partes iguais. Esta série de valores específica é mais usada em educação e psicologia. Alguns escritores preferem usar o termo centil ao invés de percentil. Em inglês Percentile.

**PERCENTIL:** O  $p^{mo}$  percentil de uma lista é o número tal que o  $p$  por cento dos elementos da lista estão abaixo dele.

**PERCENTIL:** O percentil  $i$  é o valor abaixo do qual fica  $i$  % da distribuição.

**PERCENTIL (OU CENTIL):** Uma das 100 divisões de uma distribuição de frequências; ou um centésimo desta distribuição. Palavras assemelhadas: decil que é um décimo da distribuição de frequência, quintil que corresponde a um quinto, quartil que equivale a um quarto e o tercil a um terço.

**PERCENTIL:** Cada um dos 99 valores da variável que dividem uma distribuição de frequências em 100 intervalos iguais. O percentil é representado pelo símbolo  $P_r$ , em que  $r$  representa a ordem percentil. Se, num grupo de estudantes, 60 % obtêm escores inferiores a 80, diz-se que o 60º percentil é 80 e que 60 é a ordem percentil correspondente a 80. Em inglês percentile.

**PERCENTIL:** Qualquer percentil ( $k$ ) é o valor ao qual  $k$  % dos valores, no conjunto de dados, são menores que o valor do  $k^a$  percentil e  $(1-k)$  % dos valores são maiores.

**PERCENTIL:** É um conceito comum em estatística, mas parece não ter uma definição universal. Uma das definições diz que o  $p$ -ésimo percentil ou de ordem  $p$  de uma amostra é um valor tal que  $p$  % dos elementos têm valores menores que ele e  $(100 - p)$  % têm valores maiores. Percentis de ordem 25, 50 e 75 são denominados **quartis** e, mais especificamente, primeiro, segundo e terceiro quartis. Os símbolos usuais são  $Q_1$ ,  $Q_2$  e  $Q_3$  respectivamente.

**PERCENTIL:** Os 99 valores que dividem dados ordenados em 100 grupos com aproximadamente 1% dos valores em cada grupo.

**PERCENTUAL CORRETAMENTE CLASSIFICADO:** Ver razão de sucesso.

**PERCENTUAL DE RISCO ATRIBUÍVEL NOS EXPOSTOS (RA%<sub>(exposto)</sub>):** Termo que significa uma resposta à seguinte questão, entre os que apresentam o fator de risco, que porcentagem do risco total ocorre devido a esse fator de risco?

**PERCENTUAL GLOBAL DE CONCORDÂNCIA:** Percentual total de observações que se encontram nas células a e d em uma tabela 2 x 2.

**PERCEPTRON MULTICAMADAS:** A forma mais popular de rede neural, contém pelo menos uma camada oculta de interconexões entre as interconexões de entrada e as de saída.

**PERDIDOS AO ACASO (MAR-MISSING AT RANDOM):** Classificação de dados perdidos aplicável quando valores perdidos de Y dependem de X, mas não de Y. Quando dados perdidos são MAR, dados observados para Y são uma amostra verdadeiramente aleatória para os valores de X na amostra, mas não uma amostra aleatória de todos os valores de Y devido a valores perdidos de X.

**PERÍODO DE REPOUSO:** Expressão utilizada nas pesquisas em medicina que diz respeito ao espaço temporal no qual ocorre uma pausa temporária da aplicação de um tratamento, droga ou tratamento medicamentoso em pessoas que estão sendo ensaiados.

**PCA:** Ver análise de componentes principais.

**PERCENTIL (PERCENTILE):** Refere-se a centésima parte de uma distribuição de valores. Por exemplo o quinto percentil corresponde a um vigésimo elemento da série, e também a um décimo percentil que é corresponde ao decil, etc.

**PERMUTAÇÃO (PERMUTATION):** O arranjo de objetos discretos. Diferente de combinação, a permutação é qualquer rearranjo de objetos. Assim, para o conjunto {1,2,3}, há seis permutações possíveis: {1,2,3}, {1,3,2}, {2,1,3}, {2,3,1}, {3,2,1}, {3,1,2}. Contudo, tudo isso representa a mesma combinação dos três objetos 1, 2, 3. Existem n! permutações de n objetos, mas apenas  $\frac{n!}{X!(n-X)!}$  combinações de n

objetos tirados X por vez. O símbolo “!” é o operador matemático para fatorial. Ver também coeficiente binomial e combinação.

**PLANILHA ELETRÔNICA (SPREADSHEET):** Um arquivo de computador em que cada linha representa uma única observação e cada coluna, uma única variável medida ou observada. A forma padrão na qual os dados ecológicos ambientais são colocados em meios eletrônicos. Ver também arquivo simples.

**PLATICÚRTICA (PLATIKURTIC):** Uma distribuição de probabilidades é dita platicúrtica se ela tiver relativamente muitos valores no centro da distribuição e caudas relativamente finas. A curtose de uma distribuição platicúrtica é maior que 0. Comparar com leptocúrtica.

**PLOT (PLOT):** Uma ilustração de uma variável ou um conjunto de dados. Tecnicamente, plots são elementos de gráficos.

**PODER (POWER):** A probabilidade de corretamente rejeitar uma hipótese nula estatística falsa. É calculado como  $1 - \beta$ , onde  $\beta$  é a probabilidade de cometer um erro Tipo II.

**PORCENTAGEM (PERCENTAGE):** Termo que significa literalmente, de 100, vem do Latim per centum, é a frequência relativa de um conjunto de observações multiplicada por 100. As porcentagens podem ser calculadas a partir de uma tabela de contingência, mas a análise estatística desta tabela sempre é baseada nas contagens de frequência relativa.

**PERFEITAMENTE CORRELACIONADOS:** Duas grandezas X e Y dizem-se perfeitamente correlacionadas se existe entre elas uma relação da forma  $Y = aX + b$ , onde a e b são duas constantes; sua correlação será 1 se a for positivo, e -1 se a for negativo.

**PERFIL EPIDEMIOLÓGICO:** Desafio aos poderes públicos estaduais e municipais para que, por meio de conhecimento dos problemas de saúde-doença, de suas respectivas populações, seja efetuado o planejamento das ações de saúde e sua avaliação, assegurando o uso racional dos recursos, via controle pela sociedade, sob a égide do princípio de equidade: saúde, direito de todos. Só faz sentido falar de perfil epidemiológico quando associado às classes sociais, condição essencial para a reprodução que é o perfil reprodutivo.

**PERGUNTA CONTIGENTE:** Questão feita apenas a alguns respondentes, determinada pelas respostas deles a outra pergunta. Por exemplo, todos os respondentes podem ser indagados se pertencem à uma determinada empresa e só os que responderem sim seriam indagados com que frequência vão a reuniões e confraternizações. Esta última seria uma pergunta contingente.

**PERGUNTAS ABERTAS:** Perguntas que não têm categorias pré-existentes de respostas e, portanto permitem ao entrevistado responder com suas próprias palavras.

**PERGUNTAS ABERTAS:** Itens de questionário que pedem aos respondentes dar as próprias respostas em suas próprias palavras.

**PERGUNTAS ABERTAS:** Perguntas que pedem aos respondentes que respondam com suas próprias palavras.

**PERGUNTAS COM MÚLTIPLAS FINALIDADES:** Perguntas que, de forma inadequada, pedem respostas para duas ou mais questões ao mesmo tempo.

**PERGUNTAS DE ASPECTOS GERAIS:** Perguntas nas quais o entrevistado é solicitado a acrescentar quaisquer informações, comentários ou opiniões referentes ao assunto do questionário que não foram necessariamente abordados no corpo principal dele.

**PERGUNTAS DE FILTRAGEM:** Perguntas que requerem que alguns entrevistados sejam eliminados de determinadas perguntas subsequentes ou da participação na pesquisa.

**PESQUISA:** Trabalho previamente planejado que segue determinados princípios básicos, através do qual a ciência busca obter respostas para solução de problemas preliminarmente identificados e estabelecidos.

**PERGUNTAS DE MÚLTIPLAS ESCOLHAS:** Perguntas que pedem aos respondentes para optar entre uma lista de mais de duas respostas.

**PERGUNTAS DE RESPOSTAS ESCALONADAS:** Perguntas de múltipla escolha nas quais as opções são designadas para captar a intensidade das respostas dos respondentes.

**PERGUNTAS DICOTÔMICAS:** Perguntas que pedem aos respondentes que escolham entre duas respostas.



**PERGUNTAS FECHADAS:** Perguntas que fornecem uma lista fixa de alternativas de resposta e pedem que o entrevistado selecione uma ou mais como indicativa da melhor resposta possível.

**PERGUNTAS FECHADAS:** Itens de questionário onde o respondente escolhe entre respostas padronizadas. Exemplo: Qual é o seu sexo? [ ] masculino [ ] feminino.

**PERGUNTAS FECHADAS:** Perguntas que pedem para que os respondentes que escolham entre duas ou mais respostas.

**POPULAÇÃO:** É o conjunto de todas as unidades que possuem característica em comum que desejamos medir ou inferir. Estas unidades podem ser pessoas, domicílios, bancos, universidades, etc.. Em muitas ocasiões o termo universo ou universo estatístico é utilizado no lugar de população. Repare que na definição de população não há nada que estabeleça que esta deva ser formada por uma grande quantidade de unidades, mas por todas as que estamos interessados em investigar, pode-se ter uma população com três elementos ou com um bilhão de unidades, por exemplo.

**PERÍODO DE LATÊNCIA:** Aquele compreendido entre exposição e manifestação da doença. É usado em relação às doenças infecciosas.

**PERIODICIDADE:** Padrão cíclico numa lista de unidades de amostragem, tais como o mesmo número de apartamentos em cada andar, o mesmo número de casas em cada quarteirão dentre outros. Se o período cíclico corresponder ao intervalo de amostragem numa amostra sistemática, pode-se selecionar uma amostra não-representativa, como selecionar apenas apartamentos perto do túnel de lixo ou somente casas na esquina.

**PERMUTABILIDADE:** Seja  $K=\{k_1, \dots, k_n\}$  uma permutação de  $\{1, \dots, n\}$ . Quantidades aleatórias  $X_1, \dots, X_n$  são permutáveis se as  $n!$  permutações  $(X_{k_1}, \dots, X_{k_n})$  têm a mesma distribuição  $n$ -dimensional. Em inglês Exchangeability.

**PERMUTAÇÃO:** Seja  $D$  um conjunto com  $d$  elementos chamamos de permutação a todo arranjo com  $d$  elementos, retirados de  $D$ . Por exemplo: i) Seja  $A$  um conjunto com os elementos  $\{a, b, c\}$ . As permutações de  $A$  são:  $\{(a,b,c); (a,c,b); (b,a,c); (b,c,a); (c,a,b); (c,b,a)\}$ ; ii) Quantos anagramas a palavra oba possui? As permutações da palavra dada são:  $\{(oba); (oab); (bao); (boa); (abo); (aob)\}$ . Para o cálculo de permutações por fatorial temos que antes definir fatorial:  $n! = n \cdot (n - 1) \cdot (n - 2) \cdot (n - 3) \dots$ . 3.2.1 Por exemplo, quantas são as possíveis formações de 5 pessoas em fila indiana? É determinado por cinco fatorial, isto é,  $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$ .

**PERMUTAÇÕES (ARRANJOS):** Número de maneiras diferentes de escolher  $j$  objetos de um grupo de  $n$  objetos, onde cada ordenação distinta dos objetos escolhidos é levada em conta.

**PÓS-MULTIPLICAÇÃO E PÓS-PRODUTO (POSTMULTIPLICATION AND POSTPRODUCT):** Em álgebra de matrizes ou álgebra linear, é o produto ou a multiplicação  $BA$  de duas matrizes quadradas  $A$  e  $B$ .

**POSIÇÃO (LOCATION):** O local em uma distribuição de probabilidades onde a maioria das observações pode ser encontrada. Comparar com dispersão.

**PRÉ-MULTIPLICAÇÃO E PRÉ-PRODUTO (PREMULTIPLICATION AND PREPRODUCT):** Em álgebra de matrizes, é o produto  $AB$  de duas matrizes quadradas  $A$  e  $B$ .

**PRECISÃO (PRECISION):** O nível de concordância entre um conjunto de medidas sobre o mesmo indivíduo. Também é usada na inferência bayesiana para indicar a quantidade igual a  $\frac{1}{\text{variância}}$ . Comparar com acurácia.

**PRIMEIRO EIXO PRINCIPAL (FIRST PRINCIPAL AXIS):** A nova variável, ou eixo, que resulta de uma análise de componentes principais e explica mais variabilidade nos dados que qualquer outro eixo. Também é chamado de primeiro componente principal.

**PROBABILIDADE (PROBABILITY):** A proporção de resultados em particular obtidos a partir de um dado número de tentativas, ou a proporção esperada para um número infinito de tentativas.

**PROBABILIDADE A POSTERIORI (POSTERIOR PROBABILITY):** A probabilidade de uma hipótese, conforme os resultados de um experimento e qualquer conhecimento prévio. O resultado do Teorema de Bayes. Comparar com probabilidade a priori.

**PADRONIZAÇÃO:** A transformação de uma variável aleatória normal que subtrai sua média e divide por seu desvio-padrão para gerar uma variável aleatória normal padrão.

**PARÂMETRO:** Uma grandeza desconhecida que pode variar sobre um conjunto de valores. Parâmetros ocorrem em distribuições de probabilidades e em modelos estatísticos, tais como modelos de regressão.

**PARÂMETRO DE LOCALIZAÇÃO:** Um parâmetro que define um valor central em uma amostra ou em uma distribuição de probabilidades. A média e a mediana são parâmetros de localização.

**PARES ASSOCIADOS:** Em um experimento fatorial fracionário, quando certos não podem ser estimados independentemente, eles são chamados de pares associados.

**PCR:** Uma razão de capacidade de processo com o numerador igual à diferença entre os limites de especificações do produto e o denominador igual a seis vezes o desvio-padrão do processo. É dito medir a capacidade potencial do processo porque a média do processo não é considerada. Veja capacidade de processo, Razão de capacidade de processo, Estudo de capacidade de processo e  $PCR_k$ . Algumas vezes denotada como  $C_p$  em outras referências.

**$PCR_k$ :** Uma razão de capacidade de processo com o numerador igual à diferença entre o alvo do produto e o limite de especificação mais próximos e o denominador igual à três vezes o desvio-padrão do processo. É dito medir a capacidade real do processo porque a média do processo é considerada. Veja capacidade de processo, razão de capacidade de processo, estudo de capacidade de processo e PCR. Algumas vezes denotada como  $C_{pk}$  em outras referências.

**PERCENTIL:** O conjunto de valores que divide a amostra em 100 partes iguais.

**PERMUTAÇÃO:** Uma sequência ordenada dos elementos em um conjunto usado para determinar o número de resultados em eventos e espaços amostrais.

**PLANEJAMENTO (OU EXPERIMENTO) COMPLETAMENTE ALEATORIZADO:** Um tipo de planejamento experimental em que os tratamentos ou fatores do planejamento são atribuídos às unidades experimentais em uma maneira aleatória. Em experimentos planejados, um planejamento completamente aleatorizado resulta correndo todas as combinações de tratamentos em uma ordem aleatória.



**PLANEJAMENTO COM BLOCOS COMPLETOS ALEATORIZADOS:** Um tipo de planejamento de experimentos em que tratamentos ou níveis do fator são atribuídos aleatoriamente a blocos.

**PLANEJAMENTO COMPOSTO CENTRAL (PCC):** Um planejamento de superfície de resposta de segunda ordem em k variáveis, consistindo em um fatorial com dois níveis, 2k corridas axiais e um ou mais pontos centrais. A porção fatorial com dois níveis de um PCC pode ser um planejamento fatorial fracionário, quando k é grande. O PCC é o planejamento mais amplamente usado para ajustar um modelo de segunda ordem.

**PLANEJAMENTO ORTOGONAL:** Veja ortogonal.

**PLANEJAMENTO ROTACIONÁVEL:** Em um planejamento de experimentos que foram desenvolvidos para trabalhar bem no ajuste de superfícies de resposta. Eles são geralmente planejamentos para ajustar um modelo de primeira ou segunda ordens. O planejamento composto central é um planejamento de superfície de resposta largamente utilizado.

**PONTO PERCENTUAL:** Um valor particular de uma variável aleatória, determinado a partir de uma probabilidade expressa como uma porcentagem. Por exemplo, o ponto 5 % superior da variável aleatória normal padrão é  $z_{0,05} = 1,645$ .

**POPULAÇÃO:** Qualquer coleção finita ou infinita de unidades ou objetos individuais.

**POSTO:** No contexto de dados, o posto de uma simples observação é seu número ordinal quando todos os dados são ordenadas de acordo com algum critério, tal como sua magnitude.

**POTÊNCIA:** A potência de um teste estatístico é a probabilidade de o teste rejeitar a hipótese nula quando a hipótese nula é, na verdade, falsa. Assim, a potência é igual a um menos a probabilidade do erro tipo II.

**PREVISÃO:** O processo de determinar o valor de uma ou mais grandezas estatísticas em algum ponto futuro no tempo. Em um modelo de regressão, prever uma resposta y para algum conjunto especificado de variáveis regressoras ou preditoras leva também a um valor previsto, embora possa não haver elemento temporal no problema.

**PROBABILIDADE A PRIORI (PRIOR PROBABILITY):** A probabilidade de uma hipótese antes de um experimento ser conduzido. Isso pode ser determinado pela experiência prévia, intuição, opinião de especialistas ou revisão da literatura. Comparar com probabilidade a posteriori.

**PROBABILIDADE COMBINADA DE FISHER (FISHER'S COMBINED PROBABILITY):** Metodologia usada para determinar se os resultados de múltiplas comparações são significantes, sem recorrer ao ajuste de taxa de erro por experimento. Ela é calculada como a soma dos logaritmos de todos os valores de P multiplicada por -2. Esse teste estatístico é distribuído como uma variável aleatória de quiquadrado com graus de liberdade igual a 2 vezes o número de comparações.

**PROBABILIDADE CONDICIONAL (CONDITIONAL PROBABILITY):** É a probabilidade de um evento compartilhado quando os eventos simples associados não são independentes. Se o evento B já aconteceu e a probabilidade de observar o evento A depende do resultado de B, então dizemos que a probabilidade de A é condicional a B. Representamos isso como  $P(A | B)$ , e a calculamos como:  $P(A | B) = \frac{P(A \cap B)}{P(B)}$ .

**PROBABILIDADE DE CAUDA (TAIL PROBABILITY):** A probabilidade de obter não apenas os dados observados, mas também possíveis dados mais extremos, mas que não foram observados. Também chamada de valor de P ou valor de probabilidade e calculada usando a função de distribuição cumulativa. Ver também valor de probabilidade.

**PROBABILIDADE EXATA (EXACT PROBABILITY):** A probabilidade de obter somente o resultado observado. Comparar com probabilidade de cauda.

**PROCESSOS DE MARKOV (MARKOV PROCESS):** Uma seqüência de eventos cuja probabilidade de um evento ocorrer depende somente da ocorrência de outro evento imediatamente predecessor.

**PRINCÍPIO DE VEROSSIMILHANÇA:** Esse princípio estabelece que a informação acerca de um modelo dado por um conjunto de dados está completamente contida na verossimilhança.

**PRINCÍPIOS BÁSICOS DA EXPERIMENTAÇÃO:** São conceitos teóricos utilizados no planejamento dos experimentos agrícolas, biológicos, industriais, zootécnicos, médicos, ecológicos, ensaios usando animais, etc. com a finalidade de que as conclusões obtidas se tornem válidas para a população de interesse. Estes princípios são a repetição, a casualização ou aleatorização e o controle local.

**PROBABILIDADE:** Uma medida numérica entre 0 e 1 atribuída a eventos em um espaço amostral. Números altos indicam que o evento é mais provável de ocorrer. Veja axiomas de probabilidade.

**PROBABILIDADE CONDICIONAL:** A probabilidade de um evento dado em que o experimento aleatório produz um resultado em outro evento.

**PROCESSO SEIS-SIGMA:** Originalmente usado para descrever um processo com a média no mínimo seis desvios-padrão a partir dos limites de especificação mais próximos. Agora, ele tem sido usado para descrever qualquer processo com uma taxa de defeito de 3,4 itens por milhão.

**PROPAGAÇÃO DO ERRO:** Uma análise de como a variância da variável aleatória, que representa a saída de um sistema, depende das variâncias das entradas. Uma fórmula existe quando a saída é uma função linear das entradas e a fórmula é simplificada se as entradas são consideradas independentes.

**PROJEÇÕES:** A projeção é um termo usado em econometria e análise de regressão que se refere-se à estimação do valor de uma variável dependente, dado o valor atual ou projetado da(s) variável(is) independentes (s). Quando a projeção é baseada em uma valor estimado ou projetado ao invés de um valor atual da variável independente, tem-se uma projeção condicional. A predição é, muitas vezes, utilizada alternativamente ou complementando a projeção. Outras vezes, a predição refere-se à estimação de um valor dentro da amostra da variável dependente. Projetar, então, refere-se à estimação de um valor futuro da variável dependente. Os erros de projeção surgem por causa das seguintes causas: i) a natureza aleatória do termo erro, ii) as estimativas não viesadas dos parâmetros serem iguais aos verdadeiros parâmetros somente na média, iii) erros na projeção das variáveis independentes, e iv) especificação incorreta do modelo. A variância do erro projetado,  $\sigma_u^2$ , é dada por:

Onde n é o número de observações e  $\sigma_u^2$  é a variância de u. Um estimador não viesado da variância projetada do erro,  $\hat{\sigma}_u^2$ , é dada por:

Onde  $\hat{u}$  é uma estimativa não viesada de dada por:

Quanto maior é  $n$ , menor é  $(ou)$ ,  $(ou)$ , e a diferença entre  $e$ . O valor de  $e$  é encontrado pela substituição do valor atual ou projetado de  $n$  na equação de regressão estimada. Isto é,  $e$ . O intervalo de confiança de 95% da projeção,  $e$ , é dado por:  $e$ . Onde  $t$  refere-se à distribuição  $t$  com  $n - 2$  graus de liberdade.

**PROJEÇÕES:** Uma projeção refere-se à estimação do valor da variável dependente,  $Y_p$ , dado o valor atual ou projetado da variável independente,  $X_p$ . A variância do erro projetado,  $\sigma_p^2$ , é dada por:

Onde  $n$  é o número de observações e  $\sigma_u^2$  é a variância de  $u$ . Desde que  $\sigma_u^2$  geralmente não é conhecido, usa-se  $s^2$  como uma estimativa não viesada de  $\sigma_u^2$ , tal que a variância estimada do erro projetado,  $s_p^2$ , é:

O intervalo de confiança de 95% para a projeção,  $Y_p$ , é:

Onde  $e$  e  $t$  refere-se à distribuição  $t$  com  $n - 2$  graus de liberdade.

**PROPRIEDADE DE ADITIVIDADE DA VARIÁVEL QUI-QUADRADO  $\chi^2$ :** Se duas variáveis aleatórias independentes  $X_1$  e  $X_2$  estiverem distribuídas como qui-quadrado, com graus de liberdade iguais a  $v_1$  e  $v_2$  respectivamente, então  $Y = X_1 + X_2$  é uma variável aleatória qui-quadrado, com  $u = v_1 + v_2$  graus de liberdade. Pode-se generalizar para qualquer número de variáveis aleatórias independentes qui-quadrados.

**PROPRIEDADE DE FALTA DE MEMÓRIA:** Uma propriedade de um processo de poisson. A probabilidade de uma contagem em um intervalo depende somente do comprimento do intervalo e não do ponto inicial do intervalo. Uma propriedade similar se mantém para uma série de tentativas de Bernoulli. A Probabilidade de um sucesso em um número especificado de tentativas depende somente do número de tentativas e não da tentativa inicial.

**PROPRIEDADE DE UNICIDADE DA FUNÇÃO GERADORA DE MOMENTO:** Refere-se ao fato de que variáveis aleatórias com a mesma função geradora de momento têm a mesma distribuição.

**PROPRIEDADE REPRODUTIVA DA DISTRIBUIÇÃO NORMAL:** Uma combinação linear de variáveis aleatórias normais independentes é uma variável aleatória normal.

**PROCESSO DE POISSON:** Um experimento aleatório com eventos que ocorrem em um intervalo e satisfaz as seguintes suposições. O intervalo pode ser particionado em subintervalos tal que seja zero a probabilidade de mais de um evento ocorrer em um subintervalo; a probabilidade de um evento em um subintervalo ser proporcional ao comprimento do subintervalo e o evento em cada subintervalo é independente dos outros subintervalos.

**PERTENCE A:** Símbolo matemático dado por  $\in$ .

**PESQUISA:** É a ação metódica para se buscar uma resposta; busca; investigação.

**PESQUISA:** Abordagem disciplinada ao estudo de fenômenos naturais. Busca sistemática, objetiva, e precisa de solução para um problema bem definido.

**PESQUISA:** A solicitação de informações verbais dos entrevistados pelo uso de várias técnicas de entrevista.

**PESQUISA (CARACTERÍSTICAS):** De modo geral, a pesquisa operacional pode ser caracterizada como a aplicação de métodos científicos, técnicas e instrumentos, a problemas envolvendo a operação de sistemas, de forma a fornecer àqueles no controle das operações, soluções otimizadas aos problemas.

**PESQUISA (RELEVÂNCIA):** A pesquisa relevante é aquela que é imediatamente útil, prática e relacionada com os problemas diários da sociedade.

**PESQUISA APLICADA:** É o tipo de estudo sistemático motivado pela necessidade de resolver problemas concretos.

**PESQUISA-AÇÃO:** Tipo de pesquisa aplicada onde o objetivo básico é o de solucionar, diferentes problemas coletivos no qual os cientistas e sujeitos envolvidos da pesquisa o façam de modo cooperativo e participativo.

**PESQUISA APLICADA:** Pesquisa que visa resolver determinado problema pragmático, uma melhor compreensão do mercado, uma determinação de porque uma estratégia ou tática falhou, uma redução da incerteza na tomada de decisões por parte da administração.

**PESQUISA APLICADA (RESULTADOS):** Os resultados de pesquisas aplicadas são: i) são voltadas para um grupo específico de pessoas numa situação específica onde eles podem ou não ser aplicados em outro lugar; ii) não precisam relacionar com uma área maior do conhecimento ou com outras pesquisas; iii) têm imediatas e óbvias implicações para a prática.

**PESQUISA (AUTO-CORREÇÃO):** A superioridade da pesquisa sobre outros métodos para obtenção de conhecimento está em que ela é auto-corretiva. Os erros e as inadequações serão revistos e corrigidos por outros pesquisadores quando: i) a pesquisa seja publicada; ii) o veículo ou revista científica seja o mais pertinente possível ao tópico de pesquisa; iii) a língua usada na publicação seja acessível ao maior número de cientistas que trabalham com o tópico da pesquisa; iv) os resultados apresentados sejam considerados interessantes pelo pequeno grupo de especialistas no tópico de pesquisa.

**PESQUISA AUTO-SELECIONADA:** Pesquisa em que os próprios entrevistados decidem se devem ser incluídos.

**PESQUISA AVALIATIVA:** Pesquisa para determinar a eficiência e a eficácia de programas específicos.

**PESQUISA APLICADA:** Modalidade de pesquisa que objetiva solucionar problemas ou necessidades concretas e urgentes. Nesse tipo de pesquisa, os problemas costumam ser definidos diretamente pela instituição ou órgão financiador público ou privado que financia a pesquisa. Opõe-se à pesquisa básica.

**PESQUISA BÁSICA:** Inclui investigação original para o avanço do conhecimento científico. A primeira meta do pesquisador é conseguir conhecimento ou entendimento mais completo do assunto em estudo, em vez de tentar fazer aplicações do novo conhecimento.

**PESQUISA CASUAL:** Pesquisa destinada a determinar se uma mudança numa variável provavelmente causou mudança observada em outra.

**PESQUISA BÁSICA:** Tipo de pesquisa cujo objetivo principal é o do avanço do conhecimento científico, sem haver preocupação com a aplicação imediata dos resultados a serem obtidos. Contrapõe-se à pesquisa aplicada.

**PESQUISA CORRELACIONAL:** Tipo de levantamento no qual o pesquisador não influencia ou tenta não influenciar nenhuma variável, mas apenas as mede e procura por relações ou correlações entre elas.

**PESQUISA, DELINEAMENTO DE (REQUISITOS):** Os requisitos do delineamento de uma pesquisa são: i) hipótese a priori; ii) critérios a priori que podem ser usados para medir; iii) isolamento e controle das variáveis sendo pesquisadas; iv) métodos de mensuração e verificação das variáveis em estudo.

**PESQUISA, PRINCÍPIOS DE:** Os princípios de uma pesquisa são os seguintes: i) busque conhecer tanto quando possível sobre o objeto de pesquisa; ii) prove, se possível, que o objeto existe na área a ser pesquisada; iii) use o mais eficiente método de detecção; iv) assegure-se de que verá o objeto de busca se o encontrar; v) assegure-se de que você não veria o objeto se ele não estivesse lá; vi) busque sistematicamente em vez de casuisticamente; vii) se possível, encontre um meio de determinar a direção e a distância aproximada em cada ponto de busca; viii) em problemas multidimensionais é comumente necessário encontrar um caminho unidimensional; ix) se possível, marque o ponto-de-partida e registre o caminho realmente seguido; x) use um procedimento convergente; xi) busque primeiramente no lugar mais provável; xii) distribua o tempo disponível, os recursos, e esforços em proporções adequadas e em regiões diferentes; xiii) leve em consideração a probabilidade de não perceber o objeto quando o encontrar; e xiv) leve em consideração qualquer efeito que o procedimento de busca possa ter sobre o objeto.

**PESQUISA CIENTÍFICA:** Processo que conduz a expansão e consolidação do conhecimento científico.

**PESQUISA BIBLIOGRÁFICA:** Termo usado para definir a investigação em que se utiliza apenas da análise documental.

**PESQUISA CIENTÍFICA (PRESSUPOSTOS):** Há certos pressupostos fundamentais para a pesquisa científica, os quais são: i) eventos obedecem às leis naturais, isto é, o universo é ordenado; ii) condições idênticas produzem resultados idênticos; e iii) válidas inferências podem ser feitas pelo processo de indução.

**PESQUISA CONFUNDIDA:** Pesquisa confundida ou inválida é aquela quando se observa que existem razões para crer que as diferenças obtidas na variável dependente teriam ocorrido, mesmo que a exposição à variável independente não tivesse sido manipulada.

**PESQUISA DESCRITIVA:** É aquela que se limita a uma descrição pura e simples de cada uma de suas variáveis, isoladamente sem que sua associação ou interação com as demais sejam examinadas.

**PESQUISA DE AÇÃO:** É um método analítico de solucionar um problema surgido no curso da ação educacional, com a colaboração dos agentes afetados pelo problema.

**PESQUISA DE CAMPO:** Qualquer pesquisa realizada em ambiente não controlado ou natural, ou seja, em campo, como por exemplo, pesquisa com plantas e animais em regime extensivo.

**PESQUISA DE CAMPO:** Termo usado com o significado de coleta de dados ou de inquérito.

**PESQUISA DE CAMPO:** Um método de pesquisa social que envolve a observação direta de fenômenos sociais em seus próprios ambientes. Uma forma de pesquisa qualitativa.

**PESQUISA DE COMPORTAMENTO DE CONSUMIDORES:** Estudo envolvendo observação de compradores ou consumidores, pessoalmente ou por meio de videotape, em diversificados ambientes de shopping centers.

**PESQUISA DE LEVANTAMENTO:** Pesquisa na qual o entrevistador interage com os respondentes para obter fatos, opiniões e atitudes.

**PESQUISA DE MARKETING:** Planejamento, coleta e análise de dados relevantes para as decisões de marketing, fazendo e comunicando os resultados dessa análise à administração.

**PESQUISA DE MARKETING:** Tipo de pesquisa aplicada que tem por objetivo a obtenção de dados de maneira empírica para solucionar problemas ou questões referentes ao mercado consumidor de produtos e serviços de uma companhia.

**PESQUISA DE OBSERVAÇÃO:** Pesquisa descritiva que monitora as ações dos respondentes sem interação direta.

**PESQUISA DE VIZINHOS PRÓXIMOS:** É o procedimento para seleção e localização de pontos vizinhos em relação ao ponto a ser interpolado. Está claro que a seleção deve levar em consideração a distância da amostra em relação ao ponto a ser interpolado, pois a sua influência é inversamente proporcional à sua distância. Porém, nem sempre os pontos de dados não estão dispersos regularmente sobre a área em estudo, fazendo com que a simples seleção de amostras mais próximas pode levar a uma amostragem enviesada.

**PESQUISA EDUCACIONAL:** Esforços planejados e sistemáticos para entender o processo educacional e, por meio de entendimento, aumentar sua eficiência.

**PESQUISA ELEITORAL DE ATAQUE:** Estilo de coleta de dados no qual zelosos cabos eleitorais satirizam um candidato para fazer com que os eleitores apóiem outro candidato.

**PESQUISA EPIDEMIOLÓGICA:** Aquela que se realiza por meio de produção de dados decorrentes de observação e eventual quantificação da doença ou eventos e eventual quantificação da doença ou eventos relacionados à saúde e os fenômenos correlatos, como os possíveis determinantes ou efeitos dos processos saúde-doença.

**PESQUISA DE MERCADO:** Técnica ou tipo de pesquisa utilizada para estudar a demanda de produtos, serviços ou opiniões de consumidor visando fornecer à instituição de interesse informações que auxiliem no planejamento e nas tomadas de decisões.

**PESQUISA ESTATÍSTICA:** É qualquer informação estatística retirada de uma população ou amostra, podendo ser por meio de censo ou amostragem.

**PESQUISA ETNOGRÁFICA:** Envolve pesquisa de campo e requer contextualização; a interpretação de resultados no contexto de coleta de dados. Provê descrições científicas de sistemas sociais, processos e fenômenos, em seus contextos específicos.

**PESQUISA EXPERIMENTAL, DELINEAMENTO CLÁSSICO:** Tipo de pesquisa onde há controle das condições experimentais por parte do pesquisador.

Tempo 1	Tempo 2	Tempo 3	Tempo 4
Randomização dos indivíduos			
Grupo 1	Preteste (O <sub>1</sub> )	Experimental	Posteste (O <sub>2</sub> )



<b>Grupo 2</b>	<b>Preteste</b> (O <sub>3</sub> )	<b>Controle</b>	<b>Posteste</b> (O <sub>4</sub> )
----------------	--------------------------------------	-----------------	--------------------------------------

**Há duas amostras randomizadas:** uma amostra com tratamento experimental (O<sub>1</sub> x O<sub>2</sub>) e uma, sem tratamento (controle) (O<sub>3</sub>, O<sub>4</sub>).

**PESQUISA EXPLICATIVA:** É aquela que procura estudar o nexos, a associação entre duas ou mais variáveis.

**PESQUISA DESCRITIVA:** Modalidade de pesquisa na qual o investigador descreve apenas o fenômeno observado, sem realizar inferência estatística de causa e efeito entre as características estudadas.

**PESQUISA EXPLORATÓRIA:** Pesquisa preliminar para esclarecer a natureza exata do problema a ser resolvido.

**PESQUISA EX POST FACTO:** Neste tipo de pesquisa não se tem controle direto sobre as variáveis independentes, pois elas já ocorreram ou não podem ser manipuladas. As inferências sobre relacionamentos entre a variável independente e a dependente não são feitas por intervenção direta, mas por variação concomitante. Situação de pesquisa em que as variáveis independentes já ocorreram. Neste tipo de pesquisa, o pesquisador começa com observações de uma variável dependente. Em seguida, faz um estudo retrospectivo de possíveis relacionamentos e efeitos. Pesquisa na qual o pesquisador não tem controle direto das variáveis independentes por que suas manifestações já ocorreram. O pesquisador não manipula quaisquer variáveis; tenta encontrar relacionamentos entre elas, já existentes no contexto.

**PESQUISA FUNDAMENTAL (CARACTERÍSTICAS):** i) Tipo de formulação de novas generalizações, princípios e teorias; ii) Grande importância aos procedimentos de amostragem como base para generalizações; iii) Preparação detalhada de um projeto, antes de se fazer a pesquisa, e estrita aderência a ele; iv) Relato suficientemente pormenorizado, de maneira a permitir a repetição do estudo; v) Frequentemente exige uso de estatística, amostragem, testes, e de procedimentos experimentais.

**PESQUISA GENÉTICA (CIÊNCIAS SOCIAIS):** Tipicamente este tipo de pesquisa é usado para definir padrões de crescimento ao longo do tempo. Entretanto, freqüentemente é reduzido a simples mensuração de status num dado momento. Quando aplicada para fins de detecção ou derivação de normas em grupo, tende a mascarar variações individuais. O estudo dessas variações é seu principal aspecto.

**PESQUISA HISTÓRICA:** É o processo de investigação crítica dos eventos passados para produzir uma descrição precisa e interpretação desses eventos. O fluxograma abaixo representa os passos metodológicos da pesquisa histórica.

**PESQUISA INTERDISCIPLINAR (CARACTERÍSTICAS):** i) Leva em consideração aspectos do assunto pertinentes a mais de uma disciplina científica; ii) Comumente se funda numa necessidade urgente, prática ou percebida, com o objetivo de uma aplicação; iii) O interesse da pesquisa está nas características dos sujeitos ou coisas pesquisados, e não numa população teórica; iv) O planejamento ou projeto inclui hipóteses e métodos sujeitos a modificações durante a execução; v) É importante o treinamento dos pesquisadores em dinâmica de grupo.

**PESQUISA E DESENVOLVIMENTO (P & D):** Refere-se ao trabalho que as instituições públicas e privadas realizam com o objetivo de desenvolver produtos e serviços.

**PESQUISA OPERACIONAL:** Aplicação de métodos científicos para a melhor gestão de sistemas organizados governamentais, comerciais e industriais. Distingue-se da engenharia operacional por focar sistemas nos quais o comportamento humano é importante. Estudo da eficácia do comportamento humano, para que se faça o melhor uso possível de recursos escassos para servir a determinados fins. Nele se combinam a observação, o experimento, a dedução e a indução. Seu objetivo é ajudar os diretores de indústria ou dos serviços públicos a tomar decisões.

**PESQUISA POR AMOSTRAGEM:** Ver amostragem.

**PESQUISA POR AMOSTRAGEM:** Pesquisa realizada entrevistando-se uma pequena parcela de uma grande população por meio de aplicação de um conjunto de procedimentos sistemáticos, científicos e ordenados, com a finalidade de fazer generalizações precisas a respeito dessa população.

**PESQUISA POR TELEFONE:** Coleta de informações por meio de entrevistas por telefone que envolvem um entrevistador treinado e entrevistados selecionados.

**PESQUISA EXPERIMENTAL:** Modalidade de pesquisa que tem por objetivo investigar as relações de causa e efeito entre as variáveis estudadas. Neste processo há o completo controle por parte do pesquisador das situações experimentais ou dos fatores denominados de tratamentos.

**PESQUISA PELO CORREIO:** Questionários impressos distribuídos pelo correio a uma amostra pre-determinada de entrevistados, que são solicitados a preencher o questionário e em seguida devolver este pelo correio ao pesquisador.

**PESQUISA PROGRAMÁTICA:** Consiste de um grupo de estudos interconectados de tal forma que levam à crescente profundidade e abrangência de teoria e do delineamento do estudo.

**PESQUISA PROGRAMÁTICA:** Pesquisa feita para desenvolver opções de marketing por meio de análise de oportunidade, atitudes do consumidor ou estudos sobre o uso de produtos.

**PESQUISA PURA OU BÁSICA:** Pesquisa que visa expandir as fronteiras do conhecimento e não resolver determinado problema pragmático.

**PESQUISA QUALITATIVA:** Dados de pesquisa não sujeitos à análise quantitativa ou qualitativa.

**PESQUISA QUANTITATIVA:** Estudos que usam análise matemática.

**PESQUISA SECUNDÁRIA:** Uma média dos dados coletados, que consiste em reunir e analisar os dados já obtidos e aproveitáveis.

**PESQUISA EX POST FACTO:** Tipo de pesquisa em que o pesquisador não utiliza observações experimentais próprias, isto é, estudos que usam dados obtidos de pesquisas anteriormente realizadas.

**PESQUISA SELETIVA:** Pesquisa para fazer a escolha entre várias alternativas viáveis identificadas pela pesquisa programática.

**PESQUISA HISTÓRICA:** Tipo de pesquisa muito comum usada na área das ciências sociais, cujo objetivo é estudar eventos, procedimentos e instituições do passado para investigar o seu efeito na atual população.

**PESQUISA PARTICIPANTE:** Tipo de pesquisa que utilizam técnicas como entrevistas, interação pesquisadores, extensionistas e agricultores com o objeto pesquisado.



**PESQUISAS POSTAIS AD HOC:** Questionário para determinado projeto enviados a nomes e endereços selecionados sem contato prévio com o pesquisador.

**PESQUISADOR SENIOR:** É um profissional capaz de definir e especificar um tema de pesquisa, comandar e administrar uma equipe, conduzir, terminar e redigir uma pesquisa considerada de bom nível dentro dos padrões usuais.

**PESO AO NASCER:** Peso de um animal por exemplo, tomado nas primeiras 24 horas de vida. Elevados pesos ao nascimento tendem a estar correlacionados com problemas de parto.

**PESO DISCRIMINANTE:** Peso cujo tamanho é determinado pela estrutura de variância das variáveis originais nos grupos da variável dependente. Variáveis independentes com grande poder discriminatório geralmente têm grandes pesos, e as que apresentam pouco poder discriminatório geralmente têm pesos pequenos; no entanto, a multicolinearidade entre as variáveis independentes provoca exceções a essa regra. É também chamado de coeficiente discriminante.

**PESSOA-TEMPO:** Unidade de tempo usada em estudos nos quais as pessoas são observadas por períodos desiguais. Por exemplo, quando uma pessoa é observada por 3 meses, outra por 14 meses e outra por 7 meses, a quantidade de tempo pode ser expressa como sendo de 24 pessoas-mês ou 2 pessoas-ano.

**PRODUTO ESCALAR (SCALAR PRODUCT):** Em álgebra de matrizes, é o produto de um vetor linha  $1 \times n$  e um vetor coluna  $n \times 1$ .

**PROPORÇÃO (PROPORTION):** Frequência relativa. O número de observações de uma determinada categoria, dividido pelo número total de observações. As proporções podem ser calculadas a partir de tabelas de contingência, mas a análise estatística dessa tabela é sempre baseada na frequência de contagem original.

**PROPORÇÃO DE VARIÂNCIA EXPLICADA (PVE) (PROPORTION OF VARIANCE EXPLAINED-PEV):** A quantidade de variabilidade nos dados explicada por cada fator da análise de variância. É equivalente ao coeficiente de determinação do modelo de regressão.

**PVE:** Ver proporção de variância explicada.

**PEV:** Variância do erro de predição dos valores genéticos.

**P IGUAL OU MENOR DO QUE 0,05:** Critério usualmente empregado para indicar que as diferenças são significativas; o acaso é explicação pouco provável para as diferenças encontradas; um fator ou conjunto de fatores pode ser estar determinado as diferenças.

**PESQUISA METODOLÓGICA:** Pesquisa em que a finalidade é estudar e desenvolver métodos e técnicas científicas, bem como equipamentos de investigação científica.

**PICTOGRAMA:** Um gráfico no quais os dados são representados por desenhos ou imagens.

**PICTOGRAMAS:** São construídos a partir de figuras representativas da intensidade do fenômeno. Este tipo de gráfico tem a vantagem de despertar a atenção do público leigo, pois sua forma é atraente e sugestiva. Os símbolos devem ser auto-explicativos. A desvantagem dos pictogramas é que apenas mostram uma visão geral do fenômeno, e não de detalhes minuciosos.

**PICTOGRAMAS:** São gráficos onde se utilizam figuras ou símbolos alusivos ao problema em estudo.

**PIRÂMIDE POPULACIONAL:** Diagrama de populações sob a forma piramidal, podendo representar grupos etários, proporções sexuais, locais de residências, estados civis e outras variáveis.

**PLACEBO:** De placere, agradar; procedimento ou produto inerte ministrado com fins sugestivos ou morais; influência psicológica que o profissional de saúde ou o tratamento, seja um produto ou procedimento, exerce sobre o paciente; em alguns trabalhos de investigação, de avaliação de uma intervenção, é o tratamento aplicado a indivíduos do grupo-controle e que ignoram se estão no grupo experimental ou controle, de modo a se investigar o valor real da intervenção, tornando possível descontinuar, do resultado final, o efeito inespecífico decorrente do mero ato de intervir.

**PLACEBO:** Agente farmacologicamente inativo dado a um paciente como substitutivo de um agente ativo, para garantir que a resposta do paciente é explicada pela droga e não pelo fato de se supor tratado. Conhecido também como tratamento testemunha ou controle.

**PLANEJAMENTO:** É uma tentativa de antecipar e ordenar decisões que terão de ser tomadas, visando atingir algum conjunto de objetivos específicos.

**PLANEJAMENTO:** É o estabelecimento de caminhos para alcançar os objetivos almejados.

**PLANEJAMENTO COMPLETAMENTE ALEATORIZADO:** Processo em um experimento segundo o qual cada elemento tem a mesma chance de pertencer a qualquer uma das diversas categorias ou tratamentos.

**PLANEJAMENTO EM BLOCOS ALEATORIZADO:** Planejamento em que se obtém uma medida para cada tratamento em cada um dos vários indivíduos emparelhados de acordo com características similares.

**PLANEJAMENTO RIGOROSAMENTE CONTROLADO:** Planejamento de experimento em que todos os fatores são forçados a ser constantes para que os efeitos de fatores estranhos sejam eliminados.

**PLANEJAMENTO:** Como levantar informações, que dados deverão ser obtidos, qual levantamento a ser utilizado, o cronograma de atividades, os custos envolvidos, dentre outros. É uma das fases do método estatístico.

**PLANIFICAÇÃO DO PROBLEMA:** Consiste na determinação de um processo para resolver o problema e, em especial, como obter informações sobre a variável em estudo.

**PLANO COMPLEXO DE AMOSTRAGEM:** É o que envolve mais de um processo de amostragem.

**PLANO DE AMOSTRAGEM MÚLTIPLA:** Uma forma de amostragem de aceitação na qual mais de uma amostra ou estágio é usado. Com base no número de itens defeituosos encontrados na amostra, uma decisão será tomada de aceitar o lote, rejeitar o lote, ou continuar a amostragem.

**PLOTAGEM DE RETÂNGULOS:** Um sumário gráfico de dados. Um retângulo, desenhando a partir do primeiro e do terceiro quartis, mostra a posição de 50 % dos dados do meio. Linhas pontilhadas, chamadas costeletas, estendendo-se a partir das extremidades da caixa, mostram a posição dos valores de dados maiores do que o terceiro quartil e menores do que o primeiro quartil. As posições dos pontos fora da curva também são assinaladas.

**P MAIOR QUE 0,05:** Critério usualmente empregado para indicar para as diferenças não são significativas; o acaso pode ser a explicação para as diferenças encontradas.

**PODER:** Probabilidade de rejeitar corretamente a hipótese nula quando a mesma é falsa, ou seja, de encontrar corretamente um suposto relacionamento quando ele existe. Determinando como uma função i) do nível do nível de significância estatística (alfa) dado pelo pesquisador para um erro tipo I, ii) do tamanho da amostra utilizada na análise e iii) do tamanho do efeito examinado.

**PODER:** Probabilidade de uma relação significativa ser encontrada se ela realmente existir. Complementa o nível de significância alfa ( $\alpha$ ), mais amplamente usado.

**PODER:** Probabilidade de identificar um efeito de tratamento quando ele realmente existe na amostra. O poder é definido como  $1 - \beta$  (ver beta). É determinado como uma função de i) nível de significância estatística ( $\alpha$ ) dado pelo pesquisador para um erro Tipo I, ii) tamanho da amostra usada na análise e iii) tamanho do efeito em exame.

**PODER DE UM TESTE ESTATÍSTICO:** É a probabilidade dada por  $1 - \beta$ , de que o pesquisador rejeite uma hipótese de nulidade  $H_0$  quando na realidade ela for falsa.

**PODER DO TESTE:** Denomina-se poder de um teste estatístico a probabilidade de aceitar a hipótese nula quando ela é verdadeira. A probabilidade de se cometer erro do tipo II, isto é, aceitar a hipótese nula quando ela é falsa é representada por  $\beta$ , então o poder do teste será determinado por  $1 - \beta = P(\text{aceitar } H_0 / H_0 \text{ é verdadeira})$ , isto é o poder do teste é definido como a probabilidade de não se cometer erro do tipo II. Em inglês **power of a statistical test**.

**PODER DO TESTE:** Capacidade de um teste estatístico de rejeitar a hipótese de nulidade quando de fato ela é falsa.

**PODER DO TESTE:** Corresponde a  $1 -$  probabilidade de rejeitar a hipótese nula quando esta é falsa e indica a probabilidade de decisão correta baseada na hipótese alternativa. Geralmente é interpretado como a chance de detectar uma real diferença entre as médias ou proporções. Por exemplo, um poder de 80 % significa que, se de fato houver alguma diferença, haverá uma probabilidade de 80 % de detectá-la.

**PODER DO TESTE: Probabilidade de rejeitar uma hipótese nula falsa.**

**PODER ESTATÍSTICO:** A probabilidade de não se cometer um erro tipo II.

**PODER ESTATÍSTICO DO TESTE (OU POTÊNCIA ESTATÍSTICA DO TESTE):** A probabilidade de detectar um resultado significativo, se ele de fato existe. Ver erro tipo II.

**PODER DO TESTE DE HIPÓTESE:** É probabilidade de se rejeitar uma hipótese  $H_0$  quando ela for na realidade falsa, o qual é dado por  $1 - \beta$ , onde beta ( $\beta$ ) é o erro do tipo II.

**POISSON ( $\lambda$ ):** Distribuição de Poisson, ou modelo para descrever eventos raros.

**POLÍGONO DE FREQUÊNCIA:** Diagrama que mostra a forma da distribuição de frequências; as frequências são representadas como ordenadas contra os valores da variável como abscissas e os topos vizinhos das ordenadas unidos. Em inglês Frequency polygon.

**POLÍGONO DE FREQUÊNCIA:** Diagrama que ilustra uma distribuição de frequência por meio de segmentos retilíneos em lugar de barras.

**POLÍGONO DE FREQUÊNCIAS:** Representação gráfica de uma distribuição de frequências resultante da união das intersecções dos pontos médios das classes nas abscissas e das frequências respectivas nas ordenadas. Em inglês frequency polygon.

**POLÍGONO DE FREQUÊNCIA:** Representação gráfica de uma distribuição de dados por meio de segmentos retilíneos unidos.

**POLÍGONO DE FREQUÊNCIA (GRÁFICO):** Instrumento gráfico somente para dados de intervalos. Uma linha liga os pontos que representam o ponto médio do intervalo de classe no eixo horizontal e a frequência de resposta no eixo vertical.

**POLÍGONO DE FREQUÊNCIA ACUMULADA:** É traçado marcando-se as frequências acumuladas sobre perpendiculares ao eixo horizontal, levantadas nos pontos correspondentes aos limites superiores dos intervalos de classe.

**POLÍGONO DE FREQUÊNCIAS:** São gráficos com aspecto de linhas quebradas. Constroem-se unindo por segmentos de reta os pontos médios das bases superiores dos retângulos de um histograma.

**POLIMÓRFICAS, CURVAS:** Diz-se das curvas de índice de sítio, quando as curvas não possuem uma relação de proporcionalidade constante entre si. As curvas polimórficas que não se interceptam são chamadas de disjuntas quando a curva de pelo menos um dos sítios intercepta a curva de outro, as curvas são chamadas polimórficas não-disjuntas.

**POLINÔMIO:** Transformação de uma variável independente para representar uma relação curvilínea com a variável dependente. Incluindo um termo quadrado ( $X^2$ ), um único ponto de inflexão é estimado. Um termo cúbico estima um segundo ponto de inflexão. Termos adicionais de potência superior também podem ser estimados.

**PONDERAÇÃO:** Procedimento usado em conexão com amostragem, por meio de qual as unidades selecionadas com probabilidades desiguais recebem pesos, de forma a tornar representativa a amostra selecionada.

**PONDERADORES:** São os pesos resultantes de uma técnica de estimativa também krigagem ordinária ou inverso da potência da distância, aplicados na equação da média ponderada para cálculo do teor médio ou variância.

**PONTO AMOSTRAL:** Geralmente definido como o centro de uma parcela de área variável como, por exemplo, no método de Prodan, ou como o centro de uma unidade amostral não-superficial, como no método de quadrantes e na amostragem por enumeração angular. Um ponto amostral pode ser também um local onde se faz uma avaliação qualitativa, como por exemplo classificar a ecounidade de um ponto na floresta.

**PONTO AMOSTRAL:** É o resultado de um experimento. O mesmo que evento elementar ou simples. Conhecido também como ponto amostra. Em inglês outcome.

**PONTO AMOSTRAL:** Um elemento de um espaço amostral. Um ponto amostral representa um resultado experimental. Em inglês outcome.

**PONTO AMOSTRAL:** Elemento do espaço amostral. Em inglês outcome.

**PONTO DE ALAVANCA:** Tipo de observação influente definido por um aspecto da influência chamado de alavanca. Essas duas observações são substancialmente diferentes em uma ou mais variáveis independentes, de modo que afetam a estimação de um ou mais coeficientes de regressão.

**PONTO DE ALAVANCA:** É o ponto que indica que uma observação se destaca das demais no espaço das variáveis explanatórias ou independentes. Cabe ressaltar que um ponto de alavanca não tem que ser necessariamente um outlier de regressão. Quando um ponto de alavanca é também um outlier de regressão, ou seja, quando um ponto é um valor discrepante tanto dentre os valores da variável dependente quanto dentre os valores das variáveis independentes, terá grande influência sobre os ajustes clássicos. Sendo assim esta grande influência reforça a ideia da busca de um método que seja robusto em relação a um conjunto de dados que seja contaminado, isto é, que possua valores aberrantes. Em inglês leverage.

**PONTO DE ALAVANCAGEM:** Uma observação que tem um impacto substancial sobre os resultados da regressão devido às suas diferenças de outras observações em uma ou mais das variáveis independentes. A medida mais comum de um ponto de alavancagem é o valor de projeção, contido na matriz de projeção.

**PONTO DE BITTERLICH:** Amostragem por enumeração angular.

**PONTO DE CORTE:** Determinação do melhor ponto de corte para um teste diagnóstico, ou seja, aquele que mais se aproxime do chamado Padrão Ouro. Em inglês Cutoff point.

**PONTO DE CORTE:** O truncamento artificial de um processo de amostragem em um ponto em que fica aparente a existência de dados suficientes para o objetivo em vista. Em inglês Cutoff.

**PONTO DE INFLUÊNCIA:** Ponto que afeta fortemente o gráfico de uma reta de regressão.

**PONTO DE MASSA EM ZERO:** Algumas distribuições de variáveis contínuas só valem para valores maiores a zero, porém, em situações reais este valor pode ser observado com uma frequência significativamente diferente de zero, neste caso se diz que o zero tem massa de probabilidade. Em inglês Mass point at zero.

**PONTO DE RUPTURA:** É uma medida global de robustez. Ele mede qual seria a maior porcentagem de contaminação que um estimador poderia suportar e ainda assim fornecer informação confiável sobre o parâmetro considerado. No contexto de regressão, a versão amostral finita do ponto de ruptura significa a menor fração de outliers dados atípicos com o potencial de distorcer por completo as estimativas. Um estimador robusto adequado então, é aquele que possui ponto de ruptura igual a 0,50, ou seja, é o estimador que resiste a um percentual de até 50 % dos dados contaminados. Um valor de ponto de ruptura maior que 0,50 não faria sentido pois se mais da metade das observações fosse de valores discrepantes, não seria mais possível diferenciar quais seriam os dados ruins.

**PONTO FORA DA CURVA:** Um valor de dado extraordinariamente grande ou pequeno.

**PONTO IDEAL:** Ponto em um mapa perceptual que representa a combinação mais preferida de atributos percebidos de acordo com os respondentes. Uma suposição importante é que a posição do ponto ideal

relativa aos outros objetos no mapa perceptual definiria a preferência relativa, de modo que objetos mais distantes do ponto ideal deveriam ser menos preferidos.

**PONTO MÉDIO:** Semi-soma do maior e menor valor. Ou ainda é a média aritmética entre os limites inferior e superior de um determinado intervalo de classe, numa distribuição de frequências. Ou seja:

$$P.M. = X_i = \frac{L_i + L_s}{2}$$

**PONTO MÉDIO:** Ponto interior de uma classe, equidistante dos seus limites reais. Seu valor é igual à metade da soma desses limites. Em inglês central point.

**PONTO MÉDIO:** O ponto médio de uma classe.

**PONTO MÉDIO DE CLASSE:** É o ponto que divide o intervalo de classe em duas partes iguais.

**PONTOS INFLUENTES:** São pontos que se destacam dos demais em uma análise de regressão, e exercem efeito no ajuste da reta de quadrados mínimos.

**PONTOS MÉDIOS:** Pontos médios das classes em uma tabela de frequências.

**POPULAÇÃO:** É um conjunto de seres com uma dada característica em comum e com interesse para o estudo.

**POPULAÇÃO:** A totalidade de todos os elementos, sujeitos, ou membros que possuem um especificado conjunto de uma ou mais características comuns que a define. Em estatística inferencial, a população é o grupo referente ao qual se faz inferência.

**POPULAÇÃO:** Conjunto de todos os elementos de interesse.

**POPULAÇÃO:** Totalidade dos indivíduos que apresentam uma característica de interesse. Em inglês Population.

**POPULAÇÃO:** Grupo de pessoas, objetos ou fatos que se pretende estudar, quase sempre por meio de uma amostra. Diz-se também universo. Em inglês population.

**POPULAÇÃO:** Toda questão de pesquisa define um universo de objetos aos quais os resultados do estudo deverão ser aplicados. A população alvo, também, chamada população estudada, é composta de elementos distintos possuindo um certo número de características comuns, pelo menos uma. Essa característica comum deve delimitar inequivocamente quais os elementos que pertencem à população e quais os que não pertencem. Estes elementos, chamados de unidades populacionais são as unidades de análise sobre as quais serão recolhidas informações. Por exemplo, todos os alunos de graduação que estão matriculados em alguma disciplina de estatística na Universidade Federal Rural do Semi-árido (UFERSA), em Mossoró, RN, durante o ano de 2007.

**POPULAÇÃO (AMOSTRAGEM):** No ponto de vista de amostragem existem dois conceitos de população: i) População Alvo é o objeto de estudo, podendo ser uma população biológica ou uma área de estudo, ii) População Estatística é um conjunto matemático, formado por valores numérico ou categóricos, relativo a todas as unidades amostrais. O tamanho da população estatística pode ser finito ou infinito.



**POPULAÇÃO:** É a família de medidas de uma característica específica obtida de todas as unidades de amostragem possíveis de serem selecionadas do universo. A população representa um conjunto de medidas de um único atributo de um universo. Assim pode-se ter várias populações de medidas de diferentes atributos de um mesmo universo..

**POPULAÇÃO OU UNIVERSO:** O grupo total de pessoas das quais as informações são necessárias.

**POPULAÇÃO:** Coleção completa e total dos elementos a serem estudados. Em inglês population.

**POPULAÇÃO:** Conjunto de elementos de interesse com pelo menos uma característica em comum. Em inglês population.

**POPULAÇÃO:** Especificação do universo a ser amostrado. Por exemplo, podemos especificar o universo brasileiros como todos os adultos brasileiros morando em residências a partir de 1º de agosto de 2000. Em inglês population.

**POPULAÇÃO:** Qualquer conjunto de unidade como por exemplo, residentes em um estado, trabalhadores rurais, clientes e recém-nascidos. Em estatística, tem o sentido de universo, cujas características são investigadas, em geral, pelo estudo de amostras. Em epidemiologia, este mesmo conceito é mantido. É importante diferenciar dois tipos de população, exemplificados a seguir, em estudos de mobilidade, de modo a julgar a possível presença de um viés de seleção: a população usualmente definida em bases territoriais, como em um município ou bairro, e a mobilidade preparada a partir de residentes de uma localidade é dita populacional ou em base populacional; nela estão representados os sadios e os doentes, sejam eles clientes ou não de um profissional de saúde. A de uma instituição de saúde, a morbidade observado entre os indivíduos que demandam por atendimento em estabelecimento de saúde é dita institucional; nestes casos, há um filtro de seleção, que faz com que alguns doentes se tornem clientes e outros não; para certos propósitos e generalizações, as estatísticas assim preparadas são viciadas.

**POPULAÇÃO:** É a conjunto de todas as unidades que possuem pelo menos uma característica em comum que desejamos medir. Estas unidades podem ser pessoas, domicílios, bancos, universidades, etc. Em muitas ocasiões o termo universo é utilizado no lugar de população. Repare que na definição de população não há nada que estabeleça que esta deve ser formada por uma grande quantidade de unidades, mas por todas as que estamos interessados em investigar.

**POPULAÇÃO OU UNIVERSO:** Conjunto de todos os indivíduos, itens, eventos, membros ou unidades, circunscritos em uma área e em determinado tempo. O tamanho da população é simbolizado pela letra N.

**POPULAÇÃO ABERTA:** É aquela que se renova ao mesmo tempo pela natalidade e por contingentes de imigração, e se reduz pela mortalidade e por contingentes de emigração. Opõe-se a população fechada.

**POPULAÇÃO ABERTA:** É o universo que altera seu tamanho durante o período da pesquisa, em decorrência de nascimentos, óbitos e fenômenos migratórios.

**POPULAÇÃO AUSENTE:** De uma dada circunscrição territorial, é a constituída pelo total das pessoas que possuem domicílio habitual ou legal dentro de suas divisas mas que se acham fora dele no instante do recenseamento. Opõe-se a população presente.

**POPULAÇÃO DE REFERÊNCIA:** População hipotética, para qual os dados da investigação são generalizados; por exemplo, todas as gestantes.

**POPULAÇÃO DO SURVEY:** Lista ou quase lista específica de elementos dos quais se seleciona uma amostra de survey. Exemplo: nossa população pode ser por exemplo eleitores em São Paulo em 2006, mas nossa população do survey dos últimos se seleciona de fato a amostra pode ser, por exemplo, eleitores em São Paulo que constam da lista de eleitores registrados em 18 de setembro de 2006. Em última análise, as amostras que selecionamos são representativas das populações do survey das quais as amostras são tiradas.

**POPULAÇÃO EFETIVA:** É a que tem existência atual, em face de um dado critério de inclusão. São finitas. Exemplo: o conjunto dos brasileiros do sexo masculino, maiores de 21 anos. Opõe-se a população hipotética.

**POPULAÇÃO EXPERIMENTAL:** População acessível, que serve de objeto à investigação; por exemplo, as gestantes de um hospital.

**POPULAÇÃO FECHADA:** É a que não sofre movimento migratório. Opõe-se a população aberta.

**POPULAÇÃO FECHADA:** É o universo que não altera seu tamanho durante o período de investigação.

**POPULAÇÃO FINITA:** População de indivíduos que são finitos em número, ou ainda onde o número de elementos da amostra (n) corresponde a pelo menos 5 % do número de elementos da população (N) sob estudo, isto é que  $\frac{n}{N} \geq 0,05$ . Em inglês Finite population.

**POPULAÇÃO HIPOTÉTICA:** É a que, não tendo existência atual, pode ser concebida como resultante de um dado processo. Pode ter tamanho infinito. Exemplo: a população dos pés de milho tratados pelo adubo. Opõe-se a população efetiva.

**POPULAÇÃO INFINITA:** Tipo de população que possui a propriedade de ser infinita para algum processo limitador ou, em um contexto de amostragem, pode possuir a propriedade para alguma estratégia de amostragem, ou ainda onde o número de elementos da amostra (n) corresponde a menos de 5 % do número de elementos da população (N) sob estudo, isto é que  $\frac{n}{N} < 0,05$ . Em inglês Infinite population.

**POPULAÇÃO PADRÃO:** É a distribuição de idades em uma população fixa, distribuição essa tomada como base para o cálculo de vários coeficientes ajustados.

**POPULAÇÃO PRESENTE:** De uma dada circunscrição territorial é a constituída pelo total das pessoas que possuem domicílio habitual ou legal dentro de suas divisas e que se acham ali no instante do recenseamento.

**POPULAÇÃO TOTAL:** A população teórica para qual o pesquisador quer generalizar as constatações do estudo.

**POPULAÇÃO ÚTIL:** Uma definição operacional da população geral que é representativa e a partir da qual o pesquisador é capaz de identificar como completa uma lista de seus membros.

**PORTANTO:** Símbolo matemático dado por ∴.

**PORQUE:** Símbolo matemático dado por ∴.

**PÓS-CODIFICAÇÃO:** O processo de codificar respostas a perguntas abertas ou outras perguntas que não foram codificadas no processo de pré-codificação.

**PÓS-TESTE:** Experimentos, medidas da variável dependente depois da administração do estímulo ao grupo experimental.



**POSTO:** Posição numérica de um elemento em um conjunto de dados ordenados.

**POSTO DA MATRIZ X:**  $r(X)$ , o mesmo que rank da matriz.

**POSTULADOS DE EUCLIDES:** i) Dois pontos quaisquer podem ser unidos por um segmento de reta. ii) Qualquer segmento de reta pode ser estendido indefinidamente em uma linha reta. iii) Dado qualquer segmento de reta um círculo pode ser traçado tendo como raio o segmento e como centro um dos pontos finais. iv) Todos os ângulos retos são congruentes. v) Se duas linhas são desenhadas de tal maneira a interceptar uma terceira e de tal forma que a soma dos ângulos internos em um lado é menor do que dois ângulos retos, então as duas linhas inevitavelmente devem se interceptar naquele lado se estendidas o suficiente. O quinto postulado também pode ser formulado como: para cada linha  $l$  e para cada ponto  $P$  que não pertence à linha, existe uma única linha  $m$  que passa por  $P$  e é paralela a  $l$ . Em inglês Euclid's postulates.

**POTÊNCIA ESTATÍSTICA DO TESTE:** O mesmo que poder estatístico.

**POVOAMENTO:** É uma área contígua que contém um número de árvores relativamente homogêneo ou tem uma série de características em comum. Os povoamentos florestais são caracterizados pela espécie puros ou mistos, pela idade coetâneo ou disetâneo e pelo regime de manejo alto fuste e talhadia.

**Pr:** Em estatística é a abreviatura para probabilidade.

**PRECISÃO:** Refere-se ao grau de dispersão existente em um conjunto de dados. A precisão é quantificada como a recíproca da variância. Em inglês Precision.

**PRECISÃO:** O mesmo que reprodutibilidade.

**PRECISÃO:** Indica a dispersão dos resultados em torno de um valor de referência, tomado como correto. É em geral associada ao desvio padrão ou erro padrão. É, portanto, um conceito intimamente ligado ao de incerteza, pois à medida que a incerteza aumenta a precisão se degenera. A precisão está, também, intimamente associada à subdivisão de escala e/ou sensibilidade dos aparelhos. Se não for possível ter mais algarismos significativos a precisão fica limitada e, em certos casos, leituras constantes ou invariáveis não indicam o mesmo valor da grandeza medida. A idéia da precisão fica associada também aos conceitos de repetibilidade e reprodutibilidade, que são casos especiais de precisão, de tipo interlaboratorial.

**PRECISÃO:** Capacidade de um instrumento fornecer o mesmo resultado ou um resultado muito semelhante em mensurações repetidas.

**PRECISÃO ABSOLUTA:** É a precisão especificada diretamente e na mesma unidade da estimativa que se pretende calcular.

**PRECISÃO RELATIVA:** É a precisão especificada não diretamente como precisão absoluta, mas sim proporcionalmente como porcentagem em relação ao verdadeiro valor.

**PRECISÃO AMOSTRAL:** É a proximidade entre os valores das estatísticas obtidas de várias amostras do mesmo tamanho e da mesma população.

**PRÉ-CODIFICAÇÃO:** A determinação de códigos numéricos a cada categoria de resposta no momento em que o questionário é preparado na sua forma final de aplicação.

**PREDIBILIDADE:** É o critério do cientista para validade. Se nossas formulações teóricas do que causa suicídios, tempestades, secas, ondas de crime e depressões econômicas estão corretas, devemos ser capazes de prever onde e quando esses fenômenos ocorrerão, e o número de predições corretas devem definitivamente depender do acaso.

**PREDIÇÃO:** Determinação do valor de uma variável aleatória para uma ou mais observações com base em um modelo probabilístico. Exemplos de predição: altura de árvores individuais por meio de relação hipsométrica, volume de árvores individuais por meio de equação de volume, produção atual de um povoamento por meio de uma equação de produção, produção futura ou passada de um povoamento por meio de um modelo de crescimento. Conhecido também como estimação.

**PREDIÇÃO:** A estimativa de escores em uma variável a partir de informações sobre uma ou mais variáveis. É a estimativa de uma variável a partir do conhecimento de outra. Aumenta-se a segurança da predição à proporção que a correlação entre a variável preditora e a variável critério aumenta.

**PREDIÇÃO ESTATÍSTICA:** Geralmente é uma conversão da medida de associação em uma equação de regressão pela qual uma variável dependente pode ser predita a partir de uma ou mais variáveis independentes. Quanto mais alto o grau de associação mais adequada é a predição.

**PREDITOR LINEAR ÓTIMO:** Preditor de valores futuros de um processo estocástico em forma de combinação linear de observações passadas e que minimiza o erro médio quadrático de previsão. Em inglês Optimum linear predictor.

**PREDITOR LINEAR:** Expressão da forma  $\sum_j x_j \beta_j$  usada para prever alguma quantidade desconhecida. Em inglês Linear predictor.

**PRECISÃO:** Conceito usado em uma pesquisa para indicar o nível de erro ou afastamento do valor estimado em relação ao parâmetro populacional que se está investigando através de resultados obtidos a partir da análise de uma amostra. Por exemplo, em uma pesquisa de opinião pública, num estudo estatístico fixado o tamanho amostral, os termos confiança, fiducial ou ainda fidedignidade e precisão são conceitos opostos. Não é mesmo que exatidão.

**PREFERÊNCIA:** Implica que os Objetos são julgados pelo respondente em termos de relações de domínio; ou seja, os estímulos são ordenados em preferência com relação a alguma propriedade. Ordenação direta, comparações aos pares e escalas de preferência freqüentemente são usadas para determinar preferências de respondentes.

**PREMISSAS:** São proposições que vão servir de base para uma conclusão.

**PRESSUPOSTO:** Termo que indica as condições a priori utilizadas na análise de dados na área de estatística, sendo um conceito que mostra uma condição dada como certa ou pelo menos, quase garantida.

**PRESSUPOSTO DE INDEPENDÊNCIA:** Também significa que os erros, resíduos ou desvios devidos ao efeito de fatores não controlados devem ser independentes. Isto implica em que os efeitos dos tratamentos não sejam correlacionados. Significa também que o resultado de uma observação não interfere no resultado de outra observação. Quando os fatores experimentais ou tratamentos referem-se a doses crescentes este pressuposto não é atendido, ocasião em que a análise da variância (ANOVA) deve ser feita seguindo o ajuste de modelos de regressão.

**PRESSUPOSTOS PARAMÉTRICOS:** i) A mensuração da variável dependente, a variável cujos dados estão sendo analisados, está, pelo menos, numa escala intervalar. ii) Os escores que que é a variável dependente são selecionados de uma distribuição populacional normalmente distribuída. Este pressuposto é exigível somente quando a amostra é pequena, isto é, constituída de menos de 30 elementos. iii) Quando se estudam duas ou mais populações, elas devem ter variância homogênea.

**PRÉ-TESTE:** Medição de variáveis dependentes que são realizadas antes do experimento propriamente dito, com a finalidade de dar suporte à condução do ensaio em si. Por exemplo, a aplicação de um teste de germinação, envelhecimento precoce ou de quebra de dormência em pesquisas aplicadas com sementes na agronomia, ou ainda o uso de amostra piloto em pesquisas ou levantamentos amostrais para determinação de tamanho ou dimensionamento amostral.

**PSEUDO VARIÂNCIA:** Seja  $Y$  uma variável aleatória com função de distribuição  $G$ , seja  $\Theta$  a função de distribuição gaussiana padrão, e

$0 < p < \frac{1}{2}$ . A pseudo variância a  $px100\%$  de  $Y$  defini-se por:  $v(p) = \left( \frac{G^{-1}(1-p) - G^{-1}(p)}{\Phi^{-1}(1-p) - \Phi^{-1}(p)} \right)^2$ . Nesta expres-

são,  $v(p)$  é o quadrado do quociente de uma dispersão inter-quantis da distribuição de  $Y$  pela dispersão inter-quantis correspondente para a distribuição gaussiana padrão. Observe que se  $Y$  tiver distribuição normal ou gaussiana então  $v(p) = \text{var}(Y)$  para todo  $p$ . Para distribuições simétricas de caudas pesadas,  $v(p)$  tende a crescer quando  $p \rightarrow 0$ . Assim se  $v(p)$  for aproximadamente constante para diversas escolhas de  $p$ , tem-se uma indicação de que  $Y$  é aproximadamente normal ou gaussiana.

**PRESSUPOSTOS PARA PESQUISA:** Suposições ou hipótese que devem atendidas tais como: i) Eventos obedecem às leis naturais, isto é, o universo é ordenado; ii) Condições idênticas produzem resultados idênticos; iii) Válidas inferências podem ser feitas pelo processo de indução.

**PANORAMA GERAL:** Consultar metanálise.

**PARÂMETROS:** Uma característica da população, como média, mediana, moda ou desvio padrão, que indica uma característica específica de uma distribuição.

**PAREAMENTO ARTIFICIAL:** Dois animais diferentes que foram pareados quanto a quaisquer variáveis que possam influenciar a resposta.

**PAREAMENTO NATURAL:** Irmãos de ninhada, ou alguma outra associação biológica, representam o material experimental.

**PASSO DE AMOSTRAGEM:** Termo usado na amostragem probabilística sistemática para indicar o salto que se dá em um determinado número  $k$  de elementos da amostra, quando se faz a retirada dos elementos da população ordenada para compor a amostra. Vale lembrar que a técnica de amostragem probabilística sistemática é aquela em que a metodologia de aplicação e condução é a seguinte: Considere uma população com  $N$  elementos onde  $N = kn$ , e onde  $k$  é um número inteiro. Considere também que a população está ordenada de 1 a  $N$ , formando o sistema de referências, e  $n$  o tamanho da amostra desejada, define-se a quantidade  $K = \frac{N}{n}$ , a que se dá o nome de intervalo de amostragem, passo ou intervalo de seleção e

admitindo-se que  $K$  seja um número inteiro, faz-se então um sorteio ente os números 1, 2, 3, ...,  $K$ , podendo ser obtido, por exemplo, o valor  $i$ , que será chamado início casual. Nestas condições com apenas este

sorteio, toda a amostra fica selecionada, sendo composta das unidades amostrais que tenham recebido os seguintes números ou ocupam a posição correspondente.

$i$   
 $i + K$   
 $i + 2K$   
 $i + 3K$   
 $\vdots$   
 $i + (n - 1)K$

Sendo assim temos que uma unidade  $x$  é então selecionada aleatoriamente conforme uma amostra aleatória simples AAS entre as  $k$  primeiras unidades do sistema de referências. As unidades seguintes que farão parte da amostra são obtidas a partir da primeira unidade selecionada em intervalos de comprimento  $k$ , isto é  $n$  é dado por  $n = \{x; x + k; x + 2k; x + 3k; \dots; x + (n - 1)k\}$ .

**PERCENTIS:** Os valores da variável que dividem a frequência total em 100 partes iguais.

**PICTOGRAMA:** Tipo de gráfico ou diagrama, normalmente utilizado para mostrar a frequência de distribuição de um conjunto categórico de dados, em que a frequência em uma categoria é indicada por alguma medida como por exemplo, altura, número de imagens repetidas de uma representação ilustrada de um objeto relevante.

**PLACEBO:** Uma substância farmacologicamente inerte de aparência idêntica àquela do medicamento testado, que dissocia o efeito farmacológico do tratamento de qualquer elemento sugestivo que é o efeito placebo imposto pela administração do medicamento.

**PODER (DE UM TESTE):** A probabilidade de que um teste terá que rejeitar a hipótese nula quando esta hipótese é falsa; é a chance de descobrir que o efeito real de um tratamento de um determinado tamanho é estatisticamente significativo.

**PODER-EFICIÊNCIA:** Uma maneira de comparar o poder de um teste paramétrico e aquele de seu equivalente não-paramétrico. Indica quanto o tamanho da amostra do teste não-paramétrico precisa ser aumentado para torná-lo tão potente quanto o seu equivalente paramétrico.

**POLÍGONO DE FREQUÊNCIA RELATIVA CUMULATIVA:** Um diagrama mostrando uma distribuição de frequência relativa cumulativa. É formado pela união de frequências relativas cumulativas, as quais correspondem aos pontos centrais das classes.

**POLINOMIAL:** Uma relação não-linear ou curvilínea expressa pela equação de regressão, em que o grau do polinômio é determinado pelo poder das variáveis explicativas; por exemplo, para a variável explicativa,  $x$ , a equação quadrática inclui o termo  $x^2$ , a equação cúbica inclui o termo  $x^3$ , dentre outros.

**PONTO DE INFLUÊNCIA:** Uma observação que altera os valores de um ou mais parâmetros das estimativas caso sejam omitidos de uma análise de regressão.

**PONTOS PERCENTUAIS:** O ponto percentual superior ou percentil é o valor da variável que apresenta  $p\%$  da distribuição da variável à direita dela. O ponto percentual inferior apresenta  $p\%$  à esquerda. Para um teste de duas caudas consideramos a porcentagem da distribuição em ambas as caudas, ou seja, à direita e à esquerda dos pontos percentuais relevantes.

**POPULAÇÃO:** A coleção completa finita, real ou infinita, hipotética de unidades observacionais, com pelo menos uma característica em comum.

**POPULAÇÃO HIPOTÉTICA (OU INFINITA):** A população não existe, mas pode ser conceitualizada, por exemplo, a população de animais que podem receber um novo tratamento.

**POPULAÇÃO REAL (OU FINITA):** Os indivíduos na população realmente existem e estão prontamente disponíveis ao pesquisador, sendo o contrário da população hipotética.

**POSTOS:** Os números sucessivos, começando por 1, designados aos valores das observações em um conjunto de dados que foram disponibilizados em ordem crescente ou decrescente. Desse modo, o posto da menor observação é 1, da menor observação seguinte é 2, dentre outros.

**PRECISÃO:** Refere-se ao quão bem as observações repetidas concordam entre si.

**PREFERÊNCIA PELO DÍGITO:** Quando há um elemento de julgamento envolvido nas leituras dos instrumentos certos dígitos, entre 0 e 9, são mais comumente escolhidos do que outros; isso varia de um indivíduo para outro.

**PREVALÊNCIA, DE UMA DOENÇA:** O número ou percentual de casos da doença que há em determinado momento que é a prevalência pontual ou em um período de tempo definido que é a prevalência do período.

**PROBABILIDADE SUBJETIVA:** Visão pessoal da probabilidade, conhecida como uma medida de força da crença de um indivíduo de que um determinado evento irá ocorrer.

**PROBABILIDADES POSTERIORES:** A crença de um indivíduo relacionada a um resultado particular (por exemplo, que um animal tem alguma doença); é quantificada após realizar um experimento ou conduzir um teste utilizando dados recentes do estudo para atualizar a probabilidade prévia. O termo é integrante da inferência Bayesiana.

**PROBABILIDADES PRÉVIAS:** A crença de um indivíduo quanto a um resultado específico, como por exemplo, que um animal tem uma doença; o seu valor é assumido antes de realizar um experimento ou um teste. O termo é integrante da inferência Bayesiana.

**PROCEDIMENTOS ROBUSTOS:** Um teste de hipótese é robusto se as probabilidades dos erros Tipo I e Tipo II são fortemente influenciadas quando as hipóteses nas quais se baseiam o teste não são atendidas.

**PROJETOS COM TAMANHO DE AMOSTRA FIXO:** Delineamento experimental no qual o tamanho da amostra é predeterminado, antes da coleta dos dados.

**PROPORÇÃO:** A proporção do número de eventos de interesse em relação ao número total de eventos.

**PROPORÇÃO DA TAXA DE INCIDÊNCIA, PTI:** É a proporção de duas taxas de incidência. Também conhecida como taxa relativa.

**PROPORÇÃO DE PROBABILIDADE (PP):** A proporção de duas probabilidades, normalmente a probabilidade de doença no grupo exposto ao fator dividida pela probabilidade de doença no grupo não-exposto; uma estimativa o risco relativo quando a doença é rara.

**PROPORÇÃO DE RISCO:** A razão de dois riscos.

**PROTOCOLO CONSORT:** Fornece aos autores, na forma de checklist e fluxograma, diretrizes de como devem relatar um teste clínico aleatório.

**PROPORÇÃO AMOSTRAL:** Simbolizado por  $\hat{p}$  ou  $f$ , se lê p chapéu, p circunflexo ou p estimador, e que serve para estimar a proporção populacional P.

**PROPORÇÃO POPULACIONAL (P):** É um valor resultante do cociente entre o número de casos favoráveis e o número de casos possíveis na população. Por exemplo: proporção de eleitores que votarão no candidato XYZ nas próximas eleições.

**PROCEDIMENTOS:** Parte da pesquisa que se refere a metodologia a ser utilizada numa investigação científica no qual são descritos todos os mecanismos de obtenção e análise de dados.

**PROBABILIDADE:** A chance de ocorrência de um evento particular.

**PROBABILIDADE A PRIORI:** Uma maneira de avaliar a probabilidade de um resultado com base apenas em um modelo teórico. Desse modo, pode ser denominada modelo de probabilidade.

**PROBABILIDADE DE DUAS CAUDAS:** O valor de P resultante de um teste de duas caudas em que todas as probabilidades de interesse estão contidas em ambas as caudas ou seja, a soma das áreas da cauda esquerda com a cauda direita da distribuição de probabilidade.

**PROBABILIDADE DE EXPOSIÇÃO:** A proporção da probabilidade de estar exposto à probabilidade de não estar exposto.

**PROBABILIDADE DE UMA CAUDA:** É o valor de P resultante de um teste de uma cauda, de modo que toda a probabilidade de interesse esteja contida em uma única área da cauda da distribuição teórica.

**PROBABILIDADE PÓS-TESTE:** É a probabilidade posterior de um evento, por exemplo, o resultado da doença obtida utilizando-se a melhor evidência recente, o resultado do teste diagnóstico para atualizar a probabilidade pré-teste ou prévia.

**PROBABILIDADE PRÉ-TESTE:** A probabilidade prévia de uma doença por exemplo, sua prevalência na população, que é atualizada utilizando-se as melhores evidências atuais por exemplo, resultado de um teste diagnóstico para obter a probabilidade pós-teste ou posterior da doença.

**PROVA DE SIGNIFICÂNCIA DE UM VALOR PARAMÉTRICO:** É um teste de hipótese em que se  $\hat{\theta}$  é um estimador de  $\theta$ , obtido pela utilização de um particular método de estimação, e se  $\hat{\theta} \in \phi(\theta; \sigma_\theta^2)$ , pode-se verificar as seguintes hipóteses:

$$\begin{cases} H_0 : \theta = k \\ H_1 : \theta \neq k \end{cases} \text{ ou } \begin{cases} H_0 : \theta \leq k \\ H_1 : \theta > k \end{cases} \text{ ou } \begin{cases} H_0 : \theta \geq k \\ H_1 : \theta < k \end{cases}$$

No primeiro caso tem-se considerado o teste bicaudal ou bilateral, e nos demais, unicaudal ou unilateral.

**PROVA DE ADERÊNCIA:** É aquela onde a hipótese nula  $H_0$  diz respeito ao tipo de distribuição aleatória da população da qual a amostra ou as amostras foi obtida, isto, é, se amostra  $x_n \in X_N$ . A Hipótese nula seria  $H_0 : X_n \in \phi$ , contra a alternativa  $H_1 : X_N \in \phi$ .



**PROVA DE HOMOGENEIDADE:** É o tipo de prova ou teste de hipótese em que existe a suspeita de que a amostra observada tem contribuição de mais de uma população. A prova objetiva determinar se essas populações são distintas ou não.

**PROVA DA CASUALIZAÇÃO:** É o tipo de prova em que existe o interesse em saber se a ordem dos elementos da amostra obtida é devida ao acaso ou a algum fator sistemático. Essas provas visam o processo de amostragem e não características da população, como exemplo de tipos desta prova podem ser citados os testes de Wallys e Moore, teste não-paramétrico de Kendall, dentre outros. (PEREIRA et al. 1980).

**PRÉ-TESTE (EXPERIMENTO):** Em experimentos, medidas da variável dependente antes da administração do estímulo ao grupo experimental.

**PRÉ-TESTE:** Apresentação piloto de um questionário.

**PRÉ-TESTE:** Uma implementação em pequena escala do rascunho do questionário, usada para avaliar seus fatores mais importantes, com clareza, abrangência e aceitabilidade.

**PRÉ-TESTE (PESQUISA DE SURVEY):** Teste de elementos e um desenho de pesquisa, como itens de questionário, técnicas de amostragem, dentre outros.

**PREVALÊNCIA:** Número ou proporção de pessoas portadoras de um evento em um particular momento. O coeficiente de prevalência é este número expresso em unidade de população.

**PREVALÊNCIA:** Estudo epidemiológico de ocorrência de casos de determinada doença, constatados em um determinado momento. Os estudos relativos à prevalência são denominados de transversais.

**PREVALÊNCIA:** Casística de morbidade que se destaca por seus valores maiores do que zero sobre os eventos de saúde ou não-doença. É um termo descritivo da força com que subsistem as doenças nas coletividades. Ver coeficiente de prevalência.

**PREVALÊNCIA:** Número de pessoas que, em uma população definida, tem uma doença ou condição específica em um determinado ponto do tempo, que corresponde, geralmente, ao momento em que um inquérito é realizado.

**PREVALÊNCIA DO PONTO (CASOS PREVALENTES):** Número de pessoas, em uma população definida, que tinham uma doença ou condição específica em um ponto particular no tempo, geralmente na época em que a pesquisa foi feita.

**PREVALÊNCIA INSTANTÂNEA:** Ver prevalência pontual.

**PREVALÊNCIA LÁPSICA:** É a prevalência que abrange um lapso de tempo mais ou menos longo e que não concentra a informação em um dado ponto desse intervalo. Sinônimo: prevalência por período. Ver coeficiente de prevalência lápsica.

**PREVALÊNCIA MOMENTÂNEA:** Ver prevalência pontual.

**PREVALÊNCIA NO PERÍODO:** Número de pessoas que tinham uma doença específica, em qualquer época, durante um intervalo de tempo especificado. A prevalência por período é a soma da prevalência no ponto no começo de um intervalo, mas a incidência durante o intervalo. Devido ao fato de a prevalência por período combinar incidência e prevalência, deve ser usada com extremo cuidado.

**PREVALÊNCIA NO PERÍODO:** Número de pessoas que tiveram a doença em qualquer momento durante o intervalo de tempo especificado. É igual à soma da prevalência no ponto no início do intervalo mais a incidência durante o intervalo.

**PREVALÊNCIA NO PONTO:** Número de casos na população estudada em um determinado ponto no tempo.

**PREVALÊNCIA PONTUAL:** Prevalência medida pela frequência da doença ou pelo seu coeficiente em um ponto definido no tempo, seja o dia, a semana, o mês ou o ano. Sinônimo: prevalência pontual e prevalência momentânea. Ver coeficiente de prevalência.

**PREVALÊNCIA POR PERÍODO:** Ver prevalência lápsica.

**PROPOSIÇÃO:** Termo usado para indicar uma afirmação, enunciado, juízo, postulado ou uma suposição relativa a população, que pode ser considerada como verdadeira ou falsa.

**PREVISÃO:** Em geral, previsão é o processo de determinar a magnitude de variáveis estatísticas em alguns pontos de tempo futuros. Em inglês Prediction.

**PREVISÃO:** Cálculo da magnitude que uma quantidade assumirá em um ponto do tempo futuro. Em inglês Forecasting.

**PRIMEIRA LEI DE MENDEL:** Lei que emergiu do trabalho devido ao monge austríaco Gregor Mendel no século XIX que deduziu importantes princípios genéticos baseados nos seus experimentos com ervilhas. Mendel estudou sete fenótipos da ervilha, cada qual determinado por um único gene. Essa lei afirma que para cada locus, cada genitor transmite um alelo, materno ou paterno, independentemente para cada indivíduo na prole. O princípio básico desta lei afirma que os genes permanecem intactos e distintos. Conhecida também como lei da segregação.

**PRIMEIRO TEMPO DE PASSAGEM:** Conceito importante em processos estocásticos onde, se  $x_N=a$  e  $a \neq 0$  atingir  $x_N=a$  pela primeira vez define  $N$  como o primeiro tempo de passagem desde a origem ao estado  $a$ . Para as diversas realizações de um processo estocástico  $N$  é uma variável aleatória. Em inglês First passage time.

**PRINCÍPIO DA GENERALIZAÇÃO:** Diz que o teor em um ponto não amostrado é igual ao teor na porção amostrada ou não, podendo neste caso ser igual ao teor de um depósito semelhante.

**PRINCÍPIO DA MULTIPLICAÇÃO:** Princípio que afirma: se um experimento comporta  $m$  resultados possíveis, e um segundo experimento comporta  $n$  resultados possíveis, então os dois experimentos combinados comportam  $mn$  resultados possíveis desde que os dois sejam independentes.

**PRINCÍPIO DA VARIAÇÃO GRADUAL:** Ou lei linear, diz que o teor entre dois pontos adjacentes de amostragem varia linear e gradualmente, sendo possível determinar o teor em qualquer ponto dentro do intervalo de amostragem a partir de uma simples relação linear através da interpolação linear.

**PRINCÍPIO DE VEROSSIMILHANÇA:** Toda a informação de um experimento  $X$  está contida na função de verossimilhança. Em inglês Likelihood principle.

**PRINCÍPIO DO CÁLCULO DE RESERVAS:** Matematicamente o cálculo de reservas, nada mais é que a integração numérica da função teor expressa em peso por unidade de volume dentro do depósito de volume  $V$ :  $R = \int T(v) dv$ . O trabalho de avaliação de reservas resumiria-se no cálculo da integral no



domínio do depósito que resultaria no valor procurado. Contudo, isto não é assim simples, pois depende de conhecer a função  $T(v)$ , ou seja, a descrição numérica do teor no depósito de volume  $V$ . Conhecer a função  $T(v)$  implica em conhecer a cada ponto do depósito o valor do teor e, além disso, a função que a descreve matematicamente. Na prática isto é impossível tanto pela densidade de amostragem exigida, como, também pela dificuldade de se encontrar uma função que se ajuste perfeitamente aos dados. Dada a impossibilidade da resolução direta da integral, os métodos existentes para valiação de reservas, sejam eles clássicos ou computacionais, resolvem a seguinte equação, equivalente a anterior, porém para o caso discreto:

$$R = \sum_{i=1}^n T_i V_i$$

**onde:**  $T_i$  é o teor do  $i$ -ésimo bloco;  $v_i$  é o volume do  $i$ -ésimo bloco, com:

$$V = \sum_{i=1}^n v_i$$

Agora a equação do cálculo de reservas resume-se na somatória do produto teor  $X$  volume de cada um dos  $n$  blocos do depósito. Os teores médios podem ser determinados pelos métodos de cálculos de reservas existentes.

**PRINCÍPIO DOS PONTOS MAIS PRÓXIMOS:** Diz que o teor em um ponto não amostrado é igual ao do ponto mais próximo, dando origem a tradicional zona de influência de pontos de amostragem que é exatamente igual a meia distância entre os pontos de dados.

**PRINCÍPIO FUNDAMENTAL DA CONTAGEM:** Suponha o fato de que se possa fazer  $n$  escolhas independentes com:

$m_1$  maneiras de fazer a escolha 1,

$m_2$  maneiras de fazer a escolha 2,

.....

$m_n$  maneiras de fazer a escolha  $n$ .

Então existem  $m_1 \cdot m_2 \cdot \dots \cdot m_n$  maneiras diferentes de fazer a sequência inteira de escolhas. Em inglês fundamental counting principle.

**PRINCÍPIOS DE INTERPRETAÇÃO:** São a base dos métodos clássicos de avaliação de reservas. Os três princípios: da variação gradual ou lei linear; dos pontos mais próximo e da generalização, são utilizados para avaliação de painéis, seções e blocos com base nos dados de amostragem, ou seja, os teores das amostras são estendidos para essas áreas, assumindo um teor interpretado em pontos não amostrados. O erro de estimativa decorrente desta interpretação é função do princípio utilizado, mas principalmente da densidade de amostragem.

**PRIORIS HIERÁRQUICAS:** Este conceito tem como princípio dividir a especificação da distribuição a priori em estágios. Além de facilitar a especificação esta abordagem é natural em determinadas situações experimentais. A distribuição a priori de  $\theta$  depende dos valores dos hiperparâmetros  $\phi$  e pode-se escrever  $p(\theta|\phi)$  ao invés de  $p(\theta)$ . Além disso, ao invés de fixar valores para os hiperparâmetros pode-se especificar uma distribuição a priori  $p(\phi)$  completando assim o segundo estágio na hierarquia. Assim, a distribuição a priori conjunta é simplesmente  $p(\theta, \phi) = p(\theta|\phi)p(\phi)$  e a distribuição a priori marginal de  $\theta$  pode

ser então obtida por integração como  $p(\theta) = \int p(\theta, \phi) d\phi = \int p(\theta|\phi)p(\phi) d\phi$ . A distribuição a posteriori conjunta fica assim:  $p(\theta, \phi|x) \propto p(x|\theta, \phi)p(\theta|\phi)p(\phi) \propto (x|\theta)p(\theta|\phi)p(\phi)$ , neste caso a distribuição dos dados depende somente de  $\theta$ ,  $x$  e  $\phi$  são independentes.

**PRIORI DE JEFFREYS:** Distribuição a priori que consiste na raiz quadrada do determinante da matriz de informação de Fisher. Em inglês Jeffreys prior.

**PRIORI VAGA:** Distribuição a priori que não fornece muita informação sobre o parâmetro de interesse. Em inglês Vague prior.

**PRIORIS HIERÁRQUICAS:** Uma boa estratégia para especificar a distribuição a priori é dividir a situação experimental em estágios ou em hierarquias. Desta forma, a especificação a priori é feita em duas fases: i) estrutural, para a divisão dos estágios; ii) subjetiva, para a especificação quantitativa de cada estágio, que terá seu conjunto de hiperparâmetros. Em inglês Hierarchical priors.

**PRIORIS NÃO INFORMATIVAS:** São distribuições a priori que permitem fazer inferência baseadas em informação subjetiva mínima, produzindo uma análise neutra que permite que a evidência experimental tenha uma forte influência sobre os resultados a posteriori. Exemplos destas prioris são a priori de Jeffreys, definida como a raiz quadrada do determinante da matriz de informação de Fisher, e a priori uniforme. Em inglês Non-informative priors.

**PROSPECÇÃO:** Termo usado em geologia, estatística ou medicina para designar uma análise de dados preliminar.

**PROBABILIDADE:** Propriedade de acontecimento futuro cuja realização não é certa. Na área de saúde, pode ser conceituada como medida da incerteza, dos fenômenos biológicos. Proporção na qual o numerador é o número de vezes que ocorre um evento, e o denominador, esse mesmo número somado ao número de vezes que não ocorre esse evento.

**PROBABILIDADE:** Fração cujo numerador é dado pelos eventos favoráveis e cujo denominador expressa todos os eventos possíveis. Em inglês probability.

**PROBABILIDADE:** Estudo dos fenômenos que dependem do acaso.

**PROBABILIDADE:** Probabilidade pode ser definida como a frequência relativa de um evento.

**PROBABILIDADE:** É um número pertencente ao intervalo  $[0; 1]$  e que mede o grau de certeza na ocorrência de um evento. Ou ainda é aquele número que expressa a chance de ocorrência de determinado evento aleatório, casual ou estocástico. Em inglês probability.

**PROBABILIDADE:** É o quociente entre o número de casos favoráveis e o número total de casos possíveis em uma experiência. A probabilidade de obter o número 4 no lançamento de um dado sem defeito é  $\frac{1}{6}$ . Em inglês probability.

**PROBABILIDADE:** Medida da possibilidade de ocorrência de um evento; expressa como um número entre 0 e 1. Em inglês probability.

**PROBABILIDADE:** É uma medida normalizada  $P\{A\}$  definida no conjunto das partições do conjunto  $S$ , o espaço amostral, sendo o acontecimento  $A$  um subconjunto de  $S$  ( $A \subseteq S$ ). Em inglês probability.

**PROBABILIDADE:** É um escala de mensuração usada para descrever a chance de ocorrência de um valor específico que é o evento de uma variável aleatória. Em inglês probability.

**PROBABILIDADE:** A probabilidade de ocorrência de um evento é dada pelo número esperado de vezes que este evento ocorra, em relação ao número total de eventos. Em inglês probability.

**PROBABILIDADE:** Um conceito básico que pode ser considerado indefinível, expressa de alguma forma um grau de crença, ou o limite de frequência em uma série aleatória infinita. Em inglês probability.

**PROBABILIDADE-ABORDAGEM CLÁSSICA:** Definição do espaço de probabilidade como conjunto de todos os resultados possíveis de um experimento aleatório em que se supões que todos sejam igualmente prováveis.

**PROBABILIDADE ANTERIOR OU PROBABILIDADE PRÉ-TESTE:** Probabilidade de um determinado paciente estar doente, estimada antes da realização de um teste de laboratório ou de outro tipo de teste, com base na prevalência estimada dessa doença entre os pacientes que apresentam sinais e sintomas semelhantes.

**PROBABILIDADE ARITMÉTICA:** Ou gráfico de probabilidade aritmética tem em sua ordenada a escala de probabilidade aritmética desenhada de tal modo que se uma curva acumulativa for de uma distribuição normal, ela será representada por uma reta.

**PROBABILIDADE A POSTERIORI:** É a que tem possível causa de um evento aleatório, dada uma realização desse evento. Distingue de probabilidade a priori. A distinção entre probabilidade a posteriori e a priori, sem o uso dessas expressões, foi já estabelecida por Bernoulli (1713) e as expressões foram introduzidas por Laplace (1812).

**PROBABILIDADE A PRIORI:** É a que possui a possível causa de um evento aleatório, antes, ou na ignorância, da realização desse evento. Distingue de probabilidade a posteriori.

**PROBABILIDADE COM FREQUÊNCIA RELATIVA:** Enfoque da probabilidade de um evento como a fração de vezes que o evento ocorreria se um experimento fosse realizado um grande número de vezes.

**PROBABILIDADE CONDICIONAL:** Probabilidade de um evento, sabendo-se que algum outro evento já ocorreu.

**PROBABILIDADE CONDICIONAL:** Probabilidade de ocorrência de determinado evento, quase se sabe que outro evento ocorreu.

**PROBABILIDADE CONDICIONAL:** Sejam A e B dois eventos quaisquer, a probabilidade condicional de A, tendo certeza do acontecimento de B, denotada por  $P(A|B)$ , é dada por:  $P(A|B)=P(A,B)/P(B)$ . Em inglês Conditional probability.

**PROBABILIDADE CONDICIONAL:** A probabilidade de B dado o acontecimento A é por definição:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} .$$

**PROBABILIDADE CONDICIONAL DE  $\theta$  DADO Y:**  $P(\theta|Y) = \frac{P(\theta \cap Y)}{P(Y)}$ ,  $P(Y) > 0$ .

**PROBABILIDADE DE ERRO TIPO II:** Probabilidade de se aceitar a hipótese nula quando ela for falsa, valor simbolizado pela grega beta ( $\beta$ ).

**PROBABILIDADE FREQUENTISTA:** Um dos conceitos de probabilidade, o qual afirma que sob o enfoque frequentista, as frequências relativas de ocorrência dos resultados convergem para a probabilidade de ocorrência destes resultados quando o número n de repetições do experimento tende a ser grande. Por exemplo, a frequência de um resultado A é dada por  $f_A = \frac{n_A}{n} \Rightarrow \lim_{n \rightarrow \infty} f_A = P(A)$ , em que  $n_A$  é o número de

vezes em que ocorreu o resultado A e  $P(A)$  denota a probabilidade de A.

**PROBABILIDADE FÍSICA OU CLÁSSICA:** Um dos conceitos de probabilidade, o qual afirma que a interpretação clássica de probabilidade baseia-se no conceito de resultados igualmente prováveis. Assim, a probabilidade de um resultado ou evento é dada pela razão  $\frac{1}{n}$ , em que n é o número de resultados possíveis.

**PROBABILIDADE SUBJETIVA OU PERSONALÍSTICA:** Um dos conceitos de probabilidade, o qual afirma que segundo esta interpretação, a probabilidade que uma pessoa atribui a um possível resultado ou acontecimento de algum processo que representa seu próprio julgamento da chance de o resultado ser obtido. Assim, a probabilidade de um evento A é a medida do grau de confiança ou crença que a pessoa deposita em A, ou seja, é o grau de crença do indivíduo de que o evento irá ocorrer baseado em toda evidência a ele disponível, uma vez que o valor da probabilidade é apenas um julgamento pessoal, como por exemplo chover ou não chover.

**PROBABILIDADE LÓGICA:** Um dos conceitos de probabilidade, o qual afirma que a interpretação lógica de probabilidade pode ser derivada em função do conhecimento científico que é de domínio público, ou seja do conhecimento comum a todos.

**PRIORI NÃO INFORMATIVA:** Conforme Jeffreys (1961), citado por Resende (2002), existem situações em que o conhecimento sobre determinado fenômeno é vago ou inexistente. Nestes casos, a distribuição a priori é dita vaga, difusa ou não informativa, não precisa ou de variância alta, significando que a densidade a priori reflete ignorância. Assim, a priori deste tipo não informativa poderia ser representada por  $f(\theta)\alpha$  constante;  $-\infty < \theta < \infty$ . Uma priori deste tipo poderia ser especificada através de uma distribuição com variância tendendo ao infinito (GIANOLA; FERNANDO, 1986) e, neste caso, maior ênfase ou peso serão dados a verossimilhança (dados experimentais ou amostrais). Entretendo, para inferência sobre uma variância  $\sigma^2$  prioris da forma mencionada tornam-se informativas (JEFFREYS, 1961) e uma priori não informativa poderia ser representada por  $f(\sigma^2)\alpha \frac{1}{\sigma^2}$ ;  $0 < \sigma < \infty$ . Para prioris não informativas, supõem-se inicialmente distribuições uniformes como representante de situações onde não se dispõem de informação inicial, ou seja,  $f(\theta)\alpha$  constante ( $\theta$  variando na reta) fato que implica não favorecer qualquer valor particular de  $\theta$  (BAYES, 1763).

**PROBABILIDADE DE TRANSIÇÃO:** Na teoria de processos estocásticos, a probabilidade de transição é a probabilidade condicional, que um sistema estabelecido em  $E_j$  mudará para  $E_k$  em algum tempo designado posteriormente. Em inglês Transition probability.

**PROBABILIDADE DE UM EVENTO:** Número de resultados que correspondem a esse evento, dividido pelo número total de resultados possíveis.

**PROBABILIDADE-ENFOQUE SUBJETIVO:** Enfoque da probabilidade de um evento como uma estimativa pessoal da viabilidade de ocorrência do evento.

**PROBABILIDADE POSTERIOR:** Estimativa revisada da probabilidade de um determinado paciente ter a doença, após a realização de um teste diagnóstico que é a probabilidade pós-teste ou após uma intervenção.

**PROBABILIDADE PROPORCIONAL AO TAMANHO (PPT):** Tipo de amostra por conglomerados em múltiplas etapas nas quais conglomerados não são selecionados com probabilidades iguais mas com probabilidades proporcionais a seus tamanhos, medidas pelo número de unidades a serem sub-amostradas. Ver amostras MIPS.

**PROBABILIDADE SUBJETIVA:** Conjectura ou estimativa de uma probabilidade baseada no conhecimento de circunstâncias relevantes.

**PROBABILIDADES POSTERIORES:** Probabilidades revisadas de eventos, baseadas em informação adicionais.

**PROTOCOLO DE PESQUISA:** Refere-se às etapas, procedimentos ou passos usados ou a serem seguidos em uma pesquisa científica.

**PADRONIZAÇÃO:** Conjunto de operações capazes de garantir a um teste precisão, validade, normas padronizadas, instruções para sua aplicação, avaliação e interpretação.

**PARÂMETRO:** Valor relativo a uma população.

**PERCENTIL:** Cada um dos 99 valores da variável que dividem uma distribuição de frequências em 100 intervalos iguais. O percentil é representado pelo símbolo  $P_r$ , em que  $r$  representa a ordem percentil. Se, num grupo de estudantes, 60 % obtêm valores inferiores a 80, diz-se que o 60º percentil é 80 e que 80 é a ordem percentil correspondente.

**PERMUTAÇÕES:** Permutações simples de  $m$  elementos são todos os agrupamentos de  $m$  elementos sem repetição que se podem formar com os elementos dados, de modo que cada agrupamento se diferencie de outro pela ordem de seus elementos.

**PESQUISAS HISTÓRICAS:** Tipo de pesquisa cujo objetivo é reconstruir o passado, em termos relativamente precisos e objetivos, para explicar fatos atuais, como por exemplo, estudos das influências da colonização portuguesa na cultura camponesa do Nordeste (RICHARDSON et al., 2009).

**PESQUISAS EXPLORATÓRIAS:** Tipo de pesquisa cujo objetivo é conhecer as características de um fenômeno para procurar, posteriormente, explicações das causas e conseqüências de dito fenômeno, como por exemplo, estudar as reações de um grupo de alunos à aplicação de um novo método de estudo. (RICHARDSON et al., 2009).

**PESQUISAS DESCRITIVAS:** Tipo de pesquisa cujo objetivo é descrever sistematicamente um fenômeno ou área de interesse. Dita descrição deve ser detalhada e objetiva, como por exemplo, pesquisas demográficas, estudo de opinião, análise de documentos, perfis ocupacionais, inventários de interesse, etc. (RICHARDSON et al., 2009).

**PESQUISAS EXPLICATIVAS-ENQUETES:** Tipo de pesquisa cujo objetivo é investigar possíveis relações causa-efeito, observando as conseqüências de um fenômeno em amostras relativamente grandes, como mais de 200 casos, como por exemplo, comparar as características por sexo de pacientes de um hospital psiquiátrico (RICHARDSON et al., 2009).

**PESQUISAS EXPLICATIVAS-EXPERIMENTOS:** Tipo de pesquisa cujo objetivo é investigar possíveis relações causa-efeito, submetendo um ou mais grupos experimentais a um ou mais tratamentos, e comparando os resultados com um ou mais grupos de controle que não receberam o tratamento, como por exemplo, investigar os efeitos de uma droga tranquilizante em pessoas hipercinéticas, o qual é um grupo de transtornos caracterizados por início precoce, habitualmente durante os cinco primeiros anos de vida, falta de perseverança nas atividades que exigem um envolvimento cognitivo, e uma tendência a passar de uma atividade a outra sem acabar nenhuma, associadas a uma atividade global desorganizada, incoordenada e excessiva. Os transtornos podem se acompanhar de outras anomalias. As crianças hipercinéticas são frequentemente imprudentes e impulsivas, sujeitas a acidentes e incorrem em problemas disciplinares mais por infrações não premeditadas de regras que por desafio deliberado. Suas relações com os adultos são frequentemente marcadas por uma ausência de inibição social, com falta de cautela e reserva normais. São impopulares com as outras crianças e podem se tornar isoladas socialmente. Estes transtornos se acompanham frequentemente de um déficit cognitivo e de um retardo específico do desenvolvimento da motricidade e da linguagem. As complicações secundárias incluem um comportamento dissocial e uma perda de autoestima. (RICHARDSON et al., 2009; <http://www.psiqweb.med.br/>, 2016).

**PESQUISAS EXPLICATIVAS-QUASE EXPERIMENTOS:** Tipo de pesquisa cujo objetivo é aproximar as condições do experimento em situações reais que não permitem controlar as variáveis relevantes. O pesquisador deve conhecer as limitações desse método, em relação à validade interna e externa do plano de pesquisa, inclui-se nesta categoria a grande maioria dos experimentos de campo e pesquisas operacionais realizadas em contextos reais nos quais apenas é possível o controle parcial das variáveis. Por exemplo, um estudo da eficácia de dois métodos de ensino aplicado em duas turmas diferentes, sem distribuição aleatória de ditas turmas. (RICHARDSON et al., 2009).

**PESQUISAS EXPLICATIVAS-ESTUDO DE CASOS:** Tipo de pesquisa cujo objetivo é analisar detalhadamente o passado, presente e as intenções sociais de uma unidade social, como um indivíduo, um grupo, uma instituição ou comunidade, por exemplo, o estudo de um grupo de adolescentes detentos por consumo de drogas, ou as características socioeconômicas de um conjunto habitacional (RICHARDSON et al., 2009).

**PESQUISA-AÇÃO:** Tipo de pesquisa cujo objetivo é incentivar transformações sociais de um grupo, com participação direta de seus membros em todas as etapas da pesquisa. O pesquisador realiza o trabalho de acordo com os interesses e necessidades do grupo, como por exemplo, estudos dos problemas do transporte público em uma determinada comunidade, ou ainda desenvolvimento de formas de participação comunitária para enfrentar problemas de saúde pública. (RICHARDSON et al., 2009).

**POLÍGONO DE FREQUÊNCIA:** Representação gráfica de uma distribuição de frequências resultante da união das intersecções dos pontos médios das classes nas abcissas e das frequências respectivas nas ordenadas.

**PONTO MÉDIO:** Ponto inferior de uma classe, equidistante dos seus limites reais. Seu valor é igual à metade da soma desses limites.

**POPULAÇÃO:** É o conjunto formado pelas medidas que se fazem sobre elementos do Universo.

**PRECISÃO:** Refere-se à dispersão de uma conjuntos de mensuração; quanto menos a dispersão, melhor a precisão.



**PROBABILIDADE A PRIORI:** Dada uma experiência aleatória uniforme, definida num espaço amostral ou de possibilidades  $\Omega$ , define-se como a probabilidade de ocorrer um evento E, contido em  $\Omega$ , como sendo o quociente entre o número de elementos do evento E e o número de elementos do espaço amostral  $\Omega$ .

**PROBABILIDADE A POSTERIORI:** Trata-se da probabilidade avaliada, empírica. Ela tem por objetivo estabelecer um modelo adequado à interpretação de uma certa classe de fenômenos observados mas não todos.

**PROMÉDIO:** Termo genérico que expressa qualquer medida de tendência central. Exemplos: a média, a mediana e a moda de uma escala.

**PROBABILIDADES PRÉVIAS:** Estimativas iniciais das probabilidades de eventos.

**PROBABILÍSTICO:** Seguir as leis da probabilidade ao formatar um padrão descritivo ou associação entre variáveis. Por exemplo, amostragem probabilística gera amostras que dão estimativas próximas, mas não perfeitas, das características da população da qual foram selecionadas. Podemos também observar que crianças de famílias desfeitas mais provavelmente será delinquentes de que outras, apesar de esta não ser uma relação perfeita.

**PROBIT:** Se U for uma variável definida no intervalo [0; 1], então a transformação que associa a cada valor u, neste intervalo, um valor v no intervalo  $(-\infty; +\infty)$  é denominada de probit se  $V = \Phi^{-1}(u)$ , onde F é Função de Distribuição da Normal Padrão. Em inglês probit.

**PROBLEMA:** É o marco referencial inicial de uma pesquisa. É a dúvida inicial que lança o pesquisador ao seu trabalho de pesquisa.

**PROBLEMA (O QUE DEVE SER):** O problema deve tratar, comumente de forma quantitativa, das relações de fenômenos naturais, ou mais especificamente, das condições que controlam os fatos observáveis e eventos. Os resultados de seu estudo devem servir de base para predição e controle dos fenômenos observados. O problema deve ser circunscrito, definido e específico, mas deve estar conectado com um campo maior. O problema deve ter como principal objetivo a obtenção de fatos novos. O problema deve ser adequado ao tratamento que lhe seja possível dar.

**PROBLEMA, DEFINIÇÃO DO:** No processo de conceituação ou definição e estabelecimento do problema, é importante considerar: i) adequada identificação do tópico; ii) elementos constituintes e abrangências; iii) literatura relacionada e terminologia; iv) fontes de evidência e metodologia e v) pressupostos básicos.

**PROBLEMA DE DECISÃO ADMINISTRATIVA:** As providências administrativas necessárias para resolver um problema de pesquisa de marketing.

**PROBLEMA DE DECISÃO MÚLTIPLA:** Problema de escolha de hipótese ou decisão dentre um conjunto de k alternativas exclusivas e exaustivas baseados em observações de variáveis aleatórias. Em inglês Multiple decision problem.

**PROBLEMA DE PESQUISA DE MARKETING:** Uma declaração das informações específicas necessárias para um tomador de decisões, a fim de ajudar a resolver um problema de decisão administrativa.

**PROBLEMA DOS TRÊS CORPOS:** É o problema de investigar o comportamento de três corpos que se atraem mutuamente tais como o sol, a terra e a lua e a estabilidade dos seus movimentos. Em inglês three body problem.

**PROBLEMA, SELEÇÃO DE:** Os seguintes fatores devem ser considerados na seleção de um problema de pesquisa: i) novidade e significância; ii) curiosidade intelectual e interesse; iii) aptidão e características pessoais; iv) fontes de consulta e técnicas; v) apoio logístico e cooperação; vi) riscos e custos e vii) tempo.

**PROCEDIMENTOS:** Arranjos feitos para a coleta de dados e métodos de analisar relacionamentos.

**PROCEDIMENTO DE BAYES EMPÍRICO:** Termo proposto por Robbins (1956) para denotar a aproximação a um procedimento ótimo de decisão de Bayes que pode ser derivado do uso de dados previamente coletados pelo mesmo procedimento de seleção para uma população em estudo. Em inglês Empirical Bayes procedure.

**PROCEDIMENTO DE OTIMIZAÇÃO:** Procedimento de agrupamento não-hierárquico que permite a redesignação de objetos do agrupamento originalmente designado para um outro agrupamento com base em um critério de otimização geral.

**PROCEDIMENTOS HIERÁRQUICOS:** Procedimentos de agrupamentos stepwise que envolvem uma combinação ou divisão dos objetos em agrupamentos. Os dois procedimentos alternativos são os métodos aglomerativo e divisivo. O resultado é a construção de uma hierarquia, ou estrutura em árvore, ou seja, um dendrograma, que representa a formação dos agrupamentos. Tal procedimento produz N-1 soluções de agrupamento, onde N é o número de objetos. Por exemplo, se o procedimento aglomerativo começar com cinco objetos em grupos separados, ele mostrará como quatro agrupamentos, e em seguida três, dois e finalmente um agrupamento, são formados.

**PROCEDIMENTOS NÃO-HIERÁRQUICOS:** Procedimentos que produzem apenas uma solução de agrupamento para um conjunto de sementes de agrupamentos. Ao invés de usar o processo de construção em forma de árvore encontrado nos procedimentos hierárquicos, as sementes de agrupamentos são empregadas para reunir objetos dentro de uma distância pré-especificada das sementes. Por exemplo, se quatro sementes de agrupamentos são especificadas, apenas quatro agrupamentos são formados. Os procedimentos não-hierárquicos não produzem resultados para todos os possíveis números de agrupamentos, como é feito com um procedimento hierárquico.

**PROCESSO DE DADOS PERDIDOS:** Qualquer evento sistemático externo ao respondente como erros na entrada de dados ou problemas na coleta de dados ou qualquer ação por parte do respondente como a recusa a responder uma questão que conduz a dados perdidos.

**PROCESSO DE MÉDIAS MÓVEIS:** Seja  $\{\varepsilon_t\}$  um processo discreto puramente aleatório com média zero e variância  $\sigma_\varepsilon^2$ . Um processo  $\{X_t\}$  é chamado de processo de médias móveis de ordem q, ou MA(q), se  $X_t = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \dots + \beta_q \varepsilon_{t-q}$ . Não é difícil verificar que

$$E\{X_t\} = 0 \quad e \quad Var(X_t) = (1 + \beta_1^2 + \beta_2^2 + \dots + \beta_q^2) \sigma_\varepsilon^2. \quad \text{Além disso, como } Cov(\varepsilon_t, \varepsilon_s) = \sigma_\varepsilon^2 \text{ para } t = s \text{ e}$$

$$Cov(\varepsilon_t, \varepsilon_s) = 0 \text{ para } t \neq s, \text{ a função de autocovariância é dada por}$$



$$\gamma(K) = Cov(X_t, X_{t+k})$$

$$\gamma(K) = Cov(\varepsilon_t + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q}, \varepsilon_{t+k} + \beta_1 \varepsilon_{t+k-1} + \dots + \beta_q \varepsilon_{t+k-q})$$

$$\gamma(K) = \begin{cases} 0 & k > q \\ \sigma_\varepsilon^2 \sum_{j=0}^{q-k} \beta_j \beta_{j+k} & k = 0, \dots, q \\ \gamma(-K) & k < 0 \end{cases}$$

**PROCESSO ESTOCÁSTICO:** Coletânea de variáveis aleatórias  $\{X_t\}$  onde t assume valores num intervalo T. Em muitos casos  $x_t$  é uma observação no tempo t e T é um intervalo de tempo. Um PE (Processo Estocástico) pode ser classificado por meio de : i) espaço de estados S, ii) do parâmetro de indexação t e iii) da relação de dependência entre as variáveis aleatórias  $X_t$ . O espaço de estados S é o conjunto dos possíveis valores de cada variável aleatória. Assim se for discreto, diremos que o processo é de estado discreto. Se S = R, diremos que o processo é de estado contínuo. Quanto ao parâmetro um PE pode ser discreto se T = {0, 1, 2, ...} ou contínuo se T = [0, inf.) Quanto a relação de dependência um PE pode ser: i) estritamente estacionário; ii) com incrementos independentes; iii) com incrementos independentes estacionários; iv) com incrementos ortogonais; v) um Martingale; vi) de Markov; vii) independente; viii) amplamente estacionário. Em inglês stochastic process.

**PROCESSAMENTO ANALÍTICO ON-LINE (OLAP):** Aplicação de base de dados que permite a usuários ver, manipular e analisar dados multidimensionais ou multivariados.

**PROCESSO DE ENTRADA/SAÍDA:** Uma grande classe de processos estocásticos, nos quais o estado  $\xi(t)$  no tempo t é resultado de duas variáveis aleatórias x, y;  $\xi(t)=x(t)-y(t)$ , x e y são chamadas, respectivamente, de entrada e saída. Os processos de filas e processos epidêmicos são exemplos. Em inglês Input/output process.

**PROCESSO DE MARKOV:** Processo aleatório ou estocástico cuja distribuição de probabilidade condicional para o estado em um instante futuro, dado o estado do presente, não é afetado pelo conhecimento adicional da história passada do processo. Em inglês Markov process.

**PROCESSO DE MOVIMENTO BROWNIANO:** Processo estocástico aditivo de uma variável real  $x_t$  definida no tempo t para a qual  $x_t - x_s$  tem distribuição normal com média zero e variância  $\sigma^2|t-s|$ , onde  $\sigma^2$  é uma constante. Em inglês Brownian motion process.

**PROCESSO DE POISSON:** Um processo estocástico que pode ser considerado como um caso específico de um Processo de Nascimento em que o parâmetro  $\lambda_n$  é uma constante  $\lambda$ . Em inglês Poisson process.

**PROCESSO DE VIDA E MORTE:** Processo estocástico que tenta descrever o crescimento e a queda de uma população de membros que podem morrer ou criar novos indivíduos. Em inglês Birth and death process.

**PROCESSO DETERMINÍSTICO:** Processo estocástico sem erro de previsão; aquele em que o passado determina completamente o futuro do sistema. Em inglês Deterministic process.

**PROCESSO ESTACIONÁRIO:** O processo estocástico  $\{x_t\}$  é estritamente estacionário se a distribuição multivariada de  $x_{t(1)+h}, x_{t(2)+h}, \dots, x_{t(n)+h}$  é independente de h para todo conjunto finito de valores  $t(1)+h, \dots, t(n)+h, t(1), t(2), \dots, t(n)$ . O processo é chamado estacionário, em sentido amplo, se existem média e variância que não dependem de t. Em inglês Stationary process.

**PROCESSO ESTATISTICAMENTE ESTÁVEL:** Processo com apenas uma variação natural e sem nenhum padrão, ciclo ou pontos raros.

**PROCESSO ESTOCÁSTICO:** Diz-se do processo que depende, ou resulta, de uma variável aleatória.

**PROCESSO ESTOCÁSTICO:** É um fenômeno que varia em algum grau, de forma imprevisível, à medida que o tempo passa.

**PROCESSO ESTOCÁSTICO:** Situações em que são feitas observações quanto a um período de tempo, situações essas influenciadas por efeitos aleatórios ou de azar, não só em um único instante, mas por todo intervalo de tempo ou sequencia de tempos que se está a considerar.

**PROCESSO EVOLUCIONÁRIO:** Qualquer processo estocástico ou aleatório não estacionário. A distribuição de probabilidade associada ao processo não é independente do tempo. Em inglês Evolutionary process.

**PROCESSO MARKOVIANO MÚLTIPLO:** Processo estocástico em que as probabilidades de transição dependem de valores prévios em mais de um ponto. Em muitos contextos a expressão é equivalente a processo auto-regressivo. Em inglês Multiple Markov process.

**PROCESSOS DE COMPARAÇÃO MÚLTIPLA:** Processos que permitem identificar que médias são diferentes, após concluir que pelo menos três médias não são.

**PROCESSOS MULTIVARIADOS:** Uma classe de processos estocásticos envolvendo mais de uma variável aleatória. Também chamados de processos simultâneos ou vetoriais que não devem ser confundidos com processos multidimensionais que estão relacionados com a presença de mais de um parâmetro. Em inglês Multivariate processes.

**PRODUTO ESCALAR:** O produto escalar de dois vetores  $u$  e  $v$  é definido pela equação dada por  $u \cdot v = |u| \cdot |v| \cdot \cos \theta$ , onde  $|u|$  é a norma de  $u$  e  $\theta$  é o ângulo entre os dois vetores. Em inglês dot product.

**PRODUTÓRIO:** Símbolo matemático representado pela letra do alfabeto grego Pi maiúscula, dada por  $\Pi$ , o qual serve para simbolizar o produto ou multiplicação de n valores não nulos de uma série.

**PRODUTO DE KRONECKER DE MATRIZES:** O produto  $A \otimes B$  de uma matriz A  $m \times m$  e B  $n \times n$  é a matriz  $mn \times mn$  cujos elementos são o produto de termos, um de A e um de B. Em inglês Kronecker product of matrices.

**PRODUTO DE HADAMARD DE MATRIZES:** Produto simbolizado por  $\#$ , símbolo semelhante ao utilizado para denotar o cardinal de um conjunto de valores.

**PRODUTO INTERNO BRUTO (PIB):** Medida de valor total de todos os seus bens e serviços produzidos no país.

**PRODUTO INTERNO BRUTO NOMINAL:** Produto interno bruto avaliado em moeda corrente, não corrigida pela inflação. Ver produto interno bruto real.

**PRODUTO INTERNO BRUTO REAL:** Valor que o produto interno bruto teria se todos os preços tivessem permanecido os mesmos que em um ano específico chamado ano-base; é uma medida do PIB que leva em conta a inflação.

**PRODUTO NACIONAL BRUTO (PNB):** Medida de produção nacional em relação ao produto interno bruto; inclui a produção por elementos de propriedade de cidadãos do país, mas não a produção por elementos possuídos no exterior.

**PRODUTO NACIONAL LÍQUIDO (PNL):** Medida de economia nacional, igual ao produto nacional bruto menos a depreciação.

**PROGNOSE:** Predição da produção futura de um povoamento florestal.

**PROGRAMAÇÃO INTEIRA:** Uma forma de programação matemática onde as soluções assumem valores inteiros. Em inglês Integer programming.

**PROGRAMAÇÃO LINEAR:** Procedimento usado para maximizar, ou minimizar, uma função linear de variáveis que estão sujeitas a restrições expressadas em termos lineares. Em inglês Linear programming.

**PROGRESSÃO ARITMÉTICA (P.A.):** Tipo de sucessão numérica de grande importância prática em que os acréscimos absolutos são constantes em qualquer período.

**PROGRESSÃO DE UM PROCESSO EPIDÊMICO:** Denominação da primeira fase de uma ocorrência epidêmica que corresponde ao crescimento progressivo da incidência, representando pelo ramo ascendente da curva epidêmica.

**PROGRESSÃO GEOMÉTRICA (P.G.):** Tipo de sucessão numérica de grande importância prática em que os acréscimos percentuais em todos os períodos são iguais.

**PROJEÇÕES:** Pontos definidos por retas perpendiculares de um objeto a um vetor. As projeções são usadas para determinar a ordem de preferência com representações vetoriais.

**PROJETO ANTES E DEPOIS COM GRUPO DE CONTROLE:** Um projeto verdadeiramente experimental que inclui tarefas aleatórias de assuntos ou unidades de teste para grupos experimentais ou de controle ou pré-mensuração de ambos os grupos.

**PROJETO DE COMPARAÇÃO DE GRUPO ESTÁTICO:** Projeto pré-experimental que utiliza um grupo experimental e um de controle. Entretanto, os indivíduos ou unidades de teste não são designados aleatoriamente aos dois grupos e não são feitas quaisquer pré-mensurações.

**PROJETO DE ESTUDO DE CASO ÚNICO:** Projeto pré-experimental sem controle de grupo e com mensuração e com mensuração posterior apenas.

**PROJETO DE PESQUISA:** O plano a ser seguido para atender aos objetivos de marketing; a estrutura ou arcabouço para resolver um problema específico.

**PROJETO DE SÉRIES DE TEMPO INTERROMPIDAS:** Pesquisa na qual o tratamento interrompe mensurações repetidas contínuas.

**PROJETOS DE SÉRIES DE TEMPO MÚLTIPLAS:** Um projeto de séries de tempo interrompidas com um grupo de controle.

**PROJETO EXPERIMENTAL:** Um teste no qual o pesquisador tem controle sobre uma ou mais variáveis independentes e as manipula. Em inglês experimental design.

**PROJETO PRÉ-EXPERIMENTAL:** Projeto que oferece pouco ou nenhum controle sobre fatores estranhos.

**PROJETO PRÉ-TESTE/PÓS-TESTE DE UM GRUPO:** Projeto pré-experimental com pré e pós mensurações, mas sem grupo de controle.

**PROJETO QUASE EXPERIMENTAL:** Estudo no qual o pesquisador carece de controle completo sobre a programação do tratamento ou precisa designar respondentes ao tratamento de maneira não aleatória.

**PROJETO DE SALOMÃO DE QUATRO GRUPOS:** Pesquisa que usa dois grupos experimentais e dois grupos de controle para controlar todas as ameaças de variáveis estranhas.

**PROJETO SOMENTE APÓS COM GRUPOS DE CONTROLE:** Projeto verdadeiramente experimental que envolve tarefas aleatórias de assuntos ou unidades de teste para grupos experimentais e de controle, mas nenhuma pré-mensuração da variável dependente.

**PROJETO VERDADEIRAMENTE EXPERIMENTAL:** Pesquisa com uso de um grupo experimental e um grupo de controle com tarefas aleatórias de unidades de teste para ambos os grupos.

**PROMÉDIO:** Conceito familiar, porém enganoso. Geralmente, um promédio tem o sentido de representar ou resumir as características importantes de um conjunto de valores e, neste sentido, o termo inclui à mediana e também à moda. Em um sentido mais limitado, um promédio mistura todos os valores do conjunto de dados. Do inglês Average.

**PROMÉDIO:** Termo genérico que expressa qualquer medida de tendência central. Exemplos: a média, a mediana e a moda de uma escala. Em inglês average.

**PROPOSIÇÃO:** Conclusão tirada logicamente de um conjunto de suposições iniciais ou axiomas num sistema teórico.

**PROPOSIÇÕES:** Afirmativas provisórias de relacionamentos entre objetos e eventos.

**PROPOSIÇÕES LÓGICAS:** São proposições úteis e desejáveis de estrutura lógica das nossas observações.

**PROPORCIONAL A:** Símbolo matemático dado por  $\propto$ .

**PROPORÇÃO AMOSTRAL:** É uma proporção simbolizada por  $\bar{P}$  que se lê p barra, ou por  $\hat{P}$  que se lê P circunflexo ou P chapéu, serve para estimar a proporção populacional.

**PROPORÇÃO AMOSTRAL ( $\hat{P}$ ):** O símbolo entre parêntese se lê P chapéu ou P circunflexo, a qual serve para estimar a proporção populacional P.

**PROPORÇÃO POPULACIONAL (P):** É um valor resultante do cociente entre o número de casos favoráveis e o número de casos possíveis na população. Por exemplo: a proporção de eleitores que votarão no candidato José Correto da Silva nas eleições de 2016 ou a proporção anual de crianças que nascem com cardiopatia no Estado do Rio Grande do Norte.

**PROPRIEDADE DE MÍNIMOS QUADRADOS:** Propriedade segundo a qual, em uma reta de regressão, a soma dos quadrados dos desvios verticais entre os pontos amostrais e os pontos da reta de regressão é a menor possível.

**PROPRIEDADE MARKOVIANA:** Uma sequência  $X_1, \dots, X_n$  tem a propriedade markoviana se a distribuição condicional de  $X_j$  dada a sequência inteira  $X_1, \dots, X_{j-1}$  só depende dos valores mais recentes. Em inglês Markov property.

**PROTOTESE:** Denominação usada para designar uma conjectura que é testável, plausível ou que pode ser submetida aos testes de hipóteses ou de significância. Em contrapartida existem aquelas hipóteses que não podem ser verificadas devida as limitações científicas e tecnológicas da época atual.

**PROPRIEDADES DA MÉDIA:** A esperança matemática ou média de uma variável aleatória discreta ou contínua possui as seguintes propriedades: i) a média de uma constante é a própria constante:  $E[K] = K$ ; ii) a média de uma variável aleatória  $X$ , multiplicada por uma constante é igual a constante multiplicada pela média de  $X$ :  $E[KX] = KE[X]$ ; iii) a média de uma variável aleatória dividida por uma constante é igual a média da variável aleatória dividida pela constante:

$$E\left[\frac{X}{K}\right] = \frac{E[X]}{K}; \text{ iv) a média da soma ou diferença de duas variáveis aleatórias é a soma ou diferença das}$$

médias;  $E[X \pm Y] = E[X] \pm E[Y]$ ; v) a média de uma variável aleatória somada ou subtraída de uma constante é igual à média dessa variável somada ou subtraída da mesma constante:  $E[X \pm K] = E[X] \pm K$ ; vi) a média de uma variável aleatória subtraída de sua própria média é zero:  $E[X - \bar{X}] = 0$ ; vii) a média do produto de duas variáveis aleatórias independentes é igual ao produto das médias:  $E[X \cdot Y] = E[X] \cdot E[Y]$  e viii) a média do produto de duas variáveis aleatórias dependentes discretas ou contínuas é dada por:

$$E[XY] = \sum_{i=1}^n \sum_{j=1}^m X_i Y_j P(X_i, Y_j) \text{ ou } E[XY] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} XY f(X, Y) dx dy, \text{ respectivamente.}$$

**PROPRIEDADES DA VARIÂNCIA:** A variância de uma variável aleatória discreta ou contínua possui as seguintes propriedades: i) a variância de uma constante é zero:  $Var[K] = 0$ ; ii) a variância de uma variável aleatória multiplicada por uma constante é igual a variância da variável aleatória multiplicada pelo quadrado dessa constante:  $Var[KX] = K^2 Var[X]$ ; iii) a variância de uma variável aleatória dividida por uma constante é igual a variância da variável aleatória dividida pelo quadrado dessa constante:

$$Var\left[\frac{X}{K}\right] = \frac{Var[X]}{K^2}; \text{ iv)}$$

a variância de uma variável aleatória somada ou subtraída de uma constante é igual à variância da variável aleatória:  $Var[X \pm K] = Var[X]$ ; v) a variância da soma ou diferença de duas variáveis aleatórias independentes é a soma das respectivas variâncias:  $Var[X \pm Y] = Var[X] + Var[Y]$  e vi) a variância da soma ou diferença de duas variáveis aleatórias dependentes é a soma das respectivas variâncias mais ou menos duas vezes a covariância entre ambas as variáveis  $Var[X \pm Y] = Var(X) + Var(Y) \pm 2Cov(X, Y)$ ; vii)  $Var[XY] = Cov(X, Y) = E[XY] - E[X]E[Y]$  se as variáveis aleatórias  $X$  e  $Y$  forem dependentes, caso elas sejam independentes então  $Var[XY] = Cov(X, Y) = 0$ .

**PROPRIEDADES DO VARIOGRAMA:** Vide amplitude, patamar, efeito pepita e variância espacial.

**PROSPECTIVO:** Concernente ao futuro; diz-se do estudo de um grupo seguido do presente ao futuro.

**PICTOGRAMA:** Representação gráfica para variáveis qualitativas nominais ou ordinais, cujas modalidades ou atributos são expressos por figuras que representam a intensidade de ocorrência da categoria, tipo, atributo ou evento estudado.

**PROVA DE BERNOULLI:** Experimento que admite apenas dois resultados possíveis, ou seja, sucesso ou fracasso.

**PROVA DE WALKER:** Prova de significância usada na análise periodográfica, dada por Walker (1914).

**PROVA ESTATÍSTICA NÃO PARAMÉTRICA:** É uma prova cujo modelo não especifica condições sobre os parâmetros da população da qual se extraiu a amostra.

**PROVA ESTATÍSTICA PARAMÉTRICA:** É uma prova cujo modelo especifica certas condições sobre os parâmetros da população do qual se extraiu a amostra para pesquisa.

**PROXY:** Palavra inglesa que significa procuração ou delegação; é utilizada, em português, para designar o informante ou técnica em que uma pessoa responde por outro ou pelos demais; o efeito proxy, em geral, denota subnotificação da informação.

**PSEUDOPROCEDIMENTO:** Procedimento semelhante ao real, feito em um paciente com a finalidade de o paciente e, às vezes, o médico não saber o procedimento adotado, para esse paciente, foi o real.

**PSEUDO R<sup>2</sup>:** Um valor de ajuste geral do modelo que pode ser calculado para regressão logística; comparável com a medida R<sup>2</sup> usada em regressão múltipla.

**PSEUDO-VEROSSIMILHANÇA:** Termo que é usado para indicar alguma função dos dados o qual só depende dos parâmetros de interesse e comporta-se de alguma forma como se fosse uma verossimilhança genuína.

**PUPILÔMETRO:** Dispositivo que mede mudanças na dilatação da pupila.

**P-VALOR:** Valor associado a uma estatística de teste que indica a probabilidade de um valor tão, ou mais extremo que o observado, ocorrer apenas por acaso em várias repetições de um experimento.

**P-VALOR:** É a probabilidade de cometer o erro de tipo I ou de primeira espécie o qual consiste em rejeitar a hipótese de nulidade  $H_0$  quando ela é na realidade verdadeira, com os dados de uma amostra específica. Este valor é dado pelo pacote ou software estatístico, assim o comparamos com o nível de significância escolhido e tomamos a decisão. Se o p-valor for menor que o nível de significância escolhido rejeitamos  $H_0$ , caso contrário, aceitamos  $H_0$ .

**P-VALOR:** É o menor nível de significância ou probabilidade a partir da qual começa a se rejeitar a hipótese estatística de nulidade  $H_0$ .

**P-VALOR:** É a probabilidade obtida em um teste de hipótese, inferência estatísticas, variando de 0 a 1. quando o p-valor é igual ou menor que o nível alfa previamente estabelecido, rejeita-se a hipótese de nulidade.

**P-VALUE:** Corresponde ao menor nível de significância que pode ser assumido para rejeitar a hipótese nula. Dizemos que há significância estatística quando o p-value é menor que o nível de significância adotado. Por exemplo, quando  $p=0,0001$  pode-se dizer que o resultado é bastante significativo, pois este valor é muito inferior aos níveis de significância usuais. Por outro lado, se  $p=0,048$  pode haver dúvida pois embora o valor seja inferior, ele está muito próximo ao nível usual de 5 %.

**P-VALOR:** A probabilidade de ter obtido um conjunto de dados se a hipótese nula for verdadeira. Em inglês P-value.

**POLÍGONO DE FREQUÊNCIAS:** É um gráfico de análise no qual as frequências das classes são localizadas sobre perpendiculares levantadas nos pontos médios das classes.

**POPULAÇÃO:** É o conjunto de elementos que apresentam uma ou mais características em comum.

**PARÂMETRO:** É um valor desconhecido associado a uma característica da população.

**PLANEJAMENTO DE EXPERIMENTOS AGRÍCOLAS:** Na Execução de experimentos de campo o pesquisador deve antes de começar a projetar um experimento ter uma ideia clara sobre os problemas que deseja resolver e sua importância relativa. Depois deve determinar o seguinte: i) O grau de precisão necessária no experimento; ii) Providenciar os materiais, a mão-de-obra, área e dinheiro necessários para executar o experimento. Inclusive água para irrigação e outras necessidades; iii) Verificar quais os resultados considerados já comprovados por experimentos anteriores (IGUE; LASCA, 1986) e iv) Providenciar proteção e segurança da área experimental, contra possíveis roubos, ataque de pássaros, roedores ou outros animais.

**PEREQUAÇÃO:** Do latim peraequatio, distribuição uniforme. É a operação que tem por fim substituir os termos de uma série irregular de dados experimentais descontínuos pelos valores de uma função, afim de se chegar a uma série regular, ou apenas mais regular que a primitiva e, eventualmente, contínua. Ver também, Perequação mecânica.

**PEREQUAÇÃO (MECÂNICA) DE WITTSTEIN, FÓRMULA DE:**

$$u'_i = \frac{1}{25}(u_i - 4 + 2u_i - 3 + 3u_i - 2 + 4u_i - 1 + 5u_i + 4u_i + 1 + 3u_i + 2 + 2u_i + 3 + u_i + 4)$$

em que  $u_i$  representa os termos sucessivos de uma série de dados experimentais e  $u'_i$  os valores que lhes devem ser substituídos.

**PEREQUAÇÃO MECÂNICA:** É todo processo de perequação que substitui uma sucessão descontínua de valores dados por outra sucessão descontínua de valores que são funções dos termos substituídos e de um certo número de termos adjacentes, na série primitiva. Os processos de perequação são divididos em gráfica, mecânica e analítica.

**PERMIL:** Conceito que significa qualquer uma das mil partes iguais em se pode dividir a área contida entre a curva normal de frequência e o eixo das abscissas.

$P(A)$ : Probabilidade de ocorrer o acontecimento A.

$P_r^n$ : Permutações de n objetos r a r.

$P_k$ : Valor do percentil amostral de ordem k.

$\rho_{XY}$ : Coeficiente de correlação populacional entre as variáveis aleatórias X e Y.

$p$ : Convergência em probabilidade.

$\vec{P}_r^n$ : Permutação de n objetos r a r.



# Q

**Q:** Em estatística é o valor da amplitude total estudentizada ao nível de erro de 5 % ou 1 % de probabilidade.

**QM (ERRO):** Quadrado médio para o erro usado em análise de variância.

**QM (TOTAL):** Quadrado médio para a variação total usado em análise de variância.

**QUADRADO MÉDIO DOS RESÍDUOS (DESVIOS) (QMR-QMD):** O mesmo que variância.

**QM (TRATAMENTO):** Quadrado médio para tratamentos usado em análise de variância.

**QUALITATIVO:** Adjetivo sinônimo de nominal, quando se referir a variáveis.

**QUARTIL:** Quando dividimos uma distribuição de dados em quatro partes iguais, cada uma delas leva o nome de quartil. Também é dado esse nome aos três quartis, normalmente denotados como  $Q_1$ ,  $Q_2$  e  $Q_3$ , que dividem a distribuição em quatro partes. O primeiro quartil separa os 25 % de valores inferiores dos 75 % superiores, o segundo quartil corresponde ao percentil 50 e à mediana, e o terceiro quartil separa os 75 % inferiores dos 25 % superiores.

**QUASE-EXPERIMENTO:** Tipo de delineamento de pesquisa que não cumpre todos os pré-requisitos necessários para ser considerado um delineamento experimental no qual é realizado uma seleção aleatória da amostra, controle e manipulação das variáveis [VI] e mensuração eficiente das variáveis [VD].

**QUESTIONÁRIO:** Técnica estruturada para coleta de dados; Tipo de instrumento de pesquisa que consiste num conjunto de perguntas escritas que devem ser respondidas pelos sujeitos.

**QUI-QUADRADO:** Tipo de distribuição teórica de probabilidade para variáveis aleatórias contínuas cujo domínio vai de zero a infinito a qual modela distribuições empíricas de dados amplamente utilizadas na estatística. Teste estatístico paramétrico e não paramétrico que possibilita o estudo da associação entre duas ou mais variáveis que possuam nível de mensuração nominal ou ordinal. Conhecido também pelo símbolo  $\chi^2$ , qui-quadrado nos ajuda a responder se as frequências observadas em nossa amostra diferem significativamente daquelas esperadas na população. A fórmula utilizada em seu cálculo é a seguinte:

$$\chi^2 = \sum \frac{(\text{observado} - \text{esperado})^2}{\text{esperado}}$$

**QMRES:** É uma estatística que significa o quadrado médio de resíduo.

**QTL:** Locos de características quantitativas.

$q_k$  : Valor do quartil amostral de ordem K.

**QUADRADO GRECO-LATINO:** Uma extensão do delineamento em quadrado latino. Formalmente, é um arranjo em um quadrado de dois conjuntos de letras (A,B,..., e  $\alpha,\beta,\dots$ ), uma de cada em uma cela do quadrado, de forma que nenhuma letra romana aconteça mais do que uma vez na mesma linha ou coluna,

nenhuma letra Grega aconteça mais de uma vez na mesma linha ou coluna, e nenhuma combinação das duas aconteça mais do que uma vez em qualquer lugar. Em inglês Graeco-Latin square.

**QUADRADO LATINO:** Delineamento experimental em que há dois tipos de blocos, chamados linhas e colunas.

**QUADRADO MÉDIO:** Quociente de uma soma de quadrados de desvios ou de resíduos pelo número de graus de liberdade. Quando essa soma de quadrados não inclui nenhum parâmetro, mas somente efeitos do acaso, o quadrado médio é uma estimativa da variância.

**QUADRADO MÉDIO:** Grandeza, utilizada em análise de variância, igual a uma soma de quadrados dividida pelos números de seus graus de liberdade.

**QUADRADO RETICULADO:** Delineamento do grupo dos reticulados em que há linhas e colunas, como nos quadrados latinos.

**QUADRADO YOUTEN:** Tipo de delineamento em blocos incompletos equilibrados, onde cada tratamento ocupa cada posição no bloco o mesmo número de vezes. Por exemplo, com 7 tratamentos, a,b,c,-d,e,f,g, em 7 blocos a constituição do desenho se dá em colunas de 3 unidades cada. Sendo assim tem-se o seguinte plano:

<i>I</i>	<i>II</i>	<i>III</i>	<i>IV</i>	<i>V</i>	<i>VI</i>	<i>VII</i>
<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>
<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>a</i>
<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>a</i>	<i>b</i>	<i>c</i>

O conjunto de 7 blocos constitui, por assim dizer um retângulo latino. Também se diz quadrado latino incompleto, dado por Youden (1940).

**QUADRADO LATINO (LATIN SQUARE):** Um delineamento de blocos aleatorizados no qual  $n$  tratamentos são definidos em um quadrado  $n \times n$  e que cada tratamento aparece exatamente uma vez em cada linha e em cada coluna do campo.

**QUADRADO MÉDIO (MEAN SQUARE):** A média das diferenças entre o valor de cada observação e a média aritmética dos valores, elevadas ao quadrado. Também é chamado de variância, ou segundo momento central.

**QUANTIL (QUANTILE):** Corresponde a  $n$ -ésima parte de um conjunto de dados. Tipos específicos de quantis incluem os decis, percentis e quartis.

**QUARTIL (QUARTILE):** Um quarto de uma distribuição de dados. Os quartis, superior e inferior, de um conjunto de dados são os 25 % do fundo e os 25 % do topo dos dados.

**QUADRANTE:** É o método de pesquisa para localização de pontos vizinhos próximos ao ponto a ser interpolado. Por este método a região em torno do ponto a ser interpolado é subdividido em quatro quadrantes dentro dos quais são localizados e amostrados os pontos mais próximos, garantindo assim uma boa amostragem espacial.

**QUADRO DE GALTON:** Forma de demonstração que mostra a distribuição normal a partir da queda de pequenas esferas. O quadro é de autoria de Francis Galton. Em inglês Quincux.

**QUALIDADE:** Um fenômeno pode ter sua qualidade controlada quando sua variabilidade futura pode ser prevista, no mínimo dentro de limites, a partir do conhecimento da sua variabilidade passada. A previsão dentro de limites, significa que a probabilidade que o fenômeno observado irá cair dentro de dados limites, pode ser determinada aproximadamente.

**QUALIDADE DE AJUSTE:** Grau em que a matriz de entrada real ou observada que é a de covariâncias ou correlações é prevista pelo modelo estimado. Medidas de qualidade de ajuste são computadas somente para a matriz total de entrada, não fazendo qualquer distinção entre construtos ou indicadores exógenos e endógenos.

**QUALQUER QUE SEJA:** Símbolo matemático dado por  $\forall$ .

**QUANTIS:** O conjunto de  $n-1$  valores de uma variável que a divide em um número  $n$  de iguais proporções. Por exemplo,  $n-1=3$  valores divide os dados em quatro quantis, com o valor central geralmente chamado de mediana e os valores inferior e superior geralmente chamados de quartis inferior e superior, respectivamente.

**QUARTIL(IS):** Os três valores de uma variável que a divide em quatro partes iguais. O valor central é geralmente chamado de mediana e os valores inferior e superior são geralmente chamados de quartis inferior e superior, respectivamente. Veja também quantis.

**QUANTIFICAÇÃO:** Conversão de dados para formato quantitativo ou numérico por meio de codificação.

**QUANTIFICAÇÃO DE GUEDES:** Indicador que tem a vantagem de fornecer uma tradução numérica para as curvas de mortalidade proporcional, ensejando a comparação e a evolução do nível de saúde para diferentes localidades. É calculada multiplicando-se os respectivos pesos dos grupos etários (em anos): -4 (<1), -2 (1-4), -1 (5-19), -3 (20-49) e +5 ( $\geq 50$ ) – em seguida, procede-se à soma algébrica dos valores encontrados e, por último, divide-se por 10.

**QUANTIS:** De um modo geral, pode-se definir uma medida descritiva de localização, chamada quantil de ordem  $p$  ou  $p$ -quantil, indicada por  $q(p)$ , onde  $p$  é uma proporção qualquer,  $0 < p < 1$ , tal que  $100p$  % das observações sejam menores do que  $q(p)$ , por exemplo a mediana é o quantil de ordem 0,50 ou 5º decil ou ainda o 50º Percentil.

**QUANTIS:** Os quantis são uma forma de divisão da distribuição em classes com o mesmo número de elementos. Os quantis mais utilizados são: Quartis e os Percentis.

**QUANTITATIVA, ABORDAGEM:** Um elo comum em todas as variáveis da abordagem quantitativa é que nela se emprega números ou símbolos para representar fenômenos de interesse ou modelo da realidade.

**QUARTIL:** Cada um dos três valores que dividem uma distribuição de frequências em quatro partes de frequências iguais. O primeiro quartil corresponde ao 25º percentil, o segundo à mediana e o terceiro ao 75º percentil.

**QUARTIL:** São três valores que dividem um conjunto de dados em quatro partes iguais. Em inglês quartil.

**QUARTIL:** Há três valores que separam a frequência total de uma distribuição em quatro partes iguais. O valor central é chamado de mediana e os outros dois o quartil superior e o inferior respectivamente. Em inglês Quartile.

**QUARTIL:** O primeiro quartil de uma lista é o número tal que um quarto dos números da lista está abaixo dele, o terceiro quartil é o número tal que três quartos dos números da lista abaixo dele e o segundo quartil coincide com a mediana.

**QUARTIL:** Cada um dos três valores que dividem uma distribuição de freqüências em quatro partes de freqüência igual. O primeiro quartil correspondente ao 25 percentil, o segundo à mediana e o terceiro ao 75 percentil. Em inglês *quartile*.

**QUARTIL MÉDIO:** Semi-soma do primeiro e do terceiro quartis.

**QUARTIS:** Denominamos quartis os valores de uma série que a dividem em quatro partes iguais. Precisamos, portanto de 3 quartis ( $Q_1$ ,  $Q_2$  e  $Q_3$ ) para dividir a série em quatro partes iguais. Observação: O quartil 2 ( $Q_2$ ) sempre será igual a mediana da série.

**QUARTIS:** Os quartis dividem a distribuição em 4 partes iguais. Estando os dados por ordem crescente, o 1º quartil e o 3º quartil acumulam até si, respectivamente, 25 % e 75 % dos dados. Isto significa que o âmbito inter-quartil abrange as observações centrais, que correspondem a 50 % das observações totais não estando por isso incluídas 25 % das menores nem 25 % das maiores observações. As características são as seguintes: O 1º quartil é igual ao percentil 25. O 2º quartil é igual ao percentil 50. O 3º quartil é igual ao percentil 75.

**QUARTIS:** Medida de tendência central com características de que seu valor divide os termos da série em grupos de quatro partes. O segundo quartil dessa divisão coincide com a mediana. Tem sua aplicação para divisão da série em quatro partes iguais permitindo sínteses para comparações.

**QUARTIS ( $Q_1$  E  $Q_3$ ):** São os valores ou medidas estatísticas descritivas que dividem a distribuição em quatro partes iguais. Com  $n$ =número de observações da amostra.

**QUASE CERTO:** Uma probabilidade dependente de um parâmetro  $n$ , tenderá a 1 (um) quando  $n$  tender a infinito. O evento ao que se relaciona esta probabilidade é chamado de quase certo no limite. Em inglês *Almost certain*.

**QUASE-EXPERIMENTOS:** São similares aos experimentos com teste de hipótese, exceto que envolvem variáveis independentes cujos níveis são selecionados de variações já existentes ou conhecidas, em vez de serem deliberadamente manipuladas pelo pesquisador. Por isso, não se pode aludir relacionamentos causais, quando se trabalha com quase-experimentos.

**QUASE ORTOGONAL:** Característica de um delineamento de estímulos que não é ortogonal, mas no qual os desvios da ortogonalidade são pequenos e cuidadosamente controlados na geração dos estímulos. Esse tipo de delineamento pode ser comparado com outros delineamentos de estímulos com medidas de eficiência de delineamento.

**QUASE VEROSSIMILHANÇA:** O termo foi introduzido por Wedderburn (1974) para descrever funções com propriedades semelhantes da função suporte (*log-likelihood*), exceto que não corresponde a função suporte de nenhum modelo conhecido de probabilidade. Modelos de quase verossimilhança podem ser ajustados por métodos de modelos lineares generalizados, usada em modelos com superdispersão.

**QUESTIONÁRIO:** Documento com perguntas e outros tipos de itens que visam obter informações para análise. Questionários são principalmente usados em pesquisas de survey e também em experimentos, pesquisas de campo e outros modos de observação.

**QUESTIONÁRIO:** Conjunto de perguntas para gerar os dados necessários para realizar os objetivos do projeto de pesquisa.

**QUESTIONÁRIO RASCUNHO:** Um rascunho do instrumento de pesquisa preparado na conclusão do processo preliminar de coleta de dados e antes da implementação do pré-teste.

**QUESTIONÁRIO (VALIDADE):** Como instrumento de pesquisa, o questionário está sujeito a muitas ameaças contra sua validade. Pode-se dizer que não existem critérios universais para validade de questionários, pois a validade deles está em seus itens individuais assim como geral.

**QUESTIONÁRIOS (COMO INSTRUMENTO DE PESQUISA):** O instrumento mais usado para levantamento de informações, desde 1847, quando aparentemente foi empregado pela primeira vez com esse fim. Hoje o seu uso vem sofrendo um declínio; tem-se reconhecido suas limitações e tem-se buscado limitar seu uso a situações apropriadas. Comparando com a entrevista, ele permite maior abrangência e menor esforço e custo. Respostas obtidas por meio dele podem ser mais genuína que as obtidas por meio de entrevistas. Além disso, ele garante maior uniformidade das perguntas, o que favorece a tabulação das respostas. Seu grande problema está em uma freqüente baixa de retorno.

**QUESTIONÁRIOS AUTO-ADMINISTRADOS:** Tipos de questionários preenchidos pelos respondentes ou entrevistados numa pesquisa, sem a intervenção do entrevistador.

**QUESTÕES MATRICIAIS:** Formato de questionário no qual uma tabela de itens é mostrada aos respondentes, cada um sendo respondido segundo um conjunto padronizado de respostas, tais como concordo fortemente, concordo, dentre outros.

**QUI-QUADRADO:** Teste de significação estatística não paramétrico apropriado para variáveis nominais e ordinais.

**QUI-QUADRADO:** Método para padronização de dados em uma tabela de contingência que compara as freqüências reais das células da tabela com as freqüências esperadas das mesmas. A freqüência esperada de uma célula da tabela de contingência é baseada nas probabilidades marginais de sua linha e coluna que é a probabilidade de uma linha e coluna entre todas as linhas e colunas.

**QUI-QUADRADO:** O teste qui-quadrado  $\chi^2$ . Em inglês *chi-square*.

**QUI-QUADRADO:** Um meio para responder perguntas sobre dados existentes na forma de freqüência, em vez de escores ou mensurações numa escala. A pergunta que desejamos responder, quando temos dados em freqüência é se as freqüências observadas em nossa amostra desvia significativamente das freqüências teóricas ou esperadas na população.

**QUI-QUADRADO, TESTE:** Método estatístico para testar a hipótese de que dois fatores são independentes. Em inglês *chi-square test*.

**QUOCIENTE DE FORMA:** Razão do diâmetro a uma determinada altura do tronco pelo DAP.

**QUADRADO DO COEFICIENTE DE CORRELAÇÃO ( $r^2$ ):** A proporção da variância de uma variável indicada por sua relação linear com outra variável. É também conhecido com coeficiente de determinação.

**QUADRADO MÉDIO:** A soma de quadrados dividida pelo grau de liberdade; uma estimativa de variância. O mesmo que variância.

**QUADRADO MÉDIO RESIDUAL:** Consultar variância residual.

**QUARTIL INFERIOR:** O 25º percentil.

**QUARTIL SUPERIOR:** O 75º percentil.

$\chi^2_\alpha$  : Quantil 1-  $\alpha$  da distribuição qui-quadrado.

$\chi^2_{(v)}$  : Distribuição qui-quadrado com  $v$  graus de liberdade.



# R

$\mathfrak{R}$  : Método  $\mathfrak{R}$  (Regressão) de estimação de componentes de variância.

$\mathbb{R}$  : Conjunto dos números reais.

**r**: Coeficiente de correlação amostral.

**RAR**: Redução Absoluta do Risco representado por  $p_1 - p_2$ .

**$r^2$  AJUSTADO (ADJUSTED  $r^2$ )**: Quão bem um modelo de regressão se ajusta aos dados, descontando o número de parâmetros no modelo. É usado para evitar o superajuste de um modelo simplesmente incluindo variáveis preditoras adicionais. Ver também coeficiente de determinação.

**RANDOMIZAÇÃO (RANDOMIZATION)**: A atribuição de tratamentos experimentais a populações com base em tabelas de números, ou algoritmos, randômicos. Comparar com casual. O mesmo que aleatório ou ao acaso.

**RASTRO DE AUDITORIA (AUDIT TRAIL)**: O conjunto de arquivos e documentos associados que descrevem mudanças em um conjunto original de dados. Ela descreve todas as mudanças feitas no arquivo de dados original que resultaram no conjunto de dados tratado usado nas análises.

**RAZÃO DE CHANCES A POSTERIORI (POSTERIOR ODDS RATIO)**: Em uma análise bayesiana, a razão da distribuição de probabilidades a posteriori de duas hipóteses alternativas.

**RAZÃO DE CHANCES A PRIORI (PRIOR ODDS RATIO)**: A razão da distribuição de probabilidades a priori de duas hipóteses alternativas.

**RAZÃO DE POISSON (RATE PARAMETER)**: A constante que é inserida na fórmula de uma variável aleatória de Poisson. Além disso, a média, ou valor esperado, e a variância de uma variável aleatória de Poisson.

**RACIOCÍNIO (APLICAÇÃO EM PESQUISA)**: Um instrumento indispensável na produção de verdade científica, mas não pode gerar ou mesmo identificar a verdade. Entretanto, é muito útil para: i) sugerir hipóteses; ii) avaliar compatibilidade de hipóteses com conhecimentos existentes; iii) interpretar resultados de testes de hipóteses.

**RACIOCÍNIO DEDUTIVO**: Raciocínio que parte do geral, isto é, de suposições, de proposições e de fórmulas consideradas verdadeiras para o específico, ou seja, é, para elementos específicos pertencentes à categoria geral.

**RACIOCÍNIO INDUTIVO**: Raciocínio que parte do geral, isto é, dos dados para o geral, ou seja, para elementos específicos pertencentes a categoria geral.

**RACIONAL**: Razoável; que faz sentido. Racionalidade é um paradigma dentro do qual a ciência opera e que é implícito em grande parte da vida cotidiana. É importante compreender que a vida não é racional,

mas que racionalidade é uma moldura que impomos às nossas experiências e observações num esforço para lidar eficazmente com a vida.

**RAIZ CARACTERÍSTICA:** Ver Autovalor. Em inglês Characteristic root.

**RAIZ QUADRADA:** Símbolo matemático dado por  $\sqrt{X}$ .

**RAIZ LATENTE:** Ver autovalor.

**RAÍZES CANÔNICAS:** Correlações canônicas ao quadrado, o que fornece uma estimativa da quantidade de variância compartilhada entre as respectivas variáveis estatísticas canônicas otimamente ponderadas de variáveis dependentes e independentes. Também são conhecidas como autovalores.

**REGRESSÃO:** Conjunto de técnicas estatísticas que tem por objetivo determinar a capacidade que uma variável possui de prever o comportamento de outra. Por exemplo: pode-se determinar a renda de um conjunto de sujeitos que é a variável dependente a partir do conhecimento do seu nível educacional que é a variável independente. Difere da correlação na medida em que esta última não avalia a causalidade entre variáveis.

**REGRESSÃO BIVARIADA:** Processo de regressão entre duas variáveis, uma independente que é a suposta causa e outra dependente a qual é o suposto efeito.

**REGRESSÃO MÚLTIPLA:** Processo de regressão entre uma variável dependente que é o suposto efeito e duas ou mais variáveis independentes que são as supostas causas.

**REGRESSÃO MULTIVARIADA:** A regressão linear múltipla também é uma técnica multivariada cuja finalidade principal é obter uma relação matemática entre uma das variáveis que é a variável dependente e o restante das variáveis que descrevem o sistema variáveis independentes. Sua principal aplicação, após encontrar a relação matemática é produzir valores para a variável dependente quando se têm as variáveis independentes. Ou seja, ela pode ser usada na predição de resultados. Ver análise multivariada e regressão múltipla.

**RHO:** Letra grega ( $R, \rho$ ), símbolo do coeficiente de correlação linear simples de Pearson da população bivariada de pares de valores ( $X, Y$ ).

**RANDÔMICO OU RANDOMIZADO:** Aleatório, casual, estocástico, ao acaso, sorteio.

**RANDOMIZAÇÃO:** O mesmo que aleatorização e casualização. Processo de alocação, ao acaso, de unidades destinadas a compor os grupos de estudo e controle; técnica utilizada experimentais. Ver estudo randomizado

**RANDOMIZAÇÃO:** Envolve alocar os sujeitos disponíveis para um ou para outro grupo de estudo e é geralmente utilizada em estudos clínicos.

**RANDOMIZAÇÃO:** Refere-se à maneira em que fatores estranhos, que originam os erros ou resíduos são distribuídos. Usado para eliminação de tendências ou vieses. Consiste, em experimentos, na escolha de dois membros de um par, sendo um membro usado para controle, e outro para experimento.

**RANDOMIZAÇÃO:** Veja casualização.

**RAZÃO-F DE FISHER (FISHER'S F-RATIO):** É uma estatística teste-F usada na análise de variância univariada (ANAVA ou ANOVA).

**RDA:** Ver análise de redundância.

**REGRA DOS 10 (RULE OF 10):** Um princípio guia o qual diz que você deveria ter no mínimo 10 réplicas por categoria observacional ou grupo de tratamento. A regra baseia-se na experiência, e não em algum axioma ou teorema matemático.

**RANK ANALYSIS MEASUREMENT SYSTEM (RAMS):** Máquina que permite aos respondentes registrar como se sentem pelo manuseio de um dial.

**RASTREAMENTO (TRIAGEM OU SREENING):** Identificação, entre indivíduos aparentemente sadios, daqueles suspeitos de estarem enfermos ou que apresentam alto risco de danos à saúde.

**RAZÃO:** Nível de mensuração de dados; caracteriza dados que podem ser ordenados, para os quais as diferenças entre dados são significativas; há um ponto de partida zero inerente.

**RAZÃO:** Medida de frequência de um grupo de eventos relativa à frequência de um grupo distinto de eventos. É um tipo de fração em que pelo menos parte dos elementos do numerador não está contida no denominador, vale lembrar que o numerador não é subconjunto do denominador. Pode ou não ter unidades, como por exemplo: número de leitos hospitalares por 1000 pessoas da população e número de morte infantis em um ano por 1000 nascimentos vivos naquele ano ou, então em razão entre duas taxas ou proporções.

**RAZÃO AMOSTRAL:** Proporção de elementos na população selecionados para a amostra.

**RAZÃO DE CHANCES:** É a chance a favor dividida pela chance contra, ou seja, é o número de casos favoráveis dividido pelo número de casos não favoráveis. Em inglês **odds ratio**.

**RAZÃO DE CHANCES:** Ver odds ratio.

**RAZÃO DE CHANCES:** Um termo alternativo para Risco relativo. Em inglês Odds ratio.

**RAZÃO DE CHANCE (ODDS RATIO):** O mesmo que razão de chance de exposição. Comparação da presença de um fator de risco para uma doença em uma amostra de sujeitos doentes e não doentes como controle. Razão entre o número de pessoas com a doença que foram expostas a um fator de risco ( $I_e$ ) por aqueles com a doença que não foram expostos ( $I_o$ ), divididos pela razão entre aqueles sem a doença que foram expostos ( $N_e$ ) por aqueles sem doença que não forma expostos ( $N_o$ ). Assim,

$$OR = \frac{\left(\frac{I_e}{I_o}\right)}{\left(\frac{N_e}{N_o}\right)} = \frac{I_e N_o}{I_o N_e}$$

Esta medida deveria ser usada para estudos de casos controle quando queremos observar retrospectivamente a ação de fatores de risco naqueles com e sem a doença.

**RAZÃO DE DESIGUALDADE:** A comparação da probabilidade de um evento com a probabilidade de o evento não acontecer, a qual é usada como a variável dependente em regressão logística.

**RAZÃO DE MORTALIDADE PADRONIZADA:** Total de eventos observados no grupo de estudo, dividido pelo número de eventos esperados com base nos coeficiente da população padrão aplicados ao grupo de estudo. A constante multiplicadora para essa medida é igual a 100.

**RAZÃO DE MORTALIDADE PADRONIZADA (RMP):** Coeficiente geral de mortalidade observado dividido pelo coeficiente geral de mortalidade esperado. A RMP é, geralmente, multiplicada por 100, com a população padrão assumindo um valor de 100. Se é maior do que 100, a força de mortalidade é menor na população estudada que na população padrão.

**RAZÃO DE PROBABILIDADES NEGATIVA:** Razão entre a taxa de erros falso-negativos de testes e a especificidade desse teste.

**RAZÃO DE PROBABILIDADES POSITIVA:** Razão entre a taxa de erros falso-positivos de um teste e a sensibilidade desse teste.

**RAZÃO DE PRODUTOS CRUZADOS:** Ver odds ratio.

**RAZÃO DE SUCESSO:** Percentual de objetos, indivíduos, respondentes, empresas, dentre outros corretamente classificados pela função discriminante e calculada dividindo-se o número de objetos na diagonal da matriz de classificação pelo número total de objetos.

**RAZÃO DE VEROSSIMILHANÇA (RV):** Se  $x_1, x_2, \dots, x_n$  for uma amostra aleatória de uma população  $f(x; \theta_1, \theta_2, \dots, \theta_k)$  a verossimilhança desta amostra específica será:  $L = \prod_i f(x_i; \theta_1, \theta_2, \dots, \theta_k)$ . Esta terá um máximo, com respeito aos  $\theta$  no espaço de parâmetros  $\Omega$ , que pode ser escrito  $L(\Omega_0)$ . Para um subespaço  $\omega$  do espaço de parâmetro, existirá um valor máximo  $L(\omega_0)$ . A hipótese nula  $H_0$  de que uma população específica em estudo pertença ao sub-espaço  $\omega$  de  $\Omega$  pode ser testada usando-se a razão de verossimilhança

$$\lambda = \frac{L(\omega_0)}{L(\Omega_0)}, 0 \leq \lambda \leq 1 \text{ ou alguma função simples desta. Em inglês Likelihood ratio (LR).}$$

**RAZÃO PARA A MÉDIA MÓVEL:** Método que se verifica uma retração na atividade econômica de um país.

**RAZÕES CRÍTICAS:** Classe de testes de significância estatística que dependem da divisão de alguns parâmetros, ou seja, a diferença entre as médias pelo desvio padrão (EP) daquele parâmetro.

**RAZÕES DE CHANCES (RC) OU ODDS RATIO (OR):** Chance de exposição no grupo doente dividida pela chance de exposição no grupo não doente.

**REAÇÃO GALVÂNICA DA PELE (GSR):** A mensuração de mudanças na resistência elétrica da pele associada a respostas de ativação.

**REALISMO EXPERIMENTAL:** Diz-se que um estudo tem alto grau de realismo experimental, quando os tratamentos empregados numa investigação têm um impacto real sobre o sujeito, e o forçam a atender às demandas de pesquisa. O sujeito responde à manipulação de forma natural e de maneira espontânea.

**REALISMO MUNDANO:** Referem-se ao grau em que as várias dimensões do experimento por exemplo, instruções, tratamentos, operações de medida, dentre outros refletem o mundo real, eventos non-laboratoriais.

**REGRESSÃO (REGRESSION):** O método de análise usado para explorar relações de causa e efeito entre duas variáveis, cuja variável preditora é contínua. Comparar com correlação.

**REGRESSÃO GRADATIVA (STEPWISE REGRESSION):** Qualquer técnica ou tipo de procedimento de regressão que envolve avaliar múltiplos modelos com parâmetros compartilhados. O objetivo é decidir quais parâmetros são úteis ao modelo de regressão final. Os exemplos de métodos usados na regressão gradativa incluem eliminação regressiva e seleção progressiva.

**REGRESSÃO LOGÍSTICA (LOGISTIC REGRESSION):** Um modelo estatístico para estimar variáveis resposta categóricas a partir de uma variável preditora contínua.

**REGRESSÃO MÚLTIPLA (MULTIPLE REGRESSION):** Um modelo estatístico que prediz uma resposta a partir de mais de uma variável preditora. Pode ser linear ou não linear.

**REGRESSÃO QUANTÍLICA (QUANTILE REGRESSION):** Um modelo de regressão sobre um subconjunto dos dados, onde este subconjunto refere-se a um quantil definido pelo analista.

**RECÍPROCA:** Ver não recursiva.

**REGRA DOS 10:** Tipo de escolha ou decisão do pesquisador batizada de a regra dos 10, a qual estabelece que o investigador deve coletar pelo menos 10 réplicas para cada categoria ou nível de tratamento. Por exemplo, suponha que tenha determinado que possa coletar um total de 50 observações em um experimento, examinando taxas fotossintéticas entre diferentes espécies de plantas. Um bom delineamento para uma análise de variância (ANOVA) de um fator (one-way) poderia ser comparar as taxas fotossintéticas entre não mais que 5 espécies. Para cada uma, você poderá escolher aleatoriamente 10 plantas e tirar uma medida de cada. A Regra dos 10 não se baseia em nenhum princípio teórico de delineamento experimental ou de análise estatística, mas é uma reflexão de nossa experiência de campo, obtida através de muito esforço, com delineamentos que foram bem sucedidos e aqueles que não foram. Certamente, é possível analisar conjuntos de dados com menos de 10 observações por tratamento, e frequentemente se quebra a regra. Delineamentos balanceados com muitas combinações de tratamentos, mas com apenas quatro ou cinco réplicas, podem ser bastante poderosos. E certos delineamentos de uma via com poucos níveis de tratamentos podem requerer mais de 10 réplicas por tratamento, se as variâncias forem grandes. Todavia, a Regra dos 10 é um ponto de partida sólido. Mesmo se o pesquisador estabelecer o delineamento com 10 observações por nível de tratamento, é pouco provável que termine com esse número. A despeito de seus melhores esforços, os dados podem ser perdidos por uma variedade de razões incluindo a falha nos equipamentos, os desastres climáticos, as parcelas perdidas, os distúrbios ou erros humanos, a transcrição imprópria de dados e as alterações ambientais. A Regra dos 10 possibilita a coleta de dados com um poder estatístico razoável para revelar padrões.

**REGRESSÃO QUANTÍLICA:** Tipo de análise de regressão, onde os desvios são determinados em módulo e ponderados por um fator  $h$ . A regressão linear simples ajusta a linha através do centro da nuvem de pontos e é apropriada para descrever uma relação direta de causa e efeito entre as variáveis  $X$  e  $Y$ . No entanto, se a variável  $X$  atua como fator limitante, ela pode impor um teto para os valores de  $Y$ . Assim a variável  $X$  pode controlar o valor máximo para o  $Y$ , mas não ter nenhum efeito abaixo desse máximo. O resultado poderia ser um gráfico com forma de triângulo. A regressão quantílica minimiza os desvios a partir da linha de regressão, mas a função de minimização é assimétrica: desvios positivos e negativos são ponderados diferentemente. O resíduo é calculado através da seguinte equação:

$resíduo = \sum_{i=1}^n |Y_i - \hat{Y}_i| h$ . Como na regressão robusta, a função minimiza o valor absoluto dos desvios, de

forma que não seja sensível aos dados discrepantes. No entanto, a característica chave é o multiplicador  $h$ . O valor do  $h$  é o quantil a ser estimado. Se o sinal do desvio dentro do valor absoluto for positivo, ele será multiplicado por  $h$ . Se o desvio for negativo, ele será multiplicado por  $(1,0-h)$ . Essa minimização assimétrica ajusta uma linha de regressão através da região superior dos dados para um  $h$  grande e através da região inferior para um  $h$  pequeno.

**REGRESSÃO ROBUSTA (ROBUST REGRESSION):** Um modelo de regressão que é relativamente pouco sensível a dados discrepantes ou valores extremos. Exemplos incluem a regressão de mínimos quadrados aparados e os estimadores-M.

**REAMOSTRAGEM:** Estimção não-estatística do intervalo de confiança de um parâmetro por meio do uso de uma distribuição amostral empiricamente obtida. A distribuição empírica é calculada a partir de múltiplas amostras extraídas da amostra original.

**RECENSEAMENTO:** O mesmo que Censo. Mas alguns autores chama de recenseamento o processo de escolha de todos os elementos de uma população, e censo os dados provenientes do recenseamento.

**RECOMENDAÇÕES:** Conclusões aplicadas a estratégias ou táticas de marketing que enfocam a realização de um cliente em termos de vantagem de diferencial.

**RECOLHA DE DADOS:** É a primeira etapa depois de definido o problema em estudo.

**RECURSO:** É a reserva in situ que por motivos econômicos ou tecnológicos não é viável no momento atual, podendo tornar-se viável no futuro.

**REDAÇÃO (DELINEAMENTO):** É mais que um mapa literário útil para manter o autor numa trilha lógica; é uma proteção contra repetições e omissões, é um guia à estratificação do assunto. Dá unidade e coerência à redação. Como uma estrutura mecânica, o delineamento pode ser considerado simplesmente como uma lista ordenada de palavras, frases ou sentenças que determinam ou sugerem a natureza e a seqüência dos tópicos a serem discutidos.

**REDAÇÃO CIENTÍFICA:** Trabalho escrito sobre assuntos científicos e sobre vários assuntos técnicos associados com ciências. É caracterizada por certos elementos formais, tais como: seu vocabulário técnico e científico, uso de auxílios gráficos e argumentação objetiva. Transmite informação precisa e concisamente, e não tenta de nenhuma forma despertar emoção.

**REDE DE AMOSTRAS:** Conjunto de amostras interpenetrantes.

**REDE NEURAL:** Modelo preditivo não linear que aprende por treinamento. Assemelha-se à estrutura de sistemas neurais biológicos.

**REDE NEURAL:** Programa de computador que imita os processos do cérebro humano capazes de aprender de exemplos para encontrar padrões em dados.

**REDUÇÃO DE CATEGORIAS:** É a operação de combinar categorias ou intervalos de valores de uma variável para produzir um número menor de categorias. Em inglês bracketing, collapsing or grouping.

**REDUÇÃO PROPORCIONAL DE ERRO (RPE):** Método lógico para conceituar associações entre variáveis. Por exemplo, se tivermos de adivinhar a filiação político-partidária de uma amostra de pessoas, cometeremos muitos erros. Mas cometeremos menos erros se soubermos a filiação dos seus pais e apostarmos que é a mesma. A redução de erros, enquanto percentual, é o RPE.

**REDUÇÃO MULTIDIMENSIONAL:** O objetivo é transformar julgamentos de semelhança ou preferência como por exemplo, por lojas ou marcas em distâncias representadas no espaço multidimensional.

**REDUCIONISMO (INTERPRETAÇÃO):** Consideram os conceitos e variáveis de uma determinada ciência como as mais relevantes para entender um grande spectrum do comportamento humano.

**REDUÇÃO PROPORCIONAL DO ERRO:** A extensão até onde a variável independente serve para reduzir o erro na predição da variável dependente.

**REESPECIFICAÇÃO DE MODELO:** Modificação de um modelo existente com parâmetros estimados para corrigir parâmetros inadequados encontrados no processo de estimação ou para criar um modelo concorrente para comparação.

**REFUTAÇÃO:** Termo usado para indicar a modalidade de raciocínio que tem por finalidade testar a veracidade de uma tese.

**RELAÇÃO TEMPORAL:** Tipo de correlação temporal entre as variáveis estudadas em uma pesquisa conduzida em períodos contíguos. É muito usada na meteorologia quando se estuda série histórica de dados climáticos. A pesquisa onde se utiliza o fator tempo pode ser denominada de longitudinal ou transversal.

**REGIÃO CRÍTICA:** Conjunto de todos os valores da estatística de teste que determinam a rejeição da hipótese nula.

**REGIÃO CRÍTICA:** Se o valor da estatística de teste está na região crítica, a hipótese nula é rejeitada. É o mesmo que região de rejeição da hipótese  $H_0$ .

**REGIÃO CRÍTICA:** É a região sob a curva da distribuição amostral do estimador ou da estatística do teste, a qual é dimensionada de forma a que a probabilidade dessa estatística cair nessa região, dada  $H_0$ , seja igual ao erro de tipo I. O mesmo que região de rejeição da hipótese de nulidade  $H_0$ .

**REGIÃO CRÍTICA:** Um teste estatístico de hipótese é feito com base em uma divisão do espaço amostral em duas regiões exclusivas mutuamente. Se o ponto amostral cai em uma região de aceitação, a hipótese é aceita; se está localizado na outra região que é a região de rejeição, será rejeitada. Ambas são de certa forma, críticas, mas, comumente denota-se a segunda pelo termo de região crítica. Em inglês Critical region.

**REGIÃO CRÍTICA OU DE REJEIÇÃO:** A área sob a distribuição amostral que é determinada de acordo com o nível de significância do teste. Em inglês critical region or region of rejection.

**REGIÃO DE ACEITAÇÃO:** Na teoria de teste de hipóteses, a região do espaço amostral de forma que, se um ponto amostral cair dentro dela, a hipótese testada é aceita. Do inglês Acceptance region.

**REGIÃO DE REJEIÇÃO:** Conjunto de valores da estatística de teste, para os quais a hipótese nula é rejeitada.

**R<sup>2</sup>:** Coeficiente de determinação ou explicação o qual é uma grandeza usada em modelos de regressão para medir a proporção da variabilidade total na resposta considerada pelo modelo. Computacionalmente,



$R^2 = \frac{SQ_{Regressão}}{SQ_{Total}}$  e grandes valores de  $R^2$  como próximos a um são considerados bons. Entretanto, é

possível ter grandes valores de  $R^2$  e o modelo não ser satisfatório.  $R^2$  é também chamado de coeficiente de determinação ou coeficiente de determinação múltipla na regressão múltipla, expresso em porcentagem.

**$R^2$  AJUSTADO:** Uma variação da estatística  $R^2$  que compensa o número de parâmetros em um modelo de regressão. Essencialmente, o ajuste é uma penalidade pelo aumento do número de parâmetros do modelo.

**$R^2$  CANCELADO:** Medida de multicolinearidade usada em análise de regressão, construída em alguns programas computacionais eletrônicos usados para análise de regressão de mínimos quadrados, onde o pacote estatístico no computador calcula não só o  $R^2$  comum para a equação de regressão, sob estudo, mas também aqueles  $R^2$  que são obtidos omitindo-se cada uma das variáveis independentes, por sua vez, os  $R^2$  obtidos desta maneira são chamados de  $R^2$  cancelado. Assim por exemplo, para uma equação de regressão com três variáveis independentes,  $X_{2i}$ ,  $X_{3i}$  e  $X_{4i}$  tem-se teríamos três  $R^2$  cancelados, um da regressão de  $Y_i$  sobre  $X_{2i}$  e  $X_{3i}$ , outro da regressão de  $Y_i$  sobre  $X_{2i}$  e  $X_{4i}$  e outro da regressão de  $Y_i$  sobre  $X_{3i}$  e  $X_{4i}$ . Se houver um alto grau de multicolinearidade na amostra, pelo menos uma das variáveis independentes variará grandemente de acordo com a variação de uma ou mais variáveis independentes. A introdução desta variável na equação de regressão só levará, portanto a um pequeno aumento no valor de  $R^2$ . Assim um alto grau de multicolinearidade será refletido pelo fato de que a diferença entre  $R^2$  e o mais alto dos cancelamentos de  $R^2$  será pequena.

**RAZÃO DE CAPACIDADE DO PROCESSO:** Uma razão que relaciona a largura dos limites de especificação do produto e medidas de desempenho do processo. Usada para quantificar a capacidade do processo para produzir produtos dentro das especificações. Veja capacidade de processo, estudo de capacidade de processo, PCR e  $PCR_k$ .

**RAZÃO DE PROBABILIDADES:** A probabilidade é igual à razão de duas probabilidades. Em regressão logística, o logaritmo da probabilidade é modelado como uma função linear dos regressores. Dados os valores para os regressores em um ponto, a probabilidade pode ser calculada. A razão de probabilidades é a probabilidade em um ponto dividida pela probabilidade em outro ponto.

**RAZÃO DE VEROSSIMILHANÇA:** Seja  $x_1, x_2, \dots, x_n$  uma amostra aleatória proveniente de uma população  $f(x; \theta)$ . A função verossimilhança para essa amostra é

$$L = \prod_{i=1}^n f(x_i; \theta). \text{ Queremos testar a hipótese } H_0: \theta \in \omega, \text{ sendo } \omega \text{ um subconjunto dos valores possíveis}$$

$\Omega$  para  $\theta$ . Seja  $L(\Omega)$  o valor máximo de  $L$  com relação a  $\theta$  para o conjunto inteiro de valores que o parâmetro pode assumir e seja  $L(\omega)$  o valor máximo de  $L$ , com  $\theta$  restrito ao conjunto de valores dados por  $\omega$ . A hipótese nula é testada usando a razão de verossimilhança,

$$\lambda = \frac{L(\omega)}{L(\Omega)}$$

ou uma função simples dela. Valores grandes da razão de verossimilhança são consistentes com a hipótese nula.

**REGIÃO CRÍTICA:** No teste de hipóteses, essa é a porção do espaço amostral de uma estatística de teste que levará à rejeição da hipótese nula.

**REGIÃO DE ACEITAÇÃO:** Em um teste de hipóteses, uma região no espaço amostral da estatística de teste, tal que se a estatística de teste cair em seu interior a hipótese nula é aceita. O bom senso diz que melhor terminologia é aquela em que a hipótese nula não pode ser rejeitada, uma vez que rejeição é sempre uma conclusão forte e aceitação é geralmente uma conclusão fraca.

**REGIÃO DE REJEIÇÃO:** Em um teste de hipóteses, essa é a região no espaço amostral da estatística de teste que conduz à rejeição da hipótese nula quando a estatística de teste cai nessa região.

**REPLICAÇÃO:** Termo que refere-se a reprodução de uma pesquisa com a finalidade de repetir dados de respostas obtidos em experimentos realizados anteriormente, colaborando para aumentar o seu grau de validação.

**RETROSPECÇÃO:** Termo usado para indicar a avaliação de fenômenos ocorridos em períodos anteriores.

**REGRESSÃO:** A origem do termo regressão deve-se ao antropologista, meteorologista, matemático e estatístico inglês, Sir Francis Galton (1822-1911) por volta de 1885, quando pesquisava relações entre características antropométricas de sucessivas gerações de seres humanos. Uma de suas constatações era de que cada peculiaridade de um homem é transmitida aos seus descendentes, mas, em média, numa intensidade menor. Por exemplo: embora pais com baixa estatura tendam a ter filhos também com baixa estatura, estes têm uma altura média maior do que a altura média de seus pais. O mesmo ocorre, mas em direção contrária, com pais de estatura alta. Se as características permanecessem as mesmas de geração para geração, esperar-se-ia que a reta de regressão tivesse seu coeficiente angular próximo de 1. Em sua análise, Francis Galton encontrou o valor 0,516, mostrando que a reta tende para aquela paralela ao eixo de  $X$  e passando pela média ( $Y = \bar{Y}$ ). A esse fenômeno da altura dos filhos mover-se em direção à altura média de todos os homens ele chamou de regressão, e às vezes de reversão, tendo aparecido num artigo científico de 1886 (GALTON, 1886), no Journal of the Anthropological Institute, com o título Regression towards mediocrity in hereditary stature, traduzindo para o português significa Regressão para a mediocridade em estaturas hereditárias; sendo que o termo mediocridade, referido aqui, significa a média.

**REGRESSÃO MÚLTIPLA:** A regressão múltipla é o método de análise apropriado quando o problema envolve uma variável métrica dependente que se presume relacionada com duas ou mais variáveis independentes métricas. O objetivo é prever mudanças na variável dependente em resposta à mudanças nas variáveis independentes.

**REGRA DA MULTIPLICAÇÃO:** Para probabilidade, uma fórmula usada para determinar a probabilidade da interseção de dois ou mais eventos. Para técnicas de contagem, uma fórmula usada para determinar o número de maneiras para completar uma operação a partir do número de maneiras para completar as etapas sucessivas.

**REGRA DA PROBABILIDADE TOTAL:** Dada uma coleção de eventos mutuamente excludentes, cuja união é o espaço amostra, a probabilidade de um evento pode ser escrita como a soma de probabilidades das interseções do evento com os membros dessa coleção.

**REGRA DE ADIÇÃO:** Uma fórmula usada para determinar a probabilidade da união de dois ou mais eventos, a partir das probabilidades dos eventos e sua(s) interseção(ões).

**REGRAS COMPLEMENTARES:** Um conjunto de regras aplicadas aos pontos plotados nos gráficos de controle de Shewhart, que é usado para tornar o gráfico mais sensível às causas atribuídas. Veja Gráfico de controle, Gráfico de controle de Shewart.

**REGRAS WESTERN ELETRIC:** Um conjunto específico de regras complementares que foram desenvolvidas na Western Electric Corporation. Veja Regras complementares.

**REGRESSÃO:** Métodos estatísticos usados para investigar a relação entre uma variável dependente ou de resposta,  $y$ , e uma ou mais variáveis independentes,  $x$ . As variáveis independentes são geralmente chamadas de variáveis regressoras ou variáveis preditoras.

**REGRESSÃO CORRIGIDA (RIDGE REGRESSION):** Um método para ajustar um modelo de regressão tendo como objetivo superar os problemas associados com o uso do método-padrão dos mínimos quadrados, quando há um problema com multicolinearidade nos dados.

**REGRESSÃO CURVILINEAR:** Uma expressão algumas vezes usada para modelos não-lineares ou modelos polinomiais de regressão.

**REGRESSÃO SEQUENCIAL:** O procedimento Regressão Sequencial é apropriado para medir o efeito do acréscimo de variáveis a um modelo. Esta avaliação é por meio de mudanças na soma de quadrados da regressão e nas variações do  $R^2$ . O processamento inicia-se com um modelo simples composto de somente uma variável independente e, a cada passo, adiciona-se nova variável.

**REGRESSÃO EM ETAPAS:** Um método de selecionar variáveis para inclusão em um modelo de regressão. Ele opera introduzindo as variáveis candidatas uma de cada vez, como na seleção progressiva, e então tenta removê-las seguindo cada etapa progressiva.

**REGISTRO ANEDÓTICO:** É o registro de acontecimentos significativos em relação a um aluno ou turma como um todo. Este registro pode ser feito informalmente num caderno ou fichas especiais. Em qualquer dos casos é necessário que se separe a descrição do fato da opinião do observador.

**REGRESSÃO LOGÍSTICA:** Um modelo de regressão que é usado para modelar uma resposta categórica. Para uma resposta binária (0,1), o modelo assume que o logaritmo da razão de probabilidades tendendo para zero e um é linearmente relacionado às variáveis regressoras.

**RESPONDENTE:** Indivíduo que é selecionado para responder um inventário ou um questionário, testar um produto, etc. na realização de uma pesquisa.

**RELAÇÃO DE DEFINIÇÃO:** Um subconjunto de efeitos em um planejamento fatorial fracionário que define os pares associados no planejamento.

**REIFICAÇÃO (REIFICATION):** Conversão de um conceito abstrato em um objeto material.

**RÉPLICA (REPLICATE):** Uma observação, elemento ou tentativa individual. A maioria dos experimentos possui réplicas múltiplas de cada categoria ou grupo de tratamento.

**REPLICAÇÃO (REPLICATION):** O estabelecimento de múltiplas parcelas, observações ou grupos dentro do mesmo experimento ou tratamento observacional. Idealmente, a replicação é obtida pela aleatorização.

**RESÍDUO (RESIDUAL):** A diferença entre um valor observado e seu valor previsto por um modelo estatístico.

**RESÍDUOS NORMALIZADOS (STUDENTIZED RESIDUALS):** Ver as definições ou os conceitos de resíduos ou erros padronizados.

**RESÍDUOS PADRONIZADOS (STANDARDIZED RESIDUALS):** Um erro ou resíduo transformado em um escore-Z.

**RESULTADO (OUTCOME):** Do inglês outcome, é o resultado, geralmente de uma observação ou tentativa.

**RESULTADO DISCRETO (DISCRETE OUTCOME):** Um resultado ou observação que pode assumir valores independentes ou inteiros.

**ROTAÇÃO (ROTATION):** Em uma análise fatorial, rotação é uma combinação linear de fatores comuns que são facilmente interpretados. Ver rotação oblíqua, rotação ortogonal e rotação varimax.

**ROTAÇÃO OBLÍQUA (OBLIQUE ROTATION):** Uma combinação linear dos fatores comuns em uma análise fatorial que resulta em fatores comuns que podem ser correlacionados uns com os outros. Comparar com rotação ortogonal e rotação varimax.

**ROTAÇÃO ORTOGONAL (ORTHOGONAL ROTATION):** A combinação linear dos novos fatores comuns de uma análise fatorial que resulta em fatores comuns que não têm correlação uns com os outros. Comparar com rotação oblíqua e rotação varimax.

**REPLICAÇÃO:** A execução independente de um experimento mais de uma vez.

**REPLICATAS:** Uma das repetições independentes de uma ou mais combinações de tratamentos em um experimento.

**RESÍDUO:** Geralmente é a diferença entre o valor observado e o valor previsto de alguma variável. Por exemplo, em regressão, um resíduo é a diferença entre o valor observado da resposta e o correspondente valor previsto obtido por meio de modelo de regressão.

**RESÍDUO NA FORMA DE STUDENT:** Na regressão, o resíduo na forma de student é calculado pela divisão do resíduo normal pelo seu desvio-padrão exato, produzindo um conjunto de resíduos escalonados que têm, exatamente, desvio padrão igual a um.

**RESÍDUO PADRONIZADO:** Na regressão, o resíduo padronizado é calculado pela divisão do resíduo normal pela raiz quadrada da média residual dos quadrados. Isso produz resíduos escalonados que têm, aproximadamente, variância igual a um.

**RESOLUÇÃO:** Uma medida da severidade de associação em um planejamento fatorial fracionário. Comumente consideramos planejamentos com resolução III, IV e V.

**RESPOSTA (VARIÁVEL DE):** A variável dependente em um modelo de regressão ou a variável observada de saída em um experimento planejado.

**RESULTADO:** Um elemento de um espaço amostral.

**REGISTRO METODOLÓGICO:** Uma agenda de notas detalhadas e seqüenciais no tempo sobre técnicas investigativas usadas durante uma investigação humanística, com especial atenção às distorções que determinada técnica possa ter introduzido.

**REGRA:** Um guia, método ou comando que diz ao pesquisador o que deve fazer.

**REGRA DA ADIÇÃO:** Regra para determinar a probabilidade de ocorrência, em uma única prova, do evento A, ou do evento B, ou de ambos os eventos.

**REGRA DA MULTIPLICAÇÃO:** Regra que permite determinar a probabilidade do evento A ocorrer em uma prova ou ensaio aleatório e do evento B ocorrer em uma segunda prova ou experimento casual.

**REGRA DE PARADA:** Algoritmo para determinar o número final de agrupamentos a serem formados. Sem qualquer regra de parada inerente à análise de agrupamentos, os pesquisadores desenvolveram diversos critérios e orientações para essa determinação. Existem duas classes de regras que são aplicadas post hoc e calculadas pelo pesquisador: i) medidas de similaridade e ii) medidas estatísticas adaptadas.

**REGRA DAS COMBINAÇÕES:** Regra para determinar o número de combinações distintas de elementos selecionados.

**REGRA DAS PERMUTAÇÕES (ARRANJOS):** Regra que permite determinar o número de permutações ou arranjos distintos de objetos selecionados.

**REGRA DE BAYES:** Modelo de distribuição de probabilidade condicional onde se calcula a probabilidade a posteriori do evento (A) dada a ocorrência do evento (B)  $[P(A|B)]$ , em função do conhecimento a priori da probabilidade de ocorrência do evento B, desde que o evento A tenha ocorrido  $[P(B|A)]$ . Para dois eventos, alguns programas estatísticos podem apresentar o chamado crivo ou screening test, enquanto que para três ou mais eventos outros programas dispõem do modelo da regra de Bayes.

**REGRA DE BAYES:** Regra que indica como calcular a probabilidade condicional  $Pr(B|A)$  desde que se conheçam  $Pr(A|B)$  e  $Pr(A|B^c)$ .

**REGRA DE CINCO ITENS:** Uma técnica de análise exploratória de dados que usa cinco números para sintetizar os dados: o menor valor, o primeiro quartil, a mediana, o terceiro quartil e o maior valor.

**REGRA DE COMPOSIÇÃO:** Regra usada para combinar atributos para produzir um julgamento de valor ou utilidade relativa para um produto ou serviço. Para fins de ilustração, suponha que uma pessoa seja solicitada a avaliar quatro objetos. Considera-se que a pessoa avalia os atributos dos quatro objetos e cria algum valor relativo geral para cada um. A regra pode ser simples, como a criação de um peso mental para cada atributo percebido e a soma de pesos para um escore geral no modelo aditivo ou pode ser um procedimento mais complexo que envolva efeitos de interação.

**REGRA DE DECISÃO:** Seja A o espaço de ações possíveis e  $\Omega$  o espaço de resultados possíveis de um experimento. Uma regra de decisão  $\delta$  é uma função definida em  $\Omega$  com valores em A. Em inglês Decision rule.

**REGRA DE QUADRATURA GAUSSIANA:** Em análise numérica, uma regra de quadratura é uma aproximação da integral de uma função, geralmente estabelecido como um somatório com pesos dos valores assumidos pela função em pontos específicos dentro do domínio de integração. Uma regra de quadratura

gaussiana de n pontos, chamada assim em homenagem a Carl Friedrich Gauss, é uma regra de quadratura construída para produzir um resultado exato para polinômios de grau  $2n-1$  ou menor para uma escolha adequada dos pontos  $x_i$  e pesos  $w_i$  para  $i = 1, \dots, n$ . O domínio de integração de tal regra é por convenção tomado como  $[-1, 1]$ , de modo que a regra é expressa como

$$\int_{-1}^1 f(x) dx \approx \sum_{i=1}^n w_i f(x_i).$$

Pode ser mostrado que os pontos usados para avaliar a função são exatamente as raízes de um polinômio pertencente a uma classe de polinômios ortogonais.

**REGRA DO FATORIAL:** Regra que afirma que n objetos diferentes podem ser dispostos em n! maneiras diferentes.

**REGRA EMPÍRICA:** Regra que utiliza o desvio-padrão para proporcionar informação sobre dados com distribuição em forma de sino.

**REGRA FUNDAMENTAL DA CONTAGEM:** Regra que afirma que, para uma determinada sequência de dois eventos, dos quais o primeiro pode ocorrer de m maneiras, e o segundo de n maneiras, os eventos em conjunto podem ocorrer de mn maneiras.

**REGRA PRÁTICA:** Regra segundo a qual o intervalo de um conjunto de dados tem amplitude aproximadamente igual a 4 desvios-padrão (4s).

**REGRAS DE ASSOCIAÇÃO:** Regras baseadas na correlação entre atributos ou dimensões de elementos de dados.

**REGRAS DE CORRESPONDÊNCIA:** São definições semânticas conectando o modelo teórico e o mundo da realidade. No seu melhor, as regras assumem a forma de definições operacionais rigorosas que especificam as conexões entre os conceitos e as observações relevantes.

**REGRESSÃO:** É uma técnica estatística de ajuste de uma equação representativa que descreva a relação funcional entre uma variável dependente e uma variável independente. O ajuste da função é feito pelo método dos mínimos quadrados. A regressão pode ser: linear simples, linear múltipla, polinomial, exponencial, potencial, inversa, hiperbólica e hiperbólica inversa.

**REGRESSÃO:** Medida da relação entre duas variáveis. O valor de uma característica pode ser predito sabendo o valor da outra variável. Por exemplo, características de carcaça facilmente obtidas como peso de carcaça quente, espessura de gordura, área de olho de lombo e porcentagem de gordura interna são usadas para prever habilidade de corte como porcentagem de carne vendável ou músculo na carcaça em relação a porcentagem de gordura inaproveitada.

**REGRESSÃO:** Técnica estatística de ajuste de uma equação ou modelo de regressão usada para realizar previsões de uma variável dependente ou resposta Y em função do valor de uma variável independente ou explicativa X. Em inglês regression.

**REGRESSÃO:** Técnica de análise estatística. Informa, em termos matemáticos, por meio de uma equação, a relação entre duas variáveis que é a regressão simples ou mais de duas variáveis que é a regressão múltipla.



**REGRESSÃO (COMO AMEAÇA À VALIDADE):** Uma tendência que os grupos têm, especialmente os selecionados com base em escores extremos, de regressir a valores próximos da média em mensurações subseqüentes, independentemente do tratamento experimental.

**REGRESSÃO À MÉDIA:** A tendência do comportamento das pessoas se inclinar mais para a média durante o curso de um experimento.

**REGRESSÃO BINÁRIA BIVARIADA:** Regressão em que duas variáveis binárias existem como variáveis dependentes. Em inglês Bivariate binary regression.

**REGRESSÃO COM TODOS OS SUBCONJUNTOS POSSÍVEIS:** Método para selecionar as variáveis para inclusão no modelo de regressão que considera todas as combinações possíveis das variáveis independentes. Por exemplo, se o pesquisador especificou quatro variáveis independentes potenciais, essa técnica estima todos os possíveis modelos de regressão com uma, duas, três e quatro variáveis. Então, a técnica identifica o(s) modelo(s) com mais precisão de previsão.

**REGRESSÃO COM VARIÁVEIS MUDAS OU DUMMY OU VARIÁVEIS BINÁRIAS:** Variáveis binárias são aquelas em que só assumem dois valores distintos, geralmente 0 e 1. São usadas para representar variáveis qualitativas, mudanças súbitas no processo em análise, estratos de uma variável qualquer, dentre outras. A introdução dessas variáveis tornam o modelo de regressão linear uma ferramenta extremamente flexível, capaz de lidar com muitos problemas interessantes encontrados nos estudos empíricos. Na análise de regressão, a variável dependente é muitas vezes influenciada não somente pelas variáveis que pode ser facilmente quantificadas em alguma escala bem definida como, por exemplo: renda, produto, preços, custos, altura e temperatura, dentre outros, mas também por variáveis essencialmente qualitativas, como por exemplo: sexo, raça, cor, religião, nacionalidade, guerras, terremotos, greves, convulsões e mudanças na política do governo, dentre outras. Como essas variáveis qualitativas geralmente indicam a presença ou a ausência de uma qualidade ou atributo, tais como: homem ou mulher, negro ou branco, católico ou não católico, dentre outras, um método para quantificar tais atributos é construir variáveis artificiais que assumam valores de 1 ou 0. sendo que o 0 indica ausência de um atributo, e 1 indica a presença ou a posse desse atributo. Por exemplo, 1 pode indicar que uma pessoa é homem e zero pode designar uma mulher; ou 1 pode indicar que uma pessoa tem formação superior e 0 pode indicar o contrário. Um modelo de regressão com variáveis binárias pode ser escrito na seguinte forma:  $Y_j = \alpha + \gamma Z_j + \beta X_j + e_j$ , onde  $Z_j$  é considerada uma variável binária que assume o valor um ( $Z_j=1$ ) quando a observação está presente, e o valor zero ( $Z_j=0$ ) quando a observação está ausente. Em inglês Dummy variables.

**REGRESSÃO CONJUNTA:** O modelo clássico de regressão assume que a variável dependente é uma função, geralmente linear, de um conjunto de variáveis independentes. Se o modelo linear não é adequado, mudanças são feitas para permitir a inclusão de produtos cruzados de variáveis independentes, gerando assim uma regressão conjunta. Em inglês Joint regression.

**REGRESSÃO CURVILÍNEA:** Uma regressão que não é linear. Uma forma usualmente considerada é aquela para a qual a variável dependente é representada como um polinômio das variáveis independentes. Em inglês Curvilinear regression.

**REGRESSÃO CURVILÍNEA:** Forma de análise de regressão que estuda relações não-lineares entre variáveis.

**REGRESSÃO DEFASADA:** Uma regressão em que os valores da variável dependente e pelo menos uma das variáveis independentes estão defasadas com relação às outras. Em inglês Lag regression.

**REGRESSÃO DE UM PROCESSO EPIDÊMICO:** Refere-se à última fase na evolução de uma epidemia, quando os valores de incidência decrescem, chegando a um patamar endêmico ou a uma incidência nula que é a erradicação.

**REGRESSÃO EM DIREÇÃO À MÉDIA:** O fato de indivíduos, com valores extremos de um parâmetro, em uma primeira avaliação, quando examinados em um parâmetro, em uma primeira avaliação, quando examinados em uma segunda ocasião, tenderem a apresentar valores não tão extremos e mais próximos da média de distribuição do parâmetro.

**REGRESSÃO EM DIREÇÃO À MÉDIA:** Pacientes escolhidos para participar de um estudo precisamente porque possuíam uma medida extrema de qualquer variável têm alta probabilidade de apresentar uma medida mais próxima da média em uma ocasião posterior, por ocasiões que não estão relacionadas com o tipo de eficácia do tratamento que receberam. A regressão em direção à média e também conhecida como efeito da regressão estatística.

**REGRESSÃO (PAYOFF) ESPERADO:** Valor esperado no resultado de uma decisão.

**REGRESSÃO EXPONENCIAL:** Ajuste da função exponencial, cujo modelo é dado pela equação  $Y = ae^{bx}$ . A variável dependente (y) deve ser sempre positiva.

**REGRESSÃO HIPERBÓLICA:** Ajuste da função hiperbólica ao conjunto de dados experimentais:  $Y = \frac{1}{(a+b.X)}$ . A restrição para esse ajuste é que a variável dependente deve ser diferente de zero.

**REGRESSÃO HIPERBÓLICA INVERSA:** Ajuste da função hiperbólica inversa aos pontos de dados:  $Y = \frac{X}{(ax+b)}$ . Esta função exige que as variáveis independente e dependente sejam diferentes de zero.

**REGRESSÃO INVERSA:** Permite ajustar a função inversa:  $Y = a + \frac{b}{X}$ . Para esta função, a variável independente (x), deve ser diferente de zero.

**REGRESSÃO LINEAR:** Técnica estatística de ajuste de um conjunto de pares de valores X, Y a equação de uma reta. Em inglês linear regression.

**REGRESSÃO LINEAR:** Método para encontrar a reta que mais se aproxima de um conjunto de pontos.

**REGRESSÃO LINEAR:** Modelo de regressão que testa se a relação entre duas variáveis pode ser representada por uma linha reta.

**REGRESSÃO LINEAR SIMPLES:** Análise de regressão envolvendo uma variável dependente e uma variável independente, na qual a relação entre variáveis é aproximada por uma linha reta.

**REGRESSÃO LINEAR SIMPLES:** Teste estatístico que determina o modelo estimador dos valores de Y a partir dos escores de X.

**REGRESSÃO LINEAR SIMPLES:** Mesmo que ajuste de reta.

**REGRESSÃO LINEAR SIMPLES:** Método de análise de relação entre uma variável independente e uma variável dependente.



**REGRESSÃO LINEAR PONDERADA:** Em diversas situações, ao analisarmos os resíduos de um modelo de regressão linear, ao visualizarmos que estes não apresentam a característica de variância constante, temos uma das suposições do modelo não atendidas. Quando isso acontece, diz-se que o modelo apresenta heterocedasticidade nos erros ou resíduos, ou ainda que o modelo é heterocedástico. Alguns efeitos causados por essa falha na suposição do modelo são: Os erros padrões dos estimadores, obtidos pelo *método dos mínimos* quadrados ordinários, são incorretos e, portanto a inferência estatística não é válida. Não podemos mais dizer que os estimadores de *mínimos* quadrados ordinários são os melhores estimadores de variância mínima para  $\beta$ , embora ainda possam ser não viciados. O uso de pesos se justifica por causa da seguinte situação: Suponha que a variância não seja constante, isto é,  $Var(Y_i) = \sigma_i^2$ , para  $i = 1, 2, 3, \dots, n$ . Toma-se, por exemplo, pesos de forma que  $\omega_i \propto \frac{1}{\sigma_i^2}$ ,  $i = 1, 2, 3, \dots, n$ . Com isso, as estimativas de mínimos

quadrados ponderados (MQP) tem erros padrão menores do que as estimativas de mínimos quadrados ordinários (MQO). Como é sabido, as estimativas de mínimos quadrados ordinários (MQO) são incorretos, em relação as estimativas de mínimos quadrados ponderados (MQP).

**REGRESSÃO LINEAR MÚLTIPLA:** Permite obter uma relação funcional de uma variável dependente (y) para um conjunto de variáveis independentes  $\{x_i, i=1,m\}$ :

$Y = C_0 + C_1X_1 + C_2X_2 + \dots + C_mX_m$ . **Onde:**  $c_0, c_1, \dots, c_m$  são os coeficientes;  $x_1, x_2, \dots, x_m$  são as variáveis independentes.

**REGRESSÃO LINEAR MÚLTIPLA:** Teste estatístico que determina o modelo estimador dos escores da variável Y a partir dos valores de duas ou mais variáveis preditivas:  $X_1, X_2, \dots, X_n$ .

**REGRESSÃO MÚLTIPLA:** É uma coleção de técnicas estatísticas para construir modelos que descrevem de maneira razoável relações entre várias variáveis explicativas de um determinado processo. A diferença entre a regressão linear simples e a múltipla é que na múltipla são tratadas duas ou mais variáveis explicativas.

**REGRESSÃO LINEAR MÚLTIPLA:** Método utilizado em análises multivariáveis quando a variável dependente e todas as variáveis independentes são contínuas.

**REGRESSÃO DE CUMEEIRA:** Algumas análises de regressão podem apresentar problemas na inversão da matriz de correlação simples, em consequência, principalmente, de altas correlações entre as variáveis independentes do modelo (multicolinearidade). Neste caso existe uma solução matemática proposta por Hoerl e Kennard (1970a) e Hoerl e Kennard (1970b), em que se adiciona uma constante à diagonal principal da matriz de correlação. Esta opção faz com que seja montada uma tabela de resíduos contendo as seguintes informações: i) número das observações, ii) valores da variável dependente observados, e iii) valores da variável dependente estimados. Além disso, será gerado um gráfico mostrando os desvios padronizados em relação às observações. Na análise de regressão linear múltipla ocorre alguns problemas tais como; i) nem sempre é possível calcular a inversa da matriz  $X'X^{-1}$ , isto geralmente ocorre quando as variáveis de entrada são intercorrelacionadas. Se a intercorrelação é grande existe multicolinearidade, isto, é, as linhas da matriz  $X'X$  não são linearmente independentes. Para minimizar o efeito da multicolinearidade o pesquisador pode aplicar a regularização de Tikhonov e assim a regressão passa a ser chamada de regressão de cumeeira (em inglês ridge regression). A equação usada para calcular os estimadores dos parâmetros  $\beta$  é dada por:  $\beta = X'X^{-1}X'y$ , é reescrita da seguinte forma  $\beta = (X'X + \lambda I)^{-1}X'y$ . Basicamente, com a regularização de Tikhonov, nós somamos uma constante a diagonal principal da

matriz de modo a tentar torná-la inversível, ou seja,  $\beta = (X'X + \lambda I)^{-1}X'y$ , onde  $0 \leq \lambda \ll 1$  é uma constante de valor pequeno e  $I$  é uma matriz identidade de ordem  $(p+1)$ .

**REGRESSÃO LOGÍSTICA:** Forma especial de regressão na qual a variável dependente é não-métrica, dicotômica, binária. Apesar de algumas diferenças, a maneira geral de interpretação é semelhante à da regressão linear.

**REGRESSÃO LOGÍSTICA:** Modelo linear generalizado com ligação logística. Em inglês Logistic regression.

**REGRESSÃO LOGÍSTICA:** Forma de regressão múltipla em que a variável depende é expressa em forma de categorias: sadio/doente, vivo/morto. As variáveis independentes podem ser contínuas ou categóricas.

**REGRESSÃO LOGÍSTICA:** O modelo de regressão logística é semelhante ao modelo de regressão linear. No entanto, no modelo logístico a variável resposta  $Y_i$  é binária. Uma variável binária assume dois valores, como por exemplo,  $Y_i = 0$  e  $Y_i = 1$  denominados fracasso e sucesso, respectivamente. Neste caso, sucesso é o evento de interesse.

**REGRESSÃO LOGÍSTICA SIMPLES:** Teste estatístico de regressão no qual a variável dependente Y é sempre binária, assumindo valores de 1 (sucessos) e 0 (insucessos). Por outro lado, a variável independente X é, em geral, binária, é múltiplo, dois (2) ou mais. A variável dependente é sempre binária.

**REGRESSÃO MÚLTIPLA:** Estudo das relações lineares entre três ou mais variáveis.

**REGRESSÃO MÚLTIPLA:** Método estatístico para analisar a relação entre várias variáveis independentes é uma variável dependente.

**REGRESSÃO MÚLTIPLA:** Um método usado para relacionar um grupo de variáveis a uma única variável. Muito usada em aplicações psicotecnológicas. Por exemplo, para prever escores-médios de alunos a partir dos seus escores em vestibular; a equação de regressão resultante pode ser usada como um instrumento de seleção para subsidiar decisões sobre admissão de alunos à universidade.

**REGRESSÃO MÚLTIPLA:** Regressão de uma variável dependente em mais de uma variável independente ou preditora. Em inglês Multiple regression.

**REGRESSÃO MÚLTIPLA:** Modelo de regressão no qual uma equação é desenvolvida para estimar valores de uma variável dependente por meio de conhecimento de valores de diversas variáveis independentes.

**REGRESSÃO MÚLTIPLA:** Modelo de regressão com duas ou mais variáveis independentes. Ou seja, é uma técnica estatística que pode ser usada para analisar a relação entre uma única variável dependente que é o critério e várias variáveis independentes ou preditoras. O objetivo da análise de regressão múltipla é usar as variáveis independentes cujos valores são conhecidos para prever os valores da variável dependente selecionada pelo pesquisador. Cada variável independente é ponderada pelo procedimento da análise de regressão para garantir máxima previsão a partir do conjunto de variáveis independentes. Os pesos denotam a contribuição relativa das variáveis independentes para a previsão geral e facilitam a interpretação sobre a influência de cada variável em fazer a previsão, apesar de a correlação entre as variáveis independentes complicar o processo interpretativo. Na análise multivariada, por exemplo, o conjunto de variáveis independentes ponderadas forma a variável estatística de regressão, uma combinação linear das variáveis independentes que melhor prevê a variável dependente.

**REGRESSÃO MÚLTIPLA:** É o método de análise apropriado quando o problema de pesquisa envolve uma única variável dependente métrica considerada relacionada a duas ou mais variáveis independentes métricas. O objetivo da análise de regressão múltipla é prever as mudanças na variável dependente como resposta a mudanças nas variáveis independentes. Esse objetivo é alcançado, com frequência, por meio da regra estatística dos mínimos quadrados. Sempre que o pesquisador estiver interessado em prever a quantidade ou magnitude da variável dependente, a regressão múltipla será útil. Por exemplo, despesas mensais com jantares fora de casa que é a variável dependente podem ser previstas a partir de informações referentes a renda familiar, tamanho da família e idade do chefe da família que são as variáveis independentes.

**REGRESSÃO MULTIVARIADA:** A regressão linear múltipla também é uma técnica multivariada cuja finalidade principal é obter uma relação matemática entre uma das variáveis que é a variável dependente e o restante das variáveis que descrevem o sistema que são as variáveis independentes. Sua principal aplicação, após encontrar a relação matemática é produzir valores para a variável dependente quando se têm as variáveis independentes. Ou seja, ela pode ser usada na predição de resultados. Obviamente, a soma das contribuições de diversas variáveis para uma determinada predição pode também ser feita usando as componentes principais, pois as mesmas têm a vantagem de poder ser tratadas de modo completamente independente. Portanto, é possível também fazer regressão linear múltipla das componentes principais.

**REGRESSÃO NÃO PARAMÉTRICA:** A diferença básica entre um modelo de regressão paramétrico e um outro não paramétrico envolve a quantidade de suposições que o economista estaria disposto a fazer e qual peso ele pretende dar aos dados por si mesmos. Um modelo paramétrico assume que  $g(X)$  é uma função desconhecida em um número finito de parâmetros, de forma que a tarefa consistiria em estimar os parâmetros desconhecidos, por exemplo por mínimos quadrados. Em um modelo não paramétrico a relação funcional entre as duas variáveis vive num espaço de funções muito mais amplo: assumimos apenas que  $g(X)$  esta num espaço de funções seguindo algumas restrições convenientes e buscamos uma combinação linear de funções desse espaço que aproximem bem  $g(X)$ . Dentre os métodos utilizados para regressão não paramétrica destacam-se o método de Kernel e o via Splines. Este primeiro será objeto deste trabalho, que consistirá em uma aplicação simples do estimador de núcleo conhecido como estimador de Nadaraya-Watson. Na regressão não-paramétrica a sua característica distintiva é a ausência completa ou quase completa de conhecimento a priori a respeito da forma da função que está sendo estimada. Sendo assim, mesmo que a função continue a ser estimada a partir do ajuste de parâmetros livres, o conjunto de formas que a função pode assumir que é a classe de funções que o modelo do estimador pode prever é muito amplo. Como consequência, vai existir um número elevado de parâmetros como por exemplo, quando comparado ao número de dados de entrada-saída para treinamento, os quais não mais admitem uma interpretação física isolada.

**REGRESSÃO LINEAR MÚLTIPLA:** A regressão que mede a relação entre duas variáveis torna-se uma regressão múltipla quando ela é estendida para incluir mais do que uma variável independente ( $X_1, X_2, X_3, X_4, \dots$ ) na tentativa de explicar a variável dependente  $Y$ . Enquanto as apresentações gráficas tornam-se mais difícil, a regressão múltipla conduz a uma forma que é uma extensão da regressão simples. O modelo matemático desse tipo de regressão quando existem quatro variáveis explicativas é o seguinte:  $Y = a + b X_1 + c X_2 + d X_3 + e X_4 + e_1$ .

**REGRESSÃO NÃO LINEAR:** Modelo de regressão cuja equação é não linear com relação aos parâmetros, ou não linear com respeito as variáveis. Em inglês non linear regression.

**REGRESSÃO PASSANDO PELA ORIGEM:** Em muitas situações a função de regressão de regressão linear simples de duas variáveis assume forma dada pela equação a seguir.  $Y_i = \hat{b} \cdot X_i + e_i$ , isto é uma situação em que o termo de intercepto  $\hat{a}$ , está ausente do modelo de regressão. Sendo assim muitas pesquisas em ciências agrárias, biológicas e outras, é conhecido o fato de que o valor de  $Y$  é igual a zero quando o valor de  $X$  é zero. Neste caso, é interessante considerar as situações em que o intercepto é igual a zero, em consequência direta desse fato. Esse tipo de modelagem é conhecida como regressão pela origem, pelo fato de a regressão passar pelo ponto (0,0). É importante salientar que a decisão de utilizar esse modelo deve ser tomada com alguns cuidados e somente em situações práticas particulares. Algumas razões podem ser apontadas para isso. O fato de a soma de resíduos não resultar em zero e, em alguns casos particulares, a soma de quadrados desses resíduos superar a soma de quadrados do total corrigido, proporcionando  $r^2$  negativos, são razões que devem ser consideradas. Porém, uma razão que merece ser mencionada refere-se ao caso do modelo de regressão ser empírico e local. Se a faixa de valores de variável  $X$  estiver afastada de zero, convém utilizar o intercepto para melhoria do ajustamento. Deve-se tomar o devido cuidado ao interpretar o intercepto. Na realidade, estando o intercepto longe da origem, essa interpretação é até mesmo pouco relevante, pois o erro do valor ajustado para  $X=0$  será muito elevado, de qualquer maneira. É importante destacar que o estimador da variância neste caso está associado a  $V = n-1$  graus de liberdade em vez dos  $V = n-2$  graus de liberdade do caso do modelo completo.

**REGRESSÃO PARCIAL:** Modelo de regressão no qual uma variável dependente é estimada com base numa variável independente, enquanto uma terceira variável se mantém constante.

**REGRESSÃO PASSO A PASSO:** Utilização de diferentes combinações de variáveis até atingir o melhor modelo; usado em regressão múltipla.

**REGRESSÃO POLINOMIAL:** Regressão que contém potências de uma variável independente. O preditor tem a forma  $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots$ . Em inglês Polynomial regression.

**REGRESSÃO POLINOMIAL:** É o ajuste de uma curva polinomial de grau  $K$ :  $Y = C_0 + \sum_{k=1}^K C_k X^k$ .

**REGRESSÃO POLINOMIAL:** Os modelos de regressão polinomial tem um amplo uso na pesquisa econométrica relacionada as funções de produção de custo e também nas pesquisas em geral. Como exemplo, podemos citar uma curva de segundo grau ou parábola, a qual é representada matematicamente pela seguinte equação:  $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2$  a qual se chama função quadrática, ou mais comumente polinômio de segundo grau na variável  $X$ , pois o maior expoente de  $X$  indica o grau do polinômio, por exemplo ainda, se  $X^3$  fosse acrescentado a função anterior, então ele seria um polinômio de terceiro grau, e assim por diante. O modelo estocástico da equação anterior pode ser escrito como mostrado a seguir,  $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + e_i$ , a qual é chamada de regressão polinomial de segundo grau. O modelo matemático da regressão polinomial de  $k$ -ésimo grau geral pode ser escrita como mostrado na seguinte equação  $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \dots + \beta_k X_i^k + e_i$ . Deve-se observar que nestes tipos de regressões polinomiais há apenas uma variável explicativa no lado direito, mas ela aparece com vários expoentes, tornando-se assim modelos de regressão múltipla. A propósito, note que se admitirmos que  $X_i$  seja fixo ou não estocástico, os termos de  $X_i$  com expoente também se tornam fixos ou não estocásticos.

**REGRESSÃO POTENCIAL:** Ajusta um modelo ou função potencial ao conjunto de pontos observados:  $Y = aX^b$ . As variáveis dependente e independente devem ser positivas.

**REGRESSÃO SIMPLES:** Modelo de regressão com uma única variável independente, ou seja,  $Y_i = a + bX_i$ . Em inglês Simple Regression.

**REGRESSÃO SEGMENTADA:** O modelo de regressão segmentada consiste em duas partes: uma linha inclinada ascendente ou descendente seguida de uma linha horizontal, onde seus pontos de interseção vão determinar o ponto de quebra. Este modelo de uma inclinação é mais adequado para estimar parâmetros de crescimento. Para outros tipos de variáveis biológicas, a equação do modelo de regressão segmentada descreve duas linhas de intersecção, ambas com inclinação diferente a zero. O modelo estatístico ou função de regressão ajustado é do seguinte tipo:  $Y_i = L + U(R-XLR_i) + e_i$ ,  $i = 1, 2, \dots, n_1, n_1+1, \dots, n$  cujo  $(R-XLR_i) = 0$  para  $i \geq n_1 + 1$ , e  $n_1$  é o número de observações até o ponto de quebra, e  $n$  é o número de pares de observações. Existe outro modelo de regressão utilizado como sendo do tipo:  $Y_i = L + U(R-XLR_i) + V(XGR_i - R)$ , cujo  $(R-XLR_i) = 0$  para  $i > n_1$  e  $(XGR_i - R) = 0$  para  $i \leq n_1$ . Nestes modelos,  $L$  representa a coordenada no eixo das ordenadas e  $R$  no eixo das abcissas de um ponto de quebra (ponto de quebra) em uma curva.  $U$  é o coeficiente de inclinação de uma linha quando  $X < R$ , e no modelo de duas inclinações o  $V$  representa a inclinação de uma linha quando  $X > R$  e  $e_i$  é o componente aleatório ou resíduo. Assim pela definição  $(R-XLR)$  é zero quando  $X > R$  e  $(XGR-R)$  é zero quando  $X \leq R$ .

**REGRESSÃO STEPWISE:** Um método de selecionar a melhor série de variáveis regressoras para uma equação de regressão. Isto ocorre introduzindo-se as variáveis uma por vez ou começando com toda a série e depois os rejeitando um por vez. O critério para se aceitar ou rejeitar uma variável geralmente depende da extensão pela qual ela afeta o coeficiente de correlação múltipla. Em inglês Stepwise regression.

**REGRESSOR:** Um sinônimo para variável independente em relação a uma regressão. Em inglês Regressor.

**REGULARIZAÇÃO:** É o termo empregado na Geoestatística para a operação de composição de amostras de furos de sondagem, de diferentes tamanhos, para amostras regularizadas com o mesmo tamanho ou suporte amostral. Ver também composição de amostras.

**RELAÇÃO CAUSAL:** Relação entre duas variáveis na qual uma causa a outra. Por exemplo, pode-se afirmar que aumento na educação causa redução no preconceito.

**RELAÇÃO CAUSAL:** Relação de dependência de duas ou mais variáveis na qual o pesquisador claramente especifica que uma ou mais variáveis causam ou criam um resultado representado por pelo menos uma outra variável. Deve atender as exigências de causalidade.

**RELAÇÃO DOSE-RESPOSTA:** Existe quando um aumento na intensidade ou na duração da exposição aumenta o risco do efeito adverso. Essa relação é demonstrada quase sempre em estudos de exposições crônicas, tais como os estudos da relação entre quantidade de cigarros fumados e risco de câncer de pulmão. Em estudos que envolvam uma substância potencialmente carcinogênica, a relação dose-resposta é medida depois de ultrapassado o período usual de latência dessa substância, que é o período que decorre entre o início da exposição e o desenvolvimento do câncer.

**RELAÇÃO ESTATÍSTICA:** Relação baseada na correlação de uma ou mais variáveis independentes com a variável dependente. Medidas de associação, tipicamente correlações representam o grau de relação porque há mais de um valor da variável dependente para cada valor da variável independente.

**RELAÇÃO INEXISTENTE:** É quando a variável estudada possui um valor tão pequeno que não a torna significativa, precisando então aumentar o tamanho da amostra para torná-la significativa.

**RELAÇÃO LINEAR:** Relação entre duas variáveis descrita por uma linha reta. Relações não-lineares podem ser frequentemente linearizadas pela aplicação de uma transformação. Ver anamorforse.

**RELAÇÃO VOLUME-VARIÂNCIA:** É a relação de proporcionalidade dada pelo produto do volume e variância amostral, de tal modo que o aumento do volume das amostras é acompanhado da diminuição da variância e vice-versa.

**RELACIONAMENTO (ENTRE PROPRIEDADES):** Existem três tipos de relacionamento ou associação: i) Causa-efeito:  $X$  é causa de  $Y$  se a ocorrência de  $X$  for suficiente para a subsequente ocorrência de  $Y$ ; ii) Produto-produto: deve existir um tal contexto  $X$  que, quando  $X$  for colocado nele,  $Y$  ocorre; iii) Correlação:  $Y$  pode ser uma função de  $X$ , e não apenas de  $X$ , mas de muitas outras variáveis em combinação.

**RELACIONAMENTO LINEAR:** Um certo relacionamento entre duas variáveis que permite a colocação de uma linha reta entre os pontos do diagrama de dispersão. Os pontos dispersos se agruparão elipticamente em torno de uma linha reta e não em torno de algum tipo de curva.

**RELACIONAMENTO [RAPPORT]:** Relacionamento fácil, natural e confiante entre entrevistador e respondente, tornando mais provável que este responda perguntas mais exatas e completamente.

**RELAÇÕES ENTRE VARIÁVEIS:** Duas ou mais variáveis quaisquer estão relacionadas se em uma amostra de observações os valores dessas variáveis são distribuídos de forma consistente. Em outras palavras, as variáveis estão relacionadas se seus valores correspondem sistematicamente uns aos outros para aquela amostra de observações.

**RELAÇÕES ESTATÍSTICAS:** São relações que se podem estabelecer entre determinadas variáveis de um problema em estudo.

**ROTAÇÃO VARIMAX (VARIMAX ROTATION):** Combinação linear dos fatores comuns em uma análise fatorial que minimiza a variância total dos fatores comuns. Comparar com rotação oblíqua e rotação ortogonal.

**RUÍDO BRANCO (WHITE NOISE):** Uma distribuição de erro em que os erros são independentes e não correlacionados. É chamado de ruído branco por analogia com a luz: como a luz branca é uma mistura de todos os comprimentos de onda, o ruído branco é uma mistura de todas as distribuições de erro. Ver independentes e identicamente distribuídos (IID).

**RELATIVO:** O conceito de relativo usado em números índices se refere a quantidade total de dinheiro gasto em cada ano, em relação a certo ano base, e que varia de um ano para outro devido as variações no número de unidades compradas dos diferentes artigos e igualmente devido a mudanças nos preços unitários de tais artigos, tem-se assim três variáveis em jogo as quais são preço, quantidade e valor, sendo esta última o resultado do produto do preço pela quantidade.

**RELATÓRIO MARGINAL:** Tabela gerada por computador das frequências das respostas a cada pergunta para monitorar a entrada de códigos válidos e para corrigir o uso de padrões sequenciais.

**RELATÓRIO TÉCNICO:** Relatar ou reportar tecnicamente é expor ou exhibir, de forma ordenada, os resultados de uma missão que se recebeu. O relator precisa saber exatamente o que se quer, por que se quer, e como será usado quando o apresentar. Na preparação do relatório, o relator deve buscar responder às seguintes perguntas: i) qual a finalidade do relatório; ii) quem o lerá; iii) como será usado; iv) o que



se quer; v) quando está sendo esperado; vi) quais decisões ele hasteará; vii) o que o leitor precisa saber a fim de entender o material.

**REML:** Máxima verossimilhança restrita.

**RENDA NACIONAL:** Soma de todas as rendas em um país. É igual ao produto nacional bruto menos depreciação e impostos indiretos.

**REPETIBILIDADE ( $\rho$ ):** Correlação fenotípica entre medidas repetidas de um mesmo caráter em um mesmo indivíduo, no tempo ou espaço.

**REPETIBILIDADE:** Representa a fração da variação fenotípica que se deve a fatores genéticos e a fatores de ambiente permanente.

**REPETIÇÃO:** Sequência de dados que exibem a mesma característica; usada no teste de repetições para a aleatoriedade.

**REPLICAÇÃO:** A execução de um experimento ou pesquisa mais de uma vez para aumentar a precisão e também obter uma estimação mais próxima de um erro de amostragem. Em inglês Replication.

**REPLICAÇÃO:** Exigência de que os tamanhos das amostras sejam suficientemente grandes para reduzir os efeitos da variação amostral aleatória.

**REPLICAÇÃO:** Geralmente, duplicação de um experimento para expor ou reduzir erro. Também um termo técnico usado em relação ao modelo de elaboração, referindo-se ao resultado da elaboração no qual a relação inicialmente observada entre duas variáveis persiste quando uma variável de controle é mantida constante.

**REPETIÇÃO:** Nome que se dá a cada uma das aplicações de um tratamento ou fator experimental.

**REPETIÇÃO:** Termo que se refere à implantação de cada tratamento em mais de uma unidade experimental. Sua utilização está vinculada à necessidade de avaliação da variabilidade do material experimental.

**REPLICAÇÃO:** Repetição de observações, ou de experimentos feitos anteriormente. O menor número de observações necessárias para favorecer uma comparação entre todas as condições experimentais. O processo de replicação é considerado especialmente necessário, quando a classe em estudo não é considerada bem definida e por isso sujeita a grandes variações individuais.

**REPLICAÇÃO:** Readministração de um experimento com o intento de validar os resultados em outra amostra de respondentes.

**REPOSITÓRIO:** Coleção centralizada de metadados disponível para acesso por todas as funções do armazém de dados.

**REPRESENTAÇÃO GRÁFICA MULTIDIMENSIONAL:** O mesmo que representação gráfica multivariada.

**REPRESENTAÇÃO GRÁFICA MULTIVARIADA:** Método para apresentação de um perfil multivariado de uma observação em três ou mais variáveis. Os métodos incluem procedimentos como glifos, transformações matemáticas e até mesmo representações iconográficas, por exemplo, faces.

**REPRESENTATIVIDADE:** Qualidade pela qual uma amostra tem a mesma distribuição de características que a população da qual ela foi selecionada. Por implicação, descrições e explicações derivadas de uma análise da amostra podem representar o mesmo na população. Representatividade é aprimorada por amostragem probabilística e permite generalizabilidade e o uso de estatística inferencial.

**REPRESENTATIVIDADE DE UM SISTEMA DE VIGILÂNCIA EPIDEMIOLÓGICA:** É a capacidade que tem o sistema de representar os subgrupos da população, permitindo a detecção, em qualquer época, acuradamente, da ocorrência de um agravo de acordo com o tempo, lugar e pessoa.

**REPRODUTIBILIDADE (DE UM TESTE):** O mesmo que confiabilidade, fidedignidade, repetibilidade e precisão. Consistência de resultados, quando o exame é repelido. Capacidade do teste para produzir resultados semelhantes ou próximos quando ele é repelido, em condições semelhantes, desde que o pesquisador desconheça os resultados anteriores da aplicação do teste; por exemplo, quando dois patologistas acertam ou erram o diagnóstico, após examinarem, independentemente um do outro, a mesma lâmina; um resultado é reproduzível, confiável ou preciso quando a repetição dá o mesmo resultado; por exemplo, uma balança precisa é aquela que informa o mesmo peso, quando a pesagem de um mesmo objeto é repetida. Diferenciar de validade de um teste.

**RESIDUAL:** Diferença entre um valor esperado ou previsto e um valor observado.

**RESIDUAL DE PEARSON:** O residual de Pearson,  $r_p$  é definido pela equação  $r_p = \frac{(Y - m)}{\sqrt{V(m)}}$ , onde Y é uma variável com média m, e V(m) é a variância de m. Em inglês Pearson residual.

**RESÍDUO:** Diferença entre um valor amostral y observado e o valor de y predito com base em uma equação de regressão.

**RESÍDUO:** É a diferença  $(x - \hat{m})$  entre um valor observado (x) e a estimativa da média da população respectiva ( $\hat{m}$ ).

**RESÍDUO:** Parte de uma variável dependente não explicada por uma técnica multivariada. Associado a métodos de dependência que tentam prever a variável dependente, o resíduo representa a parte inexplicada da mesma. Os resíduos podem ser usados em procedimentos diagnósticos para identificar problemas na técnica de estimação ou para identificar relações não especificadas.

**RESÍDUO:** Medida do ajuste preditivo para uma única observação, calculada como a diferença entre os valores reais e previstos da variável dependente. Considera-se que os resíduos têm uma média de zero e uma variância constante. Eles não apenas desempenham um papel-chave para determinar se as suposições subjacentes da regressão foram atendidas, mas também servem como uma ferramenta diagnóstica na identificação de observações atípicas e observações influentes.

**RESÍDUO:** Um termo geral que denota a quantidade restante depois de alguma outra quantidade ter sido subtraída. Ocorre em uma variedade de contextos específicos. Em inglês Residual.

**RESÍDUO:** Diferença entre o valor efetivo da variável dependente e o valor predito pela equação de regressão.

**RESÍDUO (e ou  $\epsilon$ ):** Erro na previsão de nossos dados da amostra. Raramente nossas previsões serão perfeitas. Consideraremos que o erro aleatório ocorrerá, mas assumimos que esse erro é uma estimativa do verdadeiro erro aleatório na população ( $\epsilon$ ), não apenas o erro na previsão da amostra (e). Considera-se



que o erro na população que se está estimando é distribuído com uma média de valor 0 e uma variância constante que significa população homoscedástica.

**RESÍDUO ELIMINADO:** Processo para calcular resíduos no qual a influência de cada observação é removida quando se calcula seu resíduo. Isso é obtido omitindo-se a  $i$ -ésima observação da equação de regressão usada para calcular seu valor previsto.

**RESÍDUO ESTUDANTIZADO:** Forma mais comumente usada de resíduo padronizado. Difere de outros métodos de padronização no cálculo do desvio padrão empregado. Para minimizar o efeito de uma única observação atípica, o desvio padrão de resíduos usados para padronizar o  $i$ -ésimo resíduo é computado a partir das estimativas de regressão, omitindo-se a  $i$ -ésima observação. Isso é feito repetidamente para cada observação, sempre omitindo a observação dos cálculos. Essa abordagem é semelhante ao resíduo eliminado, apesar de que, nessa situação, a observação é omitida do cálculo do desvio padrão.

**RESÍDUO ESTUDANTIZADO:** É a forma mais comumente usada de resíduo padronizado. Difere de outros métodos na maneira como calcula o desvio-padrão usado em padronização. Para minimizar o efeito de uma única observação atípica, o desvio-padrão residual para a observação  $i$  é computado a partir de estimativas e regressão, omitindo-se a  $i$ -ésima observação no cálculo das estimativas de regressão.

**RESÍDUO PADRONIZADO:** Estabelecimento de uma nova escala do resíduo para uma base comum, dividindo cada resíduo pelo desvio padrão dos resíduos. Assim, os resíduos padronizados têm uma média de 0 e um desvio padrão de 1. Cada valor residual padronizado agora pode ser visto em termos de erros padrão em tamanho de amostras de médias grandes. Isso fornece um meio direto de identificar observações atípicas, como aquelas com valores acima de 1 ou 2 para níveis de confiança de 0,10 e 0,05, respectivamente.

**RESÍDUO QUADRÁTICO:** Um inteiro  $m$  é um resíduo quadrático módulo  $n$  se  $m = r^2$  módulo  $n$  para algum  $r$ . Em inglês quadratic residue.

**RESERVA COMPATIBILIZADA:** É a reserva lavrável corrigida de possíveis efeitos de diluição, bem como do fator de recuperação na usina de beneficiamento.

**RESERVA GEOLÓGICA:** Termo que significa reserva in situ e que refere-se a tonelagem total de minério acima de um determinado teor de corte preestabelecido.

**RESERVA INDICADA:** Termo que significa a tonelagem e o teor do minério computados parcialmente de medidas e amostras específicas, ou de dados da produção, e parcialmente por extrapolação até distância razoável com base em evidências geológicas.

**RESERVA INFERIDA:** Estimativa feita com base no conhecimento dos caracteres geológicos do depósito mineral, havendo pouco ou nenhum trabalho de pesquisa.

**RESERVA LAVRÁVEL (MINERÁVEL OU APROVEITÁVEL):** Refere-se àquela parte da reserva geológica que pode ser extraída sob um plano específico de lavra.

**RESERVA MEDIDA:** É a tonelagem de minério computada pelas dimensões reveladas em afloramentos, trincheiras, galerias, trabalhos subterrâneos e sondagens, e na qual o teor é determinado pelos resultados de amostragem pormenorizada, devendo os pontos de inspeção, amostragem e medida estar tão proximamente espaçados e o caráter geológico tão bem definido que as dimensões, a forma e o teor da

substância mineral possam ser perfeitamente estabelecidos. A tonelagem e o teor computados devem ser rigorosamente determinados dentro dos limites estabelecidos, os quais não devem apresentar variação superior, ou inferior a 20% da quantidade verdadeira.

**RESUMO AMOSTRAL:** Teste paramétricos ( $t$  e  $z$ ) para uma ou duas amostras, quando se dispõe apenas de informações sobre tamanho, médias e variâncias amostrais.

**RESULTADOS NEGATIVOS (DE UMA INVESTIGAÇÃO):** Os que mostram ausência de efeito de um tratamento ou efeito colateral detectável de sua aplicação ou não detectam associação entre fator de risco e doença.

**RESULTADOS POSITIVOS (DE UMA INVESTIGAÇÃO):** Aqueles que detectam efeitos de um tratamento ou efeito colateral de sua aplicação ou mostram associação entre fator de risco e doença.

**RESPONDENTE:** Pessoa que fornece dados para análise respondendo a um questionário de survey.

**RESPOSTAS ESCALONADAS:** Respostas alternativas que são apresentadas ao entrevistado numa série contínua.

**RETA DE REGRESSÃO:** É a curva ou modelo que melhor se ajusta a todos os pontos  $(x_i, y_i)$ ,  $i=1, \dots, n$ , isto é, que melhor aproxima a relação entre as duas variáveis em estudo. A equação matemática dessa reta permite-nos fazer previsões da variável dependente por meio de variável independente.

**RETA DE REGRESSÃO:** Retas que melhor se ajusta a uma coleção de pontos que representam dados amostrais emparelhados.

**RETA DE REGRESSÃO:** É a reta traçada sobre uma dada nuvem de pontos, sendo um modelo matemático que pretende descrever a relação existente entre duas variáveis unidimensionais de uma distribuição bidimensional.

**RETA DE REGRESSÃO:** Retas calculadas na análise de regressão, usada para estimar a relação entre duas grandezas, a variável independente e a variável dependente.

**RETICULADO:** Grupo de delineamentos, que abrange vários tipos: reticulado quadrado, reticulado cúbico, reticulado retangular, e quadrado reticulado. Também se chama Látice.

**RETICULADOS:** Veja Álgebra de Boole. Em inglês lattices.

**RETICULADO CÚBICO:** Reticulado em que o número de tratamentos ( $v$ ) é o cubo do tamanho de cada bloco ( $k$ ):  $v = k^3$ .

**RETICULADO QUADRADO:** Reticulado em que o número de tratamentos ( $v$ ) é igual ao quadrado do tamanho de cada bloco ( $k$ ):  $v = k^2$ .

**RETICULADO RETANGULAR:** Reticulado em que, sendo  $v$  o número de tratamentos e  $k$  o tamanho de cada bloco  $v = k(k+1)$ .

**RETORNO NA QUALIDADE:** Objetivo administrativo baseado no princípio de que a qualidade, ao ser entregue, é aquela desejada pelo mercado-alvo, a qual precisa ter um impacto positivo na lucratividade.

**RETROCRUZAMENTO:** Acasalamento da progênie gerada por duas raças com uma das raças parentais. Exemplo: vacas Nelore-Angus acasaladas com touros da raça Angus.

**RETROPROPAGAÇÃO:** O mais comum processo de aprendizado em redes neurais, no qual erros na estimação das interconexões de saída são realimentados pelo sistema e usados como indicadores como um meio para recalibrar os pesos de cada interconexão.

**RETROSPECTIVO:** Referente ao passado; que utiliza dados já coletados. O estudo de caso-controle é retrospectivo, porque volta ao passado, na tentativa de elucidar etiologia. Ver estudo de coorte retrospectivo.

**REUNIÃO:** Outra maneira de denominar a operação de união entre dois eventos ou conjuntos. Em inglês reunion.

**REVELAÇÃO:** Representação das preferências de um indivíduo dentro de um espaço comum agregado de estímulos obtido para todos os respondentes como um todo. As preferências do indivíduo são reveladas e apresentadas como a melhor representação possível dentro da análise agregada.

**REVISÃO DE LITERATURA:** A revisão ou levantamento de literatura é a localização e obtenção de documentos para avaliar a disponibilidade de material que subsidiará o tema do trabalho de pesquisa. Este levantamento é realizado junto às bibliotecas ou serviços de informações existentes.

**REVISÃO DE LITERATURA:** Parte integral da condução de uma pesquisa. Ajuda o pesquisador a esclarecer seu problema e a evitar duplicação. Quase sempre serve também para: i) formular mais adequadamente as hipóteses; ii) planejar realisticamente a pesquisa; e iii) facilitar interpretações mais profundas dos resultados.

**R<sup>2</sup> AJUSTADO:** Um valor de R<sup>2</sup> corrigido que permite que os modelos de regressão múltipla com diferentes números de variáveis explicativas sejam comparados quando se analisa a adequação; ou seja, o grau de concordância entre uma distribuição empiricamente observada e uma distribuição teórica ou matemática. Consultar coeficiente de determinação.

**REDUÇÃO DE RISCO ABSOLUTO (RRA):** A diferença entre dois riscos.

**REDUÇÃO DE RISCO RELATIVO (RRR):** A diferença nos riscos do evento de interesse entre o grupo tratados e o grupo controle, ou seja, RRA, expressa como uma proporção do risco no grupo controle.

**REGRA DE ADIÇÃO:** A probabilidade de qualquer um de dois eventos mutuamente exclusivos ocorrerem é a soma da probabilidade de cada evento.

**REGRA DA MULTIPLICAÇÃO:** Teorema onde a probabilidade de dois eventos independentes ocorrerem é o produto da probabilidade individual ou marginal de cada evento.

**REGRESSÃO À MÉDIA:** Um procedimento pelo qual os animais que podem ter sido selecionados para o estudo porque tinham medidas extremas na variável de interesse são propensos a ter medidas mais próximas da média, caso a medida se repita em uma segunda ocasião.

**REGRESSÃO DE Y EM X:** É a equação que normalmente é um modelo linear e que descreve a relação entre uma variável dependente, y, e uma variável explicativa, x.

**REGRESSÃO LOGÍSTICA:** Consultar regressão logística linear múltipla.

**REGRESSÃO LOGÍSTICA CONDICIONAL:** Uma forma de análise de regressão logística utilizada quando o estudo envolve indivíduos pareados.

**REGRESSÃO LOGÍSTICA LINEAR MÚLTIPLA:** Uma forma de análise multivariada, frequentemente denominada regressão logística. É um modelo em que o interesse está focado na dependência da variável resposta binária em várias variáveis explicativas. Essa variável resposta binária indica se um indivíduo possui ou não a característica; a regressão logística mostra a relação linear entre as variáveis explicativas e a transformação logito da proporção de indivíduos com a característica.

**REGRESSÃO LOGÍSTICA MULTINOMIAL:** Uma modificação da análise de regressão logística utilizada quando a variável resposta é nominal, com mais de duas categorias. Também denominada regressão logística policotômica.

**REGRESSÃO LOGÍSTICA ORDINAL:** Uma modificação da análise de regressão logística, utilizada quando a variável resposta é ordinal, com mais de duas categorias.

**REGRESSÃO LOGÍSTICA POLICOTÔMICA:** Consultar regressão logística multinomial.

**REPETIÇÕES:** São as unidades experimentais que recebem o mesmo tratamento.

**REPETIBILIDADE:** O quanto são similares as medidas repetidas em circunstâncias idênticas de uma determinada técnica ou de instrumento ou do mesmo observador.

**REPLICAÇÃO:** Utilizamos mais de uma medida da variável de interesse em cada indivíduo.

**REPRODUTIBILIDADE:** Consultar concordância de método.

**RISCO HIDROLÓGICO R:** Conceito oriundo do desdobramento da noção de tempo de retorno tal como aplicado em projetos de estruturas hidráulicas de controle de cheias. Considerado um quantil de referência  $X_T$ , de tempo de retorno T, o risco hidrológico é definido como a probabilidade de que  $X_T$  seja igualado ou superado pelo menos uma vez, em um período de N anos. Em geral, o quantil de referência  $X_T$  corresponde à cheia para a qual foi projetada a estrutura hidráulica, enquanto o período de N anos corresponde à sua vida útil. Uma das possíveis deduções da expressão do risco hidrológico, aqui denotado por R, remete-nos à distribuição binomial. Com efeito, a probabilidade de que pelo menos um sucesso ocorra em um período de N anos é equivalente à probabilidade do complemento, em relação a 1, de que nenhum sucesso ocorra nesse período. Portanto, usando a notação Y para o número de eventos sucessos em N anos, tem-se que

$$R = P(Y \geq 1) = 1 - P(Y = 0) = 1 - \binom{N}{0} P^0 (1 - P)^{N-1}. \text{ Se o quantil de referência } X_T \text{ tem período de retorno}$$

T, a probabilidade de um sucesso, em um ano qualquer, é igual a  $\frac{1}{T}$ . Substituindo esse resultado na equação

$$R = P(Y \geq 1) = 1 - P(Y = 0) = 1 - \binom{N}{0} P^0 (1 - P)^{N-1}$$

$$\text{tem-se que } R = 1 - \left(1 - \frac{1}{T}\right)^N.$$

Se o risco hidrológico é previamente fixado, em função da importância e das dimensões da estrutura hidráulica, bem como das consequências de seu eventual colapso para as populações ribeirinhas ou para as comunidades localizadas a jusante de sua posição no sistema fluvial, pode-se empregar a equação

$$R = 1 - \left(1 - \frac{1}{T}\right)^N$$

para determinar para qual tempo de retorno deve ser calculada a cheia de projeto, por exemplo, do vertedouro de uma barragem, cuja vida útil estimada é de N anos.

**RISCO HIDROLÓGICO R:** É um conceito que é função da definição do período de retorno e representa a probabilidade de um valor x da variável aleatória X ser excedido em pelo menos uma vez em n anos sucessivos. Exprime-se por,  $R = \frac{n}{T} = 1 - \left(\frac{1}{1-T}\right)^n = 1 - (1 - G(x))^n$ .

**RESÍDUO SCHOENFELD:** Tipo de resíduo obtido em estudos de análise de sobrevivência o qual representa a diferença entre os valores observados de covariáveis de um indivíduo com tempo de ocorrência do evento  $t_j$  e os valores esperados em  $t_j$  dado o grupo de risco  $R(t_j)$ .

**RESÍDUOS MARTINGALE:** Tipo de resíduo obtido em estudos de análise de sobrevivência o qual representa a diferença entre o número observado de eventos para um indivíduo e o esperado dado o modelo ajustado, o tempo de seguimento e o percurso observado de quaisquer covariáveis tempo-dependentes. É semelhante aos resíduos dos modelos de regressão linear devido a o valor esperado igual a zero em torno do verdadeiro e desconhecido  $\beta$ ; não são simetricamente distribuídos em torno de zero, variando de  $(-\infty, 1]$  e quando o tempo de sobrevivência é censurado o resíduo é negativo; o somatório dos resíduos observados é igual a zero; os resíduos Martingale  $M_i$  são não correlacionados, mas as estimativas  $\hat{M}$  são negativamente correlacionadas, ainda que fracamente. E diferentes dos resíduos da regressão linear: i) a soma de quadrados dos resíduos não auxilia na avaliação do ajuste global do modelo como exemplo tem-se que o melhor modelo de Cox ajustado não tem a menor soma de quadrados de resíduos martingale; ii) a distribuição dos resíduos não é aproximadamente normal; o gráfico de resíduos versus valores ajustados não funciona para resíduos martingale pois estes são negativamente correlacionados com os valores ajustados.

**RESÍDUOS SCORE:** Tipo de resíduo obtido em estudos de análise de sobrevivência os quais são úteis para verificar a influência de cada observação no ajuste do modelo. Permite a estimação robusta da variância dos coeficientes de regressão, sendo usados para dados em cluster. A influência de cada observação deve ser proporcional à  $(x_i - \bar{x})$  x resíduo, uma vez que a observação deve estar distante do valor médio de x e ter um resíduo alto para exercer uma influência significativa no ajuste do modelo. Para o modelo de Cox essa medida está relacionada aos resíduos score, cuja definição matemática baseia-se em propriedades de martingale dos processo de contagem. Para cada indivíduo i pode-se calcular a diferença entre o vetor de covariáveis estimado pelo modelo e o mesmo estimado sem o indivíduo i:  $\Delta\beta = \hat{\beta} - \hat{\beta}_{(-i)} \approx I'(\mathbf{U})^{-1} - 1'D$ , que é aproximadamente igual à matriz de resíduos score. O gráfico do resíduo score para cada covariável  $\Delta\beta_k$  versus  $x_j$  revela os pontos de influência ou mostra pontos de alavanca, ou seja, os indivíduos que influenciam fortemente a estimativa do parâmetro de cada covariável. Os resíduos score podem ser escalonados pelo erro padrão da respectiva covariável, facilitando a visualização. Para as variáveis categóricas é possível fazer gráficos do tipo boxplot, permitindo a identificação de valores extremos. Uma vantagem desses resíduos sobre os outros é que eles são definidos nos tempos de ocorrência do evento, ou seja, a vantagem, deste tipo de resíduo é que eles são definidos para todos os tempos, mesmo onde

não ocorre evento, melhorando a análise quando há muita censura. Essa característica é especialmente importante quando o percentual de censuras é alto. Uma desvantagem é que, ao contrário dos resíduos de Schoenfeld, eles não indicam o tipo de desvio da suposição de riscos proporcionais quando existe uma indicação de falta de ajuste do modelo. Os resíduos score também são utilizados para se encontrar uma estimação robusta da variância de  $\hat{\beta}$  que será utilizada quando cada indivíduo gera mais de uma observação, como no caso de eventos múltiplos.

**RESIDUAL:** Em geral, o resíduo é a diferença entre quantidades; na regressão, é a diferença entre o valor observado da variável resposta e o seu valor previsto pelo modelo.

**RETIRADOS:** Animais que, após serem incluídos no teste, falham ao aceitar o protocolo, talvez em decorrência dos afeitos colaterais ou de impossibilidade de tratamento.

**REVISÃO SISTEMÁTICA:** Um exame qualitativo claramente definido de resultados publicados e não-publicados, que contém a informação para responder questões, normalmente a respeito da eficácia de um tratamento.

**RISCO:** Utilizado em testes de sobrevivência para descrever a probabilidade instantânea de presenciar o evento de interesse, como por exemplo, morte em um determinado tempo.

**RISCO ATRIBUÍVEL:** A proporção do risco total de doença atribuível a um fator.

**RISCO DE DOENÇA:** Número de novos casos expressos como uma proporção daquelas inicialmente em risco.

**RISCO RELATIVO:** A proporção de dois riscos, normalmente a proporção do risco de doença no grupo exposto em relação ao risco de doença no grupo não exposto. Fornece uma medida de força da associação entre a doença e o fator de exposição. Caso o risco relativo seja homogêneo, a exposição ao fator não a chance de o animal desenvolver a doença; caso seja maior do que a unidade, há indícios de maior risco associado à exposição.

**RISCO:** Esta palavra aparece em estatística na sua forma simples, e tem um significado especial na teoria de Funções de Decisão. Onde um número de decisões possíveis tem uma função de perda associada, o risco é o custo esperado de um experimento mais o valor esperado da função de perda. Em inglês Risk.

**RISCO:** Em geral, uma palavra que implica na existência de chance ou risco. Uso especial acontece principalmente em conexão com análise do tempo de vida de sistemas físicos ou componentes. Em inglês Hazard.

**RISCO:** Termo usado em bioestatística o qual representa a probabilidade de perigo, geralmente com ameaça física ou psicológica para o homem e/ou o meio ambiente.

**RISCO:** Probabilidade de um indivíduo desenvolver um resultado tal como uma doença ou outro desfecho clínico, em um certo período de tempo.

**RISCO:** A proporção de pessoas que não estão afetadas no começo do período de estudo, mas que passam pelo evento de risco, tal como morte, doença ou acidente durante o período de estudo.

**RISCO:** É calculado como seno a proporção das pessoas que não estão afetadas no início do período de estudo, mas que sofreram o evento a que o risco se refere durante o período de estudo.



**RISCO:** É a probabilidade de ocorrência de um resultado desfavorável, de um dano ou de um fenômeno indesejado. Dessa forma, estima-se o risco ou a probabilidade de que uma doença existe, por meio de coeficientes de incidência e prevalência. Considera-se fator de risco de um dano toda característica ou circunstância que acompanha um aumento de probabilidade de ocorrência do fato indesejado, sem que o dito fator tenha que intervir necessariamente em sua casualidade.

**RISCO ABSOLUTO:** O mesmo que incidência.

**RISCO ATRIBUÍVEL:** Sinônimo: fração atribuível ou fração etiológica. É a proporção de um dano à saúde, entre os expostos, que é devida ou atribuída a um fator de risco; ou seja, informa o quanto da distribuição do dano à saúde é devido ao fator de risco, em foco. O cálculo é feito por simples subtração, entre dois coeficientes ou proporções de expostos e não expostos ao fator de risco, usualmente expressos por taxas de incidência ou de mortalidade. A diferença informa sobre o risco em excesso, ou a fração atribuível ao fator de risco.

**RISCO ATRIBUÍVEL (RA):** É a parcela do risco total para um desfecho específico que se pode atribuir a uma determinada exposição. O mesmo que diferença de risco.

**RISCO ATRIBUÍVEL:** É um risco ou incidência adicional de doença após uma exposição, além daquele experimentado por pessoas não expostas. Corresponde à incidência da doença em pessoas expostas menos a incidência em pessoas não expostas. É a incidência adicional de doença relacionada à exposição, levando em conta a incidência basal da doença presumivelmente devido a outros fatores. O mesmo que diferença de risco.

**RISCO ATRIBUÍVEL:** É a parte da incidência de um dano à saúde que é devida ou atribuída a uma da exposição. Sua computação é feita pela subtração entre dois coeficientes ou proporções de expostos e não expostos, usualmente expressos por taxas de incidência ou de mortalidade, apontando a diferença sobre o risco em excesso, ou a fração atribuível ao fator de risco. Sinônimos: fração atribuível, fração etiológica. O mesmo que diferença de risco.

**RISCO ATRIBUÍVEL NA POPULAÇÃO:** Medida do excesso de incidência de doença numa comunidade que se associa com um fator de risco, resultante do produto do risco atribuível pela prevalência do fator de risco na população. Representa a contribuição do fator de risco para as taxas globais de doença em populações, e não apenas em indivíduos expostos.

**RISCO ATRIBUÍVEL NA POPULAÇÃO:** Indicador para verificar o risco atribuível a um fator de risco na população, considerando-se nos cálculos a frequência com que este fator de risco existe na população. Há diversos modos de computá-lo:

$$RAP = \frac{it - In}{In}$$

$$RAP = \frac{(Ie - In)P}{It}$$

$$RAP = \frac{P(RR - 1)}{P(RR - 1) + 1}$$

Onde, It = Incidência na população total; In = Incidência em não-expostos; Ie = Incidência em expostos; P = Prevalência do fator de risco na população e RR = Risco relativo.

**RISCO ATRIBUÍVEL NA POPULAÇÃO:** É o risco na população total menos o risco no grupo não exposto. O risco atribuível na população responde à seguinte pergunta: na população geral, o risco total para a doença em pauta é devido ao fator de risco ou exposição de interesse.

**RISCO ATRIBUÍVEL POPULACIONAL:** A proporção de um dano à saúde, na população, que é devida ou atribuída a um fator de risco. Leva em conta a prevalência do fator de risco, na população; por exemplo, 80 % de casos de câncer de pulmão são atribuídos ao fumo.

**RISCO COMPETITIVO:** Um evento que retira um indivíduo de estar sujeito a outro resultado clínico. Por exemplo, o óbito por acidente de trânsito afasta a possibilidade de óbito por neoplasia, o que dificulta a interpretação de investigações sobre os efeitos de radiações.

**RISCO DE BAYES:** O valor minimizado do risco esperado em uma Solução de Bayes. Do inglês Bayes risk.

**RISCO RELATIVO:** Testa estatístico simbolizado por RR, representando a razão entre o coeficiente de incidência de determinado evento em pessoas expostas a uma doença, por exemplo e coeficiente de incidência do mesmo evento em indivíduos não expostos.

**RISCO RELATIVO:** Proporção de pessoas expostas ao fator de risco que desenvolveram a doença sobre a proporção de pessoas não expostas que desenvolveram a doença. Vale 1 se as pessoas expostas e não expostas desenvolveram a doença na mesma proporção.

**RISCO RELATIVO:** É a relação entre o coeficiente de incidência referente aos expostos a fatores de risco e o coeficiente de incidência referente aos não expostos a esses mesmos fatores. O risco relativo indica quantas vezes é mais freqüente o dano nos expostos que têm o fator do que nos não expostos. Um risco relativo alto contribui para afirmar a casualidade. Essa medida é algumas vezes chamada de coeficiente de probabilidade.

**RISCO RELATIVO:** Razão entre o risco no grupo dos expostos e o risco no grupo dos não expostos, que é expressa como

$$RR = \text{Risco}_{(\text{exposto})} - \text{Risco}_{(\text{não-expostos})} = \frac{\left[ \frac{a}{(a+b)} \right]}{\left[ \frac{c}{(c+b)} \right]}$$

**RISCO RELATIVO (RR):** É a razão entre dois riscos ou a razão entre duas taxas de incidência ou de mortalidade. Corresponde ao risco da doença entre os indivíduos que tenham tido uma dada exposição. Informa quantas vezes um risco é maior que outro. Por exemplo, um risco relativo igual a 2 significa risco de duas vezes maior em um grupo, em relação a outro.

**RISCO RELATIVO (RR):** Proporção de doentes entre todos os expostos a um fator relevante de risco dividido pela proporção de doentes entre aqueles não expostos ao mesmo fator. Isto é usado em estudos de coorte onde os com e sem a doença são acompanhados para se saber que indivíduos se tornam doentes.

**RISCO RELATIVO AJUSTADO:** Um risco relativo em que uma ou mais variáveis confundidas foram neutralizadas. Quando nenhuma variável é neutralizada, trata-se de um risco relativo bruto ou não ajustado.

**RIQUEZA DE ESPÉCIES:** Número de espécies presente em determinado número de indivíduos. É a forma mais simples de estimar a diversidade de espécies de uma comunidade. Grandeza adimensional.



**ROBUSTEZ:** Um teste é considerado robusto se os pontos de significância de um teste variam pouco se a população se afasta substancialmente da normalidade. Em um sentido mais geral, um procedimento estatístico é descrito como robusto se não for sensível às suposições iniciais a qual é dependente. Em inglês Robustness.

**ROL:** É o arranjo dos dados brutos em ordem de frequência crescente ou decrescente.

**ROL:** É a tabela obtida após a ordenação dos dados em ordem de frequência crescente ou decrescente.

**ROTAÇÃO FATORIAL:** Processo de manipulação ou de ajuste dos eixos fatoriais para conseguir uma solução fatorial mais simples e pragmaticamente mais significativa.

**ROTAÇÃO FATORIAL OBLÍQUA:** Rotação fatorial computada de modo que os fatores extraídos são correlacionados. Ao invés de restringir arbitrariamente a rotação fatorial a uma solução ortogonal, a rotação oblíqua identifica o grau em que cada fator está correlacionado.

**ROTAÇÃO FATORIAL ORTOGONAL:** Rotação fatorial na qual os fatores são extraídos de modo que seus eixos sejam mantidos em 90 graus. Cada fator é independente, ou ortogonal, em relação a todos os outros. A correlação entre fatores é determinada como 0.

**ROTINA DE CHECAGEM DE ERROS:** Programas de computador que aceitam instruções do usuário para fazer uma checagem à procura de erros lógicos nos dados.

**RRR:** Redução do risco relativo representado por  $1-RR$ .

**$r_q$  :** Valor da amplitude amostral interquartil.

**REPERCURSÃO:** Na análise de regressão linear simples repercurssão do  $i$ -ésimo valor,  $x_i$ , é dada por

$$h_i = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{j=1}^n (x_j - \bar{x})^2},$$

em que se passa a designar  $h_i$  por  $h_i$ . A variância do  $i$ -ésimo resíduo é dada por,  $\text{var}(r_i) = \sigma^2(1 - h_i)$ . Consequentemente, há um padrão induzido nos resíduos através da repercurssão dos diferentes dados do conjunto de valores. A variância de um resíduo  $r_i$  é pequena quando a repercurssão é grande, como acontece quando  $x_i$  se afasta de  $\bar{x}$ . Assim, quando um modelo se ajusta, os pontos cujas abscissas se afastam mais de  $\bar{x}$  tendem a aproximar-se mais da reta de regressão.

**RUÍDO:** Termo conveniente para uma série de distúrbios aleatórios emprestados da teoria de som. Em teoria de comunicação, o ruído resulta na possibilidade de que um sinal enviado,  $x$ , seja diferente do sinal recebido  $y$ , tendo este último uma distribuição de probabilidade que depende de  $x$ . Em inglês Noise.

**ROL:** É a mais simples organização numérica. É a ordenação dos dados em ordem crescente ou decrescente.

**REPRESENTATIVIDADE:** Corresponde à possibilidade de manter as mesmas características presentes na população.

**$r^2$  :** Coeficiente de determinação amostral.

**$R^2$ :** Medida do quanto uma equação de regressão linear simples pode explicar variações na variável dependente; seu valor situa-se entre 0 que é o ajuste pobre e 1 o qual mostra um ajuste perfeito. Conhecido também como coeficiente de determinação ou de explicação.



# S

**S:** Espaço amostral, também simbolizado pela letra do alfabeto grego ômega maiúscula que é dada por  $\Omega$ .

**S:** Desvio padrão amostral ou raiz quadrada da variância.

**s:** Valor do desvio padrão amostral simbolizado pela letra s minúsculo.

**SAÚDE PÚBLICA:** É a ciência e a arte de evitar doenças, prolongar a vida e desenvolver a doença física, mental e a eficiência, por meio de esforços organizados da comunidade, para o saneamento do meio ambiente, o controle de infecções na comunidade, a organização de serviços médicos e paramédicos para diagnóstico precoce e o tratamento preventivo de doença, e o aperfeiçoamento da máquina social que irá assegurar a cada indivíduo, dentro da comunidade, um padrão de vida adequado à manutenção da saúde.

**SAZONALIDADE:** Em análise de séries de tempo, a parte do movimento atribuída ao efeito das estações do ano. Em inglês Seasonality.

**SAZONALIDADE:** Propriedade segundo a qual o fenômeno considerado é periódico e se repete sempre na mesma estação ou sazão do ano. A mortalidade por doenças respiratórias em crianças menores de 1 ano, por exemplo, no Rio grande do Sul, ocorre em maior proporção nos meses mais frios do ano: junho e julho.

**SAS:** Sigla de Statistical Analysis System. Programa de computador utilizado para cálculos e análises estatísticas de dados, bem como outros tipos de análise gráfica e de dados. Ver planilha eletrônica de dados.

**SIGMA:** Letra grega que pode significar, quando maiúscula ( $\Sigma$ ), uma somatório de valores e, quando minúscula ( $\sigma$ ), o desvio-padrão de uma população.

**SOCIOMETRIA:** Conjunto de técnicas ou métodos cujo objetivo é o de mensurar atitudes e comportamentos de vínculo, repulsa, interação ou padrões de comunicação entre os indivíduos num grupo.

**SPSS:** Sigla de Statistical Package for the Social Sciences. Programa de computador largamente utilizado para cálculos e análises estatísticas de dados, notadamente nas ciências humanas.

**STATISTICA:** Programa de computador largamente utilizado para cálculos e análises estatísticas de dados.

**SCANNERS:** Dispositivos que leem os códigos UPC em produtos e que produzem informações instantâneas sobre vendas.

**SDFBETA:** Ver DFBETA.

**SDFFIT:** Ver DFFIT.

**SE ... ENTÃO:** Símbolo matemático dado por  $\rightarrow$ .

**SEGUNDA LEI DE MENDEL:** Lei que emergiu do trabalho devido ao monge austríaco Gregor Mendel no século XIX que deduziu importantes princípios genéticos baseados nos seus experimentos com ervilhas. Mendel estudou sete fenótipos da ervilha, cada qual determinado por um único gene. Essa lei afirma que

genes localizados em diferentes loci são transmitidos independentemente. Conhecida também como lei da distribuição independente.

**SELEÇÃO:** Sobrevivência diferencial ou reprodução não-aleatória de classes de entidades fenotipicamente diferentes.

**SELEÇÃO ALEATÓRIA:** Método para selecionar unidades de uma amostra de forma que cada amostra possível tenha uma probabilidade fixa e determinada de ser selecionada. Em inglês Random selection.

**SELEÇÃO ALEATÓRIA:** Seleção de elementos amostrais de tal forma que todos os elementos disponíveis para seleção tenham a mesma chance de ser escolhidos.

**SELEÇÃO ALEATÓRIA:** Seleção de amostras pelo uso de métodos de probabilidade. Cada elemento na população amostrada deve ter uma chance conhecida não zero de seleção.

**SELEÇÃO DE MODELO:** A seleção de modelos é uma tarefa de muita importância em estatística. Uma situação real pode ser representada por meio de vários modelos que passam a competir entre eles. A seleção de modelos se encarrega de criar mecanismos para a escolha do modelo que melhor representa os dados obtidos a partir de um problema real. Em inglês Model choice.

**SELEÇÃO DE MODELOS:** Ver: Model choice. Em inglês Model selection.

**SELEÇÃO DOS REGRESSORES (STEPWISE REGRESSION) Constitui na seleção das variáveis independentes na regressão linear múltipla, no sentido de remover aquelas cujo p-valor é superior ao nível alfa previamente estabelecido. Dois métodos são utilizados:** i) Método Progressivo (Forward Selection) e ii) Modelo Regressivo (Backward Selection).

**SELEÇÃO FORWARD:** Método de seleção de variáveis numa regressão linear que consiste em ajustar o modelo inserindo as covariáveis uma a uma. Em inglês Forward selection.

**SELECIONADORES:** Perguntas usadas para fazer a triagem de respondentes apropriados.

**SEMENTES DE AGRUPAMENTO:** Centróides iniciais ou pontos de partida para agrupamentos. Esses valores são selecionados para iniciar procedimentos de agrupamento não-hierárquico, nos quais os agrupamentos são construídos em torno desses pontos pré-especificados.

**SEMI AMPLITUDE:** Para um conjunto de valores,  $x_1, \dots, x_n$  arrumados em ordem de magnitude, a semi amplitude é definida como:

$$\frac{(X_1 + X_2)}{2}. \text{ Em inglês Mid-range.}$$

**SELEÇÃO BACKWARD STEP-DOWN:** É o procedimento de escolher um modelo de regressão ótimo em que, começando com todas as variáveis explicativas no modelo, removemos todas elas sequencialmente, iniciando com a variável que menos contribui, até que a retirada de uma variável aumente significativamente a variância residual.

**SELEÇÃO FORWARD STEP-UP:** O processo de seleção de um modelo de regressão ótimo; iniciamos com a variável explicativa que mais contribui para a variação explicativa de  $y$  e incluímos mais variáveis na equação, progressivamente, até que a adição de uma variável não melhora significativamente a situação.

**SELEÇÃO ALEATÓRIA:** O processo pelo qual os indivíduos são escolhidos para serem incluídos em uma amostra aleatória.

**SELEÇÃO DE TODOS OS SUBCONJUNTOS:** O procedimento para determinação de um modelo de regressão ótimo que avalia todos os modelos possíveis para as várias combinações das variáveis explicativas de interesse.

**SELEÇÃO GRADATIVA:** O procedimento para determinar um equação de regressão ótimo; essencialmente é um procedimento de aprimoramento que se inicia com uma variável, com adição de mais variáveis sucessivamente; permite que as variáveis da equação sejam excluídas do modelo de acordo com um critério estatístico definido especificado no programa de computador.

**SENO INVERSO:** Consultar transformação arco seno ou angular.

**SENSIBILIDADE:** A efetividade de um diagnóstico ou teste de triagem na identificação de animais com a doença. É a proporção de positivos verdadeiros identificada pelo teste como positiva.

**SÉRIE AUTO-REGRESSIVA:** Uma série temporal em que o valor de uma observação depende do número de observações precedentes.

**SÉRIE DE MOMENTOS:** Uma longa série de medidas realizadas em vários pontos sucessivos no tempo. Normalmente, as observações sucessivas são dependentes, de modo que a magnitude de um valor influencia a magnitude do próximo, ou seja, temos uma série auto-regressiva.

**SÉRIE DE MOMENTOS CRUZADOS:** Consultar modela multinível.

**SERVIÇOS DIAGNÓSTICOS:** Análises de amostras de animais para beneficiar o monitoramento da saúde e o diagnóstico da doença.

**SIGNIFICÂNCIA ESTATÍSTICA:** O resultado de um teste de hipótese é estatisticamente significativo quando a decisão é rejeitar a hipótese nula; a significância estatística deve ser contrastada com a importância biológica.

**SOMA DOS QUADRADOS:** A soma dos desvios quadrados de cada observação da média; é utilizada nos cálculos dos quadrados médios na análise de variância.

**SENSIBILIDADE:** Em ensaios clínicos, a proporção de positivos corretamente identificados pelo teste clínico. Em inglês Sensitivity.

**SENSIBILIDADE:** Proporção de testes positivos, entre os doentes. Diferenciar do termo denominado de especificidade.

**SENSIBILIDADE:** Capacidade de um teste em detectar uma doença quando ela está presente.

**SENSIBILIDADE DE UM SISTEMA DE VIGILÂNCIA EPIDEMIOLÓGICA:** É a capacidade que o sistema de vigilância tem de detectar casos verdadeiros do evento sob vigilância. Um sistema de vigilância com 100 % de sensibilidade detectará todos os agravos que ocorrem na população. Um sistema que não tenha um sensibilidade ala ainda pode ser útil para a determinação de tendências.

**SENSIBILIDADE DE UM TESTE:** Percentagem de indivíduos com o evento investigado, como por exemplo determinada afecção ou se o tuberculínico é positivo (+), dentre outros.



**SEPARADOR DE FREQUÊNCIAS:** É um tipo de tabela que permite ter uma percepção imediata do aspecto global dos dados sem perda da informação contida na coleção dos dados inicial.

**SEPARATRIZES:** Além das medidas de posição que estudamos, há outras que, consideradas individualmente, não são medidas de tendência central, mas estão ligadas à mediana relativamente à sua característica de separar a série em duas partes que apresentam o mesmo número de valores.

**SEPARATRIZES:** São outras medidas de posição que englobam a própria mediana, os decis, os quartis e os percentis.

**SEQUÊNCIA:** Em inglês run.

**SEQUÊNCIA TEMPORAL:** Ordem apropriada casual de eventos.

**SCREE PLOT (SCREE PLOT):** Um gráfico de diagnóstico usado para decidir quantos componentes, fatores ou eixos são necessários reter em uma ordenação. Ele tem formato de encosta de montanha ou declive de pedra; os componentes úteis são os mais altos na encosta e os sem utilidade são os escombros, embaixo.

**SELEÇÃO PROGRESSIVA (FORWARD SELECTION):** Um método de regressão gradativa (stepwise regression) que começa com um modelo que possui apenas um dentre vários parâmetros possíveis. Os parâmetros adicionais são testados e inseridos sequencialmente um de cada vez. Ver também regressão gradativa e comparar com eliminação regressiva.

**SEMIMÉTRICA (SEMI-METRIC):** Um tipo de medida de distância ou similaridade que satisfaz três propriedades: i) a distância mínima é igual a 0 e a entre objetos idênticos é também igual a 0; ii) a distância entre dois objetos não idênticos é positiva; iii) as distâncias são simétricas, ou seja, a do objeto a até um objeto b é a mesma que do objeto b até o a. Comparar com métrica e não métrica.

**SILOGISMO (SYLLOGISM):** Uma dedução, ou conclusão, lógica, por exemplo, este é um cavalo, é derivada de duas premissas, por exemplo, i) todos os cavalos são brancos; ii) este objeto é branco. Os silogismos foram inicialmente descritos por Aristóteles. Embora sejam lógicos, eles podem levar a conclusões errôneas, por exemplo, e se o objeto branco na verdade for uma batata?.

**SIMÉTRICO (SYMMETRIC):** Um objeto cujos lados são imagens-espelho uma da outra. Uma matriz é dita simétrica se um espelho for alinhado ao longo de sua diagonal. Os valores de linha e coluna acima da diagonal devem ser iguais aos valores de coluna e linha abaixo da diagonal.

**SOMA DOS PRODUTOS CRUZADOS (SPC) (SUM OF CROSS PRODUCTS-SCP):** A soma do produto da diferença entre cada observação e seu valor esperado. A soma dos produtos cruzados dividida pelo tamanho amostral é a covariância.

**SOMA DOS QUADRADOS (SQ) (SUM OF SQUARES-SS):** A soma das diferenças ao quadrado entre o valor de cada observação e a média aritmética de todos os valores. Usada extensivamente em cálculos da análise de variância (ANOVA).

**SOMA DOS QUADRADOS DOS RESÍDUOS (RESIDUAL SUM OF SQUARES):** O que é deixado ou não é explicado por um modelo de regressão ou análise de variância (ANOVA). É calculado como a soma dos desvios quadrados de cada observação até a média de todas as observações em seus grupos de

tratamento, então essas somas são calculadas sobre todos os grupos de tratamento. Também chamada de variação do erro ou variação residual.

**SOMA DOS QUADRADOS E PRODUTOS CRUZADOS (SQPC) (SUM OF SQUARES AND CROSS PRODUCTS-SSCP):** Uma matriz quadrada com o número de linha e número de colunas igual ao de variáveis em um conjunto de dados multivariado. Os elementos da diagonal dessa matriz são a soma dos quadrados de cada variável e os elementos fora da diagonal são a soma dos produtos cruzados de cada par de variáveis. As matrizes SQPC são usadas para calcular estatísticas teste na análise de variância multivariada (MANOVA).

**SERENDIPIDADE:** Refere-se a importantes conceitos ou descobertas que aconteceram acidentalmente, ou de outra forma, sem ser propositalmente buscados.

**SÉRIE CATEGÓRICA OU ESPECIFICATIVA:** É uma série estatística na qual o elemento variável é o fenômeno estudado, mantendo-se fixos o tempo e o local de observação.

**SÉRIE CRONOLÓGICA, TEMPORAL, EVOLUTIVA OU HISTÓRICA:** É uma série estatística na qual o tempo varia, mantendo-se fixos o local e o fenômeno estudado.

**SÉRIE DE CASOS:** Conjunto de pacientes com um mesmo diagnóstico ou submetidos a uma mesma intervenção. Trata-se, em geral, de série consecutiva de pacientes, vistos em um hospital ou em outra instituição de saúde, durante um certo período. Não há grupo-controle interno, ou seja, composto simultaneamente. A comparação é feita com controles externos. Para alguns, a série de casos deve incluir um mínimo de 10 casos.

**SÉRIE DE FOURIER:** Série de fatores de seno e cosseno da forma  $a_j \cos(2\pi j/\lambda_j)$ ,  $j=0,1,\dots$ . Em inglês Fourier series.

**SÉRIE CRONOLÓGICA:** É a série estatística em que os dados são observados segundo a época de ocorrência. É também conhecida por série temporal, evolutiva ou histórica.

**SÉRIE ESPECÍFICA:** É a série estatística em que os dados são agrupados segundo a modalidade de ocorrência.

**SÉRIE GEOGRÁFICA:** É a série estatística em que os dados são observados segundo a localidade de ocorrência. É também chamada de série de localização.

**SIGMA:** É a palavra que expressa a letra do alfabeto grego que corresponde ao nosso s. A letra minúscula simboliza o desvio padrão e a letra maiúscula significa o somatório.

**SÉRIE DE FOURIER:** As séries trigonométricas infinitas formadas por seno e/ou cosseno são chamadas séries de Fourier dada pela equação:

$$\frac{a_0}{2} + \sum_{m=1}^{\infty} \left( a_m \cos\left(\frac{m\pi x}{L}\right) + b_m \text{sen}\left(\frac{m\pi x}{L}\right) \right).$$

No conjunto de pontos onde ela converge, ela define uma função  $f$ , cujos valores em cada ponto  $x$  é a soma da série para aquele valor de  $x$ . Diz-se então que esta série é a série de Fourier de  $f$ . Em resumo, qualquer função  $f(x)$  pode, segundo Fourier, ser escrita na forma da soma de uma série de funções seno e cosseno da seguinte forma geral:

$$f(x) = a_0 + a_1 \sin(x) + a_2 \sin(2x) + a_3 \sin(3x) + \dots + b_1 \cos(x) + b_2 \cos(2x) + b_3 \cos(3x) + \dots$$

Os pontinhos nessa equação indicam que os termos tipo seno e cosseno podem se estender indefinidamente, se necessário, para melhor representação da função original  $f(x)$ . Resta achar uma forma de calcular os coeficientes  $a_0, a_1, a_2, \dots, b_1, b_2$ , dentre outros, de cada termo da série. Esses coeficientes, como vemos, são as amplitudes de cada onda componente do desenvolvimento em série.

**SÉRIE DE MACLAURIN:** Seja  $f$  uma função com derivadas de todas as ordens em algum intervalo contendo  $a$  como um ponto interior. Então, a série de Taylor gerada por  $f$  em  $x = a$  é

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!} (x-a)^k = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!} (x-a)^2 + \dots + \frac{f^{(n)}(a)}{n!} (x-a)^n + \dots$$

A série de Maclaurin gerada por  $f$  é

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} x^k = f(0) + f'(0)x + \frac{f''(0)}{2!} x^2 + \dots + \frac{f^{(n)}(0)}{n!} x^n + \dots,$$

a série de Taylor gerada por  $f$  em  $x = 0$ .

A idéia proposta por Maclaurin era supor que uma função poderia ser escrita na forma de uma série de potências, ou seja,

$$f(x) = a_0 + a_1 x + a_2 x^2 + \dots + a_k x^k + \dots$$

Então restando determinar os coeficientes  $a_k$  adequadamente. Substituindo  $x$  por  $0$  tem-se:

$$f(0) = a_0 + a_1 \cdot 0 + a_2 \cdot 0^2 + \dots + a_k \cdot 0^k + \dots = a_0$$

Influenciado pelos trabalhos de Newton sobre o cálculo infinitesimal, Maclaurin observou que, nas condições enunciadas,

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + \dots + a_k x^k + \dots \Rightarrow f(0) = a_0$$

$$f'(x) = a_1 + 2a_2 x + 3a_3 x^2 + 4a_4 x^3 + \dots + k a_k x^{k-1} + \dots \Rightarrow f'(0) = 1 \cdot a_1$$

$$f''(x) = 2 \cdot a_2 + 3 \cdot 2 a_3 x + 4 \cdot 3 a_4 x^2 + \dots + k(k-1) a_k x^{k-2} + \dots \Rightarrow f''(0) = 2 \cdot 1 \cdot a_2$$

$$f'''(x) = 3 \cdot 2 \cdot a_3 + 4 \cdot 3 \cdot 2 a_4 x + \dots + k(k-1)(k-2) a_k x^{k-3} + \dots \Rightarrow f'''(0) = 3 \cdot 2 \cdot 1 \cdot a_3$$

...

Genericamente:

$$f^{(k)}(0) = k! a_k$$

Ou ainda:

$$a_0 = f(0)$$

$$a_k = \frac{f^{(k)}(0)}{k!}$$

A forma geral da Série de Maclaurin é, então, dada por

$$f(x) = f(0) + \frac{f'(0)}{1!} x + \frac{f''(0)}{2!} x^2 + \dots + \frac{f^{(k)}(0)}{k!} x^k + \dots$$

$$\text{Ou } f(x) = \sum_{k=0}^{+\infty} \frac{f^{(k)}(0)}{k!} x^k$$

Observe que para ser possível a expansão em Série de Maclaurin: A função tem de estar definida em  $x = 0$ ; a série deve ser convergente.

Ao polinômio gerado pelo truncamento da Série de Maclaurin no termo de grau  $n$  dá-se o nome de polinômio aproximador de Maclaurin de grau  $n$ :

$$p_n(x) = f(0) + \frac{f'(0)}{1!} x + \frac{f''(0)}{2!} x^2 + \dots + \frac{f^{(n)}(0)}{n!} x^n.$$

**SÉRIE DE TAYLOR:** Também conhecido por polinômio de Taylor, trata-se de uma série infinita de uma função, a qual o valor dessa função pode ser calculado algebricamente por uma soma infinitamente aproximada ou exata, dependendo do valor e do tamanho da série.

$$f(x) = f(a) + \frac{f'(a)}{1!} (x-a) + \frac{f''(a)}{2!} (x-a)^2 + \frac{f'''(a)}{3!} (x-a)^3 + \frac{f^{(4)}(a)}{4!} (x-a)^4 + \dots$$

Com essa função, depende do conceito de derivadas, pode-se usá-la como um algoritmo para descobrir o valor de qualquer função, até daquelas mais complicadas. A seguir alguns exemplos de funções, por meio de polinômio de Taylor.

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

$$a^x = 1 + \frac{x \ln a}{1!} + \frac{(x \ln a)^2}{2!} + \frac{(x \ln a)^3}{3!} + \dots$$

$$\ln x = 2 \left[ \frac{x-1}{x+1} + \frac{1}{3} \left( \frac{x-1}{x+1} \right)^3 + \frac{1}{5} \left( \frac{x-1}{x+1} \right)^5 + \dots \right] \text{ para } x > 0$$

$$\text{sen } x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots$$

$$\text{cos } x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots$$

$$\text{tg } x = x + \frac{1}{3} x^3 + \frac{2}{15} x^5 + \frac{17}{315} x^7 + \dots, \text{ para quaisquer } x.$$

**SÉRIE ESPECÍFICA:** O caráter variável é apenas o fato ou espécie. Também é chamada de série categórica.

**SÉRIE ESTATÍSTICA:** É qualquer tabela que apresenta a distribuição de um conjunto de dados estatísticos em função da época, do local ou da espécie.

**SÉRIE GEOGRÁFICA:** Apresenta como elemento variável o fator geográfico. A época e o fato ou espécie são elementos fixos. Também é chamada de espacial, territorial ou de localização.

**SÉRIE GEOGRÁFICA:** É uma série estatística na qual o local ou fator geográfico é variável, mantendo-se fixos o tempo e o fenômeno observado.

**SÉRIE ORDENADA:** Conjunto de valores de uma variável que possui uma sequência natural no tempo ou no espaço. Em inglês Ordered series.

**SÉRIE TEMPORAL:** Identifica-se pelo caráter variável do fator cronológico. O local e a espécie ou fenômeno são elementos fixos. Esta série também é chamada de histórica ou evolutiva.

**SÉRIES HOMÓGRADAS:** São aquelas em que a variável descrita apresenta variação discreta ou descontínua. Podem ser do tipo temporal, geográfica ou específica.

**SÉRIE TEMPORAL:** Uma série temporal é uma série de observações ordenadas de uma característica quantitativa de um fenômeno individual ou coletivo considerado em diferentes pontos de tempo. Embora não seja essencial, é comum para esses pontos serem equidistantes no tempo. A qualidade essencial da série é a ordem de observações de acordo com a variável do tempo. Em inglês Time series.

**SÉRIES CONJUGADAS:** Também chamadas de tabelas de dupla entrada. São apropriadas à apresentação de duas ou mais séries de maneira conjugada, havendo duas ordens de classificação: uma horizontal e outra vertical.

**SQPC:** Ver soma dos quadrados e dos produtos cruzados.

**SUBCONJUNTO (SUBSET):** Uma coleção de objetos que também é parte de um grupo, ou conjunto, maior de objetos. Ver também, conjunto.

**SERENDIPIDADE:** Conceito utilizado para indicar a obtenção de resultados em pesquisas científicas de forma acidental. Por exemplo, a descoberta do raio-X em 1895 aconteceu quando o médico alemão Wilhelm Conrad Roentgen estava pesquisando o comportamento dos raios catódicos. Ao cobrir o tubo com papelão preto, ele descobriu que alguns raios invisíveis iluminavam uma tela especial que estava há três metros de distância. Logo depois ele percebeu que os raios-X como os nomeou, podiam ser capturados por placas fotográficas e que eles passavam facilmente pela pele e músculos humanos, mas eram bloqueados por metais e ossos.

**SHAPIRO-WILK:** Teste empregado para verificar a normalidade dos dados amostrais.

**SI:** Sistema Internacional de Unidades.

**SIC:** Significa assim. Utiliza-se da mesma forma que *ipsis litteris* ou *ipsis verbis*.

**SIGILO:** Descreve um survey no qual pesquisadores conhecem ou podem conhecer as respostas de respondentes específicos do survey, mas garantem que elas permanecerão em segredo.

**SIGMA:** É a palavra que expressa a letra do alfabeto grego que corresponde ao nosso s, que lembra a soma. A letra minúscula simboliza o desviopadrão e a letra maiúscula significa somatório. Em inglês sigma.

**SIGNIFICÂNCIA CLÍNICA:** Um resultado é significativo ou importante do ponto de vista clínico se o intervalo de confiança para o parâmetro de interesse difere do valor especificado na hipótese nula  $H_0$  por uma quantidade considerada clinicamente importante ou relevante do ponto de vista prático.

**SIGNIFICÂNCIA CLÍNICO-EPIDEMIOLÓGICA:** Refere-se ao impacto que os resultados de uma pesquisa podem produzir quando adicionada ao corpo de conhecimentos que embasam uma conduta médica. Quanto maior o impacto, maior a significância clínica.

**SIGNIFICÂNCIA DE UM SISTEMA DE VIGILÂNCIA EPIDEMIOLÓGICA:** Qualidade que deve ser usada como um princípio orientador dos sistemas de vigilância. Os sistemas simples são fáceis de entender e implementar, custam menos e facilitam a flexibilidade. A simplicidade é determinada pela própria estrutura do sistema de vigilância epidemiológica, levando-se em conta a quantidade e o tipo de informação necessária para estabelecer o diagnóstico, o número e os tipos de fontes de notificação, quantidade de boletins utilizados, os métodos de distribuição dos informes, o tempo gasto em análise de dados dentre outros.

**SIGNIFICÂNCIA ESTATÍSTICA:** Um resultado é estatisticamente significativo sempre que um teste de significância produz um valor de p menor do que o valor estabelecido para alfa, que convencionalmente é igual a 0,05. A implicação de uma significância estatística com um alfa de 0,05 é a de que o acaso poderia produzir a diferença observada na comparação entre dois grupos não com a seqüência maior do que 5 vezes em 100. Isso significa aceitar que o acaso não é o responsável pelo efeito observado.

**SIC:** Nomenclatura que significa segundo informações coletadas, usada para descrever de maneira fiel e original uma informação.

**SIGNIFICÂNCIA ESTATÍSTICA:** A probabilidade de que as diferenças, encontradas em estudos comparativos, sejam triviais e devidas ao acaso; a probabilidade de rejeitar a hipótese nula. Ver P, valor P ou p-value.

**SIGNIFICÂNCIA ESTATÍSTICA DE UM RESULTADO:** A significância estatística de um resultado é uma medida estimada do grau em que este resultado é verdadeiro no sentido de que seja realmente o que ocorre na população, ou seja, no sentido de representatividade da população. Mais tecnicamente, o valor do nível-p que é a probabilidade que mede a significância e significa o risco de se rejeitar a hipótese estatística quando ela é verdadeira, representa um índice decrescente da confiabilidade de um resultado. Quanto mais alto o nível-p, menos se pode acreditar que a relação observada entre as variáveis na amostra é um indicador confiável da relação entre as respectivas variáveis na população. Especificamente, o nível-p representa a probabilidade de erro envolvida em aceitar o resultado observado como válido, isto é, como representativo da população. Por exemplo, um nível-p de 0,05  $\left(\frac{1}{20}\right)$  indica que há 5 % de probabilidade

de que a relação entre as variáveis, encontrada na amostra, seja um acaso feliz. Em outras palavras, assumindo que não haja relação entre aquelas variáveis na população, e o experimento de interesse seja repetido várias vezes, poder-se-ia esperar que em aproximadamente 20 realizações do experimento haveria apenas uma em que a relação entre as variáveis em questão seria igual ou mais forte do que a que foi observada naquela amostra anterior. Em muitas áreas de pesquisa, o nível-p de 0,05 é costumeiramente tratado como um limite aceitável de erro.

**SIGNIFICÂNCIA PRÁTICA:** Método de avaliar resultados da análise multivariada baseado em suas descobertas substanciais, em vez de sua significância estatística. Enquanto a significância estatística determina o resultado pode ser atribuído ao acaso, a significância prática avalia se o resultado é útil, isto é, substancial o bastante para garantir ação.



**SIGNIFICÂNCIA:** Grau de importância ou relevância das conclusões obtidas em uma pesquisa científica.

**SILOGISMO:** É a forma do raciocínio dedutivo. Pode ser visto também como o principal conteúdo da lógica. Consiste de duas premissas e uma conclusão: fenômenos elétricos são acompanhados por distúrbios magnéticos; o relâmpago é um fenômeno elétrico; portanto, o relâmpago é acompanhado por distúrbios magnéticos.

**SILOGISMO:** Modelo, às vezes usado para raciocínio na lógica clássica, em que um conjunto de afirmações leva logicamente a uma conclusão. Exemplo: i) Manuais contêm palavras. ii) Este livro é um manual. iii) Portanto, este livro deve conter palavras.

**SIMETRIA:** Um termo antigo para simetria em uma distribuição de frequência. É a medida daquela simetria. Em inglês Skewness.

**SIMÉTRICO:** Propriedade de dados cuja distribuição pode ser dividida em duas metades que são aproximadamente imagens uma da outra, obtidas traçando-se uma reta vertical pelo meio.

**SIMÉTRICA:** Uma distribuição é simétrica e relação a uma reta se as metades da distribuição em cada lado da reta são imagens-espelho uma da outra; por exemplo, uma distribuição normal é simétrica em relação à sua média.

**SIMILARIDADE:** Ver similaridade entre objetos.

**SIMILARIDADE:** Ver dados de similaridades.

**SIMILARIDADE:** Medida do grau de semelhança entre dois objetos. O inverso é a dissimilaridade, que se relaciona ao conceito de distância.

**SIMILARIDADE ENTRE OBJETOS:** A correspondência ou associação de dois objetos baseada nas variáveis da variável estatística de agrupamento. A similaridade pode ser medida de duas formas. Primeiro, é uma medida de associação, com coeficientes de correlação positivos maiores representando maior similaridade. Segundo, a proximidade entre cada par de objetos pode avaliar a similaridade, onde medidas de distância ou de diferença são empregadas e as menores distâncias ou diferenças representam maior similaridade.

**SIMULAÇÃO:** Processo que se comporta de forma similar a algum experimento, de modo a produzir resultados similares.

**SIMULAÇÃO:** É o exercício de uma imitação flexível de processos e resultados com a finalidade de esclarecer ou explicar os mecanismos subjacentes envolvidos. A pesquisa simulada tem grande utilidade nos primeiros estágios da geração de uma teoria.

**SIMULAÇÃO CONDICIONAL:** A simulação do valor  $Z_{sc}\{x\}$  que honra os valores amostrados nos pontos  $x_s$ , dentre uma infinidade de realizações possíveis  $\{Z_s(x), s=1,\dots\}$  da função aleatória  $Z(x)$ :  $Z_{sc}(X) = Z_0(X_\alpha), \forall X$ .

**SIMULAÇÃO MONTE CARLO:** O método de Monte Carlo (MMC) é um método estatístico utilizado em simulações estocásticas com diversas aplicações em áreas como a física, matemática e biologia. O método de Monte Carlo tem sido utilizado há bastante tempo como forma de obter aproximações numéricas de funções complexas. Este método tipicamente envolve a geração de observações de alguma distribuição de probabilidades e o uso da amostra obtida para aproximar a função de interesse. As aplicações mais

comuns são em computação numérica para avaliar integrais. A idéia do método é escrever a integral que se deseja calcular como um valor esperado. O nome Monte Carlo surgiu durante o projeto Manhattan na Segunda Guerra Mundial. No projeto e de construção da bomba atômica, Ulam, von Neumann e Fermi consideraram a possibilidade de utilizar o método, que envolvia a simulação direta de problemas de natureza probabilística relacionados com o coeficiente de difusão do neutron em certos materiais. Apesar de ter despertado a atenção desses cientistas em 1948, a lógica do método já era conhecida há bastante tempo. Por exemplo, existe um registro de um artigo escrito por Lord Kelvin dezenas de anos antes que já utilizava técnicas de Monte Carlo em uma discussão das equações de Boltzmann. Fonte: Revista Mundo Project Management (Mundo PM). Em inglês monte carlo simulation.

**SIMULAÇÃO:** Criação de matrizes de entrada de múltiplos dados com base em parâmetros especificados que refletem variação na distribuição dos dados de entrada. Em inglês Simulation.

**SIMULADOR DE ESCOLHA:** Procedimento que permite ao pesquisador avaliar muitos cenários do tipo o que aconteceria se, incluindo a preferência por possíveis configurações de produto ou serviço ou as interações competitivas entre estímulos considerados como componentes de um mercado. Logo que as utilidades parciais conjuntas tenham sido estimadas para cada respondente, o simulador de escolha analisa um conjunto de estímulos de perfil completo e prevê escolhas individuais e agregadas para cada estímulo no conjunto.

**SINERGIA:** Está presente quando o impacto que dois ou mais fatores têm sobre um indivíduo ou sobre uma população é maior do que a soma dos efeitos de cada fator em separação.

**SINGULARIDADE:** O caso extremo de colinearidade ou multicolinearidade no qual uma variável independente é perfeitamente prevista ou seja, uma correlação de  $\pm 1,0$  por uma ou mais variáveis independentes. Modelos de regressão não podem ser estimados quando existe uma singularidade. O pesquisador deve omitir uma ou mais as variáveis independentes envolvidas para remover a singularidade.

**SISTEMA:** Conjunto de dados ordenados, interdependentes e relacionadas entre si. Também chamado de rol de dados numéricos.

**SISTEMA:** Um sistema pressupõe interdependência, relacionamentos, e uma finalidade. Daí ser um sistema, um conjunto de partes interrelacionadas e interdependentes e organizadas para atingir um conjunto de metas e objetivos.

**SISTEMA OU MODELO DE EQUAÇÕES SIMULTÂNEAS:** Um sistema ou modelo de equações simultâneas refere-se ao caso em que a variável dependente, em uma ou mais equações, é também uma variável explicativa, em alguma outra equação do sistema. Isto é, não são somente os Ys determinados pelos Xs, mas alguns dos Xs são por sua vez determinados pelo Ys, assim que os Ys e os Xs são determinados conjunta ou simultaneamente.

**SISTEMA DE AMOSTRAGEM:** Refere-se a como as unidades amostrais são selecionadas para um levantamento ou seja, ao delineamento amostral do levantamento. Exemplo de sistemas de amostragem: amostragem aleatória simples, amostragem estratificada, amostragem sistemática, dentre outros.

**SISTEMA DE EQUAÇÕES DE CO-KRIGAGEM:** O estimador da co-krigagem utilizado na geoestatística é obtido como uma combinação linear de duas variáveis correlogramadas como:  $Z_{CK}^* = \sum_{\alpha=1}^n \eta_{\alpha} Z_{a_{\alpha}} + \sum_{\beta=1}^m \lambda_{\beta} Z_{b_{\beta}}$



Os ponderadores  $\{h_a, a=1, n\}$  e  $\{l_b, b=1, m\}$  são obtidos da resolução de um sistema de  $(n+m+2)$  equações:

$$\sum_{\alpha=1}^n \eta_{\alpha} \cdot \bar{C}(va_{\alpha}, va_{\beta}) + \sum_{\alpha=1}^n \eta_{\alpha} \cdot \bar{C}(va_{\alpha}, vb_{\beta}) + \mu_1 = \bar{C}(Va_{\beta}, v) \quad \text{para } \beta = 1, n$$

$$\sum_{\alpha=1}^n \eta_{\alpha} \cdot \bar{C}(va_{\alpha}, va_{\beta}) + \sum_{\alpha=1}^m \lambda_{\alpha} \cdot \bar{C}(vb_{\alpha}, vb_{\beta}) + \mu_2 = \bar{C}(V, vb_{\beta}) \quad \text{para } \beta = 1, m$$

$$\sum_{\alpha=1}^n \eta_{\alpha} = 1 \quad \text{e} \quad \sum_{\alpha=1}^m \lambda_{\alpha} = 1.$$

**SISTEMA DE EQUAÇÕES DE KRIGAGEM DISJUNTIVA:** O estimador de krigagem disjuntiva é calculado como:  $Z_{DK}^* = \sum_{k=0}^k C_k H_{kv}$ , onde:  $C_k$  são os coeficientes calculados usando as informações da distribuição

dos dados, ou seja:

$$C_k = \frac{1}{K!(2\pi)^{\frac{1}{2}} \sum_{j=1}^j W_j f(Y_j) H_k(Y_j) \cdot \exp\left[-\frac{Y_j^2}{2}\right]}$$

Onde,  $y_j$  são as abscissas da integração de Hermite;  $w_j$  são os fatores de ponderação correspondentes;  $f(y_j)$  são os valores correspondentes de  $y$  na distribuição original  $Z$ .  $H_{kv}$  são as coordenadas do estimador do bloco, calculadas pelos  $n$  pontos de dados vizinhos normalizados  $\{y_a, a=1, n\}$ . As coordenadas  $H_{kv}$  são calculadas pela seguinte equação:  $H_{kv} = \sum_{\alpha=1}^n b_{\alpha} H_k(Y_{\alpha})$ . Os coeficientes  $b_{\alpha}$  são determinados pela resolução de um sistema de  $n$  equações lineares:

$$\sum_{\alpha=1}^n b_{\alpha} (\rho_{\alpha\beta})^k = \sum_{\alpha=1}^n (\rho_{\alpha v})^k, \quad \text{onde: } (r_{ab})^k \text{ é o coeficiente de correlação entre } y_a \text{ e } y_b \text{ elevado a potência interna}$$

$k$ ;  $(r_{av})^k$  é o valor médio dos coeficientes de correlação entre uma amostra  $y_a$  e o domínio do bloco  $V$ .

**SISTEMA DE EQUAÇÕES DE KRIGAGEM DA MÉDIA:** Resulta da minimização da variância entre as médias  $\text{Var}[m-m^*]$ , ou seja,  $\sum_{\beta=1}^n \lambda_{\beta} C(V_{\alpha}, V_{\beta}) - \mu = 0$ , para  $\alpha = 1, \dots, n$ .

**SISTEMA DE EQUAÇÕES DE KRIGAGEM ORDINÁRIA:** Resulta da minimização da função variância, restrito à condição de não enviesamento, por meio da técnica dos multiplicadores de Lagrange:  $\sum_{\beta=1}^n \lambda_{\beta} \bar{C}(V_{\alpha}, V) - \mu = \bar{C}(V_{\alpha}, V_{\beta})$ , para  $\alpha = 1, \dots, n$  e  $\sum_{\beta=1}^n \lambda_{\beta} = 1$ .

**SISTEMA DE EQUAÇÕES DE KRIGAGEM TRANSITIVA:** Resulta do efeito da minimização da função variância da krigagem transitiva:  $\sum_{\beta=1}^n \lambda_{\beta} \rho(X_{\alpha} - X_{\beta}) = \rho(X_{\alpha})$  para  $\alpha = 1, \dots, n$ , onde:  $r(x_a - x_b)$  é a função de

autocorrelação medida para a distância  $(x_a - x_b)$ ;  $r(x_a)$  é a função de autocorrelação da amostra  $a$  para o ponto  $a$  ser estimado  $x_0$ . Isto só é viável após a translação do sistema de coordenadas para  $x_0$ .

**SISTEMA DE EQUAÇÕES DE KRIGAGEM UNIVERSAL:** Resulta da minimização da função variância de krigagem universal, agora sujeito a  $k$  condições de restrição, por meio da técnica dos multiplicadores de Lagrange:

$$\sum_{\beta=1}^n \lambda_{\beta} \bar{C}(V_{\alpha}, V_{\beta}) - \sum_{i=1}^k \mu_i b_{V_{\alpha}} = \bar{C}(V_{\alpha}, V), \quad \text{para } \alpha = 1, \dots, n \quad \text{e} \quad \sum_{\beta=1}^n \lambda_{\beta} b_{V_{\alpha}} = b_V.$$

**SISTEMA DE INFORMAÇÃO GEOGRÁFICA (GIS):** Sistema, baseado em computador, que coleta dados secundários ou primários para gerar mapas que exibem visualmente as respostas às perguntas da pesquisa.

**SISTEMA DE OBSERVAÇÃO:** Usado para registrar comportamentos pré-selecionados numa tentativa de quantificar comportamentos na situação sendo observada.

**SISTEMA DE SUPORTE DE DECISÕES (DSS):** Um sistema interativo e personalizado de mapeamento de informações destinado a ser iniciado e controlado por tomadores de decisão.

**SISTEMA DE VIZINHANÇA:** Considerar um grafo conexo  $G$  com vértices indexados pelo conjunto  $I = \{1, \dots, m\}$ ,  $m \leq \infty$ . Um sistema  $\Delta = \{\delta_i \in I\}$  de subconjuntos de  $G$  é um sistema de vizinhança se para todo  $i \in I$ : (a)  $i \notin \delta_i$ , (b)  $j \in \delta_i$  implica que  $i \in \delta_j$ . Em inglês Neighborhood system.

**SISTEMAS DE SUPORTE DE DECISÃO (DSS):** Sistemas interativos desenvolvidos para dar aos usuários acesso a dados para questões ad hoc ou mal definidas, para relatórios pré-especificados, por exemplo, relatos de exceção e para análises estruturadas procedimentos baseados em regras, estatísticas multivariadas, redes neurais ou outros modelos.

**SISTEMA DINÂMICO:** Um sistema dinâmico fornece uma forma de descrever como o estado de um sistema evolui para outro. O estudo remonta ao trabalho de Poincaré sobre o problema dos três corpos ou mesmo antes com Huygens sobre o pêndulo composto. Em inglês dynamical system.

**SISTEMA INTERNACIONAL DE UNIDADES:** Convenção internacional, cuja abreviação internacional é SI, que define as unidade de medidas, bem como seus símbolos e abreviações, para as diferentes grandezas físicas. O SI foi estabelecido durante uma série de reuniões da Conferência Geral de Pesos e Medidas que se iniciaram no final do século XIX. O SI reconhece duas classes de unidades: unidades básicas e unidades derivadas. As unidades básicas se referem a sete grandezas: comprimento em metro, massa em quilograma, tempo em segundo, intensidade de corrente elétrica em ampère, temperatura termodinâmica em kelvin, quantidade de matéria em mol e intensidade luminosa em candela. As unidades secundárias são formadas pela combinação das unidades básicas. Cada grandeza física é designada por uma única unidade do SI, mesmo que essa unidade possa ser expressa sob diferentes formas.

**SISTEMAS OPERACIONAIS:** Sistemas de informação que mantêm as funções básicas da organização como contabilidade, balanço, processamento de encomendas, para suas operações diárias.

**SÍTIO:** Conceito que engloba todos os fatores abióticos do ambiente que determinam a qualidade de um dado local para sustentar o crescimento de árvores e, portanto, que condicionam a produção florestal de um dado povoamento. É geralmente descrito quantitativamente pelo **índice de sítio**.

**SKINNERIANISMO:** A ciência squinneriana é boa principalmente por que iniciou o estabelecimento de uma tecnologia confiável e ruim no sentido de que afastou a atenção e esforço para desenvolvimento de tecnologias alternativas que derivam de outras perspectivas teóricas.

**SMALIAN, FÓRMULA DE:** Fórmula de cubagem cuja aproximação do volume de uma tora resulta do produto do comprimento da tora pela média das áreas transversais nas extremidades da tora. A fórmula produz resultados exatos somente para o cilindro e o parabolóide ordinário que é o parabolóide quadrático.

**SOCIOGRAMA:** O sociograma é uma técnica sociométrica que apresenta sob a forma de um gráfico as várias relações entre os sujeitos que formam um grupo, ou seja, uma ferramenta que permite medir ou

avaliar as relações sociais entre os integrantes de um grupo humano, onde os respectivos elementos se conhecem, têm objetivos em comum e se influenciam mutuamente.

**SOBREMPELHAMENTO (OVERMATCHING):** Situação prejudicial que resulta de procedimentos inadequados no pareamento: uma ou mais variáveis não deveria(m) ou não precisaria(m) ser pareadas(s).

**SOBREVIDA, ANÁLISE DE:** Conjunto de técnicas estatísticas para estudar a evolução de um evento, no tempo, e fazer inferências sobre os efeitos do tratamentos e sobre outras influências.

**SOBREVIVÊNCIA:** Refere-se às árvores que sobrevivem na amostra durante o período entre duas medições de uma parcela permanente, isto é, às árvores que são medidas em duas ocasiões sucessivas. A sobrevivência pode ser definida em termos absolutos com base no número de árvores por unidade de área, ou ainda com base na área basal, volume, biomassa, dentre outros. Quando se estuda o crescimento de uma floresta, o crescimento das árvores sobreviventes é geralmente identificado como crescimento bruto.

**SOFTWARE:** Programas de computação, manuais relacionados e demais documentos necessários para usar tais programas.

**SOFTWARE DE APRESENTAÇÃO:** Software para computadores pessoais que fornece plataformas fáceis de usar para a criação de relatórios e apresentações eficientes.

**SÓLIDOS DE REVOLUÇÃO:** São figuras geométricas de 3 dimensões que são geradas pela revolução de uma função matemática ao redor de um dos eixos no plano cartesiano. Como são gerados por revolução ao redor de um eixo, os sólidos de revolução têm sempre secção transversal circular. Assumindo que a revolução será realizada ao redor do eixo das ordenadas ou eixo-y, certas funções geram sólidos geométricos conhecidos. Uma reta paralela ao eixo das ordenadas gera um cilindro, enquanto que uma reta que intercepte o eixo das ordenadas gera um cone. Parábolas côncavas com vértice posicionado sobre o eixo das ordenadas gera a família dos parabolóides. Já as semi-parábolas convexas com vértice posicionado sobre o eixo das abscissas ou eixo-x geram a família dos nelóides. Quando as funções matemáticas revolucionadas interceptam os eixos do planos cartesiano gera-se o sólido geométrico perfeito. Quando o eixo das ordenadas não é interceptado, gera-se um sólido geométrico com a ponta truncada, que é designado como tronco: tronco de cone, tronco de parabolóide e tronco de nelóide.

**SOLUÇÃO DE BAYES:** Na terminologia das funções de decisão estatísticas, a solução de Bayes é a função de decisão que minimiza o risco médio relativo a alguma distribuição de probabilidade. Em inglês Bayes solution.

**SOLUÇÃO DEGENERADA:** Solução MDS (escalonamento multidimensional) que é inválida por causa de inconsistências nos dados ou muito poucos objetos comparados com a dimensionalidade da solução. Mesmo que o programa de computador possa indicar uma solução válida, o pesquisador deve desconsiderar a solução degenerada e examinar os dados em busca da causa. Esse tipo de solução normalmente é representado como um padrão circular de resultados ilógicos.

**SOMATÓRIO:** Representa, de forma abreviada, uma soma.

**SOFISMA:** Argumento ou raciocínio concebido com o objetivo de produzir a ilusão da verdade, que, embora simule um acordo com as regras da lógica, apresenta, na realidade, uma estrutura interna inconsistente, incorreta e deliberadamente enganosa. argumentação que aparenta verossimilhança ou veracidade,

mas que comete involuntariamente incorreções lógicas; paralogismo. Argumento aparentemente válido, mas que, na realidade, é ilógico, inválido; argumento enganoso, ardiloso. Ocasionalmente os veículos de comunicação têm sido relacionados ao uso do sofisma, exemplos incluem a falta de embasamento veiculado no telejornalismo e programas de entretenimento que usam um discurso agradável para forçar a audiência a fazer o que eles querem.

**SELEÇÃO DE VARIÁVEIS:** Problema de selecionar um subconjunto de variáveis para um modelo a partir de uma lista de candidatos que contém toda ou quase toda informação útil a respeito da resposta nos dados.

**SELEÇÃO PROGRESSIVA:** Um método de selecionar variáveis na regressão, em que variáveis são inseridas, uma de cada vez, no modelo até que nenhuma outra variável que contribua significativamente para o modelo possa ser encontrada.

**SEM REPOSIÇÃO:** Um método de selecionar amostras, em que itens não são repostos entre sucessivas seleções.

**SÉRIES TEMPORAIS:** Um conjunto de observações ordenadas tomadas em pontos no tempo.

**SIGNIFICÂNCIA:** No teste de hipóteses, um efeito é dito significativo se o valor da estatística de teste estiver na região crítica.

**SOMA DOS QUADRADOS DA REGRESSÃO:** A porção da soma dos quadrados total atribuída ao modelo que foi ajustado aos dados.

**SOMA DOS QUADRADOS DO ERRO:** Em análise de variância, essa é a porção da variabilidade total responsável pelo componente aleatório nos dados. Geralmente, é baseada na replicação de observações em certas combinações de tratamentos no experimento. É algumas vezes chamada de soma residual dos quadrados, embora este seja realmente um termo melhor para usar somente quando a soma dos quadrados é baseada no que sobra de um processo de ajuste de modelo e não na replicação.

**SOMA DOS QUADRADOS DO TRATAMENTO:** Em análise de variância, essa é a soma dos quadrados que considera a variabilidade na variável de resposta devido aos tratamentos diferentes que tenham sido aplicados.

**SOMA DOS QUADRADOS DO RESÍDUO:** Veja soma dos quadrados do erro.

**SUBGRUPO RACIONAL:** Uma amostra de dados selecionados de modo a, na medida do possível, incluir fontes casuais de variação e excluir fontes especiais de variação.

**SUPERFÍCIE DE RESPOSTA:** Quando uma resposta  $y$  depende de uma função de  $k$  variáveis quantitativas  $x_1, x_2, \dots, x_n$ , os valores da resposta podem ser vistos como uma superfície em  $k+1$  dimensões. Essa superfície é chamada de superfície de resposta. A metodologia de superfície de resposta é um subconjunto de planejamento de experimentos voltados para aproximar essa superfície com um modelo e para usar o modelo resultante para otimizar o sistema ou processo.

**SUPERPOSIÇÃO:** Quando um experimento fatorial é corrido em blocos e os blocos são muito pequenos para conter uma replicata completa do experimento, pode-se correr uma fração da replicata em cada bloco; porém, isso resulta em perda de informação em alguns efeitos. Esses efeitos são ligados ou superpostos

com os blocos. Em geral, quando dois fatores são variados de modo que seus efeitos não podem ser determinados separadamente, seus efeitos são ditos superpostos.

**SOMATÓRIO:** Símbolo matemático dado por  $\sum$  que é a letra do alfabeto grego sigma maiúscula, que corresponde a letra S do alfabeto latino e lembra soma.

**SOMA DIRETA DE MATRIZES:** Símbolo matemático dado por  $\oplus$ .

**SOMA DE QUADRADOS DA REGRESSÃO (SSR):** Soma das diferenças quadradas entre a média e valores previstos da variável dependente para todas as observações. Representa a quantia de melhoramento na explicação da variável dependente atribuível à(s) variável(is) independente(s). SSR é abreviação para sum of squares regression.

**SOMA DE QUADRADO DE X (SQX):** Termo usado em estatística como soma de quadrado de x, não confundir com o quadrado da soma de

$x \left[ \sum_{i=1}^n X_i \right]^2$ . Essa quantidade é determinada conforme a seguinte expressão:

$$SQX = \sum_{i=1}^n X_i^2 - \frac{\left( \sum_{i=1}^n X_i \right)^2}{n} \text{ ou ainda } SQX = \sum_{i=1}^n X_i^2 - n[\bar{X}]^2.$$

**SOMA DE QUADRADO DE Y:** Termo usado em estatística como soma de quadrado de y, não confundir com o quadrado da soma de y  $\left[ \sum_{i=1}^n Y_i \right]^2$ . Essa quantidade é determinada conforme a seguinte expressão:

$$SQY = \sum_{i=1}^n Y_i^2 - \frac{\left( \sum_{i=1}^n Y_i \right)^2}{n}, \text{ ou então, } SQY = \sum_{i=1}^n Y_i^2 - n[\bar{Y}]^2.$$

**SOMA DE QUADRADOS DE ERROS:** Grandeza que indica o grau do erro aleatório na análise de variância e da regressão.

$$SQY = \sum_{i=1}^n Y_i^2 - \frac{\left( \sum_{i=1}^n Y_i \right)^2}{n}$$

**SOMA DE QUADRADOS DE REGRESSÃO:** Grandeza que indica até que ponto uma equação de regressão pode explicar as variação na variável dependente.

**SOMA DE QUADRADOS DOS ERROS (SSE):** Soma dos erros de previsão ou resíduos ao quadrado em todas as observações. É usada para denotar a variância nas variáveis dependentes ainda não explicadas pelo modelo de regressão. A frase correta seria: é usada para denotar a variância na variável dependente ainda não explicada. Se nenhuma variável independente é empregada para previsão, ela se transforma nos quadrados dos erros, usando a média como o valor previsto, e assim se iguala à soma total dos quadrados. SSE é a abreviação para sum of squared errors.

**SOMA DE PRODUTO XY (SPXY):** Termo usado em estatística como soma de produto X Y, não confundir com o produto das somas de

$X \text{ e } Y \left[ \sum_{i=1}^n X_i \sum_{i=1}^n Y_j \right]$ . Essa quantidade é calculada conforme a equação abaixo:

$$SPXY = \sum_{i=1}^n \sum_{j=1}^n X_i Y_j - \frac{\sum_{i=1}^n X_i \sum_{i=1}^n Y_j}{n}, \text{ ou então,}$$

$$SPXY = S_{XY} = \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) = \sum_{i=1}^n X_i Y_i - n\bar{X}\bar{Y} = \sum_{i=1}^n \sum_{j=1}^n X_i Y_j - \frac{\left( \sum_{i=1}^n X_i \sum_{j=1}^n Y_j \right)}{n}$$

**SOMA DOS QUADRADOS:** Soma dos quadrados dos desvios dos elementos de uma lista, em relação a uma grandeza especificada. Na análise de variância ou na regressão, utilizam-se várias somas de quadrados.

**SOMA DOS QUADRADOS DEVIDO À REGRESSÃO (SSR):** A variação explicada pela regressão.

**SOMA DOS QUADRADOS DOS TRATAMENTOS:** Na análise da variância, grandeza que indica até que ponto as diferenças nas observações podem ser explicadas pelo fato de as observações se originarem de diferentes grupos.

**SOMA TOTAL DE QUADRADOS:** Soma dos quadrados dos desvios de todos os números de uma lista, contados a partir da média, usada na análise de variância e na regressão.

**SOMA TOTAL DE QUADRADOS (TSS):** Quantia total de variação existente a ser explicada pelas variáveis independentes. Esse ponto de referência é calculado somando-se as diferenças quadradas entre a média e valores reais para a variável dependente em todas as observações. TSS é a abreviação para total sum of squares. Conhecida também como soma de quadrado total (SQT).

**SOMATÓRIO, NOTAÇÃO DE:** Notação que utiliza a letra maiúscula grega  $\Sigma$  (sigma) para indicar soma de valores de uma série; por exemplo,  $\Sigma X$  significa somar todos os valores de X.

**SONDAGEM:** Procedimento de estudo de uma parte da população voltado, sobretudo, para pesquisas de opinião e de mercado. Sinônimo: Amostragem.

**SONDAR:** Técnica usada em entrevista para conseguir resposta mais completa a uma questão. É uma frase ou pergunta não-diretiva, usada para encorajar o respondente a aprofundar uma resposta. Exemplos: mais alguma coisa? e como assim?.

**SONDAGEM:** É o estudo estatístico que se baseia numa parte da população, isto é, numa amostra que deve ser representativa dessa população.

**SOROEPIDEMIOLOGIA:** Ver inquérito sorológico em epidemiologia.

**SREENING:** O mesmo que rastreamento.

**STANDARD:** Parâmetro em que nos baseamos para fazer comparações, se puder ser produzido e usado por outros, de forma que permite a diferentes pesquisadores conseguirem uma base comum para conferência ou verificação. O uso de Standards em experimentos facilita a conexão da pesquisa com o resto da ciência.

**SQ(ERRO):** Soma de quadrados representando a variabilidade que se supõe comum a todas as populações consideradas; usado na análise de variância.



**SQ(TOTAL):** Medida da variação total em torno de em todos os dados amostrais combinados, usado em análise de variância.

**SQ(TRATAMENTO):** Medida da variação entre as médias amostrais; usado na análise da variância.

**SQE:** Soma de quadrados dos resíduos que são os erros de ajuste.

**SQE:** Soma de quadrados dos desvios explicados pela regressão.

**SQT:** Soma de quadrados dos desvios totais.

**SUAVIZAÇÃO EXPONENCIAL:** Um método usado em séries de tempo para suavizar ou prever uma série. Em inglês Exponential smoothing.

**SUB-AMOSTRA:** Amostra de amostra.

**SUB-AMOSTRAGEM:** Processo de formação de uma sub-amostra.

**SUBCONJUNTO:** O conjunto A é subconjunto do conjunto B se todos os elementos dos de A são também elementos de B.

**SUFICIÊNCIA:** É uma propriedade de um estimador definida pelo matemático Inglês Ronald Aylmer Fisher (FISHER, 1922). Um estimador  $t$  é considerado suficiente para um parâmetro  $\theta$  se a distribuição de uma amostra  $x_1, x_2, \dots, x_n$  dado  $t$ , não depende de  $\theta$ . A distribuição de  $t$  contém, então, toda a informação da amostra relevante para a estimação de  $\theta$  e o conhecimento de  $t$  e sua distribuição de amostragem é suficiente para dar tal informação. Em inglês Sufficiency.

**SUFICIÊNCIA CONJUNTA:** Estimadores  $t_1, \dots, t_k$  são conjuntamente suficientes para os parâmetros  $\theta_1, \dots, \theta_k$  se a função de verossimilhança pode ser expressa como  $L_1(t_1, \dots, t_k; \theta_1, \dots, \theta_k)L_2(x_1, \dots, x_n)$  onde  $L_2$  não depende de  $\theta_1, \dots, \theta_k$ , apesar de depender de outros parâmetros. Em ingles Joint sufficiency.

**SUJEITO DE PESQUISA:** É o elemento que faz parte de uma pesquisa e nele se obtém as variáveis respostas, podendo ser representado por pessoas, animais, instrumentos, objetos, cidades, bairros, plantas, sementes, estados, países, etc.

**SUJEITO DE PESQUISA:** É o indivíduo que participa de uma pesquisa. Pode ser voluntário, involuntário ou avoluntário. O participante voluntário sabe que está sendo investigado, mas tomou uma decisão consciente de que os ganhos envolvidos em sua participação compensam possíveis perdas medidas em termos de tempo gasto, invasão de privacidade, dentre outros. O participante involuntário é aquele que se sente coagido a dispensar seu tempo numa investigação experimental e considera esse dispêndio injustificável. Frequentemente explicitam má vontade e tentam arruinar o estudo. O participante avoluntário é o indivíduo que, sem saber, entra numa situação experimental, é exposto a vários tratamentos experimentais, e depois testado. Somente no fim da pesquisa, ele fica sabendo que foi participante.

**SUJEITO DO ESTUDO:** Termo genérico que designa um indivíduo que participa de um estudo. A vantagem do termo, em relação ao termo paciente, é o fato de evitar a conotação de doença útil nos casos em que são estudadas pessoas sadias. Veja participantes.

**SUMÁRIO EXECUTIVO:** A parte do relatório de pesquisa que explica por que a pesquisa foi realizada, o que foi encontrado e o que essas descobertas significam e quais providências, se for o caso, a administração deve tomar.

**SUBJETIVISMO:** Conceito ou enfoque na qual o indivíduo atribui notas, argumentos, ou escores em geral baseado na opinião pessoal em estatística experimental não paramétrica aplicada nas ciências sociais e naturais, em análise sensorial de alimentos, em avaliação fitopatológica, em concursos, etc., ou quando se obtém valores de probabilidade de ocorrência de eventos aleatórios em situações do cotidiano onde se aplique o conceito subjetivo de probabilidade.

**SUPER DISPERSÃO:** O termo super dispersão significa que a variância da variável observada excede a variância nominal para a distribuição usada. Não prática é comum encontrar situações em que existe super dispersão. A incidência e o grau de super dispersão encontrada depende do campo de aplicação. Em inglês Over-dispersion.

**SUPERFÍCIE DE FREQUÊNCIA:** Termo análogo a bivariado da curva de frequência. Em inglês Frequency surface.

**SUPERFÍCIE DE RESPOSTA:** É a representação geométrica de uma equação de regressão com duas ou mais variáveis independentes.

**SUPERVISIONADO:** Processo de aprendizado que emprega uma amostra de treinamento e fornece feedback para a rede neural, no que tange a erros nas interconexões de saída.

**SUPORTE:** Percentual da amostra total para o qual uma regra de associação é válida. Suporte indica a substancialidade do grupo.

**SUPORTE:** É a unidade básica de amostragem sobre a qual a variável é medida. Esta unidade básica pode ser definida sobre um ponto, área ou volume centrado em  $X$ .

**SUPOSIÇÃO:** Sinônimo de conjectura ou hipótese. Refere-se a uma sentença estabelecida a priori pelo pesquisador relativa a população e aceita como verdadeira até que as evidências experimentais apontem em sentido contrário.

**SUPRA:** Significa acima, referindo-se a nota imediatamente anterior.

**SURVEY LONGITUDINAL:** Desenho de estudo compreendendo coleta de dados em diferentes ocasiões no tempo, em contraste com estudos transversais.

**$s^2$  : Valor da variância amostral. Ou quadrado médio dos desvios ou dos resíduos, ou média de quadrados de desvios.**

**$\sigma_x$  : Desvio padrão da variável aleatória X. Tanto faz  $D. P. = \sigma_{(x)} = \sigma_x = \sigma$ .**

**$\sigma_x^2$  : Variância da variável aleatória X. Tanto faz  $Var(x) = V(x) = \sigma_{(x)}^2 = \sigma^2$ .**







**TABELA:** É um quadro que resume um conjunto de dados dispostos segundo linhas e colunas de maneira sistemática.

**TABELA CONTINGENCIAL:** Uma apresentação tabular da relação entre duas variáveis.

**TABELA DE ANOVA:** Tabela que resume os resultados de um cálculo de análise de variância.

**TABELA DE CONTINGÊNCIA:** Tabela de frequências observadas em que as linhas correspondem a uma variável de classificação e as colunas correspondem a outra variável de classificação; chamada também tabela de dois critérios.

**TABELA DE CONTINGÊNCIA:** Formato para mostrar relações entre variáveis sob a forma de distribuições percentuais. Contingência é sinônimo de dependência.

**TABELA DE CONTINGÊNCIA:** Usada para determinar se a distribuição de uma variável é condicionalmente dependente ou contingente da outra variável.

**TABELA DE CONTINGÊNCIA:** Tabulação cruzada de duas variáveis não métricas ou categóricas na qual as entradas são as frequências de respostas que caem em cada célula da matriz. Por exemplo, se três marcas foram avaliadas sobre quatro atributos, a tabela de contingência marca por atributo seria uma tabela com três linhas e quatro colunas. As entradas seriam o número de vezes que uma marca foi avaliada como tendo um atributo.

**TABELA DE DUPLA ENTRADA:** Veja tabela de contingência.

**TABELA DE FREQUÊNCIA:** Tabela que mostra quantas observações estão contidas em cada categoria.

**TABELA DE FREQUÊNCIA:** Tabela desenhada para mostrar a distribuição de frequências da ocorrência de uma característica dada de acordo com alguns intervalos de classe específicos. Em inglês Frequency table.

**TABELA DE FREQUÊNCIA:** Relação de categorias de valores juntamente com suas frequências correspondentes.

**TABELA DE FREQUÊNCIA:** Tabela que mostra o número de respostas a cada pergunta do questionário de pesquisa.

**TABELA DE FREQUÊNCIA:** A tabela de frequências é uma forma de representação da frequência de cada valor distinto da variável. Juntamente com as frequências, esta poderá incluir frequências relativas, frequências acumuladas e frequências relativas acumuladas. De uma forma geral, a tabela de frequências é usada para variáveis categóricas, uma vez que no caso de uma variável contínua a maior parte dos valores terá frequência 1, o que não resolve o problema inicial de se resumir a informação. Neste caso, pode-se agrupar os valores da variável contínua em intervalos, transformando-a numa variável categórica e assim ter mais sentido a tabela de frequências respectiva.

**TABELA DE FREQUÊNCIAS:** São tabelas onde se apresentam os dados por classes e as frequências respectivas.

**TABELA DE FREQUÊNCIAS ACUMULADAS:** Tabela de frequências em que cada classe e cada frequência representa os dados acumulados até aquela classe, inclusive.

**TABELA DE FREQUÊNCIAS RELATIVAS:** Variante da tabela básica de frequências, em que a frequência de cada classe é dividida pela frequência total.

**TABELAS DE FREQUÊNCIA:** São representações tabulares nas quais os escores se apresentam em correspondência com suas repetições, com frequências dispostas em valores absolutos e/ou em percentuais, podendo haver agrupamento de dados em classes previamente definidas.

**TABELA DE CONTINGÊNCIA:** A tabela de contingência apresenta as frequências das categorias de uma variável em função das categorias de outra variável.

**TABELA DE RESULTADOS (PAYOFF TABLE):** Tabela que apresenta os resultados em diferentes estados da natureza na teoria da decisão.

**TABELA DE TABULAÇÃO CRUZADA:** Ver tabela de contingência.

**TABELA DE VOLUME:** Tabela que apresenta o volume de árvores individuais em função de medidas da árvore, geralmente o diâmetro à altura do peito (DAP), altura e quociente de forma. Antes que o uso de calculadoras e computadores eletrônicos se tornasse corrente, os modelos de predição de volume eram ajustados, mas no lugar de apresentara equação matemática do modelo, que constitui uma equação de volume, construía-se uma tabela. Por tradição, os nomes dos tipos de tabelas de volume são usados para designar os tipos de equações de volume.

**TABELA PRIMITIVA OU DADOS BRUTOS:** É uma tabela ou relação de elementos que não foram numericamente organizados. É difícil formarmos uma ideia exata do comportamento do grupo como um todo, a partir de dados não ordenados.

**TABELAS DE CONTINGÊNCIA:** Tabelas nas quais se dispõem as frequências esperadas e observadas de duas ou mais amostras, cada uma com duas ou mais categorias, em tabelas de L linhas e C colunas, ou h linhas e k colunas, tais como 2 x 2, 3 x 2, 2 x 3, 3 x 3, 5 x 7, 8 x 9, dentre outras.

**TABELA DA VIDA:** Tabela da vida é um dos instrumentos estatísticos mais antigo. Tem sido usado por empresas desde o século XVII para estimar características associadas à distribuição dos tempos de falha.

**TABULAÇÃO:** Apresentação dos dados estatísticos sob a forma de tabelas ou quadros. Em inglês tabulation.

**TABULAÇÃO CRUZADA:** Um sumário tabular dos dados para duas variáveis. As classes para uma variável estão representadas pelas linhas; as classes para outra variável estão representadas pelas colunas.

**TABULAÇÃO CRUZADA:** Exame das respostas a uma pergunta com relação às respostas a uma ou mais perguntas.

**TABULAÇÃO CRUZADA DE TRÊS VIAS:** O método pelo qual a relação entre três variáveis é apresentada numa série de tabelas contingenciais.

**TAL QUE:** Símbolo matemático dado por  $\mid$  . Ou dois pontos  $:$  .

**TAMANHO DA AMOSTRA:** É o número de elementos que constituem uma dada amostra.

**TAMANHO DA AMOSTRA:** Número de elementos da amostra.

**TAMANHO DE AMOSTRA:** Número de unidades experimentais do ensaio, geralmente determinado por meio de cálculo, mas que também pode ser obtido de algum outro critério como, por exemplo, estudando o que é usual na arca ou recrutando as unidades disponíveis. Número de pacientes envolvidos em um estudo ou número de pacientes que deverá ser envolvido em um estudo.

**TAMANHO DE AMOSTRA:** Número de unidades amostrais utilizadas numa amostragem ou levantamento. Para métodos de área fixa, deve ser indicado o tamanho da parcela. Nos métodos de amostragem por parcelas de área variável ou unidades amostrais não-superficiais, basta indicar o número de pontos amostrais incluídos no levantamento.

**TAMANHO AMOSTRAL PARA DETECÇÃO DE DIFERENÇAS SIGNIFICATIVAS ENTRE TRATAMENTOS:** Se a hipótese a ser testada for unilateral o tamanho mínimo adequado da amostra a ser usado no planejamento e execução da experimentação científica é:

$$N = \frac{(Z_{\alpha} + Z_{\beta})^2 \sigma_D^2}{\delta^2}$$

em que,  $Z_{\alpha}$  e  $Z_{\beta}$  são valores da função distribuição acumulada da distribuição

normal padrão, associados às probabilidades de erro tipo I ( $\alpha$ ) e erro tipo II ( $\beta$ ), em teste de hipótese unilateral.  $\sigma_D^2$  é a variância da diferença entre duas médias de tratamentos e  $\delta$  tamanho da diferença verdadeira entre duas médias, que se deseja detectar ou discriminar como significativa. Agora por outro lado se a hipótese a ser testada for bilateral o tamanho mínimo adequado da amostra a ser usado no planejamento e execução da experimentação científica é:

$$N = \frac{\left(\frac{Z_{\alpha}}{2} + Z_{\beta}\right)^2 \sigma_D^2}{\delta^2}$$

A quantidade  $(1 - \beta)$  é equivalente a P, a probabilidade de que o experimento exiba uma diferença estatisticamente significativa entre médias de tratamentos. Probabilidades de 0,80 e 0,90 são comuns e adequadas na prática. Probabilidades maiores tais como 0,95 ou 0,99, podem ser consideradas, mas os tamanhos amostrais necessários para atingir tais especificações são antieconômicos.

**TAMANHO DE EFEITO:** Medida padronizada de diferenças de grupos usada no cálculo de poder estatístico. Calculado como a diferença em médias de grupos dividida pelo desvio padrão, é então comparável, em estudos de pesquisa, a uma medida generalizada de efeito, ou seja, diferenças em médias de grupos.

**TAMANHO DO EFEITO:** Estimativa do grau em que o fenômeno estudado por exemplo, correlação ou diferença em médias existe na população.

**TABULAÇÃO:** Apresentação dos dados estatísticos sob a forma de tabelas ou quadros.

**TENDENCIOSIDADE:** Defeito de técnica de amostragem, de observação ou de cálculo que acarreta diferenças sistemáticas entre os valores da amostra e os da população.

**TEMPO DE RETORNO:** Em particular, pode-se definir o tempo de retorno, denotado por T e expresso em anos, como o valor esperado da variável geométrica tempo de recorrência, aqui representada por  $\tau$  .

Com essa definição escreve-se que  $T = E(\tau) = \frac{1}{p}$ . O tempo de retorno, portanto, não se refere a um

tempo cronológico. De fato, T é uma medida de tendência central dos tempos cronológicos, aqui denomina-se tempos de recorrência. Em outras palavras, o tempo de retorno T, associado a um certo evento de referência de um processo de Bernoulli indexado em anos, corresponde ao tempo médio necessário em anos para que o evento recorra, em um ano qualquer, e é igual ao inverso da probabilidade de que tal evento de referência ocorra. Em hidrologia, o conceito de tempo de retorno é empregado com muita frequência no estudo probabilístico de eventos máximos anuais, tais como enchentes ou alturas diárias de precipitação máximas anuais. Tais variáveis aleatórias são contínuas e, portanto, têm seu comportamento definido por funções densidade de probabilidades, Se, para a variável X dessa de ua distribuição de probabilidade for definido um quantil de referência  $x_p$ , de modo que o sucesso seja a superação de  $x_p$ , então, o tempo de retorno T, associado ao quantil de referência, corresponde ao número médio de anos necessário para que o evento  $\{X > x_p\}$  recorra uma vez, em um ano qualquer. O tempo de retorno corresponde ao inverso de  $P(X > x_p)$ . Embora o conceito de tempo de retorno esteja, geralmente, vinculado a eventos máximos anuais, ele também pode ser estendido ao estudo probabilístico de eventos mínimos anuais, tais como vazões médias mensais mínimas anuais. O processo de Bernoulli, nesse caso, é semelhante ao de máximos anuais, porém, o que determina o sucesso é o fato de o evento mínimo anual encontrar-se abaixo de certo valor limite  $x_p$ . O tempo de retorno, nesse caso, passa a ser entendido como o tempo médio, em anos, para que haja a recorrência de uma estiagem mais severa do que a definida por  $x_p$ , ou seja, a recorrência de um novo evento  $\{X < x_p\}$ , em um ano qualquer. Supondo que X represente a variável aleatória contínua, característica do evento mínimo anual em questão, verifica-se que, nesse caso, o tempo de retorno T, associado ao quantil de referência, corresponde ao inverso de  $P(X < x_p)$ , ou seja, ao inverso da função acumulada de probabilidades  $F_X(x_p)$ .

**TESTE:** Prova ou conjunto de provas tendentes a determinar, num indivíduo ou grupo de indivíduos, a presença ou grau de certo traço.

**TAXA:** Ver coeficiente.

**TAXA DE ABANDONO DO PROGRAMA DE VACINAÇÃO:** Indicador que permite avaliar e cobertura vacinal, expressando o percentual de crianças que não chegam a completar a vacinação mínima necessária para ficarem protegidas. Calcula-se da seguinte maneira:

$$\text{Taxa de abandono} = \frac{n^\circ \text{ de primeiras doses} - n^\circ \text{ de terceiras doses}}{n^\circ \text{ de primeiras doses}}$$

**TAXA DE INCIDÊNCIA:** Porcentagem de pessoas ou domicílios na população em geral que se ajusta às qualificações a serem amostradas.

**TAXA DE INFLAÇÃO:** Taxa de variação do nível médio de preços.

**TAU DE KENDALL ( $\tau$ ):** Coeficiente de correlação baseado no número de inversões em um ranking de dados ordinais quando comparado com outro. Em inglês Kendall's tau ( $\tau$ ).

**TAXA DE MELHORAMENTO GENÉTICO:** Taxa de melhoria por unidade de tempo em ano. É dependente da: i) herdabilidade das características consideradas; ii) diferencial de seleção; iii) correlações

genéticas entre características consideradas; iv) intervalo de geração no rebanho; e v) número de características selecionadas.

**TAXA DE RESPOSTAS:** Número de participantes num survey dividido pelo número selecionado na amostra, em forma de porcentagem. Também chamada de taxa de questionários completados ou, em surveys auto-administrados, de taxa de retorno, o percentual de questionários enviados e devolvidos.

**TAXA FERTILIDADE TOTAL (TFT):** Número de crianças dividido pelo número de mulheres assumidas terem vivido plenamente suas vidas reprodutivas.

**TAXONOMIA:** Classificação empiricamente obtida de objetos reais baseada em uma ou mais características. É tipificada pela aplicação de análise de agrupamentos ou outros procedimentos análogos. Essa classificação pode ser contrastada com uma tipologia.

**TAXONOMIA:** Disciplina biológica que trata da classificação dos organismos em grupo baseados no compartilhamento de características comuns.

**TABELA DE NÚMEROS ALEATÓRIOS:** Ver números aleatórios.

**TABULAÇÃO:** Atividade de organização dos dados em tabelas. Tabular os dados de uma pesquisa significa consolidar as informações numa planilha eletrônica de dados.

**TEOREMA:** Na matemática, fórmula ou conceito que pode ser demonstrado. Na lógica, refere-se a uma proposição que pode ser demonstrada por meio de processos dedutivos ou indutivos. Ideia ou parte de uma teoria.

**TEORIA DO ERRO:** Conjunto de ideias propostas no qual critica os filósofos de sua época e antecessores por valorizarem excessivamente a especulação e a dogmática teológica em detrimento da observação criteriosa dos fatos que é o empirismo. Nesta teoria, Bacon enumera quatro erros ou atitudes que deveriam ser evitados na busca pela verdade, os quais ele denominou de ídolos.

**TABELA DE CONTINGÊNCIA (CONTINGENCY TABLE):** Uma forma de organizar, mostrar e analisar dados para ilustrar relações entre duas ou mais variáveis categóricas. Os valores em uma tabela de contingência são as frequências ou contagens de observações em cada categoria.

**TAMANHO DO EFEITO (EFFECT SIZE):** A diferença esperada entre as médias de grupos que estão sujeitos a diferentes tratamentos.

**TESTES NÃO PARAMÉTRICOS:** Método de análise conhecido como de distribuição livre, já que os parâmetros dificilmente desempenham qualquer papel, as hipóteses são não paramétricas, tanto que se define um teste não paramétrico como aquele cujo modelo não especifica condições sobre os parâmetros da população da qual a amostra foi obtida. Eles são na maior parte dos casos, muito facilmente trabalhados numericamente, além do mais, esses procedimentos podem também ser aplicados a dados oriundos de escalas nominais e de posição. No entanto, um teste de distribuição livre, quando aplicado a dados de uma família de distribuição conhecida, é sempre mais fraco que o teste paramétrico correspondente. Em geral os testes de distribuição livre não levam em consideração a magnitude dos dados, além de não permitir testar interações, salvo sob condições especiais de aditividade.



**TRATAMENTO:** É cada um dos níveis de um fator ou cada uma das combinações dos níveis dos fatores, quando testando mais de um fator.

**TESTEMUNHA:** É o tratamento padrão de comparação. Pode ser a ausência do fator como uma dose zero de adubo, por exemplo, ou a aplicação usual do fator tal como um cultivar recomendado para cultivo na região, espaçamento adotado pelos agricultores, etc.. Também conhecido como controle ou placebo.

**TAXA DE ERRO POR EXPERIMENTO (EXPERIMENT-WIDE ERROR RATE):** O valor crítico verdadeiro do  $\alpha$  para o ponto em que a hipótese nula será rejeitada, após a correção de comparações múltiplas. Pode ser obtido usando os métodos ou testes de Bonferroni ou de Dunn-Sidak, dentre outros.

**TAXA DE ERRO POR COMPARAÇÃO (COMPARISONWISE):** Razão entre o número de inferências erradas (quantidades de vezes que os testes de comparações de médias rejeitou  $H_0$  erradamente) e o número total de inferências.

**TAXA DE ERRO POR COMPARAÇÃO (COMPARISONWISE):** Razão entre o número de erros tipo I (concluindo que  $\mu_i \neq \mu_j$  quando  $\mu_i = \mu_j$ ) e o número de comparações.

**TAXA DE ERRO POR EXPERIMENTO (EXPERIMENTWISE):** Razão entre o número de experimentos com no mínimo uma inferência errada (quantidades de vezes que os testes de comparações de médias rejeitou  $H_0$  erradamente) e o número total de experimentos.

**TAXA DE ERRO POR EXPERIMENTO (EXPERIMENTWISE):** Razão entre o número de experimentos com um ou mais erros tipo I (concluindo que  $\mu_i \neq \mu_j$  quando  $\mu_i = \mu_j$ ) e o número total de experimentos.

**TENTATIVA (TRIAL):** Uma réplica ou observação visual. Muitas tentativas formam um experimento estatístico.

**TENTATIVA DE BERNOULLI (BERNOULLI TRIAL):** Um experimento que tem apenas dois resultados possíveis, como presença/ausência; vivo/morto; cara/coroa, defeituoso/não defeituoso, sadio/doente, vacinado/não vacinado. Uma tentativa de Bernoulli resulta em uma variável aleatória de Bernoulli.

**TESTE DE HIPÓTESE:** Procedimento estatístico utilizado para verificar se os resultados obtidos a partir de uma amostra são válidos, ou seja, podem ser generalizados para toda a população de onde a amostra foi extraída ou não, são devidos ao acaso. Dizendo de outra forma, trata-se de um processo de decisão que utiliza a informação amostral para que se decida entre duas hipóteses. O teste de hipótese basicamente compara os resultados obtidos na amostra com os resultados teoricamente esperados, computando a probabilidade de erro na aceitação da hipótese. Há vários tipos de testes de hipótese que devem ser utilizados, dependendo de cada situação.

**TESTEMUNHA OU GRUPO CONTROLE:** É o conjunto de parcelas que, ou não recebe tratamento, ou recebe um tratamento já conhecido, sendo que a resposta da testemunha será comparada com as respostas dos grupos tratados.

**TRATAMENTOS:** São procedimentos que estão sendo comparados no experimento.

**TESTE ESTATÍSTICO:** Designação geral das técnicas estatísticas cujo objetivo básico é o de determinar a veracidade ou não de uma hipótese estatística, levando-se em conta determinado nível de significância.

**TESTE KOLMOGOROV-SMIRNOV:** Teste estatístico que verifica quão diferentes são duas distribuições de dados. Na prática de pesquisa, serve para determinar se uma variável possui distribuição normal ou gaussiana ou não normal. Esta informação é de extrema relevância, uma vez que norteará a decisão do pesquisador em se utilizar de técnicas estatísticas paramétricas se a distribuição dos dados for normal ou não paramétricas se a distribuição não for normal.

**TÉCNICA:** É a forma mais segura e ágil para se cumprir algum tipo de atividade, utilizando-se de um instrumental apropriado.

**TÉCNICA:** É o método utilizado na coleta de dados. Basicamente existem três técnicas de se colher dados para avaliação: observação, inquirição e a testagem.

**TÉCNICA:** Conjunto de procedimentos e recursos de que se serve uma ciência.

**TÉCNICA DA TERCEIRA PESSOA:** Uma maneira de se compreender os sentimentos dos respondentes pedindo-lhes que respondam por um terceiro, tais como o vizinho ou a maioria das pessoas.

**T<sup>2</sup> DE HOTELLING:** Teste para avaliar a significância estatística da diferença nas médias de duas ou mais variáveis entre dois grupos. É um caso especial de análise de variância multivariada (MANOVA) usado com dois grupos ou níveis de uma variável de tratamento.

**TÉCNICA:** Grupo de mecanismos ou metodologia científica usada no manuseio de ferramentas, de instrumentos, bem como no conjunto de reações de um sistema dinâmico, como por exemplo, no comportamento social, no comportamento humano, no caso de comportamento informacional, como também no comportamento atmosférico, etc. com a finalidade de se obter respostas satisfatórias.

**TÉCNICA DE DEPENDÊNCIA:** Classificação de técnicas estatísticas diferenciadas por terem uma variável ou um conjunto de variáveis identificado como a(s) variável(eis) dependente(s) e a(s) variável(eis) remanescente(s) como independente(s). O objetivo é a previsão da(s) variável(eis) dependente(s) pela(s) variável(eis) independente(s). Um exemplo é a análise de regressão.

**TÉCNICA DE INTERDEPENDÊNCIA:** Classificação de técnicas estatísticas nas quais as variáveis não são divididas em conjuntos dependentes e independentes, como por exemplo, a análise fatorial, todas as variáveis são analisadas como um único conjunto.

**TÉCNICA DE METADES DIVIDIDAS:** Método de se avaliar a confiabilidade de uma escala, dividindo-se por dois o conjunto total de itens de mensuração, correlacionando os resultados.

**TÉCNICAS BIVARIADAS:** Métodos estatísticos para determinar o relacionamento entre duas variáveis.

**TÉCNICAS DE CONTORNO AUTOMÁTICO:** Destinam-se a produção de mapas de isovalores em computador, a partir da malha regular amostrada ou interpolada. Assim, dada a malha regular, é possível proceder ao contorno automático por meio de duas técnicas: i) método do rastreamento em que cada curva é pesquisada ou rastreada em todas as células até que a curva encontre o ponto inicial que é a curva fechada ou a borda do mapa a qual é a curva aberta. Este processo é repetido para todas as curvas de isovalores desejadas; ii) método das células, em que as células são pesquisadas uma a uma para todas as curvas de isovalores desejadas. Encontrando-se todas as intersecções presentes na célula, estas são então traçadas e o procedimento é repetido para todas as células.

**TÉCNICAS DE PROJEÇÃO:** Maneiras de sondar os sentimentos mais profundos dos respondentes ao fazer com que projetem esses sentimentos para uma situação não estruturada.

**TECNOLOGIA SSL (SECURE SOCKET LAYER):** Sistema de codificação para computadores que retém informações confidenciais.

**TEMA:** É a matéria, objeto ou tópico que se é estudado em uma pesquisa científica, sendo de grande amplitude de alcance e mais abrangente, ao contrário de problema de uma pesquisa que é restrito.

**TEMPO GANHO POR DIAGNÓSTICO PRECOCE (LEARD TIME, DOS AUTORES INGLESES):** Quando o diagnóstico é feito precocemente; ocorre, por exemplo, em rastreamentos. Pode induzir a viés, por estimativas artificialmente altas das taxas de sobrevivência, quando doenças graves são detectadas em fase pré-clínica.

**TENDE A:** Símbolo matemático dado por  $\rightarrow$ .

**TENDÊNCIA:** É melhor conceituada como erro sistemático, inflacionando ou deflacionando consistentemente os escores num dado tratamento ou grupo controle. Como tendência não é aleatória, em geral não se cancela entre grupos.

**TENDÊNCIA:** i) refere-se a variação sistemática da variável de interesse tais como, espessura, teores, dentre outras em termos do seu aumento ou diminuição numa determinada direção. A tendência pode ser verificada pelo ajuste de superfícies de tendência; ii) é a componente de deriva presente na variável regionalizada, que causa a sua não estacionaridade. Esta componente de deriva pode ser representada por meio de polinômios de baixo grau, geralmente até três. Os resíduos, que são aleatórios, dados pela diferença entre o valor da variável original e o valor do polinômio ajustado, representa uma nova variável regionalizada, agora estacionária.

**TENDÊNCIA (ESTATÍSTICA):** Diferença consistente, persistente, da estatística em relação ao parâmetro que se quer estimar. Também dita viés ou vício, traduz a palavra inglesa bias. Evolução da variável em certo sentido e direção, em geral em função do tempo. Traduz a palavra inglesa Trend.

**TENDÊNCIA CENTRAL:** Fenômeno observado em uma distribuição de valores na qual os dados se agrupam em torno de um valor central, representativo da série, que pode ser a média, mediana, moda, quartis, etc., os quais resumem aspectos importantes desta distribuição.

**TENDÊNCIA CENTRAL:** É a medida do valor mais provável de um conjunto de observações. A tendência central pode ser expressa por meio da média, mediana ou moda.

**TENDÊNCIA (GERAL):** Propensão, inclinação, preferência pessoal preconcebida que causa efeito ou influencia a maneira pela qual uma medida, análise, avaliação ou procedimento é executado ou relatado.

**TENDÊNCIA CENTRAL:** Medida que indica o valor médio típico de uma distribuição; a média e a mediana são exemplos de medidas de tendência central.

**TENDÊNCIA CENTRAL:** Medidas de tendência central são facilmente identificáveis com médias. Nesse sentido dão uma indicação do que é típico numa distribuição. São localizadoras da distribuição, isto é, localizam a distribuição na escala de mensuração. Tendem a estar centralmente localizadas. A média, a mediana e a moda são as medidas de tendência central mais comuns.

**TENDÊNCIA CENTRAL, MEDIDAS DE:** São pontos na distribuição, usados para localizá-la. A média é a medida de tendência central mais comumente usada: a soma de todos os escores dividida pelo número de escores. A mediana é o ponto da distribuição pelo número de escores. A mediana é o ponto da distribuição, abaixo do qual se encontra a metade dos escores. A modal ou moda é o valor mais frequente na distribuição.

**TENDÊNCIA CURVILÍNEA:** Uma tendência que não é linear. Pode ser representada como um polinômio, uma expressão matemática mais complicada como uma curva logística, ou por algum processo de suavização como a média móvel. Em inglês Curvilinear trend.

**TENDÊNCIA, ESTUDOS DE:** Projeto de pesquisa que visa estudar as mudanças dentro de uma população, em vários períodos de tempo.

**TENDÊNCIA SECULAR:** Denominação utilizada para designar as mudanças da incidência de determinada doença ao longo de considerável tempo, geralmente medido em décadas.

**TENDÊNCIA SECULAR:** Comportamento das freqüências de evento epidemiológico acompanhado por anos consecutivos. A tendência secular pode ser expressa pelo coeficiente de inclinação de uma reta, obtida por regressão a partir das freqüências de casos de doenças ou do número de óbitos. Os coeficientes de inclinação positivos mostram tendência para o crescimento e os negativos, para o declínio. O coeficiente zero indica a constância do processo.

**TENDENCIOSIDADE:** Defeito de técnica de amostragem, de observação ou de cálculo que acarreta diferenças sistemáticas entre os valores da amostra e os da população. Em inglês bias.

**TENDENCIOSIDADE:** Ver erro sistemático.

**TEOR DE CORTE:** É o teor abaixo do qual não há viabilidade técnica ou econômica para se proceder a lavra do minério.

**TEOR MÉDIO:** Representa o teor de uma parte de um corpo de minério, de um bloco do depósito, determinado a partir de um conjunto de amostras de furos vizinhos, utilizando técnicas de estimativa tais como princípios de interpretação ou funções de interpolação.

**TEOREMA BINOMIAL:** É o resultado que mostra como expandir  $(x + y)^n = x^n + a_{n-1}x^{n-1}y + a_{n-2}x^{n-2}y^2 + \dots + y^n$  onde os coeficientes  $a_i$  são denominados de coeficientes binomiais. Em inglês binomial theorem.

**TEOREMA CENTRAL DO LIMITE:** Teorema que afirma que as médias amostrais tendem a ter uma distribuição normal com média  $\mu$  e desvio padrão  $\sigma$ . Conhecido também como teorema do limite central.

**TEOREMA DO LIMITE CENTRAL OU TEOREMA CENTRAL DO LIMITE:** É um tipo de teoria em forma de teorema muito útil para inferência estatística ou estatística indutiva onde afirma que Se  $X$  uma população possui uma distribuição de probabilidade com média  $\mu$  e desvio padrão  $\sigma_x$  finito, a média aritmética  $\bar{X}$ , da amostra aleatória simples de tamanho  $n$ , terá distribuição normal aproximada com média  $\mu$  e desvio padrão  $\frac{\sigma_x}{\sqrt{n}}$ , tornando-se a aproximação progressivamente melhor com o aumento do

$n$ . Este teorema não faz menção sobre a natureza da população. Isto significa que quando  $n$  é grande [ $n \geq 30$ ], a distribuição de amostragem teórica de  $\bar{X}$  será próxima a uma curva normal indiferentemente da forma da distribuição da população, onde esta pode ter qualquer um dos formatos, como por exemplo

dos tipos retangular, triangular, assimétrico, quadrado, em forma de sino ou normal, etc.. O teorema do limite central além de notável é também de grande aplicação prática, pois especifica completamente a distribuição da média  $\bar{X}$  em grandes amostras. De fato, na maioria dos casos em que o tamanho n da amostra é de cerca de 10,20 ou 25 elementos, a distribuição da média  $\bar{X}$  já é praticamente normal, mesmo em exemplos com tamanho de amostra com apenas 2 elementos a convergência já é verificada. Ou ainda pode-se raciocinar da seguinte forma, o teorema do limite central prova que a probabilidade de que a média amostral padronizada  $\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$  caia dentro de um dado intervalo converge para a probabilidade de

que a variável normal padronizada ou reduzida caia dentro daquele mesmo intervalo. Sendo que a única exigência é a de que a população tenha variância finita.

**TEOREMA CENTRAL DO LIMITE (TLC):** Seja  $X_1, X_2, \dots, X_n$  uma amostra aleatória independente e identicamente distribuída (i.i.d.) de tamanho n (isto é, n variáveis aleatórias independentes e igualmente distribuídas) de uma população com média  $\mu_X$  e variância  $\sigma_X^2$  finitas. Então, se  $\bar{X}$  é a média desta amostra, a função de distribuição da variável aleatória

$$Z = \frac{\bar{X} - \mu_X}{\frac{\sigma_X}{\sqrt{n}}}$$

estandardizada ou padronizada.

**TEOREMA CENTRAL DO LIMITE:** Teorema que afirma: A média de um grande número de variáveis aleatórias independentes, identicamente distribuídas, tem distribuição normal.

**TEOREMA DA PROBABILIDADE CONDICIONAL:** Sejam  $A \subset S$  e  $B \subset S$ , ou seja, para dois eventos quaisquer A e B, associados a um experimento aleatório E, sendo  $P(B) > 0$ , defini-se a probabilidade condicional de A, quando B tiver ocorrido (A/B), como sendo a seguinte equação:  $P(A/B) = \frac{P(A \cap B)}{P(B)}$ , se a

probabilidade de B for  $P(B) \neq 0$  [ $P(B) > 0$ ], também tem-se que  $P(B/A) = \frac{P(A \cap B)}{P(A)}$ , se  $P(A) \neq 0$  [ $P(A)$

$> 0$ ]. Isto é, a probabilidade condicional de A, dado B, é igual à probabilidade conjunta de A e B dividida pela probabilidade marginal, simples (não zero) de B, com  $P(B) \neq 0$ , se  $P(B) = 0$  não existe ( $\nexists$ ) P(A/B). Isto significa que a ocorrência do evento A está vinculada ou condicionada à ocorrência do evento B. No caso de três eventos aleatórios A, B e C, tem-se que  $P(C/A \cap B) = \frac{P(A \cap B \cap C)}{P(A \cap B)}$ , com  $P(A \cap B) \neq 0$ . Para

se entender a expressão acima, basta pensar que o espaço amostral ficou reduzido ou restrito ao evento B no caso de dois eventos, a presença da intersecção no numerador é justificada facilmente considerando que a ocorrência de A, quando se sabe que B já ocorreu, corresponde à ocorrência de A e de B simultaneamente, isto é, de sua intersecção.

**TEOREMA DO PRODUTO PARA PROBABILIDADE CONDICIONADA:** A probabilidade da ocorrência simultânea de dois eventos A e B é dada pelo produto da probabilidade de um dos eventos aleatórios, pela probabilidade condicional do outro evento, ou seja:  $P(A \cap B) = P(A) \cdot P(B/A)$   $\therefore P(A \cap B) = P(B) \cdot P(A/B)$ ,

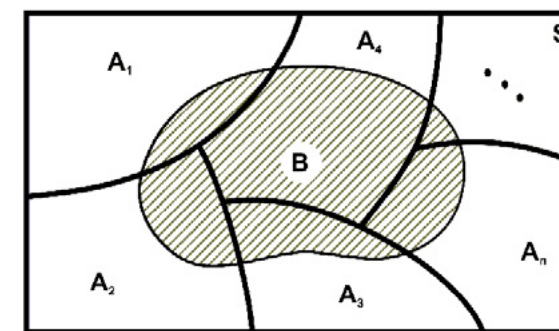
com  $P(B) \neq 0$ . Para n eventos, tem-se que:

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1) \cdot P(A_2/A_1) \cdot P(A_3/A_1 \cap A_2) \dots P(A_n/A_1 \cap A_2 \cap \dots \cap A_{n-2} \cap A_{n-1}).$$

**TEOREMA DE BAYES (BAYES THEOREM):** A fórmula para calcular a probabilidade de uma hipótese de interesse, conforme os dados observados e qualquer conhecimento prévio acerca da hipótese:

$$P(H|Y) = \frac{f(Y|H)P(H)}{P(Y)}$$

**TEOREMA DE BAYES:** Teorema de probabilidade condicional usado no cálculo de probabilidade e em estatística Bayesiana, também conhecido como teorema da revisão das probabilidades, teorema da probabilidade das causas, teorema da probabilidade a posteriori, teorema das probabilidades das hipóteses, teorema da probabilidade dos antecedentes, teorema da probabilidade total, o qual é determinado conforme é mostrado a seguir. Considere a seguinte partição de um espaço amostral S em n eventos mutuamente exclusivos  $A_1, A_2, \dots, A_n$ , e um



evento B pertencente a esse espaço amostral.

**Figura:** Diagrama de VENN mostrando a partição de um espaço amostral S juntamente com o evento B.

Inicialmente define-se o teorema da probabilidade total, conforme descrito a seguir. Suponha que os eventos  $A_1, A_2, \dots, A_n$  formem uma partição de um espaço amostral S, ou seja, os eventos  $A_i$  são mutuamente exclusivos e sua união é S. Seja B ser outro evento qualquer, então:

$B = S \cap B = (A_1 \cup A_2 \cup \dots \cup A_n) \cap B = (A_1 \cap B) \cup (A_2 \cap B) \cup \dots \cup (A_n \cap B)$ , onde os  $A_i \cap B$  são também mutuamente exclusivos.  $P(B) = P(A_1 \cap B) + P(A_2 \cap B) + \dots + P(A_n \cap B)$ . Assim pelo teorema da multiplicação  $P(B) = P(A_1) \cdot P(B/A_1) + P(A_2) \cdot P(B/A_2) + \dots + P(A_n) \cdot P(B/A_n)$ , isto é,

$$P(B) = \sum_{i=1}^n P(A_i) \cdot P(B/A_i). \text{ Sendo assim a equação do teorema de Bayes propriamente dito é obtido conforme}$$

mostrado a seguir. Para qualquer j, a probabilidade condicional de  $A_j$  dado B é definida como:

$$P(A_j/B) = \frac{P(A_j \cap B)}{P(B)}. \text{ Nesta equação, utiliza-se a expressão anterior para substituir } P(B) \text{ e } P(A \cap B) =$$

$P(A_j) \cdot P(B/A_j)$  para substituir  $P(A_j \cap B)$ , obtendo assim o teorema de Bayes

$$P(A_j/B) = \frac{P(A_j) \cdot P(B/A_j)}{P(A_1) \cdot P(B/A_1) + P(A_2) \cdot P(B/A_2) + \dots + P(A_n) \cdot P(B/A_n)}$$



O resultado dado pela equação anterior é chamado teorema de Bayes, em honra do pastor presbiteriano, filósofo e matemático inglês Thomas Bayes (1702-1761), e sua utilidade consiste em permitir aos pesquisadores calcular a probabilidade a posteriori  $P(A/B)$ , em termos das informações a priori  $P(A)$  e  $P(B)$ .

**TEOREMA DA SOMA DAS PROBABILIDADES:** Se A, B e C são três eventos aleatórios quaisquer, contidos num espaço amostral S, então a probabilidade da reunião dos três eventos é dada por:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C).$$

**Generalizando:** se  $A_1, A_2, A_3, \dots, A_n$  são n eventos aleatórios quaisquer, então:

$$P(A_1 \cup A_2 \cup A_3 \cup A_4 \cup \dots \cup A_n) = (-1)^0 P(A_1) + (-1)^1 \sum_{i < j = 2, \dots, n} P(A_i \cap A_j) + (-1)^2 \sum_{i < j < k = 3, \dots, n} P(A_i \cap A_j \cap A_k) + \dots + (-1)^{n-1} \sum_{i < j = 2, \dots, n} P(A_1 \cap A_2 \cap A_3 \cap A_4 \cap \dots \cap A_n)$$

onde este teorema é aplicado apenas em espaços amostrais finitos e discretos.

**TEOREMA DO LIMITE CENTRAL (CENTRAL LIMIT THEOREM):** O resultado matemático que permite o uso de estatística paramétrica sobre dados que não têm distribuição normal. O Teorema do Limite Central mostra que qualquer variável aleatória pode ser transformada em uma variável aleatória normal.

**TEOREMA DO LIMITE CENTRAL:** O teorema do limite central ou central do limite é aquele que mostra que, se uma variável aleatória (VA) X puder ser descrita como a soma de quaisquer n variáveis aleatórias independentes (VAs)  $(X_1, X_2, X_3, X_4, \dots, X_n)$ , então, para n suficientemente grande, essa soma X terá distribuição normal. Ou seja,  $X_1, X_2, X_3, X_4, \dots, X_n$  são variáveis aleatórias independentes, cada uma com  $E(X_i) = \mu_i$  e  $V(X_i) = \sigma_i^2$ , então, para  $n \rightarrow \infty$ , a soma:  $X = X_1, X_2, X_3, X_4, \dots, X_n$  tem distribuição normal, isto é,  $X \sim N(E(X), V(X))$  onde a esperança matemática da soma destas variáveis aleatórias é:

$$E(X) = E(X_1 + X_2 + X_3 + X_4 + \dots + X_n) = E(X_1) + E(X_2) + E(X_3) + E(X_4) + \dots + E(X_n) = \sum_{i=1, \dots, n} \mu_i$$

$$V(X) = V(X_1 + X_2 + X_3 + X_4 + \dots + X_n) = V(X_1) + V(X_2) + V(X_3) + V(X_4) + \dots + V(X_n) = \sum_{i=1, \dots, n} \sigma_i^2$$

O teorema do limite central tem uma prova razoavelmente complexa, mas pode ser aplicado em situações em que a variável aleatória de interesse X corresponde à soma de muitos efeitos aleatórios pequenos  $X_1, X_2, X_3, X_4, \dots, X_n$ . Cada parcela  $X_i$  contribui com um valor sem importância para a variação final da soma X. Por exemplo: i) o consumo de eletricidade X de uma cidade em certa época, que é a soma das pequenas procuras  $X_1, X_2, X_3, X_4, \dots, X_n$  de um grande número de consumidores; ii) a quantidade de água X em um reservatório em certo instante, que é o resultado da soma de um grande número de pequenas contribuições individuais  $X_1, X_2, X_3, X_4, \dots, X_n$ ; iii) o desvio do alvo X de um atirador com certa arma, que pode ser pensado como a soma de muitos desvios pequenos  $X_1, X_2, X_3, X_4, \dots, X_n$ , devidos a diversas causas; iv) o erro de medição X de um instrumento particular, que pode ser pensado como a soma de muitas contribuições pequenas, nenhuma das quais influenciando muito no erro final.

**TEORIA (THEORY):** Na ciência, uma teoria é uma coleção de conhecimentos aceitos que foram construídos através de repetidas observações e testes estatísticos de hipóteses. As teorias formam as bases dos paradigmas.

**TEOREMA DOS EVENTOS INDEPENDENTES:** Sejam  $A \subset S$  e  $B \subset S$ . Intuitivamente se A e B são independentes é porque  $P(A/B) = P(A/B^c) = P(A)$  e  $P(B/A) = P(B/A^c) = P(B)$ , ou seja, que as probabilidades condicionais sejam iguais às correspondentes probabilidades absolutas ou simples. A e B são eventos aleatórios estatisticamente independentes quando as probabilidades  $P(A) \neq 0$  e  $P(B) \neq 0$ , se  $P(A \cap B) = P(A).P(B)$ , pois a partir da fórmula da probabilidade condicional tem-se que

$$P(B).P(A/B) = \frac{P(A \cap B)}{P(B)}. P(B)$$

e sendo  $P(A/B) = P(A)$ , então tem-se que  $P(A \cap B) = P(A).P(B)$ . O uso da expressão anterior permite verificar que o evento impossível  $\emptyset$  é independente de qualquer evento A.

**TEOREMA DA PARTIÇÃO DE COCHRAN:** Ao decompor a soma dos quadrados de N variáveis aleatórias independentes normais unitárias  $Z_i$  em k variáveis aleatórias  $T_j$  que são cada uma delas a soma dos quadrados de  $v_j$  variáveis aleatórias normais unitárias  $U_{ji}$ , as  $T_j$  são independentes, seguindo a distribuição do qui quadrado com  $v_j$  graus de liberdade, se e só se a soma dos  $v$  for N, ou seja, temos que:  $\sum_{i=1}^N Z_i^2 = \sum_{j=1}^k T_j = \sum_{j=1}^k \sum_{i=1}^{v_j} U_{ji}^2 \Rightarrow T_j \sim \chi_{v_j}^2$ , se e somente se  $U_{ji} \sim N(0,1)$ ,  $Z_j \sim N(0,1)$  e  $\sum_{j=1}^k v_j = N$ .

**TEOREMA DA PROBABILIDADE TOTAL:** Para caracterizar e entender o teorema da probabilidade total é necessário supor que os eventos  $A_1, A_2, \dots, A_n$  formam uma partição de um espaço amostra S. Seja B um evento de S. O teorema nos diz como calcular a probabilidade de B em função da partição  $A_i$ , isto é:  $P(B) = P(A_1).P(B/A_1) + P(A_2).P(B/A_2) + \dots + P(A_n).P(B/A_n)$ . Este teorema nada mais é do que o conceito de média ponderada da Estatística aplicada a teoria da probabilidade. Em inglês total probability theorem.

**TEOREMA DE BAYES:** Para uma quantidade de interesse  $\theta$  associada a uma população representada por uma variável X que tem densidade dada por  $P(X|\theta)$ , o teorema de Bayes estabelece que  $P(\theta|X) = P(X|\theta)P(\theta)/P(X)$ . Do inglês Bayes theorem.

**TEOREMA DE BAYES:** Responde a duas questões importantes que permanecem sem resposta quando se leva em consideração apenas a sensibilidade e a especificidade: i) Se o resultado do teste for positivo, qual é a probabilidade de que o paciente tenha a doença? ii) Se o resultado do teste for negativo, qual a probabilidade de que o paciente não tenha a doença? O teorema estipula que a probabilidade de uma dada condição está presente em um indivíduo está relacionada com a prevalência dessa condição na população a que esse indivíduo pertence.

**TEOREMA DE BAYES:** Teorema usado na probabilidade condicional para revisar probabilidades em que para caracterizar e entender é necessário supor que os eventos aleatórios  $A_1, A_2, \dots, A_n$  formam uma partição de um espaço amostra S. Seja B um evento qualquer de S. O teorema mostra como calcular a probabilidade de  $A_1$  condicionada a ocorrência do evento B, isto é:  $P(A_1/B) = P(A_1).P(B/A_1) / [P(A_1).P(B/A_1) + P(A_2).P(B/A_2) + \dots + P(A_n).P(B/A_n)]$ . Em inglês Bayes' theorem.

**TEOREMA DE BLACKWELL:** Teorema referente a estimadores de variância mínima estabelecido por Blackwell (1947) e também por Rao (1949). Se um estimador de variância mínima existe, será sempre uma função de uma estatística suficiente. Em inglês Blackwell's theorem.



**TEOREMA DE LIMITE CENTRAL:** Uma distribuição de grande número de média de amostras que se aproximam de uma distribuição normal, independentemente da distribuição real da população de onde foram extraídas.

**TEOREMA DE NEYMAN-PEARSON:** Teorema usado na estatística o qual é enunciado da seguinte maneira. Seja  $X_1, X_2, \dots, X_n$  uma amostra aleatória independente e identicamente distribuída (iid) proveniente de uma população cuja função densidade de probabilidade f.d.p. ou função de probabilidade, no caso de variáveis aleatórias discretas) é  $f(x|\theta)$ , em que  $\theta \in \{\theta_0, \theta_1\}$ . Seja  $\alpha \in ]0, 1[$  dado. Seja  $k^*$  uma constante

positiva. Seja  $T^*$  a estatística de teste cuja região crítica,  $C_{T^*}$ , é dada pelo subconjunto do espaço amostral definido por

$$(x_1, x_2, \dots, x_n) \in C_{T^*} \text{ se e somente se, } \frac{L_0}{L_1} = \frac{L(\theta_0 | (x_1, x_2, \dots, x_n))}{L(\theta_1 | (x_1, x_2, \dots, x_n))} = \frac{\prod_{i=1}^n f(x_i | \theta_0)}{\prod_{i=1}^n f(x_i | \theta_1)} < k^*, \text{ sendo } \alpha \text{ a dimensão}$$

deste teste, isto é,  $\dim T^* = P(C_{T^*} | \theta = \theta_0) = \alpha$ . Então,  $T^*$  é o teste de dimensão  $\alpha$  mais potente para ensaiar a hipótese  $H_0 : \theta = \theta_0$  versus  $H_1 : \theta = \theta_1$ .

**TEOREMA DE RAO E BLACKWELL:** Teorema que fornece uma forma de construir um estimador não viciado de variância mínima a partir de um estimador não viciado e um estimador suficiente. Em inglês Rao-Blackwell theorem.

**TEOREMA DE CRAMER-RAO:** Teorema estatístico cujo enunciado é dado por. Seja  $X_1, X_2, \dots, X_n$  uma amostra aleatória independente e identicamente distribuída (iid) de uma população X cuja função densidade de probabilidade é  $f(X|\theta)$ , onde o parâmetro  $\theta$  toma valores num intervalo da reta real. Seja  $\hat{\theta} = G(X_1, X_2, \dots, X_n)$  e conseqüentemente,  $\hat{\theta} = G(x_1, x_2, \dots, x_n)$  um estimador não-enviesado de  $\tau(\theta)$ . Então, se as quatro condições de regularidade se cumprirem:

- i)  $\frac{\partial}{\partial \theta} \ln f(x|\theta)$  existe para todo o x no domínio da variável aleatória X e para todo o valor do parâmetro  $\theta$ ,
- ii)  $\frac{\partial}{\partial \theta} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \prod_{i=1}^n f(x_i | \theta) dx_1 \dots dx_n = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \frac{\partial}{\partial \theta} \prod_{i=1}^n f(x_i | \theta) dx_1 \dots dx_n,$
- iii)  $\frac{\partial}{\partial \theta} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} G(x_1, x_2, \dots, x_n) \prod_{i=1}^n f(x_i | \theta) dx_1 \dots dx_n = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} G(x_1, x_2, \dots, x_n) \frac{\partial}{\partial \theta} \prod_{i=1}^n f(x_i | \theta) dx_1 \dots dx_n,$
- iv)  $E \left[ \left( \frac{\partial}{\partial \theta} \ln f(X|\theta) \right)^2 \right] > 0$  existe, tem-se então que:  $v(\hat{\theta}) = \frac{[\tau'(\theta)]^2}{nE \left[ \left( \frac{\partial}{\partial \theta} \ln f(X|\theta) \right)^2 \right]}$ .

**TEOREMA DE TCHEBICHEV:** Teorema usado na inferência estatística onde se afirma que em qualquer conjunto de números, a fração que está a k ou menos desvio padrão da média é no mínimo  $1 - \frac{1}{k^2}$ .

**TEOREMA DE TCHEBICHEV:** Teorema que utiliza o desvio-padrão para dar informação sobre uma distribuição de dados.

**TEOREMA DE WISHART:** Seja a distribuição de probabilidade normal k-dimensional do vetor aleatório  $\xi : \{\xi_1\}$ , com  $E(\xi_1) = 0$ ,  $E(\xi_1 \xi_1) = \alpha_{ij}$ . Seja que, para uma amostra simples de tamanho n, temos que a m.a. dos  $x_i$  é  $x_i$ , e a cov. De  $x_i$  com  $x_j$  é  $s_{ij}$ , sendo que, no que se segue, só consideraremos as cov's

para as quais  $i \leq j$ . Nesse termo: Os vetores aleatórios  $\bar{x} : \bar{x}_1 : \bar{x}_2, \dots, \bar{x}_k$  e  $s^2 : s_{11}, s_{12}, \dots, s_{1k}, s_{22}, \dots, s_{kk}$  são estatisticamente independentes. O primeiro tem distribuição normal k-dimensional, caracterizada por  $E(\bar{x}_i) = E(\xi_i)$ , para todo i, e variância generalizada igual à de  $\xi$  dividida por n. O segundo tem distribuição caracterizada pela função densidade de probabilidade

$$\left[ \pi^{\frac{k^2-k}{4}} \prod_{t=1}^k \tau \left( \frac{n-t}{2} \right) \right]^{-1} \left( \frac{n}{2^k M} \right) L^{\frac{n-k-2}{2}} \exp \left[ -\frac{n}{2M} \sum_{i,j} M_{ij} s_{ij} \right]$$

Para todo  $s^2$  cujas componentes  $s_{ij}$  acima dadas formam uma matriz (L) positiva definida. Na expressão acima, M denota a v. generalizada de  $\xi$ , e L o determinante da matriz formada com os elementos  $s_{ij}$ , dado por Wishart (1928).

**TEOREMA DE REPRESENTAÇÃO DE DE FINETTI:** Um teorema usado na teoria de probabilidade que se deve a De Finetti (1937), onde se estabelece que uma amostra simples, que gera uma sequência binária permutável, pode ser sempre considerada como obtida pela amostragem binomial em que a probabilidade de P do tipo I tem uma distribuição inicial do tipo Iii, única. Em inglês de finetti's representation theorem.

**TEOREMA DO LIMITE CENTRAL:** Esse teorema afirma que, dada qualquer população com uma média e variância finita, à proporção que se aumenta o tamanho da amostra, a distribuição das médias de amostra aproxima-se da distribuição normal com média igual à média da população e variância igual à variância da população dividida pelo tamanho da amostra. Dada qualquer população com média  $\mu$  variância finita  $\sigma^2$ , à proporção que se aumenta o tamanho da amostra, sem limite, a distribuição da média da amostra aproxima-se de uma distribuição normal com média  $\mu$  e variância  $\frac{\sigma^2}{n}$ , em que n é o tamanho da amostra.

**TEOREMA DO LIMITE CENTRAL:** Na sua forma mais simples, o teorema estabelece que se n variáveis independentes têm variâncias finitas então sua soma tenderá, quando expressada em forma padronizada, à distribuição Normal quando n vai para o infinito. Conhecido também como teorema central do limite. Em inglês Central limit theorem.

**TEOREMA DO LIMITE CENTRAL:** É um dos mais importantes teoremas ou resultados usados na inferência estatística, o qual possui o seguinte enunciado. Se X que é uma população possui uma distribuição com média  $\mu$  e desvio padrão  $\sigma_x$ , a média aritmética  $\bar{X}$ , da amostra aleatória de tamanho n, terá distribuição normal aproximada com média  $\mu$  e desvio padrão  $\frac{\sigma_x}{\sqrt{n}}$ , tornado-se a aproximação

progressivamente melhor com o aumento do tamanho da amostra n. É importante lembrar que este teorema não faz menção sobre a natureza da população, isto significa que quando n é grande [ $n \geq 30$ ], a distribuição de amostragem teórica da média aritmética  $\bar{X}$  será próxima a uma curva normal indiferentemente da forma da distribuição da população que pode ter quaisquer formatos, como por exemplo, retangular, triangular, assimétrico, quadrado, em forma de sino ou normal, dentre outros. O teorema do limite central além de notável é também de grande aplicação prática, pois especifica completamente a distribuição da média  $\bar{X}$  em grandes amostras. De fato, na maioria dos casos em que o tamanho n da amostra é de cerca de 10, 20 ou 25 elementos, a distribuição por amostragem da média  $\bar{X}$  já é praticamente normal. Ou ainda podemos raciocinar da seguinte forma, o teorema do limite central prova que: a probabilidade de que a média amostral padronizada  $\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$  caia dentro de um dado intervalo converge

para a probabilidade de que a variável aleatória normal padronizada caia dentro daquele mesmo intervalo. Sendo que a única exigência é a de que a população tenha variância finita.

**TEOREMA DO LIMITE CENTRAL:** É, um dos mais importantes teoremas da estatística matemática relacionada a distribuições de frequências de amostragem e pode ser enunciada como: Se amostras aleatórias de tamanho fixo são retiradas de uma população cuja distribuição teórica é de forma arbitrária, mas com média e variância finitas, a distribuição das amostras tende mais e mais a uma distribuição normal com média  $m$  e variância  $\frac{s^2}{n}$  tanto quanto o tamanho das amostras aumenta. Se  $X_1, X_2, \dots, X_n$  são variáveis aleatórias, a média  $\bar{X}$  é  $\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$  então o valor esperado de  $\bar{X}$ , usando as proprie-

dades da média é:  $E[\bar{X}] = \frac{E[X_1] + E[X_2] + \dots + E[X_n]}{n}$  ou  $E[\bar{X}] = \frac{n\mu}{n} = \mu$ . A variância de  $\bar{X}$  é calculada

como:  $Var[\bar{X}] = Var\left[\frac{([X_1] + [X_2] + \dots + [X_n])}{n}\right]$ . Aplicando-se as propriedades da variância tem-se:

$Var = \frac{\{Var[X_1] + Var[X_2] + \dots + Var[X_n]\}}{n}$ . A distribuição de  $\bar{X}$  tem média  $m$  e variância  $\frac{s^2}{n}$  que se aproxima

da distribuição normal tanto quanto aumenta o tamanho da amostra. Na maioria dos casos a aproximação é boa a partir de 40 amostras.

**TEOREMA DO LIMITE CENTRAL:** Para amostras razoavelmente grandes, a distribuição das médias de muitas amostras é normal ou gaussiana, mesmo que os dados nas amostras individuais possam ter inclinações, curtoses ou irregularidades.

**TEOREMA DO NÚMERO PRIMO:** O teorema do número primo assegura que o número de primos  $n$  tende ao infinito tão rápido quanto  $\frac{n}{\ln(n)}$ . Em inglês prime number theorem.

**TEOREMA FUNDAMENTAL DA ÁLGEBRA:** Resultado que assegura que qualquer polinômio com coeficientes reais ou complexos tem uma raiz no plano complexo. Em inglês fundamental theorem of algebra.

**TEORIA:** Explicação abrangente de algum setor da existência, incluindo definições de elementos constituindo o que vai ser explicado, o conjunto de suposições e axiomas que servirão como ponto de partida da teoria e por fim conjunto de afirmações inter-relacionadas sobre as relações entre os elementos.

**TEORIA:** Teoria como pesquisa significa muitas coisas diferentes a diferentes pessoas. Teoria refere-se a qualquer ordenamento sistemático de idéias sobre os fenômenos de uma área de pesquisa.

**TEORIA:** Um conjunto sistemático de relações causais que fornece uma explicação consistente e abrangente de um fenômeno. Na prática, uma teoria é a tentativa de um pesquisador em especificar o conjunto inteiro de relações de dependência que explicam um conjunto particular de resultados. Uma teoria pode ser baseada em ideias geradas a partir de uma ou mais de três fontes principais: pesquisa empírica prévia; experiências passadas e observações de comportamento real, atitudes, ou outros fenômenos; e outras teorias que fornecem uma perspectiva para análise. Assim, a construção de teorias não é de domínio exclusivo de pesquisadores acadêmicos; ela tem um papel explícito para outros profissionais também. Para qualquer pesquisador, a teoria fornece um meio de tratar do quadro geral e avaliar a importância relativa de vários conceitos em uma série de relações.

**TEORIA:** Grupo flexível de conceitos sobre assunto particular de uma população. Organização sistematizada de sugestões acerca de um específico domínio do conhecimento humano.

**TEORIA:** É um conjunto de princípios e definições que servem para dar organização lógica a aspectos selecionados da realidade empírica. As proposições de uma teoria são consideradas leis se já foram suficientemente comprovadas e hipóteses se constituem ainda problema de investigação. (GOLDENBERG, 1998).

**TEORIA (REDAÇÃO):** Quando temos que expor uma teoria, o fato se complica porque não apenas temos que descrevê-la, mas também somos levados a persuadir o leitor aceitá-la. Eis dois princípios de argumentação que podem ser usados em uma redação: i) no começo do trabalho estabeleça um quadro para os argumentos, definindo a premissa em que se baseiam os argumentos e os limites a serem colocados; ii) antecipe as objeções que possam surgir na mente do leitor.

**TEORIA ANALÍTICA DOS NÚMEROS:** É o estudo da teoria dos números, ou teoria algébrica dos números, com recursos da análise, especialmente da análise complexa, tais como séries infinitas e convergência. Em inglês analytic number theory.

**TEORIA ASSINTÓTICA:** Teoria baseada em distribuições limite. Do inglês Asymptotic theory.

**TEORIA CIENTÍFICA:** É a teoria que explica e organiza fatos previamente sem ligação; sugere experimentos úteis, estimulando o progresso científico; pode ser testada, e sua pretensão de ser verdadeira está aberta à verificação; e ainda prevê o resultado de experimentos novos. Se a predição for testada, a confiança na teoria proposta aumentará. Todavia, as teorias da evolução biológica darwiniana e da relatividade einsteiniana são exemplos de teorias suficientemente robustas para se manterem intocadas, mesmo sem terem sido plenamente verificadas.

**TEORIA DA SELEÇÃO DE AMOSTRAS:** Capítulo da Estatística que tem por objeto o estudo dos diversos tipos de amostra, das técnicas de sua escolha, e dos estimadores mais convenientes segundo o caso, com o fim de caracterizar as respectivas populações originárias, com o máximo de precisão e o mínimo de custo.

**TEORIA DE DECISÃO DE BAYESIANA:** Teoria de decisão baseada no teorema do padre Thomas Bayes, Reverendo Thomas Bayes, autor do primeiro artigo de uma forma quantitativa de inferência indutiva. Nascido em 1702, filho primogênito de Ann Bayes e Joshua Bayes, falecido em Tumberge Wells, em 17 de abril de 1761. Em inglês Bayes decision theory.

**TEORIA DE DECISÃO:** Disciplina da estatística que envolve e explora a estrutura do processo de tomada de decisão. O termo é usado de forma genérica e interdisciplinar, decorrendo, possivelmente, da natureza ampla do processo de decisão. Em inglês Decision theory.

**TEORIA DE GALOIS:** É o estudo de certos grupos que podem ser associados com equações polinomiais. Se a solução de uma equação pode ou não ser escrita por meio de funções racionais e raízes quadradas, cúbicas, dentre outras, depende de certas propriedades teóricas do grupo de Galois. Em inglês Galois theory.

**TEORIA DE MENSURAÇÃO:** Tem a ver com o delineamento de regras e procedimentos cuja aplicação se destina a aumentar a probabilidade de que os acontecimentos no mundo dos conceitos corresponderá aos acontecimentos no mundo da realidade.

**TEORIA DE NEYMAN-PEARSON:** Teoria geral de teste de hipótese, atribuída a Jerzy Neyman e Egon Pearson. Está baseada na consideração de dois tipos de erros que podem ser cometidos na hora de julgar uma hipótese estatística. Em inglês Neyman-Pearson theory.

**TEORIA DINÂMICA:** A teoria dinâmica difere da teoria estática no envolvimento de maior número de relacionamentos complexos e no fato de que suas explicações contém mais interdependências. Um exemplo desse tipo de teoria é o Dynamic Theory of Adaptive Employee Behavior de March & Simon: Quanto mais baixo a satisfação de um empregado, tanto maior será a sua busca de fontes alternativas de satisfação. O comportamento de busca é estimulado pelo valor esperado de recompensa. Quanto maior o valor esperado de recompensa, tanto maior o nível de aspiração.

**TEORIA DA AMOSTRAGEM:** É construída em torno de um conceito que se um número suficiente de unidades representativas de uma população são selecionadas sem enviezamento, o valor médio das unidades irá aproximar aquele da população total.

**TEORIA DAS VARIÁVEIS REGIONALIZADAS:** Formalizada por Matheron (1963), a partir de estudos práticos desenvolvidos por Danie Gerhardus Krige (1919-2013) no cálculo de reservas nas minas de ouro do Rand na África do Sul, tem por objetivo o estudo e a representação de forma adequada das propriedades estruturais das variáveis regionalizadas para resolução de problemas de estimativa.

**TEORIA DO POTENCIAL OU HARMÔNICA:** É o estudo das funções harmônicas ou potenciais. Estas funções são soluções de equações diferenciais parciais denominadas de equações de Laplace. Estas funções são utilizadas no estudo da eletricidade, magnetismo, gravitação e fluxos de fluidos. Em inglês potential theory or harmonic analysis.

**TEORIA DOS JOGOS:** A área da matemática que trata, geralmente, da teoria da competição de dois ou mais jogadores sob um conjunto específico de regras. Em inglês Game theory.

**TEORIA DOS NÚMEROS:** É o estudo das propriedades dos números naturais. Ela inclui tópicos como o estudo dos números primos que abrange o teorema do número primo, reciprocidade quadrática, formas quadráticas, aproximação Diofantina, equação Diofantina, campos de números algébricos, o último teorema de Fermat e os métodos para provar estas afirmações. Em inglês number theory.

**TEORIA E MODELO:** Teoria é uma explicação sistemática e generalizada de uma parte da realidade cientificamente aceitável; um modelo é uma representação simplificada da realidade em vez de uma ampla explicação. Teoria é generalizada a todos os casos; modelos cobrem poucas situações específicas. Seria muito bom que pudéssemos basear nossos modelos em teoria, mas é mais provável que teorias úteis surgiram se primeiro formos capazes de construir modelos de comportamento sobre o qual estamos interessados em teorizar.

**TEORIA FREQUENTISTA DE PROBABILIDADE:** A teoria frequentista de probabilidade considera a probabilidade de um evento como o limite da frequência de ocorrência desse evento em uma série de  $n$  tentativas quando  $n$  tende ao infinito. Em inglês Frequency theory of probability. Conhecida também como conceito ou definição empírica ou a posteriori de probabilidade.

**TEORIA FREQUENTISTA:** Teoria estatística baseada na abordagem frequentista. Em inglês Frequentist theory.

**TEORIA GERAL:** Idealmente busca abranger todo ou parte substancial do fenômeno de um dado campo do discurso. É formada como uma pirâmide: na sua base estão os fatos específicos, com abstrações explicando esses fatos na área superior imediata, e generalizações cada vez mais abrangentes à proporção que se atinge o ápice.

**TEORIAS:** São meios eficientes e econômicos que servem para descrever, codificar, resumir, integrar e armazenar informações. Consistem de conceitos, proposições e leis. Permitem um entendimento mais profundo da significância dos fenômenos, antecipa relacionamentos não conhecidos e servem como guias à pesquisa. Boas teorias devem favorecer explicações mais parcimoniosas dos fatos empíricos descobertos do que qualquer outra competitiva. Validações empíricas delas devem ser viáveis.

**TEÓRICO:** Que faz parte da teoria. Sem nenhum aspecto prático real. Explicação puramente teórica.

**TESE:** É um trabalho semelhante a Dissertação, distinguindo-se pela efetiva contribuição na solução de problemas, e para o avanço científico na área em que o tema for tratado.

**TECNOLOGIA:** Termo utilizado para designar o conhecimento técnico e científico bem como suas aplicações por meio do uso de ferramentas, processos e materiais criados, utilizados ou ambos a partir de tal conhecimento. Refere-se também a um conjunto ou classes de técnicas.

**TESTE BICAUDAL (TWO-TAILED TEST):** O teste da hipótese alternativa estatística diz que a estatística do teste observada não é igual ao valor esperado sob a hipótese nula estatística. As tabelas estatísticas e os programas estatísticos geralmente proveem valores críticos e de  $P$  associados para os testes bicaudais. Comparar com teste unicaudal.

**TESTE DE BONDADE DO AJUSTE (GOODNESS-OF-FIT TEST):** Teste estatístico usado para determinar a bondade do ajuste. Ou quão bem representa um modelo de distribuição teórica de probabilidade uma amostra de valores ou uma distribuição de frequência empírica de dados. Por exemplo, os testes não paramétricos do qui quadrado, Kolmogorov Smirnov, Shapiro Wilk, etc.

**TESTE DE RANDOMIZAÇÃO (RANDOMIZATION TEST):** Testes estatísticos que recaem no remanejamento dos dados, usando bootstrap, jackknife ou outra estratégia de reamostragem. O remanejamento simula os dados que seriam coletados caso a hipótese nula estatística fosse verdadeira. Ver também métodos de Monte Carlo, bootstrap e jackknife.

**TESTE DE RAZÃO DE VEROSSIMILHANÇAS (LIKELIHOOD RATIO TEST):** Tipo de prova ou método para realizar inferência estatística através do procedimento de testar diferentes hipóteses com base na razão de suas relativas probabilidades.

**TABELA DE CONTINGÊNCIA:** Um arranjo tabular expressando a designação de membros de um conjunto de dados de acordo com duas ou mais categorias ou critérios de classificação. Como por exemplo, tabelas  $h$  por  $k$  com  $h$  linhas e  $k$  colunas, tabelas  $2 \times 2$ ,  $3 \times 4$ ,  $5 \times 6$ , etc.

**TAMANHO DA AMOSTRA:** É o número de observações ou variáveis aleatórias em uma amostra com  $n$  elementos.

**TÉCNICAS DE CONTAGEM:** Fórmulas usadas para determinar o número de elementos em espaços amostrais e eventos. Conhecida também como análise combinatória.



**TABELA DE CONTIGÊNCIA:** Uma tabela de frequências que mostra como os indivíduos são classificados em diferentes categorias de dois ou mais fatores. Conhecida também como tabela de dependência.

**TABELAS:** Um arranjo ordenado, normalmente de números ou palavras em fileiras e colunas, que exhibe um conjunto de fatos de maneira distinta e compreensiva.

**TAMANHO DA AMOSTRA:** O número de animais incluídos em uma investigação quando um subgrupo de uma população é estudado.

**TAU DE KENDALL:** Um coeficiente de correlação não paramétrico.

**TAXA:** Número de novos eventos por unidade de investigação, por unidade de tempo.

**TAXA RELATIVA:** Consultar proporção da taxa de incidência, cuja sigla PTI.

**TEMPO DE SOBREVIVÊNCIA MEDIANO:** Corresponde à probabilidade de sobrevivência de 0,5 na análise de sobrevivência.

**TENDÊNCIA:** Distorção sistemática dos dados.

**TENDÊNCIA CENTRAL:** Indica a posição que representa o meio de um grupo de observações.

**TENDÊNCIAS DE PUBLICAÇÃO:** A tendência de revistas aceitarem apenas resultados significativos para publicação e, assim, dar pesos inadequados aos efeitos de tratamento positivos da metanálise.

**TEOREMA DE BAYES:** É a base da inferência Bayesiana, a qual utiliza evidencia recente de um estudo para atualizar a probabilidade prévia de um resultado específico, como por exemplo, doença, a fim de obter a probabilidade posterior do resultado.

**TESTE ALEATÓRIO CONTROLADO (TAC):** Um teste clínico que inclui pelo menos um grupo controle e que utiliza processos aleatórios para distribuir os animais nos diferentes grupos de tratamento e no grupo controle.

**TESTE CEGO:** Em um teste clínico, o tratador de animais e/ou o pesquisador não tem conhecimento dos tratamentos específicos que os animais recebem. Também determinado teste mascarado.

**TESTE CEGO-ÚNICO:** Somente uma das partes, o tratador ou o assistente não sabe, ou seja, é cego quanto ao tratamento que o animal recebe em um teste clínico. Também chamado de teste mascarado único.

**TESTE CLÍNICO A CAMPO:** Um estudo comparativo envolvendo novos tratamentos ou medidas preventivas utilizadas em condições naturais, de campo ou semicampo.

**TESTE CLÍNICO CONTROLADO:** Um teste clínico que inclui um grupo controle.

**TESTE CRUZADO:** Tipo de delineamento onde dois ou mais tratamentos são utilizados em ordem aleatória para cada animal, indivíduo ou unidade experimental individual. O objetivo é examinar os efeitos dos tratamentos dentro dos animais, ou parcelas experimentais ao invés de entre os animais, ou unidades experimentais aumentando a precisão da estimativa da diferença entre tratamentos.

**TESTE DA AMPLITUDE MÚLTIPLA DE DUNCAN:** Um teste de comparações múltiplas das médias em três ou mais grupos independentes.

**TESTE DA SOMA DOS POSTOS DE WILCOXON:** Um teste não paramétrico utilizado para comparar as distribuições dos dados em dois grupos independentes. Gera o mesmo valor P que o teste de Mann-Whitney U e pode ser utilizado como uma alternativa não paramétrica para o teste t com duas amostras.

**TESTE DE DOIS LADOS:** Um teste de hipótese em que o valor de P é determinado relacionando o teste estatístico a ambas as caudas de uma distribuição teórica. A decisão a priori de fazer um teste de duas caudas depende da especificação da hipótese alternativa, que não indica a direção do efeito de tratamento.

**TESTE DE DUAS CAUDAS:** Consultar teste de dois lados.

**TESTE DE HIPÓTESE:** O procedimento no qual se formula e testa uma hipótese acerca da população, utilizando os dados da amostra.

**TESTE DE HIPÓTESES:** São processos que nos permitem decidir se devemos aceitar ou rejeitar um hipótese.

**TESTE DE HIPÓTESE PARA A MÉDIA (  $\mu$  ) DE UMA POPULAÇÃO INFINITA OU FINITA E AMOSTRAGEM COM REPOSIÇÃO, NORMAL, COM VARIÂNCIA (  $\sigma^2$  ) CONHECIDA [AMOSTRA GRANDE (  $n \geq 30$  ) OU PEQUENA (  $n < 30$  )] OU VARIÂNCIA (  $\sigma^2$  ) DESCONHECIDA [AMOSTRA GRANDE (  $n \geq 30$  )]:** É o teste de hipótese conduzido conforme o quadro a seguir:

HIPÓTESE DE NULIDADE $H_0$	ESTATÍSTICA TESTE	HIPÓTESE ALTERNATIVA $H_1$	REGIÃO CRÍTICA OU REGIÃO DE REJEIÇÃO DA HIPÓTESE $H_0$ [RRH <sub>0</sub> ]
$\mu = \mu_0$	$Z_{TESTE} = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$  Com $\sigma$ conhecido ou não	$\mu < \mu_0$ $\mu > \mu_0$ $\mu \neq \mu_0$	$Z \leq -Z_{(\alpha)}$ $Z \geq Z_{(\alpha)}$ $Z \leq -Z_{\left(\frac{\alpha}{2}\right)}$ $Z \geq Z_{\left(\frac{\alpha}{2}\right)}$

**TESTE DE HIPÓTESE PARA A MÉDIA (  $\mu$  ) DE UMA POPULAÇÃO INFINITA OU FINITA E AMOSTRAGEM COM REPOSIÇÃO, APROXIMADAMENTE NORMAL COM VARIÂNCIA (  $\sigma^2$  ) DESCONHECIDA E AMOSTRA PEQUENA (  $n < 30$  ):** É o teste de hipótese conduzido conforme o quadro a seguir:

HIPÓTESE DE NULIDADE $H_0$	ESTATÍSTICA TESTE	HIPÓTESE ALTERNATIVA $H_1$	REGIÃO CRÍTICA OU REGIÃO DE REJEIÇÃO DA HIPÓTESE $H_0$ [RRH <sub>0</sub> ]
$\mu = \mu_0$	$t_{TESTE} = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}}$ , com $V = n - 1 G.L.$ ,  Com $\sigma$ desconhecido	$\mu < \mu_0$ $\mu > \mu_0$ $\mu \neq \mu_0$	$t \leq -t_{(v;\alpha)}$ $t \geq t_{(v;\alpha)}$ $t \leq -t_{\left(v;\frac{\alpha}{2}\right)}$ $t \geq t_{\left(v;\frac{\alpha}{2}\right)}$



**TESTE DE HIPÓTESE PARA A MÉDIA DAS DIFERENÇAS ( $\mu_D$ ) DE DUAS POPULAÇÕES QUE NÃO SÃO INDEPENDENTES, ISTO É, AS VARIÁVEIS SÃO EMPARELHADAS (DADOS EMPARELHADOS OU PAREADOS) E COM A AMOSTRA DE DIFERENÇAS (n) PEQUENA (n < 30):**

É o teste de hipótese conduzido conforme descrito a seguir. Os resultados das duas amostras constituem dados emparelhados ou pareados, quando estão relacionados dois a dois segundo algum critério que introduz uma influência marcante entre os diversos pares, que supomos, porém, influir igualmente sobre os valores de cada par. Ou seja, a característica peculiar das amostras pareadas é que, para cada observação do primeiro grupo, há uma correspondente no segundo grupo. Na técnica conhecida como auto-emparelhamento, as medidas são tomadas em um único indivíduo, em dois pontos distintos no tempo. Um exemplo comum de auto-emparelhamento é o experimento antes depois, no qual cada indivíduo é examinado antes que um determinado tratamento seja aplicado e depois que o tratamento for completado. Um segundo tipo de emparelhamento ocorre quando um investigador casa os indivíduos de um grupo com os de um segundo grupo, de modo que os membros de um par sejam tão parecidos quanto possível em relação a características relevantes, tais como idade, sexo, etc. O emparelhamento é frequentemente empregado na tentativa de controlar as fontes estranhas de variação que poderiam, de outra maneira, influenciar os resultados da comparação. Se as mediadas são feitas no mesmo indivíduo ou elemento e não em dois elementos diferentes, uma certa quantidade de variabilidade biológica por exemplo é eliminada. Não temos de nos preocupar com o fato de um indivíduo ser mais velho do que o outro ou se um é homem e o outro mulher. A intenção do emparelhamento, portanto, é fazer uma comparação mais precisa. Pode-se diferenciar três tipos de pareamento, os quais são: auto-pareamento, pareamento natural e pareamento artificial. O auto pareamento ocorre quando o indivíduo serve como seu próprio controle, como na situação em que um indivíduo recebe duas drogas administradas em ocasiões diferentes. Outra situação, por exemplo, é aquela em que se mede o nível de colesterol de uma pessoa antes e depois de um programa de dieta ou tratamento. Finalmente, a comparação de dois órgãos no mesmo indivíduo, como braços, pernas, olhos, narinas, segundo alguma característica estudada também constitui um auto-pareamento. O pareamento natural consiste em formar pares tão homogêneos quanto possível, controlando os fatores que possam interferir na resposta, sendo que o pareamento aparece de forma natural. Por exemplo, em experimentos de laboratório pode-se formar pares de cobaias selecionadas da mesma ninhada; em investigações clínicas, gêmeos univitelinos são muito usados. No pareamento artificial escolhem-se indivíduos com características semelhantes, tais como, idade, sexo, nível socioeconômico, estado de saúde ou, em geral, fatores que podem influenciar de maneira relevante a variável resposta. Na prática os pesquisadores em geral podem encontrar dificuldades no conhecimento das características que devem ser controladas e mesmo conhecendo-se pode ser difícil formar pares homogêneos, como por exemplo, no caso em que o número de fatores é muito grande. Em muitas situações, embora desejável, torna-se difícil ou mesmo impossível a implementação do planejamento com amostras pareadas ou emparelhadas. É óbvio que, se os dados das duas amostras estão emparelhados tem sentido calcularmos as diferenças ( $d_i$ ) correspondentes a cada par de valores, reduzindo assim os dados a uma única amostra de n diferenças. Por outro lado, testar a hipótese de que a diferença entre as médias das duas populações emparelhadas seja igual a um certo valor  $d_0$ , equivale a testar a hipótese de que a média de todas as diferenças referentes às populações seja igual a  $d_0$ , o que decorre diretamente

das propriedades da média ou seja, vamos testar simplesmente a hipótese  $H_0 : \mu_d = d_0$ , contra uma hipótese alternativa  $H_1$  que poderá corresponder a um teste unilateral ou bilateral. Os exemplos 1 e 2 ilustram situações em que os dados obtidos de duas amostras são correlacionados, pareados ou ainda emparelhados. i) Exemplo 1: Quando certo caráter é medido no mesmo indivíduo, em épocas diferentes, os valores obtidos nas duas mensurações tendem a ser mais parecidos entre si do que se houvessem sido obtidos de indivíduos diferentes. Como por exemplo, a medição de taxas de crescimento de plantas de uma determinada cultura, antes e depois de se aplicar uma substância inibidora da fotossíntese. ii) Exemplo 2: As eficiências de duas rações podem ser comparadas utilizando-se vários pares de animais irmãos de uma mesma leitegada tais como suínos, por exemplo, ou de uma mesma ninhada como camundongos, por exemplo. O experimento consiste em alimentar cada membro, de cada par, com uma das rações alocada ao acaso. Indivíduos destes tipos que são irmãos germanos pelas suas semelhanças genéticas, tendem a apresentar respostas correlacionadas aos estímulos a que são submetidos. O teste de hipótese é o descrito no quadro a seguir:

HIPÓTESE DE NULIDADE $H_0$	ESTATÍSTICA TESTE	HIPÓTESE ALTERNATIVA $H_1$	REGIÃO CRÍTICA OU REGIÃO DE REJEIÇÃO DA HIPÓTESE $H_0$ [RRH <sub>0</sub> ] REJEIÇÃO DA HIPÓTESE $H_0$ [RRH <sub>0</sub> ]
$\mu = \mu_D$	$t_{TESTE} = \frac{\bar{X} - \mu_D}{\frac{S}{\sqrt{n}}}$ , com $V = n - 1$ G.L.  Com $\sigma$ desconhecido	$\mu < \mu_D$ $\mu > \mu_D$ $\mu \neq \mu_D$	$t \leq -t_{(v;\alpha)}$ $t \geq t_{(v;\alpha)}$ $t \leq -t_{(v;\frac{\alpha}{2})}$ $t \geq t_{(v;\frac{\alpha}{2})}$

**TESTE DE HIPÓTESE PARA A DIFERENÇA ENTRE AS MÉDIAS ( $\mu_1 - \mu_2$ ) DE DUAS POPULAÇÕES INFINITAS, OU FINITAS E AMOSTAGEM COM REPOSIÇÃO, COM VARIÂNCIAS ( $\sigma_1^2$  e  $\sigma_2^2$ ) CONHECIDAS [AMOSTRAS GRANDES ( $n_1 \geq 30$  E  $n_2 \geq 30$ ) OU PEQUENAS ( $n_1 < 30$  E  $n_2 < 30$ )] OU COM VARIÂNCIAS DESCONHECIDAS [AMOSTRAS GRANDES ( $n_1 \geq 30$  E  $n_2 \geq 30$ )]:** É o teste de hipótese conduzido conforme o quadro a seguir:

HIPÓTESE DE NULIDADE $H_0$	ESTATÍSTICA TESTE	HIPÓTESE ALTERNATIVA $H_1$	REGIÃO CRÍTICA OU REGIÃO DE REJEIÇÃO DA HIPÓTESE $H_0$ [RRH <sub>0</sub> ]
$\mu_1 = \mu_2$	$Z_{TESTE} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ $Z_{TESTE} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}},$ <p><math>\sigma_1</math> e <math>\sigma_2</math> conhecidos.</p> <p>Ou não conhecidos.</p>	$\mu_1 < \mu_2$ $\mu_1 > \mu_2$ $\mu_1 \neq \mu_2$	$Z \leq -Z_{(\alpha)}$ $Z \geq Z_{(\alpha)}$ $Z \leq -Z_{\left(\frac{\alpha}{2}\right)}$ $Z \geq Z_{\left(\frac{\alpha}{2}\right)}$

TESTE DE HIPÓTESE PARA A DIFERENÇA ENTRE AS MÉDIAS ( $\mu_1 - \mu_2$ ) DE DUAS POPULAÇÕES INFINITAS, OU FINITAS E AMOSTRAGEM COM REPOSIÇÃO APROXIMADAMENTE NORMAIS, COM VARIÂNCIAS ( $\sigma_1^2$  e  $\sigma_2^2$ ) DESCONHECIDAS E ESTATISTICAMENTE IGUAIS ( $\sigma_1^2 = \sigma_2^2 = \sigma^2$ ) OU SEJA, POPULAÇÕES HOMOCEDEÁSTICAS E AMOSTRAS PEQUENAS ( $n_1 < 30$  E  $n_2 < 30$ ): É o teste de hipótese conduzido conforme o quadro a seguir:

HIPÓTESE DE NULIDADE $H_0$	ESTATÍSTICA TESTE	HIPÓTESE ALTERNATIVA $H_1$	REGIÃO CRÍTICA OU REGIÃO DE REJEIÇÃO DA HIPÓTESE $H_0$ [RRH <sub>0</sub> ]
$\mu_1 = \mu_2$	$t_{TESTE} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \text{ com } V = n_1 + n_2 - 2, \text{ e}$ <p><math>\sigma_1 = \sigma_2</math> mas desconhecidos, sendo</p> $S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}, \text{ e } S_p = \sqrt{S_p^2}$	$\mu_1 < \mu_2$ $\mu_1 > \mu_2$ $\mu_1 \neq \mu_2$	$t \leq -t_{(v;\alpha)}$ $t \geq t_{(v;\alpha)}$ $t \leq -t_{\left(v;\frac{\alpha}{2}\right)}$ $t \geq t_{\left(v;\frac{\alpha}{2}\right)}$

TESTE DE HIPÓTESE PARA A DIFERENÇA ENTRE AS MÉDIAS ( $\mu_1 - \mu_2$ ) DE DUAS POPULAÇÕES INFINITAS, APROXIMADAMENTE NORMAIS, COM VARIÂNCIAS ( $\sigma_1^2$  e  $\sigma_2^2$ ) DESCONHECIDAS E ESTATISTICAMENTE DIFERENTES ( $\sigma_1^2 \neq \sigma_2^2$ ) OU SEJA, POPULAÇÕES HETEROCEDEÁSTICAS E COM TAMANHOS DE AMOSTRAS PEQUENAS ( $n_1 < 30$  E  $n_2 < 30$ ): É o teste de hipótese conduzido conforme o quadro a seguir:

HIPÓTESE DE NULIDADE $H_0$	ESTATÍSTICA TESTE	HIPÓTESE ALTERNATIVA $H_1$	REGIÃO CRÍTICA OU REGIÃO DE REJEIÇÃO DA HIPÓTESE $H_0$ [RRH <sub>0</sub> ]
$\mu_1 = \mu_2$	$t_{TESTE} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}, \text{ com}$ $V = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 - 1}}$ <p><math>\sigma_1 \neq \sigma_2</math> e desconhecidos.</p>	$\mu_1 < \mu_2$ $\mu_1 > \mu_2$ $\mu_1 \neq \mu_2$	$t \leq -t_{(v;\alpha)}$ $t \geq t_{(v;\alpha)}$ $t \leq -t_{\left(v;\frac{\alpha}{2}\right)}$ $t \geq t_{\left(v;\frac{\alpha}{2}\right)}$

O número de graus de liberdade corrigido pode ser calculado através da seguinte fórmula:

$$V = \frac{\left[\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right]^2}{\left\{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 + 1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 + 1}\right\}} - 2$$

TESTE DE HIPÓTESE PARA A VARIÂNCIA  $\sigma^2$  DE UMA POPULAÇÃO NORMAL:

É o teste de hipótese conduzido conforme o quadro a seguir:

HIPÓTESE DE NULIDADE $H_0$	ESTATÍSTICA TESTE	HIPÓTESE ALTERNATIVA $H_1$	REGIÃO CRÍTICA OU REGIÃO DE REJEIÇÃO DA HIPÓTESE $H_0$ [RRH <sub>0</sub> ]
$\sigma^2 = \sigma_0^2$	$\chi^2_{TESTE} = \frac{(n-1)S^2}{\sigma_0^2}$ <p>Com <math>V = n - 1</math></p>	$\sigma^2 < \sigma_0^2$ $\sigma^2 > \sigma_0^2$ $\sigma^2 \neq \sigma_0^2$	$\chi^2 \leq \chi^2_{(1-\alpha;v)}$ $\chi^2 \geq \chi^2_{(\alpha;v)}$ $\chi^2 \leq \chi^2_{\left(1-\frac{\alpha}{2};v\right)}$ e $\chi^2 \geq \chi^2_{\left(\frac{\alpha}{2};v\right)}$

TESTE DE HIPÓTESE PARA A RAZÃO  $\left(\frac{\sigma_1^2}{\sigma_2^2}\right)$  ENTRE AS VARIÂNCIAS ( $\sigma_1^2$  e  $\sigma_2^2$ ), DE DUAS

POPULAÇÕES NORMAIS: É o teste de hipótese conduzido conforme o quadro a seguir:

HIPÓTESE DE NULIDADE $H_0$	ESTATÍSTICA TESTE	HIPÓTESE ALTERNATIVA $H_1$	REGIÃO CRÍTICA OU REGIÃO DE REJEIÇÃO DA HIPÓTESE $H_0$ [RRH <sub>0</sub> ]
$\sigma_1^2 = \sigma_2^2$	$f_{TESTE} = \frac{\sigma_2^2}{\sigma_1^2} \cdot \frac{S_1^2}{S_2^2}$  Com $V_1 = n_1 - 1$ e $V_2 = n_2 - 1$	$\sigma_1^2 < \sigma_2^2$ $\sigma_1^2 > \sigma_2^2$ $\sigma_1^2 \neq \sigma_2^2$	$f \leq f_{(1-\alpha; V_1, V_2)}$ $f \geq f_{(\alpha; V_1, V_2)}$ $f \leq f_{(1-\frac{\alpha}{2}; V_1, V_2)}$ e $f \geq f_{(\frac{\alpha}{2}; V_1, V_2)}$

**TESTE DE HIPÓTESE PARA A PROPORÇÃO (P) DE UMA POPULAÇÃO INFINITA, NORMAL, E COM AMOSTRA GRANDE ( $n \geq 30$ ):** É o teste de hipótese conduzido conforme o quadro a seguir:

HIPÓTESE DE NULIDADE $H_0$	ESTATÍSTICA TESTE	HIPÓTESE ALTERNATIVA $H_1$	REGIÃO CRÍTICA OU REGIÃO DE REJEIÇÃO DA HIPÓTESE $H_0$ [RRH <sub>0</sub> ]
$P = P_0$	$Z_{TESTE} = \frac{\hat{P} - P_0}{\sqrt{\frac{P_0 q_0}{n}}}$  Com população infinita, e $n\hat{P} > 5$ , $n\hat{q} > 5$ , $n \geq 30$ , $\hat{P} = \frac{X}{n}$ e $q_0 = 1 - P_0$	$P < P_0$ $P > P_0$ $P \neq P_0$	$Z \leq -Z_{(\alpha)}$ $Z \geq Z_{(\alpha)}$ $Z \leq -Z_{(\frac{\alpha}{2})}$ e $Z \geq Z_{(\frac{\alpha}{2})}$

**TESTE DE HIPÓTESE PARA A DIFERENÇA ENTRE AS PROPORÇÕES ( $P_1 - P_2$ ), DE DUAS POPULAÇÕES INFINITAS, NORMAIS, COM AMOSTRAS GRANDES ( $n_1 \geq 30$  e  $n_2 \geq 30$ ), QUANDO A HIPÓTESE DE NULIDADE ( $H_0$ ) DA DIFERENÇA ENTRE AS PROPORÇÕES, SE REFERE A UM VALOR DIFERENTE DE ZERO:** É o teste de hipótese conduzido conforme o quadro a seguir:

HIPÓTESE DE NULIDADE $H_0$	ESTATÍSTICA TESTE	HIPÓTESE ALTERNATIVA $H_1$	REGIÃO CRÍTICA OU REGIÃO DE REJEIÇÃO DA HIPÓTESE $H_0$ [RRH <sub>0</sub> ]
$P_1 - P_2 = P_0$ $P_0 \neq 0$	$Z_{TESTE} = \frac{(\hat{P}_1 - \hat{P}_2) - (P_1 - P_2)}{\sqrt{\frac{\hat{P}_1 \hat{q}_1}{n_1} + \frac{\hat{P}_2 \hat{q}_2}{n_2}}}$  Com populações infinitas e $n_1 \hat{P}_1 > 5; n_2 \hat{P}_2 > 5$ ; $n_1 \hat{q}_1 > 5; n_2 \hat{q}_2 > 5$ ; $n_1 \geq 30; n_2 \geq 30$	$P_1 < P_2$ $P_1 > P_2$ $P_1 \neq P_2$	$Z \leq -Z_{(\alpha)}$ $Z \geq Z_{(\alpha)}$ $Z \leq -Z_{(\frac{\alpha}{2})}$ e $Z \geq Z_{(\frac{\alpha}{2})}$

**TESTE DE HIPÓTESE PARA A DIFERENÇA ENTRE AS PROPORÇÕES ( $P_1 - P_2$ ), DE DUAS POPULAÇÕES INFINITAS, NORMAIS, COM AMOSTRAS GRANDES ( $n_1 \geq 30$  e  $n_2 \geq 30$ ), QUANDO A HIPÓTESE DE NULIDADE ( $H_0$ ) DA DIFERENÇA ENTRE AS PROPORÇÕES, SE REFERE A UM VALOR IGUAL A ZERO:** É o teste de hipótese conduzido conforme o quadro a seguir:

HIPÓTESE DE NULIDADE $H_0$	ESTATÍSTICA TESTE	HIPÓTESE ALTERNATIVA $H_1$	REGIÃO CRÍTICA OU REGIÃO DE REJEIÇÃO DA HIPÓTESE $H_0$ [RRH <sub>0</sub> ]
$P_1 - P_2 = P_0$ $P_0 = 0$	$Z_{TESTE} = \frac{(\hat{P}_1 - \hat{P}_2) - (P_1 - P_2)}{\sqrt{\frac{P' q'}{n_1} + \frac{P' q'}{n_2}}}$  $P' = \frac{n_1 \hat{P}_1 + n_2 \hat{P}_2}{n_1 + n_2}$ e $q' = 1 - P'$	$P_1 < P_2$ $P_1 > P_2$ $P_1 \neq P_2$	$Z \leq -Z_{(\alpha)}$ $Z \geq Z_{(\alpha)}$ $Z \leq -Z_{(\frac{\alpha}{2})}$ e $Z \geq Z_{(\frac{\alpha}{2})}$

Vale salientar que se a população for finita e a amostragem for sem reposição (ASR) é necessário e deve-se utilizar o fator de correção para população finita [FCPF], mostrado a seguir, empregando sua fórmula, o qual deve ser multiplicado pelo erro padrão do estimador usado no teste. Por exemplo, se o estimador for a média da amostra, cujo erro padrão é dado por,  $S_{(\bar{x})} = \frac{S}{\sqrt{n}}$ , neste caso tem-se que usar a multipli-

cação do fator  $FCPF = \frac{(N-n)}{(N-1)}$  pelo erro padrão deste estimador, antes de aplicar o teste, ou seja,  $S_{(\bar{x})} = \frac{S}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$ .

**TESE DA IRRELEVÂNCIA DAS HIPÓTESES:** Princípio defendido pelo notável economista norte americano Friedman (1953). Para ele a irrealidade das premissas é uma vantagem positiva e para ser importante uma hipótese deve ser descritivamente falsa em suas premissas.

**TRATAMENTO:** É o método, elemento ou material cujo efeito desejamos medir ou comparar em um experimento. Por exemplo: variedades de cana-de-açúcar, adubação para a cultura do milho e muitos outros, também chamado de variável independente.

**TRATAMENTO TESTEMUNHA OU CONTROLE (PLACEBO):** É um tratamento no qual o pesquisador não está particularmente interessado, mas que pode ser necessário para revelar por comparação se os outros tratamentos são efetivos, ou não efetivos. A inclusão de testemunha permite uma divisão das unidades experimentais em dois grupos, e conseqüentemente tornar possível se detectar diferenças nas respostas.

**TRATAMENTO QUALITATIVO:** É aquele onde cada nível se encerra em si mesmo. Os fatores qualitativos podem ser nominais ou ordinais. Os nominais não guardam entre si qualquer relação de grandeza, isto é, por exemplo, sexo, raça, cultura, dentre outros. Enquanto que os ordinais, como o próprio nome diz, puderam ser ordenados, por exemplo: baixa e média infestação ou infestação severa e muito severa de determinada doença. Em todos estes casos mantém-se sempre o aspecto de qualidade e não existem valores intermediários entre os níveis. Como por exemplo podemos citar tipos de adubos: adubo orgânico, adubo mineral.

**TESTE UNICAUDAL (ONE-TAILED TEST):** Um teste da hipótese estatística alternativa de que os valores observados da estatística teste é maior ou menor e não ambos que o valor esperado sob a hipótese nula estatística. Como as tabelas de estatística normalmente dão o resultado para um teste bicaudal, o valor da probabilidade unicaudal pode ser encontrado dividindo a probabilidade bicaudal por dois. Comparar com teste bicaudal.

**TOLERÂNCIA (TOLERANCE):** Um critério para decidir quando incluir ou não variáveis ou parâmetros em um procedimento de regressão gradativa. A tolerância reduz a multicolinearidade entre variáveis candidatas.

**TOTAL DA COLUNA (COLUMN TOTAL):** A soma marginal dos valores de uma coluna em uma tabela de contingência. Ver também total da linha; total geral.

**TOTAL DA LINHA (ROW TOTAL):** A soma marginal dos valores de uma única linha de uma tabela de contingência. Ver também total da coluna e total geral.

**TOTAL MARGINAL (MARGINAL TOTAL):** A soma das frequências das linhas ou colunas de uma tabela de contingência.

**TRAÇO (TRACE):** A soma dos elementos da diagonal principal de uma matriz quadrada.

**TRATAMENTO (FATOR) QUANTITATIVO:** É aquele que é expresso na forma de intervalo e os seus níveis são simplesmente valores representativos deste intervalo, neste caso dependendo da metodologia da análise utilizada, poderão ser realizadas interpolações. Como por, exemplo, podemos citar: doses de adubo mineral potássico (0,100, 200, 300, 400 e 500 kg/ha).

**TESTE DE HIPÓTESES:** É uma regra de decisão que auxilia o pesquisador a tomar decisões para rejeitar ou não uma hipótese estatística com base nos elementos ou dados amostrais ou do experimento conduzido. Exemplo: o teste F de Fisher-Snedecor, os quais podem ser unilateral e bilateral. Nível de significância do teste de hipótese é a probabilidade de se rejeitar uma hipótese de qualidade quando na verdade ela é verdadeira.

**TESTE DE KOLMOGOROV-SMIRNOV:** Um teste que pode ser utilizado para testar se os dados são provenientes de uma distribuição particular, como por exemplo, a normal ou se os dois grupos de dados são provenientes da mesma distribuição.

**TESTE DE LEVENE:** É um teste paramétrico utilizado para investigar a igualdade da variância em dois ou mais grupos.

**TESTE DE MANN-WHITNEY U:** É um teste não-paramétrico utilizado para comparar dois grupos de observações independentes. Induz ao mesmo resultado que o teste da soma dos postos de Wilcoxon.

**TESTE DE MCNEMAR:** Um teste com base no teste de qui-quadrado que compara duas proporções em dados pareados.

**TESTE DE NEWMAN-KEULS:** Um teste de comparações múltiplas de médias em grupos independentes.

**TESTE DE POSTOS COM SINAIS DE WILCOXON:** Um teste não-paramétrico utilizado para comparar a distribuição dos dados em grupos de observações pareadas relacionadas. Pode ser empregado como uma alternativa não-paramétrica para o teste t pareado.

**TESTE DE PROPORÇÃO DA VARIÂNCIA (TESTE-F):** Um teste paramétrico com base na distribuição-F, utilizado para comparar duas variâncias.

**TESTE DE SCHEFFÉ:** Um teste de comparações múltiplas de médias de três ou mais grupos independentes.

**TESTE DE SINAL:** Um teste não-paramétrico utilizado para analisar os dados em um único grupo, ou aqueles de pares de indivíduos relacionados.

**TESTE DE HARTLEY:** Tipo de teste usado para verificar a homocedasticidade ou homogeneidade de variâncias, também conhecido como teste da razão máxima ou de F máximo. É recomendado e considerado eficaz em ensaios com no máximo 12 tratamentos e com o mesmo número de repetições.

**TESTE DE COCHRAN:** Tipo de teste usado para verificar se os tratamentos comparados em um experimento possuem variâncias homogêneas ou se existe homocedasticidade.

**TESTE DE SUCESSÕES:** Um teste de hipótese em que o valor de P é determinado submetendo ao teste estatístico uma única cauda da distribuição de probabilidade teórica. A decisão a priori para fazer um teste de uma cauda, depende da especificação da hipótese alternativa, a qual deve indicar a direção do efeito de interesse com base na impossibilidade do efeito ocorrer na outra direção.

**TESTE DE BROWN E FORSYTH:** Teste para verificar a igualdade de variâncias. Este teste faz parte das rotinas do programa estatístico Sas System.

**TESTE DE O' BRIEN:** Teste que serve para verificar se a variabilidade entre as observações de um tratamento é semelhante aos dos outros tratamentos.

**TESTE DE SEQUÊNCIAS OU DE ALEATORIEDADE:** Tipo de teste que serve para verificar a independência ou a aleatoriedade dos erros experimentais. Permite também verificar se os dados que compõem uma série tem comportamento aleatório e ainda é usado para identificar tendências ou comportamentos cíclicos.

**TESTE DE UMA CAUDA:** Consultar teste de um lado.



**TESTE DIAGNÓSTICO:** Um procedimento que diferencia animais doentes de saudáveis.

**TESTE ESTATÍSTICO:** Um valor que segue uma distribuição de probabilidade teórica e que representa a base de um teste de hipótese. Ao comparar o valor computado do teste estatístico com a distribuição de probabilidade adequada, determinamos o valor P e decidimos se temos ou não dados suficientes para rejeitar a hipótese nula.

**TRATADO:** Compêndio volumoso e didático que trata de determinado conteúdo do conhecimento humano e com a finalidade de ser uma obra extremamente abrangente.

**TENDÊNCIA:** Um efeito que distorce sistematicamente um resultado estatístico ou uma estimativa, fazendo com que ele(a) não represente a verdadeira grandeza de interesse.

**TENDÊNCIA CENTRAL:** A tendência dos dados em se agrupar em torno de algum valor. Tendência central é geralmente expressa por uma medida de localização, como a média, a mediana ou a moda.

**TENTATIVAS DE BERNOULLI:** Sequências de tentativas independentes com somente dois resultados, geralmente chamados de sucesso e falha, em que a probabilidade de sucesso permanece constante.

**TESTE ESTATÍSTICO DE WALD:** Utilizado para testar a significância a partir de zero de um coeficiente de regressão cuja estimativa é obtida pela máxima verossimilhança; segue uma distribuição Normal.

**TESTE EXATO DE FISHER:** Um teste pelo qual o valor de P é calculado em um teste de hipótese de dados apresentados em uma tabela de contingência; é preferível utilizar a aproximação qui quadrado ao teste estatístico qui quadrado quando as frequências esperadas na tabela são muito baixas.

**TESTE LOG-RANK:** Um teste não paramétrico que compara curvas de sobrevivência.

**TESTE MASCARADO ÚNICO:** Consultar teste cego-único.

**TESTE CRAMER-VON MISES:** Teste usado para verificar se a distribuição dos dados experimentais ou de levantamentos se ajustam a distribuição teórica de probabilidade normal ou gaussiana.

**TESTE DE ANDERSON DARLING:** Teste usado para verificar se a distribuição dos dados se ajustam a distribuição teórica de probabilidade normal ou gaussiana.

**TESTE DE LILLIEFORS:** Teste usado para verificar se a distribuição dos dados, dos erros, resíduos ou desvios seguem a distribuição teórica de probabilidade normal ou gaussiana.

**TESTE DE D'AGOSTINO:** Teste usado para verificar se a distribuição dos dados se ajustam a distribuição teórica de probabilidade normal ou gaussiana.

**TESTE SHAPIRO WILK:** Teste usado para verificar se a distribuição dos dados se ajustam a distribuição teórica de probabilidade normal ou gaussiana.

**TESTE DE ASSIMETRIA:** Teste usado para verificar se a distribuição dos dados se ajustam a distribuição teórica de probabilidade normal ou gaussiana.

**TESTE DE CURTOSE:** Teste usado para verificar se a distribuição dos dados se ajustam a distribuição teórica de probabilidade normal ou gaussiana.

**TESTE NÃO-PARAMÉTRICO:** Normalmente conhecido como teste livres de distribuição, ou seja, métodos que não fazem suposições acerca da distribuição de dados subjacentes.

**TESTE DE ADERÊNCIA DE QUI QUADRADO:** Teste usado para verificar se a distribuição dos dados se ajustam a distribuição teórica de probabilidade normal ou gaussiana.

**TESTE QUI-QUADRADO ( $\chi^2$ ):** Um teste não-paramétrico com base na distribuição qui-quadrado, frequentemente utilizado para comparar proporções.

**TESTE QUI-QUADRADO ( $\chi^2$ ) PARA TENDÊNCIA:** Um teste qui quadrado específico utilizado para determinar se há uma tendência nas proporções classificadas por uma variável ordinal.

**TESTE SHAPIRO-WILK W:** Um teste de normalidade do conjunto de dados.

**TESTE T:** Consultar teste t pareado e teste t não pareado; estes são testes de significância com base na distribuição t de Student.

**TESTE T COM DUAS AMOSTRAS:** Consultar teste t não pareado.

**TESTE T COM UMA AMOSTRA:** Um teste paramétrico com base na distribuição-t utilizado para testar  $H_0$ , que admite que a verdadeira média assume um valor particular.

**TESTE T NÃO PAREADO:** Um teste paramétrico com base na distribuição t de Student para comparar as médias em dois grupos independentes. É também denominado teste t com duas amostras.

**TESTE T PAREADO:** Um teste paramétrico com base na distribuição t de Student, utilizado para analisar os dados pareados de dois grupos.

**TESTE-F:** Consultar teste de proporção da variância (teste-F).

**TESTES CLÍNICOS:** Uma forma de estudo experimental em condições controladas, realizado para avaliar a eficácia de um ou mais tratamentos ou medidas preventivas, quando aplicadas a humanos ou animais.

**TESTE NÃO PARAMÉTRICO DA SOMA DOS POSTOS DE WILCOXON (WILCOXON'S RANK SUM TEST):** Teste de hipótese ou de significância usado na estatística não paramétrica para dados que não seguem uma distribuição teórica de probabilidade bem caracterizada como a normal, ou de outro tipo, isto é, os dados não precisam estar normalmente distribuídos (free distribution), o qual é um método básicos da estatística não paramétrica que têm seu análogo não paramétrico como, por exemplo, é o equivalente não paramétrico do teste t para duas médias em amostras independentes e com menos de 30 elementos. Este método não paramétrico utiliza os postos dos dados em lugar dos dados em si. Os postos são determinados colocando-se os dados em ordem crescente e atribuindo-se posto 1 para o menor valor, posto 2 o segundo menor, etc. As técnicas da estatística não paramétrica são, particularmente, adaptáveis aos dados das ciências do comportamento. A aplicação dessas técnicas não exige suposições quanto à distribuição da variável populacional. Os testes não paramétricos são extremamente interessantes para análises de dados qualitativos. Trata-se de uma extensão do teste dos sinais. É mais interessante pois leva em consideração a magnitude da diferença para cada par. O teste de Wilcoxon exige que a variável em análise seja medida em escala ordinal ou numérica, e a diferença entre duas observações, feitas no mesmo par, também possa ser ordenada. Se existem dados com o mesmo posto ou com valores iguais, o posto dos valores que são iguais é a média dos postos que lhes corresponderiam se não fossem iguais.

**TESTE NÃO PARAMÉTRICO DOS SINAIS DE WILCOXON (WILCOXON'S SIGNED RANK TEST):**

Teste de hipótese ou de significância usado na estatística não paramétrica para dados que não seguem uma distribuição teórica de probabilidade bem caracterizada como a normal, ou de outro tipo, isto é, os dados não precisam estar normalmente distribuídos (free distribution), o qual é um método básico da estatística não paramétrica que têm seu análogo não paramétrico correspondente ao teste t pareado paramétrico, ou teste t de Student para dados emparelhados ou pareados em amostras dependentes com menos de 30 elementos. Este método não paramétrico utiliza os postos dos dados em lugar dos dados em si. Os postos são determinados colocando-se os dados em ordem crescente e atribuindo-se posto 1 para o menor valor, posto 2 o segundo menor, etc. É utilizado para análise de amostras dependentes. Logo, esse teste é uma alternativa para o teste t para amostras dependentes. É aplicado em situações em que o pesquisador deseja determinar se duas condições são diferentes. O teste do sinal tem pouco poder, pois usa como informação apenas o sinal das diferenças entre pares. A única pressuposição exigida pelo teste do sinal é a de que a distribuição da variável seja contínua. Esse teste não faz qualquer suposição sobre a forma da distribuição das diferenças de médias. É útil nos trabalhos de pesquisa em que é impossível ou inviável a obtenção de uma mensuração quantitativa, mas é possível estabelecer postos em relação a cada um dos dois membros de cada par. A lógica do teste é que as condições podem ser consideradas iguais quando as quantidades de + e - forem aproximadamente iguais. Se existem dados com o mesmo posto, ou seja, dados iguais, o posto dos valores que são iguais é a média dos postos que lhes corresponderiam se não fossem iguais. As técnicas da estatística não paramétrica são, particularmente, adaptáveis aos dados das ciências do comportamento. A aplicação dessas técnicas não exige suposições quanto à distribuição da variável populacional. Os testes não paramétricos são extremamente interessantes para análises de dados qualitativos.

**TESTE NÃO PARAMÉTRICO DE KRUSKAL-WALLIS (KRUSKAL-WALLIS TEST):**

Teste de hipótese ou de significância usado na estatística não paramétrica para dados que não seguem uma distribuição teórica de probabilidade bem caracterizada como a normal, ou de outro tipo, isto é, os dados não precisam estar normalmente distribuídos (free distribution), o qual é um método básico da estatística não paramétrica que têm seu análogo paramétrico correspondente ou equivalente a análise de variância (ANOVA) para dados obtidos em ordem completamente aleatória, ou no delineamento inteiramente casualizado. Este método não paramétrico utiliza os postos dos dados em lugar dos dados em si. Os postos são determinados colocando-se os dados em ordem crescente e atribuindo-se posto 1 para o menor valor, posto 2 o segundo menor, etc. Trata-se de teste extremamente útil para decidir se k amostras ( $k > 2$ ) independentes provêm de populações com médias iguais. Esse teste só deve ser aplicado se a amostra for pequena e/ou as pressuposições, exigidas para proceder à Análise de Variância, estiverem seriamente comprometidas. Como o teste de Mann-Whitney, esse teste também condiciona que a variável em análise seja medida em escala ordinal ou numérica. Se existem dados com o mesmo posto, isto é, dados iguais, o posto dos valores que são iguais é a média dos postos que lhes corresponderiam se não fossem iguais. As técnicas da estatística não-paramétrica são, particularmente, adaptáveis aos dados das ciências do comportamento. A aplicação dessas técnicas não exige suposições quanto à distribuição da variável populacional. Os testes não paramétricos são extremamente interessantes para análises de dados qualitativos.

**TESTE NÃO PARAMÉTRICO DE FRIEDMAN-R (FRIEDMAN-R TEST):**

Teste de hipótese ou de significância usado na estatística não paramétrica para dados que não seguem uma distribuição teórica de probabilidade bem caracterizada como a normal, ou de outro tipo, isto é, os dados não precisam estar normalmente distribuídos (free distribution), o qual é um método básico da estatística não paramétrica

que têm seu análogo paramétrico correspondente ou equivalente a dados obtidos com aleatorização por blocos ou no delineamento em blocos completos casualizado. Este método não paramétrico utiliza os postos dos dados em lugar dos dados em si. Os postos são determinados colocando-se os dados em ordem crescente e atribuindo-se posto 1 para o menor valor, posto 2 o segundo menor, etc. Se existem dados com o mesmo posto ou com valores dados iguais, o posto dos valores que são iguais é a média dos postos que lhes corresponderiam se não fossem iguais. As técnicas da estatística não paramétrica são, particularmente, adaptáveis aos dados das ciências do comportamento. A aplicação dessas técnicas não exige suposições quanto à distribuição da variável populacional. Os testes não paramétricos são extremamente interessantes para análises de dados qualitativos.

**TESTE NÃO PARAMÉTRICO DO COEFICIENTE DE CORRELAÇÃO DE SPEARMAN (SPEARMAN RANK COEFFICIENT OF CORRELATION):**

Teste de hipótese ou de significância usado na estatística não paramétrica para dados que não seguem uma distribuição teórica de probabilidade bem caracterizada como a normal, ou de outro tipo, isto é, os dados não precisam estar normalmente distribuídos (free distribution), o qual é um método básico da estatística não paramétrica que têm seu análogo paramétrico correspondente ou equivalente ao teste t de Student para o coeficiente de correlação linear simples de Pearson. Os postos são determinados colocando-se os dados em ordem crescente e atribuindo-se posto 1 para o menor valor, posto 2 o segundo menor, etc. Este método não paramétrico utiliza os postos dos dados em lugar dos dados em si. Os postos são determinados colocando-se os dados em ordem crescente e atribuindo-se posto 1 para o menor valor, posto 2 o segundo menor, etc. Se existem dados com o mesmo posto ou com valores iguais, o posto dos valores que são iguais é a média dos postos que lhes corresponderiam se não fossem iguais. As técnicas da estatística não-paramétrica são, particularmente, adaptáveis aos dados das ciências do comportamento. A aplicação dessas técnicas não exige suposições quanto à distribuição da variável populacional. Os testes não-paramétricos são extremamente interessantes para análises de dados qualitativos.

**TESTES DE SIGNIFICÂNCIA REPETIDOS:** Testes de hipóteses realizados em estágios intermediários de um teste. Consultar também análises interinas.

**TESTES DE SUPERIORIDADE:** Utilizado para mostrar que dois ou mais tratamentos são estatisticamente diferentes.

**TESTES LIVRES DE DISTRIBUIÇÃO:** Consultar teste não paramétrico.

**TESTES PARAMÉTRICOS:** Analisam a hipótese acerca do(s) parâmetro(s) de uma distribuição e fazem suposição acerca da forma subjacente da distribuição das observações; por exemplo, é normal, ou há igualdade das variâncias.

**TRANSFORMAÇÃO:** Um procedimento matemático, como por exemplo, utilizando logaritmo (log) de cada valor no conjunto total de dados na tentativa de formar um novo conjunto de dados que preenche os requisitos particulares da análise; por exemplo, normalidade, relação linear ou variância constante.

**TRANSFORMAÇÃO ANGULAR:** Consultar transformação arco seno.

**TRANSFORMAÇÃO ARCO SENOS:** Uma transformação para uma proporção, p, para  $\text{sen}^{-1}\sqrt{p}$  ou  $\text{arc sen}\sqrt{p}$ . Isso torna a curva sigmoide linear e estabiliza a variância. Também conhecida como transformação angular ou transformação seno inversa.

**TRANSFORMAÇÃO LOGÍSTICA (LOGITO):** É o tipo de transformação,  $\log_e \left\{ \frac{P}{(1-P)} \right\}$ , da proporção, p, utilizada para tornar linear a curva sigmoide.

**TRIAGEM:** Um procedimento de identificação de animais assintomáticos que encontram-se sob risco de determinada doença.

**TEOREMA DO LIMITE CENTRAL:** A forma mais simples do teorema do limite central estabelece que a soma de n variáveis aleatórias distribuídas independentemente tenderá a ser normalmente distribuída quando n se tornar grande. É uma condição necessária e suficiente que nenhuma das variâncias das variáveis aleatórias individuais seja grande em comparação a sua soma. Há mais formas gerais do teorema central que permitem variâncias infinitas e variáveis aleatórias correlacionadas e há uma versão multivariada do teorema.

**TEOREMA DE BAYES:** Uma equação para uma probabilidade condicionada, como P(A|B), em termos da probabilidade condicional reversa P(B|A).

**TESTE COMPLEMENTAR:** Um teste não paramétrico para comparar duas distribuições ou verificar a independência das medidas.

**TESTE DA MÍNIMA DIFERENÇA SIGNIFICATIVA (OU TESTE LSD DE FISHER):** Uma aplicação do teste t para comparar pares de médias seguindo a rejeição da hipótese nula em uma análise de variância. A taxa de erro é difícil de calcular exatamente pelo fato de as comparações não serem todas independentes.

**TESTE DE RAZÃO DE VEROSSIMILHANÇA:** Um teste de uma hipótese nula versus uma hipótese alternativa, usando uma estatística de teste derivada de uma razão de verossimilhança.

**TESTE DE HIPÓTESES:** Qualquer procedimento usado para testar uma hipótese estatística.

**TESTE DE HOMOGENEIDADE:** Em uma tabela de contingência bidimensional r por c, ele testa se as proporções nas c categorias são as mesmas para todas as r populações.

**TESTE DE INDEPENDÊNCIA:** Em uma tabela de contingência bidimensional r por c, ele testa se as categorias nas linhas e nas colunas são independentes.

**TESTE DE KRUSKAL-WALLIS:** Um método de teste não paramétrico usado para a análise de variância como uma alternativa ao teste F.

**TESTE DE WILCOXON DA SOMA DOS POSTOS:** Um teste não paramétrico para a igualdade de médias em duas populações. Ele é algumas vezes chamado de teste de Mann-Whitney.

**TESTE DE WILCOXON DO POSTO SINALIZADO:** Um teste, livre de distribuição, da igualdade de parâmetros de localização de duas distribuições pensadas idênticas a princípio.

**TESTE DO POSTO SINALIZADO:** Um teste estatístico baseado nas diferenças dentro de um conjunto de observações pareadas. Cada diferença tem um sinal e um posto e o teste usa a soma das diferenças com relação ao sinal.

**TESTE DOS SINAIS:** Um teste estatístico baseado nos sinais de certas funções das observações e não nas suas magnitudes.

**TESTE F:** Qualquer teste de significância envolvendo a distribuição F. Os testes F mais comuns são: i) testando hipóteses acerca das variâncias ou dos desvios-padrão de duas distribuições normais independentes; ii) testando hipóteses acerca das médias dos tratamentos ou dos componentes da variância na análise de variância e iii) testando a significância da regressão ou testes nos subconjuntos de parâmetros em um modelo de regressão.

**TESTE QUI-QUADRADO:** Qualquer teste de significância baseado na distribuição qui-quadrado. Os testes qui quadrados mais comuns são: i) testando hipóteses acerca da variância ou desvio-padrão de uma distribuição normal e ii) testando a adequação do ajuste de uma distribuição teórica a dados amostrais.

**TESTE T:** Qualquer teste de significância baseado na distribuição t. Os testes t mais comuns são: i) teste de hipóteses para a média de uma distribuição normal com variância desconhecida; ii) teste de hipóteses para a média de duas distribuições normais e iii) teste de hipóteses para os coeficientes individuais de regressão.

**TESTE T COMBINADO:** Uma hipótese para comparar a média de duas populações com as variâncias consideradas iguais.

**TESTE T<sup>2</sup> DE HOTELLING PARA COMPARAR DIFERENÇA ENTRE MÉDIAS MULTIVARIADAS DE DUAS AMOSTRAS:** Tipo de teste que é uma generalização do teste t de Student univariado para dados multivariados, onde  $\bar{Y}_1$  e  $\bar{Y}_2$  são dois vetores colunas de médias de cada grupo para cada uma das variáveis, C é a matriz de variância e covariância amostral para cada grupo ( $C_1$  e  $C_2$ ),  $C_p$  é a matriz estimativa combinada de covariância dos dois grupos,  $m_1$  e  $m_2$  são os tamanhos amostrais de cada grupo, onde a estatística T<sup>2</sup> de Hotelling é calculada como:

$$C_p = \frac{[(m_1 - 1)C_1 + (m_2 - 1)C_2]}{(m_1 + m_2 - 2)} \quad \text{e} \quad T^2 = \frac{m_1 m_2 (\bar{Y}_1 - \bar{Y}_2)^T C_p^{-1} (\bar{Y}_1 - \bar{Y}_2)}{(m_1 + m_2)}.$$

Seguindo as regras de multiplicação de matrizes, esta fórmula do teste T<sup>2</sup> dá o T<sup>2</sup> de Hotelling como um escalar, ou um único número. Uma transformação linear do T<sup>2</sup>. Produz a conhecida estatística do teste F dada por:  $F = \frac{(m_1 + m_2 - n - 1)T^2}{(m_1 + m_2 - 2)n}$ ,

onde n é o número de variáveis. Sob a hipótese nula de que os vetores de médias das populações de cada grupo são iguais, isto é,  $(\mu_1 = \mu_2)$ , o F está distribuído como uma variável aleatória F, com n graus de liberdade no numerador e  $(m_1 + m_2 - n - 1)$  para o denominador, e assim o teste procede de maneira usual.

**TESTES PARA COMPARAR MÉDIAS MULTIVARIADAS DE MAIS DUAS AMOSTRAS (MANOVA SIMPLES):** Tipo de teste que usado para comparar médias multivariadas para mais de dois grupos, onde H é a matriz de somas dos quadrados e produtos cruzados (SQPC) entre grupos,  $\bar{Y}_k$  é o vetor de médias das amostras no grupo k das q observações naquele grupo,  $\bar{Y}_.$  é o vetor de médias das amostras de todos os grupos tratamentos, E é a matriz somas dos quadrados e produtos cruzados (SQPC) dentro de grupos e T é a matriz de somas dos quadrados e produtos cruzados (SQPC) total:

$$H = q \sum_{k=1}^g (\bar{Y}_k - \bar{Y}_.) (\bar{Y}_k - \bar{Y}_.)^T, \quad \bar{Y}_k = \frac{1}{q} \sum_{l=1}^q y_{kl}, \quad \bar{Y}_. = \frac{1}{kq} \sum_{k=1}^g \sum_{l=1}^q y_{kl}, \quad E = \sum_{k=1}^g \sum_{l=1}^q (y_{kl} - \bar{Y}_k) (y_{kl} - \bar{Y}_k)^T \quad \text{e}$$

$$TE = \sum_{k=1}^g \sum_{l=1}^q (y_{kl} - \bar{Y}_.) (y_{kl} - \bar{Y}_.)^T. \quad \text{Lambda de Wilk} = \Lambda = \frac{|E|}{|E + H|}, \quad \text{onde } | | \text{ denota o determinante de}$$

uma matriz.  $\text{Traço de Pillai} = \sum_{i=1}^s \left( \frac{\lambda_i}{\lambda_i + 1} \right) = \text{traço}[(E + H)^{-1} H]$ , onde s é o menor valor entre qualquer um



dos dois, ou os graus de liberdade que é o número de grupos menos os seja,  $g-1$  ou o número de variáveis  $n$ ,  $\lambda_i$  é o  $i$ -ésimo autovalor de  $E^{-1}H$ , e traço é o traço de uma matriz.

*Traço de hotelling – Lawley* =  $\sum_{i=1}^n \lambda_i = \text{traço}[(E^{-1}H)]$  e *Maior raiz de Roy* =  $\theta = \frac{\lambda_1}{\lambda_1 + 1}$ , onde  $\lambda_1$  é o maior

e primeiro autovalor de  $E^{-1}H$ . Para tamanhos amostrais grandes, o lambda de Wilk, o traço de Hotelling-Lawley e o traço de Pillai convergem para o mesmo valor de  $P$ , embora o traço de Pillai seja o mais suscetível às violações de pressupostos, como a normalidade multivariada. Cada uma dessas quatro estatísticas de teste pode ser transformada em uma estatística  $F$  e testada como na análise de variância univariada (ANOVA). Contudo, seus graus de liberdade não são os mesmos. O lambda de Wilk, o traço de Pillai e o traço de Hotelling-Lawley são calculados a partir de todos os autovalores e, então, têm um tamanho amostral  $nm$  (número de variáveis  $n$  x número de amostras  $m$ ) com  $n(g-1)$  graus de liberdade no numerador ( $g$  é o número de grupos). A maior raiz de Roy usa apenas o primeiro autovalor, então seus graus de liberdade no numerador são apenas  $n$ . Os graus de liberdade para o lambda de Wilk com frequência são fracionários, e a estatística  $F$  associada é apenas uma aproximação. A escolha entre essas estatísticas de teste da análise de variância multivariada (MANOVA) não é crucial. Elas em geral dão resultados muito semelhantes.

**TODAS AS REGRESSÕES (SUBCONJUNTOS) POSSÍVEIS:** Um método de seleção de variáveis para a regressão, que examina todos os subconjuntos possíveis de variáveis regressoras candidatas. Algoritmos computacionais eficientes têm sido desenvolvidos para implementar todas as regressões possíveis.

**TRATAMENTO:** Em planejamento de experimentos, um tratamento é um nível específico de um fator interesse. Assim, se esse fator for a temperatura, os tratamentos são os níveis específicos de temperatura usados no experimento.

**TRATAMENTO:** É a condição imposta a parcela cujo efeito procuramos medir ou comparar em um experimento.

**TESTABILIDADE:** Principal característica de uma boa hipótese. Diz-se que uma hipótese possui esta característica, quando suas explicações, ao serem formuladas em termos operacionais, são compatíveis com fatos conhecidos.

**TESTE:** Prova ou conjunto de provas tendentes a determinar, num indivíduo ou grupo de indivíduos, a presença ou grau de certo traço. Em inglês test.

**TESTE ACELERADO:** Frequentemente, o processo de experimentação, ou manufatura, deve acontecer em um período de tempo muito menor que o ciclo de vida normal do indivíduo estudado. Este tipo de testes é praticado em estudos de confiabilidade e de análise de sobrevivência. Do inglês Accelerated testing.

**TESTE A PRIORI:** Ver comparação planejada.

**TESTE BARTLETT DE ESFERICIDADE:** Teste estatístico da significância geral de todas as correlações em uma matriz de correlação.

**TESTE BICAUDAL:** O mesmo que teste bilateral.

**TESTE BICAUDAL:** Teste cujo objetivo é testar apenas se as médias ou proporções são iguais ou diferentes e não estabelecer qual delas é maior ou menor.

**TESTE BILATERAL (DUAS CAUDAS):** Teste de hipóteses em que a região crítica é dividida entre as áreas extremas direitas e a esquerda da distribuição de probabilidade.

**TESTE BILATERAL:** Um teste é dito bilateral se a região crítica estiver dividida meio a meio entre valores superiores e inferiores. Em inglês two-tailed test.

**TESTE BILATERAL:** Teste de hipótese em que a região crítica consiste de ambas as caudas da distribuição, de modo que a hipótese nula é rejeitada se o valor da estatística de teste é muito grande ou muito pequeno.

**TESTE BILATERAL:** Teste de significância ou estatístico no qual a hipótese alternativa não especifica a direção da diferença a ser detectada, com esta representação:  $\mu_1 \neq \mu_0$  podendo  $\mu_1$  ser maior ou menor que  $\mu_0$ .

**TESTE BILATERAL:** Teste estatístico que opõe, à hipótese da nulidade, a hipótese de que existe diferença entre as populações ou grupos em comparação.

**TESTE BINOMIAL (DUAS PROPORÇÕES):** Teste estatístico baseado na distribuição binomial ou na aproximação à curva normal, onde se comparam as proporções de duas amostras.

**TESTE BINOMIAL (UMA PROPORÇÃO):** Teste baseado na distribuição binomial ou na aproximação à curva normal, onde se compara a proporção de uma amostra com a do parâmetro.

**TESTE C DE COCHRAN:** Em inglês Cochran's C test.

**TESTE DA CORRELAÇÃO DE KENDALL PARA ORDENAÇÕES:** Teste não paramétrico de significância da correlação utilizada para dados ordinais.

**TESTE DA DIFERENÇA ENTRE MÉDIAS:** Um teste de significado estatístico, com a variável dependente numa escala de intervalos e apenas duas categorias da variável independente.

**TESTE DA DIFERENÇA ENTRE PROPORÇÕES:** Um teste de significado estatístico, com a variável dependente medida em porcentagem e apenas duas categorias da variável independente.

**TESTE DA EXCELÊNCIA DO AJUSTE:** Termo geral usado para descrever a comparação entre os dados reais e os resultados preditos por um modelo estatístico em particular, como os valores esperados gerados pela utilização da análise pelo qui quadrado.

**TESTE DA HIPÓTESE DE PROPORÇÕES:** Teste para determinar se a diferença entre proporções é maior do que seria de se esperar por causa de erro de amostragem.

**TESTE DA MEDIANA:** Teste estatístico não-paramétrico de distribuição livre, para duas amostras independentes, no sentido de constatar se provieram de uma população com a mesma mediana. Os dados devem ser mensurados, pelo menos, em escala ordinal.

**TESTE DA PROBABILIDADE EXATA DE FISCHER OU TESTE EXATO DE FISCHER:** Teste de significância estatística que usado para analisar dados de tabelas de contingência  $2 \times 2$ , nas quais um ou mais das contagens esperadas são muito pequenas para satisfazer as condições para o uso da análise pelo qui quadrado.

**TESTE DA SEQUÊNCIA:** Em inglês run test.



**TESTE DA SOMA DE POSTOS DE WILCOXON:** Teste não paramétrico de hipóteses, usado para comparar duas amostras independentes.

**TESTE DA SOMA DE POSTOS DE WILCOXON:** Processo de teste da hipótese de que não há diferença entre as médias de populações.

**TESTE DA SOMA DOS POSTOS DE WILCOXON:** Um nome alternativo para o teste de Mann-Whitney. Em inglês Wilcoxon's rank sum test.

**TESTE DE ADERÊNCIA:** Teste para avaliar quão bem uma distribuição de frequências observadas se ajusta a uma distribuição teórica. Em inglês goodness-of-fit test.

**TESTE DE ADERÊNCIA:** Processo estatístico que envolve a coleta de evidência e a tomada de decisão sobre se determinada hipótese deve ser aceita ou rejeitada.

**TESTE DE ADERÊNCIA:** É aquele em que se observa o ajustamento ou concordância dos escores observados aos valores teóricos esperados ou deduzidos do ponto de vista matemático. É denominado de goodness of fit na literatura inglesa. Os teste G e do qui quadrado, para uma amostra, são exemplos de testes de aderência.

**TESTE DA RAZÃO DE VEROSSIMILHANÇA:** Teste de hipótese ou de significância que compara uma hipótese  $H_0$  contra uma alternativa  $H_1$ , baseado na razão de duas verossimilhanças, cada uma derivada de  $H_0$  e  $H_1$ . Em inglês Likelihood ratio test.

**TESTE DE AMES:** É um teste rápido e usado frequentemente para estimar o potencial mutagênico de uma substância química.

**TESTE DE AMOSTRA ÚNICA:** Testes de significância para dados de intervalo e percentuais comparando os resultados de uma amostra com um padrão estabelecido.

**TESTE DE BARTLETT:** Teste aproximado da homogeneidade do conjunto de variâncias de um número de amostras independentes Normais. Do inglês Bartlett's test.

**TESTE DE BARTLETT:** É usado para verificar a homogeneidade das variâncias residuais entre os tratamentos.

**TESTE DE BEHRENS E FISHER:** Teste de significância sobre a diferença de médias de amostras aleatórias de duas distribuições Normais com variâncias diferentes. Em inglês Behrens-Fisher test.

**TESTE DE BOX:** Teste estatístico para a igualdade de matrizes de variância/covariância das variáveis dependentes ao longo dos grupos. É muito sensível, especialmente à presença de variáveis não normais. Um nível de significância de 0,01 ou menos é usado como ajuste para a sensibilidade da estatística.

**TESTE DE COCHRAN:** Teste estatístico não paramétrico ou de modelo livre de distribuição de probabilidade, onde os dados estão definidos somente com dois valores: 1 (sim=sucesso) e 0 (não=insucesso).

**TESTE DE COEFICIENTES DE VARIAÇÃO:** Teste estatístico da diferença entre dois coeficientes de variação, para verificar se os mesmos provieram da mesma população, proposto por Beiguelman (2002).

**TERMO DE ERRO:** Termo que é necessário para tornar uma equação verdadeira se a predição não for perfeita. O termo de erro é a porção da variação da variável dependente que não é explicada pelo modelo estatístico.

**TESTE DE ESFERICIDADE DE BARTLETT:** Testa a hipótese nula de que a matriz das covariâncias das variáveis dependentes ortonormalizadas é proporcional a uma matriz identidade. Em inglês Bartlett's sphericity test.

**TESTE DE ESFERICIDADE DE MAUCHLY:** Testa a hipótese nula de que a matriz populacional de correlações é uma matriz identidade. Se o valor da estatística qui quadrado obtido é significativo então a matriz de correlações a ser analisada é não aleatória. Neste caso é conveniente utilizar o teste de Humphrey e Ilgen. Em inglês Mauchly sphericity test.

**TESTE DE FRIEDMAN:** Teste estatístico de modelo livre de distribuição de probabilidade abrangendo três (3) ou mais amostras relacionadas, para comparação das respectivas médias, cujos escores são mensurados em escala nominal, ordinal ou em postos. As amostras devem ser do mesmo tamanho. Em inglês Friedman test.

**TESTE DE GABRIEL:** Procedimento de comparação simultânea de valores médios numa análise de variância. Baseada na soma de quadrados entre grupos para os  $2^k - k - 1$  subconjuntos de dois ou mais médias dos k grupos testada contra o valor crítico ou tabelado fixo da distribuição F com (k-1, n-k) graus de liberdade. Em inglês Gabriel's test.

**TESTE DE HIPÓTESE:** Processo de decisão para testar hipóteses estatísticas. Em inglês Hypothesis testing.

**TESTE DE HIPÓTESE:** Determinação de se as expectativas representadas por uma hipótese de fato existem no mundo real.

**TESTE DE HIPÓTESE:** Quando se determinam parâmetros da amostra, como por exemplo a média, é por vezes necessário saber se esse parâmetro é consistente em relação ao parâmetro que se previa para a população. A este procedimento chamamos teste de hipótese. Consideremos o seguinte exemplo: Queremos saber se uma determinada moeda é equilibrada, ou seja, quando atirada ao ar a probabilidade de sair caras ou coroas é igual a  $\frac{1}{2}$ . Inicialmente, não temos nada que nos indique o contrário. Assim, a nossa hipótese inicial, chamada hipótese nula, é que a moeda seja equilibrada. Para se testar essa hipótese, pode-se tomar uma decisão por exemplo coletando-se uma amostra de 100 lançamentos e com base no resultado delibera-se se aceita ou rejeita a hipótese nula. Suponhamos que o resultado dos lançamentos foi 45 coroas e 55 caras. Será este resultado suficientemente forte para rejeitarmos a hipótese nula? Ou será que, se a moeda realmente for equilibrada, a probabilidade deste resultado ser devido ao acaso é alta? Efetivamente, a probabilidade de obter 45 ou menos coroas em 100 lançamentos, com uma moeda equilibrada, é de aproximadamente 0.18. Este valor é demasiado elevado para rejeitar esta hipótese, ou seja, a probabilidade de este resultado ter sido um acaso, é alta. Assim devemos aceitar a hipótese nula (HN). Note-se que a afirmação é aceitar e não provar a hipótese nula (HN). O resultado apenas não foi suficientemente forte para rejeitar a HN. Suponhamos agora, que o resultado do lançamento foi 30 coroas e 70 caras. A probabilidade de se obter um resultado, tão ou mais extremo do que este, com uma moeda equilibrada é de 0.002. Agora podemos dizer que este situação é pouco provável ter acontecido e, então, rejeitar a HN que dizia que a moeda era equilibrada.

**TESTE DE HIPÓTESE (ERROS):** No teste de uma hipótese há dois possíveis erros, rejeitar uma hipótese verdadeira ou deixar de rejeitar uma hipótese falsa. Uma vez que se tomou decisão sobre a hipótese, existe a possibilidade de haver feito apenas um tipo de erro.

**TESTE DE HIPÓTESE (FINALIDADE):** O foco de um teste de hipótese é determinar se uma média de amostra foi obtida de uma dada população, especificada pela hipótese nula, ou se a média da amostra foi obtida de uma outra população alternativa.

**TESTE DE HIPÓTESE DE PESQUISA DE UMA OU DUAS DIREÇÕES:** O uso de determinadas estatísticas Z ou t para determinar o significado estatístico é ditado pela maneira pela qual é colocada a pergunta ou hipótese da pesquisa. Perguntas sobre divergências cuja direção é especificada são perguntas unidirecionais; perguntas nas quais a direção da divergência não é especificada são bidirecionais.

**TESTE DE HIPÓTESES:** Método para testar afirmações sobre populações; chamado também teste de significância.

**TESTE DE HOMOGENEIDADE:** Teste da afirmação de que diferentes populações têm a mesma proporção de determinada característica.

**TESTE DE INDEPENDÊNCIA:** Teste da hipótese nula de que, em uma tabela de contingência, as variáveis linha e as variáveis coluna não são relacionadas.

**TESTE DE KOLMOGOROV E SMIRNOV:** Teste de significância proposto por Kolmogorov (1933) e desenvolvido por Smirnov (1939a e 1939b). O teste pode ser usado como teste de bondade de ajuste e a estatística de teste pode ser usada para definir intervalos de confiança para distribuições de probabilidade desconhecidas. Em inglês Kolmogorov-Smirnov test.

**TESTE DE KRUSKAL-WALLIS:** Teste não paramétrico de hipóteses usado para comparar três ou mais amostras independentes; chamado também teste H.

**TESTE DE KRUSKAL-WALLIS:** É um tipo de teste que tem o seu uso recomendado para a análise de dados de ensaios inteiramente casualizados que não satisfaçam às hipóteses básicas da análise de variância. No entanto tem como contra indicação o seu uso na análise de dados aos quais se possa aplicar a análise de variância, que deve ser sempre preferida, quando cabível. Os resultados obtidos por meio de teste são pouco satisfatórios, por ele utilizar apenas a ordem dos dados observados, sem levar em conta seus valores absolutos. Em inglês Kruskal Wallis test.

**TESTE DE KRUSKAL-WALLIS:** Teste estatístico de modelo livre de distribuição de probabilidades, para análise de médias de K amostras independentes, do mesmo tamanho ou desiguais, cujos devem ser mensurados, no mínimo, a nível ordinal.

**TESTE DE LEHMANN:** Teste não paramétrico para as variâncias de duas amostras. Em inglês Lehmann's test.

**TESTE DE LOGRANK:** Teste de significância usado para comparar taxas de sobrevivência determinadas pela análise da tábua de vida pelo método de Kaplan Meyer.

**TESTE DE MANN WHITNEY:** Teste estatístico não-paramétrico de distribuição livre, para duas amostras independentes do mesmo tamanho ou desiguais e dados mensurados, no mínimo, a nível ordinal. É também conhecido como Wilcoxon rank sum test.

**TESTE DE MANTEL:** Comparação da distância entre duas matrizes, como, por exemplo, a composição genética entre populações e respectivas distâncias geográficas ou temporais, obtendo-se valores de r, variando de -1 a +1, denotando ou não associação matricial.

**TESTE DE McNEMAR:** Teste estatístico para dados dispostos em tabela de contingência 2 x 2, a fim de comparar proporções de dois grupos pareados, onde são observados os pares concordantes e os discordantes em relação a dois tratamentos, A e B, avaliando-se, em termos probabilísticos, somente os pares discordantes em relação a esses mesmos tratamentos: (+ -) e (- +). Em inglês McNemar test.

**TESTE DE MERCADO SIMULADO (STM) OU MERCADO DE PRÉ-TESTE:** Como alternativa aos mercados tradicionais de teste, dados de levantamentos e modelos matemáticos são usados para simular resultados de mercados de teste a um custo bem mais baixo.

**TESTE DE PARKER:** Teste de estimação do tamanho de população por meio de amostragem múltipla em universos abertos.

**TESTE DE POISSON:** Teste para uma amostra, cujos eventos são raros e ocorridos em determinado período, comparando-se, pelo modelo de distribuição de Poisson, os dados amostrais com os da população.

**TESTE DE QUI QUADRADO ( $\chi^2$ ):** Teste estatístico não paramétrico utilizado para verificar se as frequências observadas ajustam-se às esperadas com base em uma hipótese.

**TESTE DE REPETIÇÕES:** Método não paramétrico para testar aleatoriedade.

**TESTE DE SCHUMACHER E ESCHMEYER:** Teste de estimação do tamanho da população por meio de amostragem múltipla em universos fechados

**TESTE DE SIGNIFICAÇÃO ESTATÍSTICA:** Classe de computações estatísticas que indicam a probabilidade de que a relação observada entre variáveis numa amostra se dever apenas a erro de amostragem.

**TESTE DE SIGNIFICÂNCIA:** Teste que permite determinar a probabilidade de obter o valor de um teste estatístico, por exemplo, os testes, t, r, F, qui quadrado dentre outros, considerando que a hipótese nula é verdadeira.

**TESTE DE SIGNIFICÂNCIA:** Veja teste de hipóteses.

**TESTE DE SIGNIFICÂNCIA:** O mesmo que teste estatístico.

**TESTE DE SIGNIFICÂNCIA ESTATÍSTICA:** Um série de testes estatísticos que permitem ao pesquisador identificar se existem ou não diferenças verdadeiras entre variáveis.

**TESTE DE WILCOXON:** Teste não paramétrico de comparação de médias. Está baseado nos signos das diferenças dos valores. Em inglês Wilcoxon's test.

**TESTE DE WILCOXON:** Teste não paramétrico de distribuição livre para duas amostras pareadas, cujos dados devem ser mensurados, pelo menos, a nível ordinal. É também conhecido com Wilcoxon signed rank test.

**TESTE DE WILCOXON DA ORDENAÇÃO ATRIBUÍDA PARA PARES COMPARADOS:** Teste não paramétrico de significância utilizado para comparar dados ordinais pareados. É análogo ao teste t pareado para dados contínuos.

**TESTE DO COEFICIENTE DE CORRELAÇÃO DE SPEARMAN PARA DADOS ORDINAIS:** Teste que análogo ao coeficiente de correlação de Pearson, mas que é usado para dados raqueados.

**TESTE DO QUI QUADRADO:** Este teste é usado para analisar a independência de duas variáveis categóricas. Isto se consegue comparando os valores observados e os valores esperados da distribuição conjunta das duas variáveis.

**TESTE DO QUI-QUADRADO (AMOSTRAS INDEPENDENTES):** Teste estatístico para  $n$  amostras independentes, cujas proporções observadas nas diversas modalidades estão dispostas em tabelas de contingência  $l \times c$ , sendo os valores esperados deduzidos matematicamente, e onde se procura determinar se as proporções observadas nas diferentes categorias são independentes ou estão associadas. O qui quadrado apresenta uma família de distribuição de variáveis com  $(l-1)(c-1)$  graus de liberdade.

**TESTE DO QUI QUADRADO DA INDEPENDÊNCIA:** Teste de significância estatística usado para analisar dados nominais ou dicotômicos em uma tabela de contingência. O primeiro passo é determinar o valor do qui quadrado para cada célula na tabela. Isso é feito calculando o quadrado da diferença entre o valor observado (O) e o valor esperado (E) de uma célula e dividindo pelo valor esperado para aquela célula. A fórmula padrão para o qui quadrado é

$$\sum \left[ \frac{(O-E)^2}{E} \right].$$

**TESTE DO QUI QUADRADO DE MCNEMAR:** Modificações do teste do qui quadrado para uso em dados pareados. Sua fórmula é

$$\frac{(|b-c|-1)^2}{(b+c)}.$$

**TESTE DO SINAL:** É um tipo de teste que tem seu uso recomendado na análise de ensaios com apenas dois resultados possíveis sim ou não, de probabilidades iguais, por hipótese. Mas têm como contra indicação o fato de não poder ser aplicado quando os dados não representam frequências, e mais ainda também não se aplica quando não são iguais as probabilidades atribuídas por hipótese, a cada um dos dois resultados possíveis do ensaio.

**TESTE DO SINAL:** Teste de significância não paramétrico para ser utilizado com dados contínuos, ordinais ou dicotômicos. O teste do sinal é usado para determinar se, em média, um grupo mostra ou não melhores resultados para um número significativamente maior de variáveis de desfecho que um outro grupo.

**TESTE DOS SINAIS:** Teste de hipóteses não paramétrico usado para comparar amostras de duas populações.

**TESTE DO SINAL:** Processo para testar a hipótese de que não há diferença entre duas grandezas dada por postos, em lugar de valores numéricos.

**TESTE DOS POSTOS COM SINAIS DE WILCOXON:** Processo para testar a hipótese de que não há diferença entre médias de populações.

**TESTE DOS SINAIS:** Teste estatístico não paramétrico para duas amostras pareadas, onde se leva em consideração apenas o sinal das diferenças entre cada par de escores (+ ou -), independentemente, portanto, da magnitude das diferenças.

**TESTE DOS SINAIS DE WILCOXON:** Um teste não paramétrico ou de distribuição livre para testar a diferença entre duas populações utilizando amostras emparelhadas. O teste toma por base as diferenças absolutas dos pares de observações das duas amostras, ordenados de acordo com o seu valor onde cada posto ou diferença recebe o sinal da diferença original. A estatística teste é a soma dos postos positivos. Em inglês Wilcoxon's signed rank test.

**TESTE DOS SINAIS DE WILCOXON:** Teste não paramétrico de hipóteses, usado para comparar duas amostras dependentes.

**TESTE DUNCAN:** É um teste que tem recomendado o seu uso na comparação de tratamentos qualitativos como cultivares, tipos de poda, tipos de enxertia, tipo de trato cultural, dentre outros. É contra indicado na comparação de tratamentos quantitativos com mais de dois níveis, como por exemplo, dosagens de adubação potássica em cinco níveis crescentes. Apresenta como desvantagem o fato de dá resultados significativos em excesso quando o número de tratamentos é superior a dois e especialmente quando esse número é maior do que quatro.

**TESTE DUNNETT:** É um tipo de teste recomendado para ser usado na comparação de médias de tratamentos com uma testemunha ou controle também conhecido como placebo, mas não é recomendado para outros tipos de comparações, sendo assim é pouco usado na prática.

**TESTE ESTATÍSTICO:** Diz-se que foi feito um teste estatístico quando se usam dados observados e uma estatística de teste para tomar a decisão de rejeitar ou não uma hipótese e se associa a essa decisão um p-valor.

**TESTE ESTATÍSTICO (PODER):** O poder de um teste estatístico se define como a probabilidade de rejeitar a hipótese nula, quando de fato for falsa. Um dos possíveis erros no teste de hipótese é deixar de rejeitar uma hipótese falsa, é o que chamamos de erro beta ou erro tipo II. Em termos matemáticos, o poder de um teste é (1-probabilidade de se cometer um erro beta). Em geral, aumenta-se o poder de um teste estatístico com o aumento do tamanho da amostra  $n$ .

**TESTE ESTATÍSTICO, SELEÇÃO DE:** O teste estatístico específico a ser usado depende das hipóteses que serão testadas, mas também é necessário satisfazer pressupostos. Os níveis de mensuração das variáveis, especialmente a variável dependente, são considerações importantes. Se estivermos interessados no relacionamento entre duas variáveis, usamos o coeficiente de correlação. Se a variável dependente for medida numa escala intervalar e a independente for categórica, de mensuração, ordinal ou nominal, usamos análises paramétricas. Testes de hipóteses usando médias seriam exemplos do uso de análise paramétrica. Quando usamos menos que escala intervalar, ou certos pressupostos não podem ser satisfeitos, usamos análises não paramétricas.

**TESTE, EFEITO DE (AMEAÇA À VALIDADE INTERNA):** Esta ameaça à validade interna de uma pesquisa convencional de grupos ocorre quando os escores desenvolvidos num posteste são, em parte, um resultado influenciado pelo fato de terem os sujeitos feito um preteste.

**TESTE EXATO DE FISHER:** Um teste de independência entre duas variáveis categóricas dicotômicas. Foi introduzido por Fisher (1935). O teste envolve o uso da distribuição hipergeométrica para calcular a probabilidade de uma determinada combinação de totais parciais, ou seja, somas de linhas ou colunas sob a hipótese nula de independência. Em inglês Fisher's exact test.



**TESTE EXATO DE FISHER:** Teste estatístico, para duas amostras independentes cujas proporções estão dispostas em tabela de contingência  $2 \times 2$ , devendo ser escolhido quando os valores observados são pequenos, inclusive com proporções cujos escores são iguais a zero (0).

**TESTE EXPERIMENTAL:** Consiste na verificação das conseqüências empíricas de uma dada hipótese, trabalhando-se comparativamente com grupos formados aleatoriamente que se denomina grupo experimental ou grupo teste e o grupo controle ou de comparação.

**TESTE EM CENTRAL:** Local onde animais de vários rebanhos são ajuntados a fim de se avaliar diferenças em certas características de desempenho, sob condições uniformes de manejo.

**TESTE F:** É o teste básico para a análise de variância, apresenta como contra indicações a comparação de tratamentos quantitativos com mais de dois níveis, a não ser quando associado a componentes de regressão. No caso de existir uma causa de variação com apenas um grau de liberdade, este teste equivale ao teste t, ao teste de Tukey e ao teste de Duncan.

**TESTE F:** Teste de significância estatística usado na análise de variância (ANOVA). A razão que se obtém dividindo duas estimativas de variância, supostas independentes.

**TESTE F:** Teste de hipótese obtido por Snedecor (1934) cuja finalidade é comparar estimativas de variâncias. Na análise de variância aplica-se o teste F com o objetivo de testar as hipóteses estatísticas. Este teste pode ser aplicado independentemente da análise da variância mas é na análise da variância dos delineamentos experimentais que ele encontra sua maior aplicação.

**TESTE F DE FRIEDMAN:** Processo para testar a hipótese de que não há diferença entre preferências no caso de haver mais de duas possibilidades.

**TESTE F DE SNEDECOR:** O mesmo que teste F. Em inglês F test.

**TESTE H:** Veja teste de Kruskal-Wallis.

**TESTE H DE KRUSKAL-WALLIS:** Uma generalização do teste da soma de postos de Wilcoxon, para a situação em que há mais de duas populações.

**TESTE KAPPA:** Teste estatístico não-paramétrico destinado a comparar as proporções da mesma variável mensurada a nível nominal em duas ocasiões distintas. Testa-se a reprodutibilidade períodos. Os dados são dispostos em tabela de contingência  $2 \times 2$ .

**TESTE KAPPA:** Medida da extensão em que a concordância entre dois observadores é maior do que a concordância por metro acaso.

**TESTE MAIS PODEROSO:** Um teste de hipótese que é mais poderoso contra uma hipótese alternativa. Em inglês Most powerful test.

**TESTE MONOCAUDAL:** Teste cuja hipótese alternativa é uma desigualdade, ou seja, deseja-se testar se o valor observado é maior ou menor ao valor crítico correspondente à hipótese nula.

**TESTE NÃO PARAMÉTRICO:** Um teste de hipótese não paramétrico testa associações dependência/independência e modelos ao invés de parâmetros. Em inglês nonparametric test.

**TESTE NÃO PARAMÉTRICO:** É aquele em que não há pressuposto sobre modelo de distribuição nem quanto aos parâmetros, não se aplicando ao mesmo o Teorema do Limite Central. Em inglês nonparametric test.

**TESTE NÃO PARAMÉTRICO:** Assim se denomina todo teste que não postule distribuição para sua validade. Os testes não paramétricos pressupõem desconhecimento da distribuição da variável na respectiva população e qualquer nível de mensuração. Em inglês nonparametric test.

**TESTE PADRÃO:** O teste usado como padrão na avaliação da acuidade de um teste novo ou seja, de um teste em estudo. Portanto, o teste padrão faz a classificação convencional de doença e não doença.

**TESTE PARAMÉTRICO:** É aquele com pressuposto de um modelo de distribuição, tais como normal, binomial, dentre outros, cujos parâmetros assumem um desses modelos, aplicando-se-lhe o Teorema do Limite Central. Em inglês nonparametric test.

**TESTE G (AMOSTRAS INDEPENDENTES):** Teste estatístico para n amostras cujas proporções das diversas modalidades estão dispostas em tabelas de contingência  $l \times c$ , sendo os valores esperados deduzidos matematicamente, procurando-se determinar se as proporções observadas nas diferente categorias são independentes ou estão associadas. Os graus de liberdade neste teste são calculados como segue:  $(l-1) \times (c-1)$ .

**TESTE PATOGNOMÔNICO:** Teste cujo resultado, se positivo, é sinônimo de ter a doença.

**TESTE Q DE COCHRAN:** Em inglês Cochran's Q test.

**TESTE QUI QUADRADO:** Um teste de significância estatística usado para variáveis organizadas em categorias e apresentadas numa tabela contingencial.

**TESTE QUI QUADRADO:** Teste da adequação do ajuste entre a distribuição observada e a distribuição esperada de uma variável.

**TESTE QUI QUADRADO:** É um tipo de teste muito usado em estatística sendo o seu uso recomendado na aplicação principalmente em tabelas de contingência relativas a frequências observadas. No entanto tem como contras indicação o fato de não poder ser aplicado quando os dados não representam frequências, nem quando são excessivamente baixas as frequências esperadas, teóricas ou calculadas. Neste caso as frequências devem ser independentes entre si. Esse teste também se aplica à comparação de uma teoria que indique quais asa frequências esperadas, como por exemplo, as leis do Monge Gregor Mendel o pai da genética. Em inglês Chi Square test.

**TESTE t DE STUDENT:** Teste paramétrico que utiliza duas amostras independentes. Testa a diferença entre duas médias populacionais quando os desvios padrões populacionais são desconhecidos, o que ocorre na grande maioria dos casos. Em inglês t test or Student t test.

**TESTE t DE STUDENT:** O teste t de Student é um teste de hipótese para médias. No caso de se querer comparar dois grupos, a hipótese nula é que a diferença das médias é zero, isto é, não há diferenças entre os grupos.

**TESTE t:** É o teste em que o seu uso atual mais comum é o calculo do intervalo de confiança de uma média. Possui como contra indicação o uso no cálculo da diferença mínima significativa a chamada d.m.s.. Esse



teste equivale ao teste F, ao teste de Tukey e ao teste de Duncan quando a causa de variação tem apenas 1 grau de liberdade.

**TESTE t:** Teste estatístico cujo objetivo é testar a igualdade entre duas médias. O teste supõe independência e normalidade das observações. As variâncias dos dois grupos podem ser iguais ou diferentes, havendo alternativas de teste para as duas situações. Pode ser usado para testar a igualdade entre uma média amostral conhecida e uma média populacional suposta.

**TESTE t:** Testes que comparam a diferença entre duas médias.

**TESTE t:** Teste estatístico cujo objetivo é testar a igualdade entre duas médias. O teste supõe independência e normalidade das observações. As variâncias dos dois grupos podem ser iguais ou diferentes, havendo alternativas de teste para as duas situações. Neste caso, consideramos apenas o caso em que as variâncias são iguais.

**TESTE TRIANGULAR:** Tipo de teste sensorial utilizado em métodos de análise de avaliação sensorial de alimentos, o qual se presta muito bem, para verificar se há diferença entre dois produtos, ou produtos obtidos por dois tratamentos diferentes. Pode ser usado para comparar mais de dois tratamentos, sendo nesse caso mais trabalhoso, pois sempre comparamos 2 a 2. Pode ser usado para comparar uma série de produtos com um padrão. Por exemplo, o teste pode ser usado inicialmente com um júri de peritos, para verificar se há diferença e, posteriormente, para um grande número de pessoas da população, para determinar a aceitação, por parte do consumidor. É um teste para seleção de provadores, para um júri de peritos, em que se pode avaliar a sensibilidade individual a diferentes paladares.

**TESTE TUKEY:** É um teste recomendado para ser usado na comparação de médias de tratamentos qualitativos, como cultivares, tipos de poda, tipos de enxertia, tipos de ração, marca de implemento agrícola, tipo de vacina, dentre outros. Tem como contra indicação o seu uso na comparação de tratamentos quantitativos com mais de dois níveis, tais como dosagens de adubação nitrogenada em quatro níveis, quantidade de água usada em irrigação em cinco níveis, dentre outros. Esse teste equivale ao teste F, ao teste t, e ao teste de Duncan quando a causa de variação tem apenas 1 grau de liberdade.

**TESTE U DE MANN WHITNEY:** Teste não-paramétrico de significância utilizado para comparar dois grupos de dados ordinais. É análogo ao teste t de Student para dados contínuos.

**TESTE U DE MANN WHITNEY:** Teste de hipótese equivalente ao teste da soma de postos de Wilcoxon para duas amostras independentes. Em inglês Mann Whitney U test.

**TESTE UNILATERAL:** Um teste é dito unilateral quando a região crítica ou região de rejeição está situada à direita. Em inglês one-tailed test.

**TESTE UNILATERAL:** Teste de hipóteses em que a região crítica consiste de apenas uma cauda da distribuição; a hipótese nula e rejeitada somente se o valor da estatística de teste tem extremo em uma direção.

**TESTE UNICAUDAL (UNILATERAL) DIREITO:** O mesmo que teste unilateral à direita.

**TESTE UNICAUDAL (UNILATERAL) ESQUERDO:** O mesmo que teste unilateral à esquerda.

**TESTE UNILATERAL (UNICAUDAL):** Teste de uma hipótese para a qual a região de rejeição está totalmente localizada em uma cauda da distribuição da estatística de teste. Em inglês One-sided test.

**TESTE UNILATERAL (UNICAUDAL):** Teste estatístico que opõe, à hipótese da nulidade, a hipótese de que a diferença entre as populações ou grupos em comparação está apenas à direita, ou apenas à esquerda da diferença nula.

**TESTE UNILATERAL (UNICAUDAL) À DIREITA:** Teste de hipóteses em que a região crítica está localizada na área direita extrema da distribuição de probabilidade.

**TESTE UNILATERAL (UNICAUDAL) À ESQUERDA:** Teste de hipóteses em que a região crítica está localizada na área esquerda extrema da distribuição de probabilidade.

**TESTE UNILATERAL (UNICAUDAL):** Teste estatístico no qual a hipótese alternativa especifica a direção da diferença a ser detectada, assim representado:  $\mu_1 < \mu_0$  ou  $\mu_1 > \mu_0$ .

**TESTE z:** Teste de significância usado para comparar diferenças entre proporções.

**TESTES ASSINTÓTICOS:** Testes que utilizam estatísticas obtidas de distribuições limite. Do inglês Asymptotic tests.

**TESTE DE ADITIVIDADE DE TUKEY:** É aquele que verifica se os efeitos dos fatores que ocorrem no modelo matemático são aditivos.

**TESTE DE ALEATORIEDADE:** Também conhecido como teste de sequência, verifica a aleatoriedade dos erros, ou seja, a sua independência. Para comprovar a propriedade de aleatoriedade de uma amostra esse teste faz uso da análise das iterações ou seja, sequência de símbolos idênticos. Este teste, basicamente, verifica o número de iterações existentes na amostra, se o número de iterações é muito grande ou muito pequeno sugere-se falta de aleatoriedade.

**TESTE DE LILLIEFORS:** É aquele que verifica a normalidade da distribuição dos erros ou resíduos. No caso em que se deseja testar normalidade e a média e a variância não são previamente especificadas, mas sim estimadas através dos dados da amostra, deve-se utilizar o teste de Lilliefors. Este teste tem procedimento análogo ao Kolmogorov Smirnov porém utiliza uma tabela própria e mais adequada a este tipo de situação.

**TESTES ASSINTÓTICOS DE HIPÓTESE:** Testes de hipótese baseados em resultados assintóticos. Do inglês Asymptotic tests of hypothesis.

**TESTES ESTATÍSTICOS:** Testes utilizados para avaliar se os resultados encontrados podem ser imputados à chance; ou seja, se podem ser devidos ao acaso.

**TESTES LIVRES DE DISTRIBUIÇÃO:** Testes que não exigem uma distribuição particular. Veja também testes não-paramétricos.

**TESTES NÃO PARAMÉTRICOS:** Processos estatísticos para testar hipóteses ou estimar parâmetros, quando não há suposições formuladas sobre a natureza ou a forma das distribuições populacionais; chamados também testes livres de distribuição.

**TESTES NON PROJATIVOS:** São testes de lápis e papel que requerem do indivíduo respostas afirmativas, selecionando uma opção de número de possíveis respostas.

**TESTES PARAMÉTRICOS:** Processos estatísticos baseados em parâmetros populacionais para testar hipóteses ou estimar parâmetros.

**TESTE POST HOC:** Teste estatístico de diferenças médias executado depois que os testes estatísticos para efeitos principais foram realizados. Em geral, os testes post hoc não usam um único contraste, mas em vez disso testam diferenças entre todas as possíveis combinações de grupos. Ainda que forneçam informação diagnóstica abundante, eles aumentam a taxa de erro Tipo I geral fazendo múltiplos testes estatísticos e, por isso, devem usar níveis de confiança muito estritos.

**TESTES PROJETIVOS:** Envolve uma palavra, figura, ou algum estímulo, a fim de provocar uma resposta inestruturada. Pode-se pedir a um respondente para indicar o que vê num borrão ou desenvolver uma história, a partir de um quadro.

**TESTE t:** Teste para avaliar a significância estatística da diferença entre duas médias amostrais  $(\bar{X}_1 - \bar{X}_2)$ , para uma única variável dependente. O teste t é um caso especial de análise de variância (ANOVA) para dois grupos ou níveis de uma variável de tratamento.

**TESTE t (STUDENT):** Teste paramétrico para uma amostra, duas amostras pareadas ou duas amostras independentes, baseados no modelo de distribuição de Student e geralmente efetuados quando se desconhece as variâncias paramétricas. O modelo de Student constitui uma família de distribuição e está relacionado com os graus de liberdade.

**TESTE Z:** Testes paramétricos para uma amostra ou duas amostras independentes, baseados no modelo de distribuição normal e geralmente efetuados quando as variâncias das populações são conhecidas. Nos estudos estatísticos, a distribuição normal constitui um dos modelos mais importantes em termos probabilísticos.

**TESTE Z:** Teste estatístico cujo objetivo é testar a igualdade entre duas médias ou também a igualdade entre uma média conhecida numa população e uma média calculada pelo pesquisador numa amostra. O teste supõe normalidade das observações.

**TESTE Z:** Teste estatístico cujo objetivo é testar a igualdade entre uma média conhecida numa população e uma média calculada pelo pesquisador numa amostra. O teste supõe normalidade das observações.

**TESTES DE ASSOCIAÇÃO DE PALAVRAS:** Testes nos quais o entrevistador diz uma palavra e os respondentes precisam responder com a primeira coisa que lhes vêm à mente.

**TESTES DE COMPLEMENTAMENTO DE SENTENÇAS E HISTÓRIAS:** Testes nos quais os respondentes completam sentenças ou histórias em suas próprias palavras.

**TESTES DE DESENHOS ANIMADOS:** Testes nos quais os respondentes preenchem o diálogo de uma das personagens do desenho.

**TESTES EQUIVALENTES:** Dois testes são equivalentes quando têm o mesmo poder, isto é, um rejeita a hipótese  $H_0$ , quando o outro rejeita  $H_0$ , e, aceita  $H_0$  quando o outro aceita  $H_0$ .

**TIPOLOGIA:** Classificação tipicamente nominal de observações em termos de seus atributos em duas ou mais variáveis. Um exemplo de tipologia seria a classificação de jornais em urbanos liberais, rurais liberais, urbanos conservadores ou rurais conservadores.

**TIPOLOGIA:** Classificação conceitual de objetos baseada em uma ou mais características. Uma tipologia geralmente não tenta agregar observações reais, mas ao invés disso, fornece a fundamentação teórica para a criação de uma taxonomia, a qual agrega observações reais.

**TLC:** Teorema do limite central.

**TODAS AS REGRESSÕES POSSÍVEIS:** Tipo de procedimento usado na análise de regressão para a seleção de um subconjunto de variáveis regressoras, no qual temos  $k$  variáveis regressoras, e o mesmo consiste em fazer os ajustes de todos os modelos possíveis e, nesse caso, se incluímos um modelo sem variáveis regressoras, então são ajustados um total de  $2^k$  modelos, uma vez que teremos  $k$  modelos com uma variável regressora;

$$\frac{k(k-1)}{2}$$

modelos com duas variáveis regressoras, e assim por diante. Por meio de um dado critério, os modelos ajustados são comparados e o melhor deles é então selecionado. Na prática, este procedimento não é viável se o valor de  $k$  muito alto.

**TODO E QUALQUER:** Símbolo matemático dado por  $\forall$ .

**TOLERÂNCIA:** Medida de colinearidade e multicolinearidade comumente usada. A tolerância da variável  $i$  ( $TOL_i$ ) é  $1-R_i^2$ , onde  $R_i^2$  é o coeficiente de determinação para a previsão da variável  $i$  pelas outras variáveis preditoras. Os valores de tolerância que se aproximam de zero indicam que a variável é altamente prevista, ou seja, colinear com as outras variáveis preditoras.

**TOLERÂNCIA:** Medida de colinearidade e multicolinearidade comumente usada. A tolerância da variável  $i$  ( $TOL_i$ ) é  $1-R_i^2$ , onde  $R_i^2$  é o coeficiente de determinação para a previsão da variável  $i$  pelas outras variáveis independentes. À medida que o valor da tolerância se torna menor, a variável é mais bem prevista pelas outras variáveis independentes indicando que existe colinearidade.

**TOLERÂNCIA:** Proporção da variação nas variáveis independentes não explicada pelas variáveis que já estão no modelo ou função. Pode ser usada como proteção contra a multicolinearidade. Calculada como  $1 - R_i^2$ , onde  $R_i^2$  é a quantia de variância da variável independente  $i$  explicada por todas as outras variáveis independentes. Uma tolerância de 0 significa que a variável independente sob consideração é uma combinação linear perfeita de variáveis independentes já no modelo ou equação. Uma tolerância de 1 significa que uma variável independente é totalmente independente de outras variáveis que já estão no modelo.

**TÓPICO:** É a subdivisão do assunto ou do tema.

**TOPOLOGIA:** É o estudo das propriedades que não se alteram por deformações contínuas. O desenvolvimento da topologia tem grande influência em várias áreas da matemática. A topologia está dividida em: geral, algébrica e diferencial. Em inglês topology.

**TORAGEM:** Ato de seccionar o tronco de uma árvore em toras. Em sentido figurado designa o ato de subdividir o tronco de uma árvore em secções para realização da cubagem rigorosa. Dentre as formas de se realizar a toragem para cubagem rigorosa, destacam-se: i) toragem em comprimentos fixos de 1 ou 2 m, medindo-se sempre a área transversal nas extremidades e aplicando-se uma fórmula de cubagem; ii) toragem em posições fixas onde se mede com maior frequência as áreas transversais na base do tronco, onde se encontra a maior parte do volume, como por exemplo: 0 m; 0,1 m; 0,3 m; 1,0 m; 1,3 m (diâmetro

á altura do peito-DAP); 2,3 m; 3,3 m. iii) Cubagem em posições relativas, onde se mede em posições do tronco relativas à altura total da árvore, como por exemplo: 0 %, 10 %, 30 %, 50 %, 70 % e 90 %.

**TRAÇO:** Representa a quantia total de variância na qual a solução fatorial é baseada. O traço é igual ao número de variáveis, baseado na suposição de que a variância em cada variável é igual a um (1).

**TRAÇO OU TRAÇO DE UMA MATRIZ QUADRADA:** O traço de uma matriz quadrada é a soma dos elementos de sua diagonal principal. Em inglês trace.

**TRANSAÇÃO:** Ação sobre um elemento operacional de dados, como criação, modificação ou eliminação representando um único evento de negócios. Exemplos incluem a venda de um produto a um cliente, recebimento de pagamento, transferência de fundos entre contas ou transferência de produtos completos ou comprados para inventário.

**TRANSFORMAÇÃO:** Uma variável pode ter uma característica indesejável, como não-normalidade, que diminui a habilidade do coeficiente de correlação em representar a relação entre ela e outra variável. Uma transformação como calcular o logaritmo ou a raiz quadrada da variável cria uma nova variável e elimina a característica indesejável, permitindo uma medida melhor da relação. Transformações podem ser aplicadas em variáveis dependente, independentes ou ambas. A necessidade e o tipo específico de transformação podem ser baseados em motivos teóricos como a transformação de uma relação não linear conhecida ou empíricos, identificados por meios gráficos ou estatísticos.

**TRANSFORMAÇÃO ANGULAR**  $\left[ \text{Arc sen} \sqrt{\frac{X}{100}} \right]$ : É o tipo de transformação recomendável para dados

expressos em porcentagens, que geralmente seguem uma distribuição teórica de probabilidades binomial. Existem tabelas apropriadas para essa transformação, nas quais entramos diretamente com a porcentagem X e obtemos o  $\text{Arc sen} \sqrt{\frac{X}{100}}$ . Conhecida também como transformação seno inversa, arco seno ou de Bliss.

Se as porcentagens estiverem todas compreendidas numa amplitude de 30 a 70 %, torna-se desnecessária a transformação e podemos analisar diretamente os dados originais. A transformação também é desnecessária quando as porcentagens são resultantes da divisão dos valores observados nas parcelas por um valor constante, como, por exemplo, a média do tratamento testemunha ou controle, ou quando são representativas de concentração, como por exemplo, o teor de nitrogênio (N) na folha de uma árvore, a taxa de pureza da semente de milho, o teor de proteína do trigo, o teor de sacarose da cana-de-açúcar, dentre outras. Vale lembrar ainda que devem ser transformados os dados de porcentagem provenientes de dados discretos num total de casos, como, por exemplo, porcentagem de germinação que é o número de sementes germinadas dividido pelo número total de sementes, porcentagem de plantas doentes o qual é obtido pelo número de plantas doentes dividido pelo número de plantas consideradas, porcentagem de animais doentes que é o número de animais dividido pelo número total de animais do rebanho, dentre outras.

**TRANSFORMAÇÃO ANGULAR**  $\left[ \text{Arc sen} \sqrt{\%} \right]$ : Essa transformação é utilizada, quando os dados estão associados a uma distribuição teórica de probabilidade binomial, como o número de vezes que um determinado caráter aparece num total definido. É o caso, por exemplo, do número de plantas sobreviventes em relação ao número total de plantas, que deveria estar presente, se nenhuma tivesse morrido. Os dados são antes transformados em porcentagens, e, em seguida, são calculados os  $\text{Arc sen} \sqrt{\%}$ , utilizando uma tabela apropriada ou uma calculadora ou ainda um computador. Nesse tipo de transformação, todos os dados deveriam estar baseados em um número igual de observações, mas o método pode também ser utilizado

quando esse número não é rigorosamente igual. Bartlett (1947) sugere que 0% seja substituído, antes de entrar na tabela, por  $\frac{1}{4n}$ , e 100% por  $100 - \frac{1}{4n}$ , sendo n o número total de observações. Segundo Steel e Torrie (1960) a transformação raiz quadrada é recomendada para porcentagem entre 0 % e 20% ou 80% e 100%, sendo subtraídos de 100 antes da transformação  $\sqrt{X}$ . Porém, quando a amplitude de variação for grande, a transformação  $\text{Arc sen} \sqrt{\%}$  deve ser preferida. Quando é utilizada uma transformação, todas as comparações entre médias são feitas na escala transformada. Quando se achar preferível não apresentar os resultados na escala transformada, os dados finais devem ser transformados novamente para a escala original. Isto é feito elevando-se ao quadrado, no caso de  $\sqrt{X}$ , achado o antilogaritmo, no caso de  $\log X$ , e procurando o valor correspondente na tabela de arc sen %, no caso de transformação angular.

**TRANSFORMAÇÃO ARCSENO**  $\left[ \text{Arc sen} \sqrt{Y} \right]$ : Tipo de transformação usada em análise de regressão por exemplo, a qual estabiliza a variância quando os dados são proporções.

**TRANSFORMAÇÃO ARCSENO**  $\left[ \text{Arcsen} \sqrt{X} \right]$ : Tipo de transformação especialmente recomendada para dados expressos em porcentagem, e principalmente para casos em que os dados estejam nas faixas entre 0 % e 20 % ou entre 80 % e 100 %. Os valores transformados tanto podem ser expressos em graus como em radianos. Se os valores em porcentagem estiverem todos na faixa entre 20 % e 80 %, pouco ou nada se ganhará ao fazer-se a transformação. A variância dos novos valores é aproximadamente constante, sendo igual a  $\left[ \frac{821}{n} \right]$ , quando expressos em graus, e igual a  $\left[ \frac{0,25}{n} \right]$ , se em radianos (STEEL; TORRIE,

1981). Isto mostra que se pressupõe um denominador constante, mas esta transformação é comumente empregada quando os denominadores são diferentes.

**TRANSFORMAÇÃO ARCO SENO:** A transformação  $Y = \text{arc sen} \sqrt{\frac{X}{100}}$ , também denominada angular,

é recomendada para estabilizar a variância para dados de porcentagens, principalmente se estiverem no intervalo de 0 a 30 % e ou de 70 a 100 %. Com dados de porcentagem existe um número máximo, n, de observações, que serve de base para o cálculo da porcentagem. Esse valor de n deve ser o mesmo, ou aproximadamente o mesmo, para cada unidade experimental. Como exemplo desse tipo de dado podemos citar as porcentagens de controle de herbicidas aplicados em 3 diferentes épocas, sobre o crescimento de picão, que é uma planta invasora ou erva daninha.

**TRANSFORMAÇÃO DA RAIZ QUADRADA:** Uma transformação de variáveis que é usada para estabilizar a variância de dados amostrais obtidos de uma população Poisson. A transformação da raiz quadrada leva a mesma relação para a distribuição de Poisson como a transformação do arco seno para a transformação binomial. Em inglês Square root transformation.

**TRANSFORMAÇÃO DA RAIZ QUADRADA**  $\left[ \sqrt{X} \right]$ : Essa transformação é utilizada quando os dados são representados por números inteiros (contagens), em geral maiores do que 10. É também aconselhada no caso de porcentagens, referentes a contagens e variando de 0 % a 20 %, e de 80 % a 100 %. As porcentagens entre 80% e 100% deveriam ser de preferência, subtraídas de 100, antes de se fazer a transformação. A transformação  $\sqrt{X}$  é, ainda, indicada no caso de porcentagens, fora dos limites considerados acima, quando as observações estão claramente numa escala contínua. Quando as contagens são menores do que 10, e as porcentagens muito baixas, aparecendo valor zero, aconselha-se utilizar a transformação

$\sqrt{X + 0,50}$  ou  $\sqrt{X + 1,0}$ , em lugar de  $\sqrt{X}$ .



**TRANSFORMAÇÃO DE DADOS:** O efeito da assimetria na distribuição normal consiste em se obter resultados significativos em excesso para o teste F. Este teste, no entanto, será pouco afetado se os desvios da normalidade forem pequenos. Se a distribuição desviar-se do modelo ou Gaussiano normal ela poderá aproximar-se, por exemplo, da distribuição de Poisson, uma distribuição onde vale a relação variância=média ou de uma distribuição log-normal, onde temos que a variância=(média)<sup>2</sup>; ou de uma distribuição binomial, onde a variância é proporcional à quantidade [média (1-média)]. Nestes casos, a variância não mais se mantém constante, variando com as médias dos tratamentos, e o uso de uma estimativa conjunta de variância, S<sup>2</sup>, não se justifica. Uma maneira eficiente de se lidar com estes problemas é o uso da transformação de dados. Se a variância de uma variável aleatória X for uma função conhecida de sua média  $\mu$ , isto é  $\sigma^2 = F(\mu)$ , então podemos procurar uma transformação da variável X para outra, Y, isto é,  $Y = f(X)$ , tal, que a variância de Y seja constante isto é,  $\sigma_Y^2 = K$ . Escrevendo  $Y = f(X)$  pela expansão de Taylor com um termo,  $Y \cong f(\mu) + f'(\mu)(X - \mu)$ , temos que a média é  $E(Y) \cong f(\mu)$  e a variância,

$\sigma_Y^2 = E[Y - E(Y)]^2 \cong E[f'(\mu)(X - \mu)]^2 = [f'(\mu)]^2 F(\mu)$ . Por esta fórmula vemos que para tornar

$\sigma_Y^2$  aproximadamente constante, basta tomar  $[f'(\mu)]^2$  proporcional a  $\frac{1}{F(\mu)}$ . Então, se  $[f'(\mu)]^2 = \frac{1}{F(\mu)}$

e  $f'(\mu) = \frac{1}{\sqrt{F(\mu)}}$ , para determinar  $f(\mu)$  fazemos  $f'(\mu)d\mu = \frac{d\mu}{\sqrt{F(\mu)}}$ , integrando obtemos

$f(\mu) = \int \frac{d\mu}{\sqrt{F(\mu)+C}}$ . Na distribuição de Poisson, por exemplo,  $\sigma_X^2 = F(\mu) = \mu$ . Substituindo em

$f(\mu) = \int \frac{d\mu}{\sqrt{F(\mu)+C}}$  temos  $f(\mu) = \int \frac{d\mu}{\sqrt{\mu+C}} = 2\sqrt{\mu+C}$ . Verificamos que  $Y = \sqrt{X}$  é a transformação da

variável X, com distribuição de Poisson, que estabiliza a variância na escala transformada, Y. Ao invés se analisarmos a variável X, analisamos Y, com variância aproximadamente constante. Outro exmplo é o caso da distribuição log-normal, onde  $\sigma_X^2 = F(\mu) = \mu^2$ . Então,  $f(\mu) = \int \frac{d\mu}{\mu} + C = \ln|\mu| + C$ . Fica claro

que, para dados com distribuição log-normal, a transformação logarítmica é a adequada. Para o caso de variável com distribuição binomial, com  $\sigma_X^2 = F(\mu) = K\mu(1-\mu)$ , seguindo o mesmo cálculo veremos

que a função  $Y = \text{arc sen } \sqrt{X}$  é a transformação que estabiliza a variância.

**TRANSFORMAÇÃO ANGULAR (ANGULAR TRANSFORMATION):** Tipo ou modelo de transformação onde a função  $Y^* = \text{arco-seno } \sqrt{X}$ , onde o arco-seno(X) é a função que retorna o ângulo  $\theta$  para o qual o seno de  $\theta = X$ . Também chamada de transformação do arco-seno ou de transformação do arco-seno da raiz quadrada.

**TRANSFORMAÇÃO BOX-COX (BOX-COX TRANSFORMATION):** Uma família de transformações de potências definida pela fórmula  $Y^* = \frac{(Y^\lambda - 1)}{\lambda}$  (para  $\lambda \neq 0$ ) e  $Y^* = \ln(Y)$  (para  $\lambda = 0$ ). O valor do  $\lambda$  que

resulta no melhor ajuste a uma distribuição normal é usado para a transformação final dos dados. Esse valor deve ser obtido por métodos de computação intensiva.

**TRANSFORMAÇÃO DA RAIZ QUADRADA (SQUARE ROOT TRANSFORMATION):** É a função dada por:  $Y^* = \sqrt{Y}$ . É usada com mais frequência para dados de contagem, que são variáveis aleatórias de Poisson.

**TRANSFORMAÇÃO DO ARCOSENO DA RAIZ QUADRADA (ARCSINE SQUARE-ROOT TRANSFORMATION):** Ver transformação angular.

**TRANSFORMAÇÃO INVERSA (RECIPROCAL TRANSFORMATION):** Tipo de transformação em que a função  $Y^* = \frac{1}{Y}$  em geral é usada para dados hiperbólicos.

**TRANSFORMAÇÃO LOGIT (LOGIT TRANSFORMATION):** Uma transformação algébrica para converter os resultados de uma regressão logística, que produz uma curva em forma de S, em uma linha reta.

**TRANSFORMAÇÃO REVERSA (BACK-TRANSFORMED):** A consequência de mudar uma variável ou uma estimativa estatística de volta às unidades de medida originais. A transformação reversa só é necessária se as variáveis tiverem sido transformadas antes das análises estatísticas. Ver também transformação.

**TRANSFORMAÇÃO ARCO-SENO (ARCSINE TRANSFORMATION):** Ver o conceito de transformação angular.

**TRANSFORMAÇÃO DE RAIZ QUADRADA [ $\sqrt{X}$ ]:** É aquela frequentemente utilizada para dados de contagem, que geralmente seguem a distribuição teórica de probabilidade de Poisson, na qual a média é igual à variância. Como exemplos podemos citar: o número de ervas daninhas por parcela, o número de frutos de acerola contados por árvore, o número de insetos capturados em armadilhas luminosas, o número de pulgões ou ácaros contados por folha, dentre outros. Quando ocorrem zeros ou valores baixos, as transformações recomendadas são  $\sqrt{X+0,50}$  ou  $\sqrt{X+1,0}$ . Verificada a necessidade de transformação, os dados serão transformados e toda a análise estatística, ou seja, a análise de variância, testes a posteriori de comparações múltiplas e desdobramento dos graus de liberdade de tratamentos deverá ser feita com os dados transformados. Se houver interesse em ilustrar com as médias não transformadas, as mesmas não deverão ser calculadas a partir dos dados originais e sim, aplicando à média dos dados transformados a operação inversa da transformação. Por exemplo, se a transformação utilizada for a de raiz quadrada de  $X+0,50$ , ou seja  $\sqrt{X+0,50}$ , a média não transformada será calculada elevando-se a média transformada ao quadrado e subtraindo-se 0,50. Se a heterocedasticidade for do tipo irregular, para eliminá-la, podemos simplesmente eliminar os tratamentos discrepantes ou, caso isto não seja possível ou recomendável, subdividi-los em grupos e testá-los separadamente, por meio de resíduos apropriados a cada grupo.

**TRANSFORMAÇÃO LOGARÍTMICA (LOGARITHMIC TRANSFORMATION):** A função  $Y^* = \log(Y)$ , onde log é o logaritmo em qualquer base útil. Frequentemente é usada quando as médias e variâncias são correlacionadas positivamente.

**TRANSFORMAÇÃO DE DADOS:** Mudança de escala dos dados de uma variável, podendo ser efetuada pelos logaritmos, pela raiz quadrada, pela ordenação dos escores ou rank's, pelo quadrado dos valores, dentre outros.

**TRANSFORMAÇÃO DE LAPLACE:** Se uma função  $g(t)$  está relacionada a uma segunda  $f(x)$  pela equação  $g(t) = \int e^{-tx} f(x) dx$  para  $x > 0$ , então  $g(t)$  é a transformação de Laplace de  $f(x)$ . Em inglês Laplace transformation.



**TRANSFORMAÇÃO LOGARÍTMICA [log X ou Ln X]:** É o tipo de transformação utilizada quando é constatada uma certa proporcionalidade entre as médias e os desvios padrões dos diversos tratamentos. Para exemplificar esse tipo de transformação podemos considerar, o caso de contagens de insetos, de nematóides que são vermes subterrâneos, de helmintos em exames coprológicos em ovinos, dentre outros, nestas situações se a população é numerosa, as contagens serão altas tanto para o tratamento testemunha como para os tratamentos não eficientes, por exemplo, variação de 100 a 10000 insetos, de 200 a 20000 vermes, dentre outros, ao passo que, para outros tratamentos, que controlam melhor a praga estudada, a amplitude de variação será baixa, por exemplo, entre 6 e 60 insetos. Sendo assim a transformação logarítmica é aconselhável nestes casos.

**TRANSFORMAÇÃO LOGARÍTMICA [log X]:** A transformação logarítmica é utilizada principalmente quando as médias e os desvios padrões ou erros tendem a serem proporcionais, sendo, nesse caso, os coeficientes de variação aproximadamente iguais. É utilizada, ainda, quando os dados são representados por números positivos, ou por porcentagens que abrangem uma grande amplitude de variação. Quando aparece o valor zero, utiliza-se a transformação  $\log(X + 1,0)$ . A base 10 para os logaritmos é utilizada normalmente, por conveniência, porém qualquer outra base pode ser empregada.

**TRANSFORMAÇÃO LOGARÍTMICA [log (X + K)]:** A transformação logarítmica é utilizada principalmente em três situações onde ela é mais adequada: quando variâncias são proporcionais ao quadrado da média; na presença de efeitos multiplicativos na escala original que, portanto, se tornam aditivos com a transformação, e medidas em valores inteiros que cobrem uma amplitude muito grande. Teoricamente qualquer base logarítmica pode ser adotada, e até o surgimento dos computadores, a base 10 era a mais usada por conveniência ou tabelas mais facilmente disponíveis. Neste tipo de transformação, a constante K pode assumir, no geral, dois valores, 0 e 1. O valor 1 é empregado obrigatoriamente quando da ocorrência de zeros ou optativamente com a ocorrência de valores menores que 10.

**TRANSFORMAÇÃO LOGARÍTMICA:** É feita por meio da aplicação da função logaritmo aos dados originais, reduzindo o intervalo de variação dos mesmos e conseqüentemente a variância. A transformação logarítmica permite fazer a comparação de dois ou mais conjuntos de valores, pois se as dispersões forem muito diferentes, ou se existir muita assimetria, a comparação é difícil, se não impossível. Ainda segundo aquele autor, a transformação mesmo que não seja imprescindível, melhora o entendimento que se tem de um conjunto de valores. Deve-se ressaltar a importância da simetria, pois onde existem assimetrias muito acentuadas é impossível, pelo menos difícil, apresentar um valor que sirva como indicador de nível para um conjunto de valores. A transformação logarítmica reduz o intervalo de variação original grande, para números grandes, para um intervalo pequeno, enquanto para números pequenos ocorre o contrário.

**TRANSFORMAÇÃO LOGARÍTMICA:** Se os efeitos dos fatores no modelo não forem aditivos, mas sim multiplicativos, as observações estarão mais próximas da distribuição log – normal, uma distribuição onde o logaritmo da variável é que tem distribuição normal. Já é sabido que na distribuição log-normal a variância é proporcional ao quadrado da média:  $\sigma^2 = K\mu^2$ . A transformação  $Y = \ln X$  é, portanto, a transformação adequada a dados onde se observa este tipo de relação entre média e variância, resolvendo tanto o problema de heterogeneidade de variâncias como a falta de aditividade no modelo. Se ocorrerem valores de X menores que 1 podemos usar a transformação  $Y = \ln(X + 1)$ ; isto evita que se usem números negativos na análise, além de resolver o problema dos valores de X iguais a zero.

**TRANSFORMAÇÃO LOGARÍTMICA NA BASE e [log Y]:** Tipo de transformação usada em análise de regressão, por exemplo, a qual estabiliza a variância, quando esta tende a crescer à medida que Y também cresce. Em algumas situações pode também ajudar a normalizar os dados. Se a relação entre X e Y é do tipo exponencial, essa transformação introduz uma relação linear entre X e a variável transformada, log (Y).

**TRANSFORMAÇÃO LOGIT:** Transformação dos valores da variável dependente binária discreta da regressão logística em uma curva em S denominada curva logística que representa a probabilidade de um evento. Essa probabilidade é então usada para formar a razão de desigualdade, a qual atua como a variável dependente na regressão logística.

**TRANSFORMAÇÃO LOG-LOG:** Transformação de uma probabilidade P de acordo com a fórmula  $Y = -\log(-\log P)$ . Em inglês Log-log transformation.

**TRANSFORMAÇÃO NORMALIZADORA:** Transformação de variáveis com o objetivo de reduzir a distribuição original a uma normal ou aproximadamente normal. Em inglês Normalizing transformation.

**TRANSFORMAÇÃO QUADRÁTICA [Y<sup>2</sup>]:** Tipo de transformação usada em análise de regressão, por exemplo, a qual é usada para estabilizar variâncias, quando esta tende a decrescer com a média dos Y's. Normaliza os dados quando os resíduos se mostram com assimetria negativa. Lineariza os dados quando esses têm uma relação curvilínea. Se a relação entre X e Y é do tipo curvilínea, essa transformação introduz uma relação linear entre X e a variável transformada Y<sup>2</sup>. É importante lembrar que quando fazemos uma transformação na variável original Y, as estimações e previsões estão expressas em novas unidades, conforme a transformação usada. Muitas vezes o nosso objetivo é fazer estimações e previsões na escala original, portanto, este é um problema que não deve ser esquecido.

**TRANSFORMAÇÃO RAIZ QUADRADA:** A transformação  $Y = \sqrt{X}$  é recomendada principalmente, quando temos dados de contagem de eventos raros, por exemplo: número de plantas atacadas por uma doença, número de acidentes na produção de um material, dentre outros. Nestes casos a variável segue uma distribuição de Poisson. Também é utilizada para qualquer tipo de contagem, por exemplo, número de frutos produzidos por planta, ou mesmo de dados que não sejam de contagem, desde que se observe que as variâncias são proporcionais às médias dos tratamentos. Se alguns dados de contagem forem pequenos, ou iguais a zero, devemos somar 0,5 a cada valor, antes de usar a raiz quadrada. Neste caso temos a transformação  $Y = \sqrt{X + 0,5}$ . Observemos que todos os dados serão acrescidos da constante 0,5, não só os pequenos.

**TRANSFORMAÇÃO RAIZ-QUADRADA [√Y]:** Tipo de transformação usada em análise de regressão, por exemplo, a qual é usada para estabilizar a variância quando esta é proporcional à média dos Y's. Em particular, quando a variável Y for uma contagem, com distribuição de Poisson, a variável transformada, pode ser considerada com distribuição normal.

**TRANSFORMAÇÃO RAIZ-QUADRADA [√X+k]:** Tipo de transformação usada para dados que seguem uma distribuição de Poisson, isto é, que possuem média e variância iguais, os quais são melhor analisados após a aplicação da transformação raiz quadrada. Alguns exemplos de dados que seguem esta distribuição são as contagens de números de plantas daninhas ou de insetos de uma dada espécie por unidade de área, por exemplo, onde as contagens são sempre em números inteiros e baixos. Dados em porcentagem baseados em contagens e com denominador comum, situados entre 0 % e 20 % ou entre 80

% e 100 %, também podem ser transformados por raiz quadrada. A constante k adicionada ao dado pode, usualmente, assumir três valores, 0,00, 0,50 e 1,00. Quando os valores são muitos pequenos e, principalmente, se zeros estão presentes, a simples raiz quadrada (k=0) tende a supercorrigir e, portanto, k=0,50 ou k=1 devem ser empregados.

**TRANSFORMAÇÃO RECÍPROCA**  $\left[\frac{1}{Y}\right]$ : Tipo de transformação usada em análise de regressão por exemplo, a qual é usada para estabilizar a variância, no sentido de minimizar o efeito de possíveis valores muito altos de Y.

**TRANSFORMAÇÃO RECÍPROCA**  $\left[\frac{1}{Y}\right]$ : Tipo de transformação que refere-se a uma função de transformação adequada quando Y refere-se à taxa de sobrevivência. Por conseguinte,  $\left[\frac{1}{Y}\right]$  refere-se à taxa de mortalidade.

**TRANSFORMAÇÃO z DE FISHER:** É a transformação  $z = \frac{1}{2} \log_e \frac{1+r}{1-r}$ , onde r denota o coeficiente

de correlação de amostra simples de uma população normal bidimensional, dada por Fisher (1921).

**TRANSFORMAÇÕES DE DADOS:** Uma variável pode ter uma característica indesejável, como não normalidade, que diminui seu uso em uma técnica multivariada. Uma transformação, como o logaritmo ou a raiz quadrada da variável, cria uma variável transformada que é mais adequada para escrever a relação. As transformações podem ser aplicadas em variáveis dependentes, independentes ou ambas. A necessidade e o tipo específico de transformação podem ser baseados em razões teóricas, por exemplo, transformar uma relação não-linear conhecida ou razões empíricas, como por exemplo, problemas identificados por meios gráficos ou estatísticos.

**TRANSFORMAÇÃO PELA RAIZ CÚBICA**  $(\sqrt[3]{Y})$ : Outras transformações podem ser usadas para converter relações não lineares em lineares. Por exemplo, a transformação pela raiz cúbica  $(\sqrt[3]{Y})$  é adequada para medidas de massa ou volume ( $Y^3$ ), as quais relacionadas alometricamente com medidas lineares de tamanho ou comprimento corporal (Y). Em estudos que examinam as relações entre duas medidas de massa ou volume ( $Y^3$ ), como a comparação das massas cerebral e corporal, tanto a variável X quanto Y são logaritmicamente transformadas. A transformação logarítmica reduz a variação dos dados, que pode variar em várias ordens de magnitude. (GOTELLI; ELLISON, 2010).

**TRANSFORMAÇÕES MATEMÁTICAS:** Conforme Brito (2014), as transformações matemáticas dos resultados de experimentos aleatórios podem realizar-se para homogeneizar as variâncias e/ou normalizar a distribuição de variáveis respostas. As hipóteses são testadas nas variáveis transformadas através de testes não paramétricos, mas, se não for conveniente apresentar os dados na nova variável transformada, as médias podem ser transformadas de volta para a medida original. Segundo ainda o autor entre as transformações desenvolvidas para homogeneizar as variâncias, e que podem também conduzir a normalização da variável incluem-se as descritas a seguir:

Tipo de Transformação	Aplicar quando
$\sqrt{Y}$	Os valores dos $Y_i$ forem contagens de números pequenos
$\sqrt{Y} + \sqrt{Y+1}$	Os valores dos $Y_i$ forem contagens e alguns $Y_i$ forem iguais a zero
$Log(Y)$	A dispersão dos $Y_i$ é elevada, e as variâncias proporcionais às médias
$Log(Y+1)$	A dispersão dos $Y_i$ é elevada e alguns valores dos $Y_i$ forem iguais a zero
$\frac{1}{Y+1}$	Os valores dos $Y_i$ forem muito próximos de zero
$Arcsen(\sqrt{Y})$	Os valores dos $Y_i$ forem proporções ou percentagens dispersas
$Log\left[\frac{(1+Y)}{1-Y}\right]$	$-1 \leq Y \leq 1$
$(1-Y)^{\frac{1}{2}} - \frac{1}{3}(1-Y)^{\frac{3}{2}}$	$0 \leq Y \leq 1$

**TESTE DE ADITIVIDADE DE TUKEY:** A não aditividade presente nos dados resulta em heterogeneidade das contribuições das observações para variância residual, não sendo esta, portanto, uma estimativa eficiente da variância comum. Em outras palavras, a expectância  $V(e_i) = \sigma^2$  não se verifica. De fato, a variância do erro é geralmente maior do que a que seria obtida se o modelo fosse aditivo. O modelo linear aditivo pressupõe que os efeitos de tratamentos são aditivos, isto é, o valor de qualquer observação resulta da soma dos efeitos das várias causas de variação que afetam os dados. Uma forma comum de não aditividade ocorre quando os efeitos são multiplicativos. Suponha-se o caso de dois tratamentos repetidos em dois blocos. No modelo aditivo a diferença entre o bloco 1 e o bloco 2 é uma quantidade fixa, qualquer que seja o tratamento. Da mesma forma, a diferença entre o tratamento A e o tratamento B é também uma quantidade fixa, qualquer que seja o bloco. O modelo multiplicativo, a diferença entre os blocos 1 e 2 é uma percentagem fixa, independente do tratamento. Do mesmo modo, a diferença entre os tratamentos A e B é uma percentagem fixa, independentemente do bloco. O teste de Tukey para aditividade é empregado para vários objetivos, tais como auxiliar na decisão sobre a necessidade de transformação dos dados, sugerir a transformação mais conveniente e verificar se a transformação feita foi eficiente em produzir aditividade. O teste está teoricamente relacionado com uma transformação matemática dos dados experimentais para a forma  $Y = X^P$ , onde X representa dados na escala original e Y representa os mesmos dados na escala transformada. O problema consiste em encontrar o valor da potência P tal que os efeitos sejam aditivos na escala transformada  $Y = X^P$ . Se, por exemplo,  $P = \frac{1}{2}$ , isto representa a transformação dos dados em  $\sqrt{X}$ . Se  $P = -1$ , os X são transformados em  $\frac{1}{X}$ , o que implica em analisar

os recíprocos de X em vez de X. Quando P = 0 deve-se interpretar como uma transformação logarítmica porque a variável  $X^P$  se comporta como Log X quando P é muito pequeno. O argumento matemático do teste se baseia no cálculo integral. (NUNES, 1998).

**TRANSFORMAÇÃO DE BOX-COX, OU TRANSFORMAÇÃO DE POTÊNCIA GENERALIZADA:** A transformação de Box-Cox, ou transformação de potência generalizada, é na verdade um tipo de transformação da família expressa pela equação,

$$Y^* = \frac{(Y^\lambda - 1)}{\lambda} \quad (\text{for } \lambda \neq 0)$$

$$Y^* = \log_e(Y) \quad (\text{for } \lambda = 0)$$

onde  $\lambda$  é o número ou valor que maximiza a função log da verossimilhança:

$$L = -\frac{v}{2} \log_e(s_r^2) + (\lambda - 1) \frac{v}{n} \sum_{i=1}^n \log_e Y, \text{ onde } v \text{ são os graus de liberdade, } n \text{ é o tamanho da amostra, e } s_r^2 \text{ é a}$$

variância dos valores de Y transformados (BOX; COX, 1964, citados por GOTELLI; ELLISON, 2010). O valor resultante de  $\lambda$ , que resulta quando a equação  $L = -\frac{v}{2} \log_e(s_r^2) + (\lambda - 1) \frac{v}{n} \sum_{i=1}^n \log_e Y$  é maximizada, é usado na equação

$$\begin{cases} Y^* = \frac{(Y^\lambda - 1)}{\lambda} & (\text{for } \lambda \neq 0) \\ Y^* = \log_e(Y) & (\text{for } \lambda = 0) \end{cases}$$

para fornecer o melhor ajuste dos dados transformados a uma distribuição normal. A equação

$$L = -\frac{v}{2} \log_e(s_r^2) + (\lambda - 1) \frac{v}{n} \sum_{i=1}^n \log_e Y$$

deve ser resolvida iterativamente (tentando diferentes valores de  $\lambda$  até L ser maximizado), usando programas de computador. Determinados valores de  $\lambda$  correspondem às transformações já descritas. Quando  $\lambda = 1$ , a equação

$$\begin{cases} Y^* = \frac{(Y^\lambda - 1)}{\lambda} & (\text{for } \lambda \neq 0) \\ Y^* = \log_e(Y) & (\text{for } \lambda = 0) \end{cases}$$

resulta em uma transformação linear (operação de deslocamento), quando  $\lambda = \frac{1}{2}$ , o resultado é a transformação da raiz quadrada, quando  $\lambda = 0$ , o resultado é a transformação logarítmica natural e quando  $\lambda = -1$ , o resultado é a transformação inversa. Antes de partir para o problema de maximizar a equação

$$L = -\frac{v}{2} \log_e(s_r^2) + (\lambda - 1) \frac{v}{n} \sum_{i=1}^n \log_e Y, \text{ deve-se tentar transformar os dados usando transformações aritméticas}$$

simples. Se os dados têm assimetria à direita, deve-se usar as transformações mais familiares a partir da série

$$\frac{1}{\sqrt{Y}}, \sqrt{Y}, \ln(Y) e \frac{1}{Y}. \text{ Se têm assimetria à esquerda, tente } Y^2, Y^3, \text{ etc. (SOKAL; ROHLF, 1995, citados por}$$

GOTELLI; ELLISON, 2010).

**TRANSFORMAÇÃO ATRAVÉS DA FUNÇÃO DISCRIMINANTE LINEAR DE FISHER:** Campos et al. (2013), transformando os dados por meio da função discriminante linear de Fisher (FDF), para posterior análise, concluíram que este tipo de transformação mostrou-se como uma técnica viável para apurar ou detectar diferenças significativas. Ainda de acordo com estes autores, trabalhando com mudas de cafeeiro, estimaram a função discriminante linear de Fisher como uma função linear nas características da qualidade de mudas, representada por:  $FDF = b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6 + b_7X_7$ , em que  $X_i$ , com  $i = 1,2,3,4,5,6,7$  representa cada uma das sete características, e  $b_i$  com  $i = 1,2,3,4,5,6,7$ , são os coeficientes ou pesos a ser determinados pelo método proposto. Após a estimação dos coeficientes da FDF dada por  $FDF = b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6 + b_7X_7$ , substituiu-se os valores observados de cada característica para obtenção da nova variável (FDF), que explica grande parte das informações contida nas características avaliadas. Os valores obtidos dessa nova variável foram submetidos a análise de variância conforme esquema proposto para cada característica. Segundo ainda os autores para construção da função discriminante linear de Fisher foi utilizado o autovetor t, que é associado ao máximo autovalor e que maximiza a razão  $\frac{t'Ht}{t'Rt}$ , em que H e R são respectivamente as matrizes da soma de quadrados e da soma

de produtos devidos aos efeitos dos tratamentos e dos resíduos, conforme análise multivariada apresentada por Padovani e Aragon (2005).

**TRANSFORMAÇÕES DE DADOS MULTIVARIADOS:** Conforme Resende (2007), no contexto dos modelos multivariados, as variáveis correlacionadas podem ser transformadas em variáveis não correlacionadas, visando à realização de análises univariadas em vez de uma multivariada, porém sem perda de acurácia. Dentre os procedimentos de transformações, destacam-se: i) transformação canônica, que é adequada quando todos os caracteres são avaliados em todos os indivíduos, ou seja, quando as matrizes de incidência X e Z são as mesmas para todos os caracteres e ii) transformação Cholesky, quando observações de alguns caracteres são perdidas e a perda é sequencial. Segundo ainda este autor, devido à dificuldade de convergência e alto custo da análise sob modelo multivariado, o procedimento de transformação de variáveis originais em variáveis não correlacionadas geneticamente e residualmente, é muito interessante. Tal procedimento é denominado transformação canônica (THOMPSON, 1977, citado por RESENDE, 2007) e atua por meio da decomposição das matrizes de covariância genética e residual entre caracteres, usando matrizes de transformação. Sendo  $Var(y) = Var(a) + Var(e) = G + R$  e sendo G e R matrizes simétricas e positivas definidas, existe uma matriz de transformação T tal que  $TRT' = I$  e  $TGT' = H$ , em que I é uma matriz identidade e H é uma matriz diagonal. Assim, fazendo-se essas multiplicações, R se torna igual a identidade e G se torna diagonal, ou seja, não existem mais covariâncias explícitas em R e G. Multiplicando os dados originais por T, obtém-se o vetor  $y^*$  das variáveis não correlacionadas. Análises univariadas são então aplicadas sobre  $y^*$  e os resultados (u) são convertidos para a escala original, produzindo valores genéticos idênticos aqueles obtidos sob o modelo multivariado (m). As expressões de



conversão são dadas por  $\hat{b}_m = T^{-1}\hat{b}_u^*$  e  $\hat{a}_m = T^{-1}\hat{a}_u^*$ . É importante mencionar que, para a realização das análises univariadas sobre  $y^*$ , a herdabilidade é dada por  $h_a^2 = \frac{\sigma_a^2}{(\sigma_a^2 + 1)}$  e o coeficiente lambda associado ao fator shrinkage nas equações de modelo misto é dado por  $\lambda_1 = \frac{1}{\sigma_a^2}$ , visto que a variável transformada

apresenta variância residual unitária. Segundo ainda Resende (2007), outro tipo de transformação que pode ser usada é a transformação Cholesky. Tal transformação produz variáveis com correlação residual nula e variância unitária. A matriz de transformação  $L^{-1}$  é obtida por meio da decomposição Cholesky de R. Tal decomposição é dada por  $R = LL'$ , em que L é uma matriz triangular inferior. A matriz de transformação é a inversa de L. A decomposição de Cholesky é uma especialização da decomposição LU (em matriz triangular inferior L e superior U), a qual é uma consequência do método de eliminação de Gauss ou de absorção de uma linha da matriz por vez, via operações elementares sobre linhas da matriz, em processo que culmina com a triangularização. No processo de eliminação de Gauss, os elementos diagonais são denominados pivôs. A transformação de variáveis via decomposição Cholesky é equivalente ao processo de condensação pivotal (AITKEN, 1937, citado por RESENDE, 2007) usado na obtenção de variáveis canônicas.

**TRANSFORMAÇÕES LINEARES:** As transformações lineares do tipo  $\hat{y}_i = \hat{a} + \hat{b}x_i$ , permitem simplificar em certos casos, a redução dos dados relativos a uma ou mais variáveis e também a inferência estatística relativa às médias e às variâncias, incluindo a análise da variância (ANOVA). No entanto estas transformações não apresentam interesse no que se refere à aplicabilidade dos métodos estatísticos, pois não modificam a normalidade nem a não normalidade das distribuições, nem a igualdade ou desigualdade das variâncias, nem a linearidade ou não linearidade de uma eventual regressão.

**TRANSFORMAÇÃO RAIZ QUADRADA:** A transformação raiz quadrada dada por  $y = \sqrt{x}$ , permite estabilizar as variâncias sempre que há proporcionalidade entre a variância e a média da variável inicial. Este é um caso particular em que os dados tem distribuição de Poisson. Sendo assim nestas condições, temos o seguinte:  $\sigma_x^2 = k m_x$ , de tal forma que

$$\sigma_y^2 \approx k m_x \left( \frac{d\sqrt{X}}{dX} \right)_{m_x}^2 = k m_x \left( \frac{1}{2\sqrt{m_x}} \right)^2 = \frac{k}{4}. \text{ Esta transformação é útil sobretudo para as variáveis}$$

aleatórias discretas ou descontínuas, possuindo distribuições semelhantes às distribuições de Poisson. Quando os valores observados são no conjunto, relativamente baixos, a prática mostra que é preferível utilizar a transformação  $y = \sqrt{x + \frac{1}{2}}$ , ou até, segundo alguns autores como Anscombe (1948) e De Munter

(1958) a fórmula pode ser  $y = \sqrt{x + \frac{3}{8}}$ . Vale lembrar que é por intermédio de uma transformação raiz

quadrada que a aproximação para a distribuição normal é utilizada no caso das distribuições qui quadrado ( $\chi^2$ ) com mais de 30 graus de liberdade. Frequentemente esta transformação é utilizada para dados de contagem como o número de insetos mortos ou capturados em armadilhas luminosas, número de plantas com uma determinada sintomatologia, número de frutos com determinada característica ou sintoma de uma doença, número de ovos de determinado tipo, etc. Esses dados discretos, normalmente seguem a distribuição teórica de probabilidade de Poisson, na qual a média é igual a variância. Nesse caso a transformação adequada para homogeneizar as variâncias dos tratamentos, de tal sorte que os dados passem

a apresentar uma distribuição normal é a raiz quadrada. A transformação básica para os dados nas condições especificadas consiste simplesmente na extração da raiz quadrada dos dados, e em seguida procede-se a análise da variância. Se os dados apresentarem valores zero, recomenda-se acrescentar ao valor genérico X que representa a parcela, uma constante k, que normalmente assume 0,5 ou 1,0 ( $\sqrt{x+0,5}$  ou  $\sqrt{x+1,0}$ ), podendo, entretanto, assumir outros valores, desde que, logo após a transformação, se aplique os testes que comprovem a melhor homogeneidade dos tratamentos, além da redução do valor do coeficiente de variação. (SILVA; SILVA, 1999).

**TRANSFORMAÇÃO EM MODELOS DE REGRESSÃO:** Almeida (2005), afirma que quando se quer usar uma regressão curvilínea a forma mais direta é fazer uma inspeção visual nos dados. Fazer um gráfico de pontos, se o gráfico não se apresenta como linear (alguma indicação de linearidade), emprega-se então uma regressão curvilínea, ou quando há outras razões para suspeitar que as relações não são lineares, como por exemplo, fenômenos claramente modelados por leis de potência, ou lei exponencial, etc, as relações devem ser convertidas para formas lineares. Segundo o autor existem muitos tipos possíveis de regressão curvilínea, baseados numa variedade de relações entre as variáveis:  $y = bx^a$ ,  $y = \frac{a+b}{X}$ ,  $y = ab^x$ .

Existem várias outras possibilidades. Transformações para relações lineares podem ser usadas assim: use qualquer transformação que leve a representar a relação através de funções de forma linear, como logaritmos, multiplicação, divisão, etc. Por exemplo, quer se obter o modelo  $y = a + b \cdot X$ , ou uma forma similar, então emprega-se tais modificações. Na transformação usar alguma função da variável resposta y em lugar do próprio y. Regressão curvilínea é um exemplo dessa transformação

$$Y = \frac{a}{b} x^{\alpha+1}, \ln Y = \ln\left(\frac{a}{b}\right) + (\alpha+1)\ln x \text{ e } Y' = A + Bx'. \text{ Conforme ainda Almeida (2005), as técnicas tem}$$

aplicação mais geral como outras transformações lineares sendo não linear em linear.

$$Y = a + \frac{b}{x} \Rightarrow Y = a + b\left(\frac{1}{x}\right); Y = \frac{1}{(a+bx)} \Rightarrow \frac{1}{Y} = a + bx; Y = \frac{x}{(a+bx)} \Rightarrow \left(\frac{x}{y}\right) = a + bx; \text{ ou ainda outros}$$

casos tais como,  $Y = axb^x \Rightarrow \ln Y = \ln a + x \ln b$  ou ainda,  $Y = a + bx^n \Rightarrow Y = a + b(x^n)$ . O autor cita as condições de quando transformar, ou seja, quando as propriedades conhecidas do sistema medido sugerem uma transformação, quando o intervalo dos dados medidos cobrem várias ordens de grandeza e quando a hipótese de uma variância homogênea dos resíduos é violada. A transformação devido a homocedasticidade deve ser feita, se num gráfico de pontos a diferença dos resíduos versus a resposta prevista não é homogênea. Então os resíduos são ainda uma função das variáveis previsoras. A transformação da resposta pode resolver o problema. Então qual transformação deve ser usada, e assim segundo o autor verifica-se o seguinte: i) calcule o desvio padrão dos resíduos; ii) coloque num gráfico de pontos esses desvios como função da média das observações; iii) considere múltiplos experimentos para um conjunto de valores previsores; iv) verifique a linearidade, se há essa linearidade, então use a transformação logarítmica. Outros testes podem ser usados para verificar a necessidade de transformação: i) se a variância versus a média das observações medidas é linear, use uma transformação de raiz quadrada; ii) se o desvio padrão versus o quadrado médio é linear, use uma transformação inversa; iii) se o desvio padrão versus a média elevada a uma potência é linear use uma transformação de potência. O princípio geral das transformações é para uma função observada  $S = g(\bar{y})$  se  $h(y) = \int \frac{1}{g(y)} dy$  transforme para  $w = h(y)$ . Por exemplo, uma transfor-

mação logarítmica, se o desvio padrão versus a média é linear, então  $g(y) = ay$ , assim  $h(y) = \int \frac{1}{ay} dy = a \ln y$



Em estudos de regressão segundo este autor na verificação de não linearidade desenhe o gráfico de pontos, se não for linear, verifique as possibilidades curvilineares e suas transformações. O uso de uma regressão linear quando as relações entre resposta e previsores não são lineares é um erro.

**TRANSFORMAÇÃO DE DADOS OU DE VARIÁVEIS ATRAVÉS DO USO DE VARIÁVEIS CENTRADAS (CENTERED), PADRONIZADAS OU TAMBÉM ESTANDARTIZADAS (STANDARDIZED):**

Matos (1995), descreve a importância do uso de transformação de dados afirmando que em alternativa ao uso das variáveis originais (raw), podem ser usadas variáveis centradas (centered), estandardizadas (standardized) ou com norma unitária (unit length), obtidas através das transformações indicadas a seguir. Todos estes procedimentos visam compatibilizar, de algum modo, variáveis que podem ter escalas e dispersões muito diferentes. Em particular, a comparação da influência relativa das diversas variáveis, com base nos parâmetros estimados, só faz sentido se as variáveis forem normalizadas. Como se verá noutro local deste texto, os resultados obtidos depois de qualquer das transformações que se descrevem a seguir são sempre iguais aos da versão com os dados originais. Também os parâmetros têm relações simples entre si, permitindo passar facilmente de uma formulação a outra. Por exemplo, na centragem uma transformação simples consiste em centrar cada variável em relação à sua média. A variável transformada  $M_k$  obtém-se de  $X_k$  através de:  $m_{ik} = x_{ik} - \bar{X}_k$ . Semelhantemente ao que se fez para  $X$ , também aqui se define  $M = [m_1, m_2, \dots, m_p]$ . Estandartização: A estandardização corresponde a uma transformação para média nula e desvio padrão unitário de cada variável original  $X_k$ . A nova variável  $Z_k$  é obtida através de:  $z_{ik} = \frac{x_{ik} - \bar{X}_k}{s_k}$ . Neste caso, define-se  $Z = [z_1, z_2, \dots, z_p]$ . Norma unitária: Esta transformação substitui os

valores de cada variável  $X_k$  por uma nova variável  $W_k$ , obtida pela seguinte regra:  $w_{ik} = \frac{x_{ik} - \bar{X}_k}{d_k}$ . Definindo

aqui também  $W = [w_1, w_2, \dots, w_p]$ , verifica-se que a matriz  $W'W$  apresenta diagonal unitária (daí o nome da transformação). Os restantes elementos  $(W'W)_{uv}$  correspondem à correlação entre  $X_u$  e  $X_v$ . Note-se ainda que  $Z'Z = (n-1)W'W$ . A estimação de parâmetros: A estimativa não tendenciosa de  $b$  pelo método dos mínimos quadrados é dada por:  $\hat{b} = (X_a' X_a)^{-1} X_a' y$ . No caso de variáveis centradas, estandardizadas ou de norma unitária, o processo de obtenção da estimativa dos parâmetros  $b_0$  utiliza uma expressão análoga à anterior, substituindo-se  $X_a$  respectivamente por  $M$ ,  $Z$  ou  $W$ . A estimativa de  $a$  é, em todos esses casos, igual à média de  $Y$ . Os valores de  $\hat{b}_k$  obtidos se as variáveis forem centradas são iguais aos do caso geral. Para variáveis estandardizadas e de norma unitária, cada  $\hat{b}_k$  vem multiplicado respectivamente por  $s_k$  e  $d_k$  em relação ao caso geral. A menos de erros de arredondamento, os valores estimados com qualquer dos modelos são rigorosamente correspondentes. Segundo ainda o autor na análise de resíduos para verificar a adequação do ajuste do modelo de regressão o pesquisador tem que verificar que de acordo com os pressupostos da regressão, os resíduos devem distribuir-se aleatoriamente em torno de 0, tanto no modelo global como em relação a cada variável. Caso tal não se verifique, será normalmente necessário alterar o modelo, incluindo ou retirando variáveis, ou realizando alguma transformação que adeque melhor o modelo aos dados (por exemplo,  $X_k^2$  em vez de  $X_k$ ). Verificação de pressupostos: apresentam-se, a seguir, alguns testes que permitem verificar se os pressupostos em relação aos erros do modelo são verificados pelos resíduos. Trata-se de verificações a posteriori que poderão levar à revisão do modelo. Aleatoriedade: uma forma corrente de verificar a aleatoriedade dos resíduos é o teste às sequências de sinais dos resíduos, através do runs test (teste de corridas), importante sobretudo quando as observações dependem do tempo. Considerando apenas os sinais (+ ou -) dos resíduos, pela ordem em que foram recolhidos, haverá  $n_1$  sinais (+),  $n_2$  sinais (-) e  $r$  corridas (sequências máximas de sinais iguais seguidos).

Na sequência (+ - - + + + - - - + + -), por exemplo, será  $n_1=7$ ,  $n_2=6$  e  $r=6$ . Usando em seguida tabelas para o runs test, determinam-se valores críticos que ajudam a determinar, com nível de significância 5%, se a sequência é ou não aleatória. Em função de  $n_1$  e  $n_2$ , as tabelas dão dois valores (inferior e superior) que terão que enquadrar o valor de  $r$ . Caso contrário, suspeita-se de não-aleatoriedade. No caso do exemplo, os dois valores são 3 e 12, concluindo-se pela aleatoriedade, uma vez que  $3 \leq r \leq 12$ . Correlação sucessiva: a verificação de independência é usualmente feita através do teste de Durbin-Watson à correlação entre resíduos sucessivos. O teste é útil, sobretudo em dados dependentes do tempo. Heterocedasticidade: a detecção de desigualdades de variância dos erros pode ser realizada a partir de um gráfico dos resíduos  $r_i$  em função dos  $\hat{y}_i$ . Se o aspecto não for uma mancha de largura uniforme, por exemplo alargando com o aumento de  $\hat{y}_i$ , poderá ser necessário transformar  $Y$  ( $\ln Y$ ,  $\frac{1}{Y}$ , etc.) ou alterar o modelo. Um gráfico semelhante, mas dos quadrados dos resíduos, pode confirmar suspeitas e ajudar a detectar isolados. Normalidade: a verificação visual da normalidade é feita ordenando os resíduos de forma crescente, e desenhando-os em papel de distribuição normal. Se a presunção de normalidade se verificar, os resíduos deverão estar aproximadamente em linha reta. Expressão do modelo: são úteis alguns gráficos de resíduos em relação a variáveis, para verificação visual da correção da expressão do modelo. Os gráficos potencialmente mais interessantes são: resíduos em função das variáveis. Permitem verificar se é necessário transformar as variáveis ( $\ln X$ ,  $\sqrt{X}$ , etc.). Resíduos em função de produtos de variáveis. No caso de ser detectado um padrão, deve ser incluído no modelo um novo termo com o produto em causa ( $X_u X_v$ , por exemplo); Resíduos parciais. Gráfico dos resíduos obtidos sem incluir  $X_k$  em função de  $X_k$ . Permitem detectar não-linearidades que levem à transformação de  $X_k$ . Se o ajuste for bom, o gráfico tem o aspecto de uma reta com inclinação igual ao parâmetro da variável na regressão.

**TRANSFORMAÇÃO LOGIT:** Na transformação logit utiliza-se uma propriedade interessante da função logística a qual é a possibilidade dela poder ser linearizada. Denotando-se  $E(Y)$  por  $\pi$ , pois a resposta média é a probabilidade quando a variável resposta é binária. Fazendo-se a transformação:  $\pi' = \log_e \left( \frac{\pi}{1-\pi} \right)$

então, obtém-se  $\pi' = \beta_0 + \beta_1 X$ . Esta transformação é chamada de transformação logit da probabilidade  $\pi$ . A razão  $\pi/(1-\pi)$  na transformação logit é chamada de Odds (Chance). A função resposta transformada  $\pi' = \beta_0 + \beta_1 X$  é denominada como função resposta logit, e  $\pi'$  é denominada de resposta média logit. Observe em  $\pi' = \beta_0 + \beta_1 X$  que:  $-\infty \leq \pi' \leq \infty$  para  $-\infty \leq X \leq \infty$ . (AYRES et al., 2007).

**TRANSFORMAÇÃO PROBIT:** Para obter a transformação probit considere que uma função de resposta curvilínea com a mesma forma da função logística  $E(Y) = \frac{\exp(\beta_0 + \beta_1 X)}{1 + \exp(\beta_0 + \beta_1 X)}$  ou equivalente a  $E(Y) = [1 + \exp(-\beta_0 - \beta_1 X)]^{-1}$ , é obtida transformando  $\pi$  por meio da distribuição normal acumulada. Esta transformação é chamada de transformação probit. O modelo de regressão probit é menos flexível do que a regressão logística pois não pode ser diretamente aplicada com mais de uma variável preditora. A distribuição de probabilidade acumulada é dada por:  $P(T < X) = \Phi(\beta_0 + \beta_1 X)$ . As transformações logit e probit são muito utilizadas em inferência estatística para duas dimensões na interpretação dos resultados de ensaios biológicos como, por exemplo, em ajuste de modelos de regressão. (AYRES et al., 2007).

**TRANSFORMAÇÃO COMPLEMENTO LOG LOG:** Outra transformação bastante aplicada é aquela onde uma outra função diferente da logística de resposta curvilínea é usada e ela é denominada de transformação complemento log log da probabilidade  $\pi$  dada por:  $\log_e(-\log_e(1-\pi))$ . Diferentemente das transformações logit e probit esta transformação não é simétrica em torno de  $\pi = 0,5$ . (AYRES et al., 2007).

**TRANSFORMAÇÃO RECÍPROCA:** Tipo de transformação própria para dados onde a média ao quadrado e o desvio padrão são proporcionais, como nos casos em que a variável dependente é tempo de reação. Usa-se a variável  $Y = \frac{1}{x}$  ou  $Y = \frac{1}{(x+1)}$  quando alguns valores são zero.

**TRANSFORMAÇÃO DE YEO-JOHNSON:** A transformação de Yeo-Johnson criada por Yeo e Johnson (2000) se justifica pois segundo estes autores a contribuição de Box-Cox (1964) foi o maior passo na determinação de uma maneira objetiva de se efetuar transformação de dados, entretanto, como a transformação de Box-Cox é válida apenas para valores positivos de X, havia espaço para algum tipo de melhoria. Embora seja possível efetuar uma troca de parâmetros, em caso de valores negativos para utilização da transformação de Box-Cox, existe o inconveniente de tal ação afetar a teoria que suporta a definição do intervalo de confiança de  $\lambda$ . Yeo e Johnson (2000) propuseram uma nova família de transformação de dados, válida tanto para valores positivos como para valores negativos da variável X. Sua fórmula, definida como uma função  $\Psi : R \times R \rightarrow R$ , é apresentada a seguir:

$$\Psi^{(\lambda, x)} = \begin{cases} \left\{ \frac{(X+1)^\lambda - 1}{\lambda} \right\} & (X \geq 0, \lambda \neq 0) \\ \log(X+1) & (X \geq 0, \lambda = 0) \\ -\left\{ \frac{(-X+1)^{2-\lambda} - 1}{2-\lambda} \right\} & (X < 0, \lambda \neq 2) \\ -\log(-X+1) & (X < 0, \lambda = 2) \end{cases}$$

Fica claro que outros tipos de situações podem ocorrer e assim poder escolher outros tipos de transformações. De uma maneira geral, no caso de uma relação linear do coeficiente b entre o logaritmo das variâncias e o logaritmo das médias, a transformação a ser utilizada é, desde que b seja diferente de 2, a seguinte:  $Y = X^{1-\frac{b}{2}}$ . Nestas condições, tem-se, com efeito, o resultado dado por:  $\log \sigma_x^2 = b \log m_x + \log k$ , onde

$$\sigma_x^2 = k m_x^b \text{ e } \sigma_y^2 \cong k m_x^b \left( \frac{dX^{1-\frac{b}{2}}}{dX} \right)_{m_x}^2 = k \left( 1 - \frac{b}{2} \right)^2 = k'. \text{ Muitos exemplos destas transformações são mostrados}$$

particularmente por Taylor (1961). Métodos mais rigorosos foram igualmente propostos por Box e Cox (1964) e por Kruskal (1965), para realizar de maneira empírica a escolha de uma transformação ótima. Na prática corrente, estes métodos são, contudo muitos trabalhosos e não parece necessário levar de forma demasiada e longa a preocupação em utilizar uma transformação ideal. Mas independentemente das dificuldades de cálculo, o principal inconveniente das mudanças de variáveis reside, com efeito, em complicar a interpretação dos resultados. Esta dificuldade é de importância relativamente secundária para as transformações raiz quadrada e logarítmica, por exemplo, mas pode ser muito mais considerável em transformações mais complexas. Sendo assim parece razoável, limitarmo-nos, tanto quanto possível, as transformações mais simples.

**TRANSFORMAÇÕES POTÊNCIA:** De acordo com Murteira (1993), as transformações mais empregadas na análise exploratória e na descrição de dados de observação, são as transformações potência, e elas se caracterizam por ser funções elementares, pois os valores transformados calculam-se facilmente, são estritamente crescentes, contínuas e regulares, com derivadas de todas as ordens. Em consequência da monotonia, permitem preservar a ordem dos dados. Isto é, a mediana transformada é a mediana da

coleção transformada, os quartis transformados são os quartis da coleção transformada, etc. Em suma, as medidas resumo transformadas são com pequenas diferenças resultante das interpolações, as medidas resumo da coleção transformada. A forma geral das transformações potência é a seguinte:

$$\begin{cases} t_p(x) = a x^p + b, & P \neq 0 \\ t_p(x) = c \ln x + d, & P = 0 \end{cases}, \text{ onde } \ln \text{ designa o logaritmo neperiano, } a, b, c, d \text{ e } P \text{ são números reais com}$$

a restrição  $P > 0 \Rightarrow a > 0$  e  $P < 0 \Rightarrow a < 0$ . A escolha de P decorre dos objetivos da análise dos dados.

**TRANSIÇÃO EPIDEMIOLÓGICA:** Teoria que considerava a exigência de um primeiro estágio, refere-se a idade das pestilências e da fome, um segundo estágio alusivo ao declínio das pandemias e um terceiro relatando as doenças degenerativas e criadas pelo homem. Segundo esta teoria, todos os países evoluíram, passando por estes estágios, com ganhos consequentes na expectativa de vida e redução na importância das doenças infecciosas.

**TRANSFORMAÇÃO NA APLICAÇÃO DA TÉCNICA DE REGRESSÃO PONDERADA:** Segundo Charnet et al. (1999), o uso de transformação na aplicação da técnica de regressão ponderada pode ser explicada da seguinte maneira: suponha que um modelo de regressão linear múltipla (MRLM) é ajustado e os procedimentos de adequação indicam que a variância do erro não é constante. Pode-se tentar ajustar o modelo:  $y = X\beta + \varepsilon$ ,  $\varepsilon \sim N(\mathbf{0}; \sigma^2 V)$ , sendo V matriz diagonal, dada a independência entre erros, com diagonal principal igual ao vetor de variâncias dos erros, a menos do fator  $\sigma^2$ ,

$$V = \begin{bmatrix} v_1 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & v_2 & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & v_n \end{bmatrix}$$

Essa modelagem permite a definição de variâncias diferentes. Note que, se  $V = I$ , tem-se um modelo de regressão linear múltipla (MRLM). Para que se possa usar todos os resultados obtidos para o modelo de regressão linear múltipla (MRLM), será feita uma transformação na variável y que, posteriormente, será revertida para a escala original. Seja

$$W = \begin{bmatrix} \frac{\mathbf{1}}{\sqrt{v_1}} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \frac{\mathbf{1}}{\sqrt{v_2}} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \frac{\mathbf{1}}{\sqrt{v_n}} \end{bmatrix}$$

Note que,

$$WVW = \begin{bmatrix} \frac{\mathbf{1}}{\sqrt{v_1}} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \frac{\mathbf{1}}{\sqrt{v_2}} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \frac{\mathbf{1}}{\sqrt{v_n}} \end{bmatrix} \begin{bmatrix} v_1 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & v_2 & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & v_n \end{bmatrix} \begin{bmatrix} \frac{\mathbf{1}}{\sqrt{v_1}} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \frac{\mathbf{1}}{\sqrt{v_2}} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \frac{\mathbf{1}}{\sqrt{v_n}} \end{bmatrix} = I$$

Pré-multiplicando ambos os lados da equação do modelo pela matriz W, obtém-se  $Wy = WX\beta + W\varepsilon$ . Sejam:  $y^* = Wy$ ,  $X^* = WX$  e  $\varepsilon^* = W\varepsilon$ . Então, Segundo ainda Charnet et al. (1999) tem-se o modelo  $y^* = X^*\beta + \varepsilon^*$  sendo  $\varepsilon \sim N(\mathbf{0}; \sigma^2 I)$ , porque  $Var[\varepsilon^*] = WVar[\varepsilon]W = W\sigma^2 VM = \sigma^2 WVV = \sigma^2 I$ . Portanto este modelo, que será chamado de modelo transformado, é um modelo de regressão linear múltipla (MRLM). Usando os resultados já apresentados para um modelo de regressão linear múltipla (MRLM), tem-se:  $\hat{\beta}_p = (X^* X^*)^{-1} X^* y^*$ , sendo  $\hat{\beta}_p$  a notação usada para o estimador de quadrados mínimos no modelo transformado. Os valores preditos e o vetor de resíduos do modelo transformado são, respectivamente:  $X^* \hat{\beta}_p$  e  $e_p = y^* - X^* \hat{\beta}_p$ . Assim,

$$\hat{\sigma}_p^2 = \frac{(y^* - X^* \hat{\beta}_p)(y^* - X^* \hat{\beta}_p)'}{n - k}$$

análise de variância (ANOVA) que irá auxiliar na verificação da significância do modelo transformado, e os gráficos de resíduos também devem ser analisados, como, por exemplo, o gráfico dos resíduos,  $e_p$ , versus preditos  $X^* \hat{\beta}_p$ ; os gráficos de normalidade, etc. Sendo o modelo transformado adequado, intervalos de confiança e de predição referentes aos valores das variáveis regressoras, na escala transformada, dados por  $x^*$ , que pode ser, por exemplo, uma linha de  $X^*$ , são obtidos, respectivamente, por:

$$x^* \hat{\beta}_p \pm t_{\left(\frac{\alpha}{2}, n-k\right)} \sqrt{\hat{\sigma}_p^2 x^* (X^* X^*)^{-1} x^*} \quad \text{e} \quad x^* \hat{\beta}_p \pm t_{\left(\frac{\alpha}{2}, n-k\right)} \sqrt{\hat{\sigma}_p^2 [1 + x^* (X^* X^*)^{-1} x^*]}$$

intervalos de confiança (I.C.s) para a variável transformada, é necessário fazer a transformação inversa para voltar à escala original. Para tanto, deve-se multiplicar vetores por  $W^{-1}$ , que é a matriz com os inversos dos desvios padrões, ou então tem-se que dividir um escalar pelo desvio padrão correspondente (a menos do fator  $\sigma$ ). Assim, os valores preditos na escala original são:

$\hat{y} = W^{-1} \hat{y}^* = W^{-1} X^* \hat{\beta}_p = W^{-1} WX \hat{\beta}_p = X \hat{\beta}_p$ . Ou seja, se quiser obter estimativas de  $E[Y/x]$ , faz-se o produto  $x' \hat{\beta}_p$ . Intervalos de confiança para os valores das variáveis regressoras, na escala original, dados pelo vetor x, são dados por:

$$\sqrt{v} x' \hat{\beta}_p \pm \sqrt{v} t_{\left(\frac{\alpha}{2}, n-k\right)} \sqrt{\hat{\sigma}_p^2 x^* (X^* X^*)^{-1} x^*} = x' \hat{\beta}_p \pm t_{\left(\frac{\alpha}{2}, n-k\right)} \sqrt{\hat{\sigma}_p^2 \sqrt{v} x^* (X^* X^*)^{-1} \sqrt{v} x^*} =$$

$$x' \hat{\beta}_p \pm t_{\left(\frac{\alpha}{2}, n-k\right)} \sqrt{\hat{\sigma}_p^2 x' (X^* X^*)^{-1} x}, \text{ sendo que, nessa expressão, } x^* = \frac{\mathbf{1}}{\sqrt{v}} x \text{ e } \sqrt{v} \text{ relativo ao desvio}$$

padrão correspondente. Intervalos de predição para os valores das variáveis regressoras, na escala original, dados pelo vetor x, são dados por:

$$\sqrt{v} x' \hat{\beta}_p \pm \sqrt{v} t_{\left(\frac{\alpha}{2}, n-k\right)} \sqrt{\hat{\sigma}_p^2 x^* [1 + x^* (X^* X^*)^{-1} x^*]} = x' \hat{\beta}_p \pm t_{\left(\frac{\alpha}{2}, n-k\right)} \sqrt{\hat{\sigma}_p^2 [v + \sqrt{v} x^* (X^* X^*)^{-1} \sqrt{v} x^*]} =$$

$$x' \hat{\beta}_p \pm t_{\left(\frac{\alpha}{2}, n-k\right)} \sqrt{\hat{\sigma}_p^2 [v + x' (X^* X^*)^{-1} x]}. \text{ Se as variâncias dos erros são heterogêneas, pode-se contornar}$$

esse problema usando a matriz de pesos W, cujos elementos são os inversos dos desvios padrões dos erros, a menos de um fator  $\sigma^{-1}$ . Muitas vezes não se tem essa informação, isto é, não se conhece os desvios padrões. Uma solução nesse caso é usar algum tipo de estimativa para esses desvios padrões. Na prática,

são usadas várias alternativas, das quais serão citadas algumas a seguir: i) se as variâncias aumentam com o valor de certa variável regressora, X, pode-se supor que as variâncias são proporcionais aos valores de X, e então definir os elementos da diagonal principal de W como sendo  $\frac{1}{\sqrt{x}}$ ; ii) se as variâncias aumentam

com o valor de certa variável regressora, X, sendo este aumento muito acentuado a partir de certo valor de X, pode-se supor que as variâncias são proporcionais aos valores de  $X^2$ , e então definir os elementos da diagonal principal de W como sendo  $\frac{1}{x}$  e iii) se tem-se valores repetidos das variáveis regressoras, pode-se estimar as variâncias em cada subgrupo, ou então fazer uma modelagem das variâncias em função de alguma variável regressora. Quando se utiliza algum software estatístico para justar uma regressão ponderada, é necessário informar os pesos utilizados, que são os quadrados dos elementos da diagonal principal de W. Usualmente, nos softwares é usada a denominação WEIGHT ou WEIGHTS para esses pesos. É bastante comum que os softwares apresentem os intervalos de confiança e predição já na escala original. Um cuidado deve ser tomado ao fazer os gráficos de resíduos, é preciso dispor destes gráficos na escala transformada, para conferir o ajuste do modelo nessa escala transformada e muitas vezes os resíduos disponíveis já estão na escala original, neste caso é necessário multiplica-los pela raiz quadrada do peso, antes de se fazer os gráficos apropriados.

**TRANSFORMAÇÃO BOX-COX:** Box-Cox (1964) propuseram um método para a família de transformações potência que fornece: i) estrutura linear simples; ii) constância da variância do erro; iii) independência entre as observações e iv) normalidade. A transformação potência é modificada para que a variável transformada seja contínua em  $\lambda = 0$ . A expressão obtida é a seguinte:

$$Y_i(\lambda) = \begin{cases} \frac{Y_i^\lambda - 1}{\lambda}, & \text{para } \lambda \neq 0 \\ \log Y_i, & \text{para } \lambda = 0. \end{cases}$$

Então,  $Y(\lambda) = (Y_1(\lambda), \dots, Y_n(\lambda))'$  é um vetor de dimensão  $n \times 1$ , podendo-se ajustar o modelo  $Y(\lambda) = X\beta + \varepsilon$  aos dados transformados. O método de máxima verossimilhança de estimação de  $\lambda$  é constituído das três etapas abaixo: i) arbitra-se valores para  $\lambda$ . Os valores de  $\lambda$  são escolhidos num determinado intervalo. Inicialmente, o intervalo pode ser  $\lambda = \{-3, -2, -1, 0, 1, 2, 3\}$ ; ii) calcula-se, para cada valor de  $\lambda$ , o máximo da log-verossimilhança dado por,  $l_{\max}(\lambda) = -\frac{1}{2} n \log(\hat{\sigma}^2(\lambda)) + \log(J(\lambda, Y))$  onde,

$$\hat{\sigma}^2(\lambda) = \frac{Y'(\lambda)(I - X(X'X)^{-1}X')Y(\lambda)}{n}$$

$$J(\lambda, Y) = \prod_{i=1}^n \frac{\partial Y_i(\lambda)}{\partial Y_i}$$

$$J(\lambda, Y) = \prod_{i=1}^n Y_i^{\lambda-1}, \quad \forall \lambda;$$

iii) depois de calcular  $l_{\max}(\lambda)$  para os valores do intervalo, verifica-se se o gráfico de  $l_{\max}(\lambda)$  versus  $\lambda$  contém o ponto de máximo da curva. Se isto ocorrer, o procedimento está terminado e o valor de  $\lambda$  correspondente ao ponto de máximo é o estimador de máxima verossimilhança de  $\lambda$ . Caso contrário, é necessário ampliar o intervalo de variação dos valores para  $\lambda$ . Finalmente, o intervalo de  $100(1 - \alpha)\%$  de confiança para  $\lambda$  é dado por  $\{\lambda : l_{\max}(\hat{\lambda}) - l_{\max}(\lambda) \leq 1/2 \chi_1^2(1 - \alpha)\}$ , onde  $l_{\max}(\hat{\lambda})$  é a ordenada correspondente ao ponto de máximo da curva  $l_{\max}(\lambda)$  versus  $\lambda$ .



**TRANSFORMAÇÃO DE BOX-MÜLLER:** Bussab e Morettin (2002) descrevem a importante transformação de Box-Müller, da seguinte maneira: considere as variáveis aleatórias X e Y, independentes e ambas tendo distribuição normal padrão com média zero (0) e variância igual a um (1), ou seja, N(0,1). É sabido que:  $R^2 = X^2 + Y^2$  e  $tg\theta = \frac{Y}{X}$ . a densidade conjunta de x e y é,

$$f(x, y) = \frac{1}{2\pi} e^{-\frac{(x^2+y^2)}{2}}. \text{ Considere a transformação de variáveis } r = \sqrt{x^2 + y^2}, \text{ e } \theta = \arctg\left(\frac{y}{x}\right). \text{ A densidade}$$

conjunta de r e θ é obtida usando o resultado  $g(z, w) = f(g_1(z, w), g_2(z, w))|J|$ . Tem-se que  $x = \sqrt{r} \cos \theta$

$y = \sqrt{r} \sin \theta$  e o jacobiano da transformação é  $|J| = \frac{1}{2}$ . Segue-se que a densidade de r e θ é

$$f(r, \theta) = \frac{1}{2\pi} e^{-r^2} \cdot \frac{1}{2}, \quad 0 < r < \infty, \quad 0 < \theta < 2\pi. \text{ Dessa relação pode-se concluir que } R = r^2 \text{ e } \theta \text{ são inde-}$$

pendentes, com  $R^2 \sim Exp(2)$ ,  $\theta \sim U(0, 2\pi)$ . Portanto pode-se escrever que,

$$X = R \cos \theta = \sqrt{-2 \log U_1} \cos(2\pi U_2) \text{ e } Y = R \sin \theta = \sqrt{-2 \log U_1} \sin(2\pi U_2). \text{ Aqui usou-se o fato}$$

de que, se  $R^2 \sim Exp(2)$ , gerado um número aleatório (NA)  $NAU_1$ , vem que  $-2 \log U_1 \sim Exp(2)$  e se  $\theta \sim U(0, 2\pi)$ , então gerado um NA  $U_2$ , vem que  $2\pi U_2 \sim U(0, 2\pi)$ . O método de Box-Müller gera valores de duas normais padrões independentes  $Z_1$  e  $Z_2$ . Logo, se o pesquisador quiser gerar valores da distribuição conjunta de X e Y, independentes e normais, com  $X \sim N(\mu_x, \sigma_x^2)$  e  $Y \sim N(\mu_y, \sigma_y^2)$  basta considerar  $X = \mu_x + \sigma_x Z_1$  e  $Y = \mu_y + \sigma_y Z_2$ . Já na simulação de uma distribuição gama, os autores afirmam que se a variável aleatória x possui distribuição gama isto é,  $X \sim Gama(r, \beta)$ , com r inteiro, então  $X = Y_1 + Y_2 + \dots + Y_r$ , onde cada  $Y_i \sim Exp(\beta)$  e as variáveis aleatórias  $Y_i$  independentes. Logo, para gerar um valor de uma distribuição  $Gama(r, \beta)$ , com  $r > 0$ , inteiro, basta gerar r valores de uma variável aleatória de uma distribuição teórica de probabilidade exponencial de parâmetro  $\beta$  e depois somá-los.

**TRANSFORMAÇÕES DE MODELOS DE REGRESSÃO NÃO LINEAR EM MODELOS LINEARES:** São tipos de transformações introduzidas nas variáveis e nos parâmetros de modelos de regressão não linear que são modificados através de transformações ou anamorfose em modelos lineares cujo objetivo e simplificar os cálculos de obtenção das somas de quadrados, e de produtos bem como os valores das estimativas dos parâmetros da equação modelo ajustada. A tabela seguinte mostra diversos tipos de modelos linearizados com suas respectivas transformações:

	Tipo de Função	Coordenadas		Equação na Forma Linear
		Abscissa	Ordenada	
1	$y = a + bx$	x	y	$[y] = a + b[x]$
2	$y = be^{ax}$	x	log y	$[\log y] = \log b + (a \log e)[x]$
3	$y = ax^b$	log x	log y	$[\log y] = \log a + b[\log x]$

	Tipo de Função	Coordenadas		Equação na Forma Linear
		Abscissa	Ordenada	
4	$y = a_0 + a_1x + ax^2$	$x - x_0$	$\frac{y - Y_0}{x - x_0}$	$\left[\frac{y - Y_0}{x - x_0}\right] = a_1 + 2a_1x_0 + a_2[(x - x_0)]$
5	$y = a + \frac{b}{x}$	$\frac{1}{x}$	y	$[y] = a + b\left[\frac{1}{x}\right]$
6	$y = \frac{x}{(a + bx)}$	x	$\frac{x}{y}$	$\left[\frac{x}{y}\right] = a + b[x]$
7	$y = \frac{a}{(b + cx)}$	x	$\frac{1}{y}$	$\left[\frac{1}{y}\right] = \left(\frac{b}{a}\right) + \left(\frac{c}{a}\right)[x]$
8	$y = c + be^{ax}$	x	$\log \frac{\Delta y}{\Delta x}$	$\left[\log \frac{dy}{dx}\right] = \log(ab) + (a \log e)[x]$
9	$y = c + ax^b$	log x	$\log \frac{\Delta y}{\Delta x}$	$\left[\log \frac{dy}{dx}\right] = \log(ab) + (b - 1)[\log x]$
10	$y = c + \frac{b}{x - a}$	$x - x_0$	$\frac{x - x_0}{y - y_0}$	$\left[\frac{x - x_0}{y - y_0}\right] = \frac{a - x}{c - y_0} + \frac{1}{c - y_0}[x - x_0]$
11	$y = c + \frac{x}{a + bx}$	x	$\frac{x - x_0}{y - y_0}$	$\left[\frac{x - x_0}{y - y_0}\right] = (a + bx_0) + \frac{b(a + bx_0)}{a}[x]$
12	$y = d + cx + be^{ax}$	x	$\log \frac{\Delta^2 y}{\Delta x^2}$	$[y - be^x] = d + c[x]$ $\left[\log \frac{d^2 y}{dx^2}\right] = \log(a^2 b) + (a \log e)[x]$
13	$y = dc^x b^m$ , onde $m = a^x$	x	$\log \frac{\Delta^2(\log y)}{\Delta x^2}$	$\left[\log \frac{d^2(\log y)}{dx^2}\right] = \log\left[\frac{(\log b)(\log a)^2}{(\log e)^2}\right] + (\log a)[x]$
14	$y = de^{cx} + be^{ax}$	$\frac{y_{k+1}}{y_k}$	$\frac{y_{k+2}}{y_k}$	$[ye^{-cx}] = d + b[e^{(a-c)x}]$ $[\log y - a^x \log b] = \log d + (\log e)[x]$ $\left[\frac{y_{k+2}}{y_k}\right] = -e^{(a+c)\Delta x} + (e^{a\Delta x} + e^{c\Delta x})\left[\frac{y_{k+1}}{y_k}\right]$
15	$y = e^{ax}(d \cos bx + c \sin bx)$	$\frac{y_{k+1}}{y_k}$	$\frac{y_{k+2}}{y_k}$	$\left[\frac{y_{k+2}}{y_k}\right] = -e^{2a\Delta x} + (2e^{a\Delta x} \cos b\Delta x)\left[\frac{y_{k+1}}{y_k}\right]$ $\left[\frac{yc^{-ax}}{\cos bx}\right] = d + c[\tan bx]$

**Observação:** Nas equações 14 e 15,  $y_k, y_{k+1}, y_{k+2}$  são valores consecutivos para um incremento  $\Delta x$   
 Fonte: Yevjevich (1964) e Naghettini e Pinto (2007).



**TRANSFORMAÇÕES DE MODELOS DE REGRESSÃO NÃO LINEAR EM MODELOS LINEARES ATRAVÉS DE MODIFICAÇÕES OU ANAMORFOSES:** Tipos de transformações utilizadas para estimar parâmetros de modelos não lineares em equações lineares, mediante transformações logarítmicas ou inversa, como por exemplo no modelo exponencial, potencial, hiperbólico, etc.

Intervalos das estimativas dos parâmetros	Tipo de Modelo	Função	Transformações	Forma Linear	Observações
$\alpha, \beta$ e $x$ positivos, e $\alpha, x > 0$ e $\beta < 0$	Função Potência	$y = \alpha x^\beta$	$y' = \log y, x' = \log x$	$y' = \log \alpha + \beta x'$	$y > 0, x > 0$
$\beta > 0$ e $\beta < 0$	Função Exponencial	$y = \alpha e^{\beta x}$	$y' = \ln y$	$y' = \alpha + \beta x'$	$y > 0$
$\beta > 0$	Função Logarítmica	$y = \alpha + \beta \log x$ $y = \alpha + \beta \ln x$	$x' = \log x$	$y = \alpha + \beta x'$	$x > 0$
$\beta < 0$ ; $\alpha, \beta > 0$ e $x > \frac{\beta}{\alpha}$ e $\alpha > 0$ e $x > \frac{\beta}{\alpha}, \beta < 0$	Função Hiperbólica	$y = \frac{x}{\alpha x - \beta}$	$y' = \frac{1}{y}, x' = \frac{1}{x}$	$y' = \alpha - \beta x'$	$y \neq 0, x \neq 0$
-----	Função Logística	$y = \frac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}}$	$y' = \ln\left(\frac{y}{1-y}\right)$	$y' = \alpha + \beta x$	$0 < y < 1$

**TRANSITIVO:** É o modelo de variograma teórico, descrito pelas seguintes equações:

$$\gamma(h) = C_0 + p \cdot h \quad \text{para } h < a \quad \text{e} \quad \gamma(h) = C_0 + C \quad \text{para } h \geq a.$$

**TRANSPOSTA DA MATRIZ A:** Símbolo usado em álgebra matricial dado por  $A'$ .

**TRANSVERSAL, ESTUDOS (CROSS-SECTIONAL STUDY):** Projeto de pesquisa, destinado a estudar alguns fenômenos, por meio de uma amostra representativa deles, num dado momento. Estudos em que se coletam os dados, em dado momento, a partir de uma amostra aleatória de uma população geral que contém duas ou mais subpopulações, com a intenção de comparar os dados das sub-amostras ou notar tendências em tais sub-amostras.

**TRATAMENTO:** Propriedade ou característica que permite distinguir uma das outras diferentes populações; usado em análise de variância.

**TRATAMENTO:** Em estatística, regime, método, ou procedimento testado em um ensaio clínico ou experimento.

**TRATAMENTO:** Variável independente que o pesquisador manipula para verificar se existe efeito sobre a(s) variável (eis) dependente(s), como em um experimento.

**TRATAMENTO:** Variável independente que um pesquisador manipula para verificar se existe efeito sobre as variáveis dependentes. A variável de tratamento pode ter diversos níveis. Por exemplo, diferentes intensidades de apelos de publicidade podem ser manipuladas para ver o efeito sobre a credibilidade do consumidor.

**TRATAMENTO:** Ver estímulo.

**TRATAMENTO:** Em metodologia científica, designa qualquer procedimento, produto ou intervenção passível de mudar o curso dos acontecimentos, não estando, portanto, restrito a fins terapêuticos; em uma experiência, os tratamentos podem ser: a vacina para um grupo e o placebo para o outro.

**TRATAMENTO:** A variável independente que é manipulada em um experimento.

**TRATAMENTO ALOCADO:** Tratamento administrado a um paciente, com indicado no momento em que esse paciente decidiu participar do experimento.

**TRATAMENTO CONTROLE:** Droga, dispositivo, ou procedimento administrado em um ensaio clínico para servir como o padrão contra o qual os tratamentos em teste são avaliados. O tratamento controle pode ser um placebo, um pseudoprocimento, um tratamento padrão ou nenhum tratamento, dependendo do delineamento do estudo.

**TRATAMENTO PADRÃO:** Maneira amplamente aceita de tratar determinada doença ou condição.

**TRATAMENTO PROFILÁTICO:** Tratamento de um caso clínico ou de um portador, com a finalidade de reduzir o período de transmissibilidade.

**TRATO ESTATÍSTICO:** De uma cidade, é sua subdivisão geográfica, com divisas bem definidas e, usualmente, escolhidas de modo a garantir um alto grau de homogeneidade das características sociais e econômicas. Tais divisas são mantidas intactas por meio de uma série de censos, para que se possam fazer comparações. Também se diz área social e zona social.

**TREINAMENTO:** Ver aprendizado.

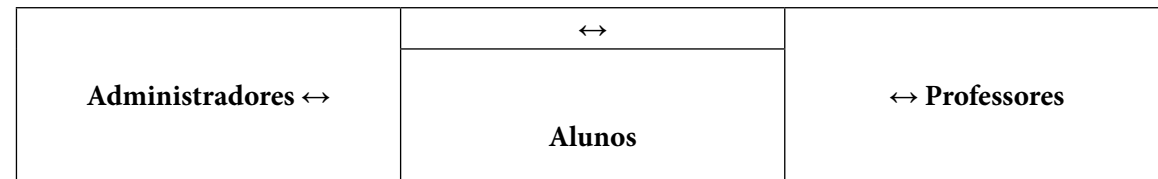
**TRENA:** Instrumento para medição de distância. Consiste em uma fita graduada em centímetros e metros enrolada em um carretel. As mais comuns são de 20, 30 e 50 metros e feitas de fibra de vidro ou aço.

**TRANSPOSIÇÃO (TRANSPOSE):** O rearranjo de uma matriz tal que as linhas da matriz original são iguais às colunas da nova qwue é a matriz transposta e as colunas da matriz original são iguais às linhas da matriz transposta. Simbolicamente, para uma matriz A com elementos  $a_{ij}$ , a matriz transposta  $A^T$  tem os elementos  $a'_{ij} = a_{ji}$ . A transposição de matrizes é indicada ou pelo sobrescrito T ou pela aspa (').

**TRATAMENTO (TREATMENT):** Um conjunto de categorias de variáveis preditoras usadas em um delineamento de análise de variância. Os tratamentos são equivalentes a fatores e são constituídos de níveis.

**TRIANGULAÇÃO:** Aplicação de pelo menos dois métodos na execução de uma pesquisa científica.

**TRIANGULAÇÃO:** É uma validação qualitativa cruzada. Avalia a suficiência dos dados, considerando fontes múltiplas de dados ou múltiplos procedimentos de coleta. É fundamentalmente uma comparação; uma comparação de informações para determinar se há ou não convergência.



**TRUNCAMENTO:** Ponto, em uma distribuição, a partir do qual os animais são retidos ou eliminados.

**TRUNCAMENTO:** Quando se planeja um estudo, interessa formulá-lo para que se verifique o acontecimento de interesse. Dessa forma, só são incluídos no estudo indivíduos a quem ocorreu ou vai ocorrer o acontecimento de interesse. A este mecanismo que consiste em excluir os indivíduos que não são relevantes para o estudo em questão, chama-se truncamento.

**TRUNCAMENTO:** Termo usado em análise de sobrevivência indicando os Indivíduos que não são incluídos por motivo relacionado à ocorrência do evento estudado.

**TRUNCAMENTO À DIREITA:** Ocorre quando só são selecionados indivíduos a quem já tenha ocorrido o acontecimento de interesse. Este tipo de truncamento é frequente em estudos que partem da ocorrência da morte para a seleção dos indivíduos.

**TRUNCAMENTO À ESQUERDA:** Neste tipo de truncamento, os indivíduos não são seguidos a partir do instante inicial relevante para o problema em estudo, mas sim a partir de um instante posterior. Por exemplo, no estudo da sobrevivência de doentes submetidos a hemodiálise, só serão incluídos doentes que se encontravam em tratamento na data predeterminada para o começo do estudo ou os que iniciaram o tratamento numa data posterior.

**T QUADRADO DE HOTTELING:** É um teste multivariado para verificar diferenças de médias para casos onde existam múltiplas variáveis dependentes intervalares e dois grupos formados por variáveis independentes categóricas. Para passar do traço de Hotelling ou traço de Lawley-Hotelling para o T quadrado deve-se multiplicar o traço por (n-l), onde n é o tamanho da amostra considerando todos os grupos e l é o número de subgrupos. Em inglês Hotelling's t-square.

**TAMANHO DE AMOSTRA PARA UM TESTE DE HIPÓTESE UNILATERAL À ESQUERDA PARA A PROPORÇÃO:**  $(H_1 : \pi < \pi_0)$ . A fórmula é dada por:

$$n = \frac{(z_\alpha \sqrt{\pi_0(1-\pi_0)} + z_\beta \sqrt{(\pi_0 - \delta)(1-\pi_0 + \delta)})^2}{\delta^2},$$

onde,  $z_\alpha$  é o valor da distribuição normal padrão que corresponde ao valor de Z acima do qual a área é igual a  $\alpha$ ,  $z_\beta$  é o valor da distribuição normal padrão que corresponde ao valor de Z acima do qual a área é igual a  $1 - poder = \beta$ ,  $\pi_0$  é o valor da proporção populacional supondo-se que a hipótese nula é verdadeira e  $\delta$  é a menor diferença a ser detectada.

**TAMANHO DE AMOSTRA PARA UM TESTE DE HIPÓTESE UNILATERAL À DIREITA PARA A PROPORÇÃO:**  $(H_1 : \pi > \pi_0)$ . A fórmula é dada por:

$$n = \frac{(z_\alpha \sqrt{\pi_0(1-\pi_0)} + z_\beta \sqrt{(\pi_0 + \delta)(1-\pi_0 - \delta)})^2}{\delta^2},$$

onde,  $z_\alpha$  é o valor da distribuição normal padrão que corresponde ao valor de Z acima do qual a área é igual a  $\alpha$ ,  $z_\beta$  é o valor da distribuição de Gauss, normal padrão ou simétrica em forma de sino, que corresponde ao

valor de Z acima do qual a área é igual a  $1 - poder = \beta$ ,  $\pi_0$  é o valor da proporção populacional supondo-se que a hipótese nula ou de nulidade é verdadeira e  $\delta$  é a menor diferença a ser detectada entre tratamentos.

**TAMANHO DE AMOSTRA PARA UM TESTE DE HIPÓTESE BILATERAL PARA A PROPORÇÃO:**

$(H_1 : \pi \neq \pi_0)$ . A fórmula é dada por:

$$n = \frac{\left( \frac{z_\alpha}{2} \sqrt{\pi_0(1-\pi_0)} + z_\beta \sqrt{(\pi_0 - \delta)(1-\pi_0 + \delta)} \right)^2}{\delta^2} \text{ e}$$

$$n = \frac{\left( \frac{z_\alpha}{2} \sqrt{\pi_0(1-\pi_0)} + z_\beta \sqrt{(\pi_0 + \delta)(1-\pi_0 - \delta)} \right)^2}{\delta^2}$$

onde,  $z_\alpha$  é o valor da distribuição normal padrão que corresponde ao valor de Z acima do qual a área é igual a  $\frac{\alpha}{2}$ ,  $z_\beta$  é o valor da distribuição normal padrão que corresponde ao valor de Z acima do qual a

área é igual a  $1 - poder = \beta$ ,  $\pi_0$  é o valor da proporção populacional supondo-se que a hipótese nula é verdadeira e  $\delta$  é a menor diferença a ser detectada. O tamanho da amostra n será o maior dos dois valores de  $n_1$  e  $n_2$  obtidos através das fórmulas anteriores.

**TAMANHO DE AMOSTRA PARA UM TESTE DE HIPÓTESE UNILATERAL PARA A COMPARAÇÃO DE DUAS PROPORÇÕES POPULACIONAIS A PARTIR DE AMOSTRAS INDEPENDENTES:**

$(H_1 : \pi_{11} > \pi_{21})$  e  $(H_1 : \pi_{11} < \pi_{21})$ . A fórmula é dada por:

$$n = \frac{(z_\alpha + z_\beta)^2 [\pi_{11}(1-\pi_{11}) + \pi_{21}(1-\pi_{21})]}{\delta^2},$$

onde,  $z_\alpha$  é o valor da distribuição normal padrão que corresponde ao valor de Z acima do qual a área é igual a  $\alpha$ ,  $z_\beta$  é o valor da distribuição normal padrão que corresponde ao valor de Z acima do qual a área é igual a  $1 - poder = \beta$ ,  $\pi_{11}$  e  $\pi_{21}$  são definidos como proporções populacionais em uma tabela de contingência 2x2, com os totais marginais da variável A fixados, e  $\delta$  é a diferença entre estas duas proporções a ser detectada. A utilização da fórmula anterior requer por parte do pesquisador, a especificação dos valores para  $\alpha$  e  $\beta$ , e para  $\pi_{11}$ ,  $\pi_{21}$  e  $\delta$ . Como  $\delta$  representa a diferença entre estas duas proporções, basta que o pesquisador especifique somente uma das duas proporções e  $\delta$ .

**TAMANHO DE AMOSTRA PARA UM TESTE DE HIPÓTESE BILATERAL PARA A COMPARAÇÃO DE DUAS PROPORÇÕES POPULACIONAIS A PARTIR DE AMOSTRAS INDEPENDENTES:**

$(H_1 : \pi_{11} \neq \pi_{21})$ . As fórmulas do tamanho da amostra será o maior valor dos dois valores obtidos de n dados por:

$$n = \frac{\left( \frac{z_\alpha}{2} + z_\beta \right)^2 [\pi_{11}(1-\pi_{11}) + (\pi_{11} - \delta)(1-\pi_{11} + \delta)]}{\delta^2} \text{ e o outro tamanho amostral será dada por:}$$

$$n = \frac{\left(\frac{z_\alpha + z_\beta}{2}\right)^2 [\pi_1(1 - \pi_1) + (\pi_1 + \delta)(1 - \pi_1 - \delta)]}{\delta^2}, \text{ onde, } z_\alpha \text{ é o valor da distribuição normal padrão que cor-}$$

responde ao valor de Z acima do qual a área é igual a  $\frac{\alpha}{2}$ ,  $z_\beta$  é o valor da distribuição normal padrão que

corresponde ao valor de Z acima do qual a área é igual a  $1 - \text{poder} = \beta$ ,  $\pi_{11}$  e  $\pi_{21}$  são definidos como proporções populacionais em uma tabela de contingência 2x2, com os totais marginais da variável A fixados, e  $\delta$  é a diferença entre estas duas proporções a ser detectada. A utilização da fórmula anterior requer por parte do pesquisador, a especificação dos valores para  $\alpha$  e  $\beta$ , e para  $\pi_{11}$ ,  $\pi_{21}$  e  $\delta$ . Como  $\delta$  representa a diferença entre estas duas proporções, basta que o pesquisador especifique somente uma das duas proporções e  $\delta$ .

#### TAMANHO DE AMOSTRA PARA UM TESTE DE HIPÓTESE UNILATERAL PARA A MÉDIA POPULACIONAL $\mu$ :

$(H_1 : \mu < \mu_0 \text{ ou } H_1 : \mu > \mu_0)$ . A fórmula é dada pela expressão:  $n = \frac{(z_\alpha + z_\beta)^2 \sigma^2}{\delta^2}$ , onde,  $z_\alpha$  é o valor da distri-

buição normal padrão que corresponde ao valor de Z acima do qual a área é igual a  $\alpha$ ,  $z_\beta$  é o valor da distribuição normal padrão que corresponde ao valor de Z acima do qual a área é igual a  $1 - \text{poder} = \beta$ , e  $\delta$  é a diferença a ser detectada.

#### TAMANHO DE AMOSTRA PARA UM TESTE DE HIPÓTESE BILATERAL PARA A MÉDIA POPULACIONAL $\mu$ :

$(H_1 : \mu \neq \mu_0)$ . A fórmula é dada pela expressão:

$$n = \frac{\left(\frac{z_\alpha + z_\beta}{2}\right)^2 \sigma^2}{\delta^2}, \text{ onde, } z_\alpha \text{ é o valor da distribuição normal padrão que corresponde ao valor de Z acima}$$

do qual a área é igual a  $\frac{\alpha}{2}$ ,  $z_\beta$  é o valor da distribuição normal padrão que corresponde ao valor de Z acima do qual a área é igual a  $1 - \text{poder} = \beta$ , e  $\delta$  é a diferença a ser detectada.

**TUBO DE PANAMÁ:** Instrumento simples utilizado na amostragem por enumeração angular em áreas planas. Consiste em um tubo metálico de 20 cm que com uma mira em uma extremidade e um orifício de visada ou ocular em outra. O tubo de panamá possui apenas um fator de área basal. Os mais comuns são os de 1, 2 ou 4 metros quadrados por hectare.

**TESTE MODIFICADO DE TUKEY:** Tipo de teste de hipótese a posteriori desenvolvido por Andrade e Ferreira (2010), para casos de homogeneidade e heterogeneidade de variâncias em delineamentos balanceados e não balanceados.

$T_{(v)}$  : Distribuição t de Student, com v graus de liberdade.

$t_\alpha$  : Quantil  $1 - \alpha$  da distribuição t de Student.







**UNIÃO:** A união ou reunião de dois conjuntos A e B, é o conjunto dos elementos ou valores que estão em A ou em B ou em ambos. Simbolicamente dado por  $\cup$ . Antigamente era conhecida como operação xícara.

**UNIÃO (UNION):** Na teoria de conjuntos ou na álgebra de eventos, são todos os elementos únicos obtidos pela combinação de dois ou mais conjuntos. Na teoria de probabilidades, é a probabilidade de dois eventos independentes ocorrerem. Em ambos os casos, a união é escrita com o símbolo  $\cup$ . Comparar com intercessão. Significa A ou B.

**UNIÃO DE EVENTOS A E B:** O evento que contém todos os pontos amostrais que estão em A, em B ou em ambos. A união é denotada por  $\cup$ .

**UNIÃO OU SOMA LÓGICA:** Símbolo matemático dado por  $\cup$ .

**UNIDADE:** Menor unidade em que o tratamento é aplicado e cuja resposta não é afetada pelas demais unidades. Unidade básica para a coleta de dados e análises. Normalmente um paciente na experimentação com seres humanos, mas também pode ser material, ou parte, desse paciente como uma amostra de sangue ou um dente, coleção de indivíduos em outros contextos como por exemplo, moradores de um domicílio, uma ala de hospital. Sinônimo de unidade experimental ou parcela experimental em experimentação ou nos ensaios clínicos e de unidade observacional em estudos observacionais.

**UNIDADE:** Indivíduo, elementos, item, membro ou unidade estatística onde se estuda uma ou mais variáveis. A unidade estatística pode ser simples: pessoa, objeto, animal, planta, protozoário, dentre outros, ou coletiva: cardume, blocos residenciais, turmas escolares, colméias, colônias bacterianas e outras.

**UNIDADE EXPERIMENTAL:** É o sujeito ao qual se aplica um dos tratamentos. Pode também ser chamado de parcela ou canteiro. Pode ser uma área de terra ou solo, um vaso, um animal, uma placa de petri, um indivíduo, a posição de montagem de um pneu, etc.

**UNIDADE EXPERIMENTAL OU PARCELA:** É a unidade que vai receber o tratamento e fornecer os dados que deverão refletir seu efeito. Por exemplo uma planta, um vaso com plantas, uma placa de petri com um meio de cultura, dentre outros.

**UNIDADE AMOSTRAL:** Unidade na qual são observadas e medidas as características quantitativas e qualitativas da população. A amostra é composta pelo conjunto de unidades amostrais. Cada unidade amostral gera uma única observação da variável de interesse. As unidades amostrais podem ser parcelas de área fixa, de área variável ou unidades amostrais não-superficiais, como pontos e linhas amostrais.

**UNIDADE AMOSTRAL:** O mesmo que unidade de amostragem. É a menor porção física onde são coletadas as informações em um levantamento amostral.

**UNIDADE DE AMOSTRAGEM:** Cada uma das partes disjuntas em que uma população é exaustivamente decomposta para que, do conjunto delas, se façam extrações, a fim de constituir uma amostra, ou um estágio de amostra. Também se diz unidade amostral.

**UNIDADE DE AMOSTRAGEM:** É a parte distinta do universo sobre a qual as medidas são feitas e indivisível em unidades elementares. Para depósitos minerais a unidade é especificada pelo responsável pela pesquisa, que inclui: tamanho e modo de retirada física da amostra, bem como o intervalo de valores a ser analisado.

**UNIDADE DE ANÁLISE:** É o elemento ou unidade física tal como uma pessoa, domicílio ou organização da população que representa o foco do estudo de pesquisa.

**UNIDADE DE OBSERVAÇÃO:** Fonte dos dados de um estudo médico. A unidade corresponde, geralmente, a um participante do estudo.

**UNIDADE DE OBSERVAÇÃO:** Unidade da qual dados são coletados. Tipicamente, é a unidade de análise, mas as duas podem diferir, como quando entrevistas são feitas em residências que são as unidades de observação e dados coletados sobre os membros individuais que são as unidades de análise.

**UNIDADE ELEMENTAR:** Um dos indivíduos que, agregado, compõe a população. A menor unidade fornecedora de informação que, com agregação adequada, leva à propriedade populacional em estudo. Em inglês Elementary unit.

**UNIDADE ESTATÍSTICA:** É cada um dos elementos da população.

**UNIDADE EXPERIMENTAL:** Veja unidade ou parcela.

**UNIDADE FINAL DE AMOSTRAGEM:** Denominação dada às unidades amostrais da população da qual se extrai a amostra de último estágio, na amostragem em estágios múltiplos.

**UNIDADE OBSERVACIONAL:** Veja unidade.

**UNIDADE PRIMÁRIA DE AMOSTRAGEM:** Denominação dada às unidades amostrais da população da qual se extrai a amostra de primeiro estágio, na amostragem em estágios múltiplos.

**UNIDADES DE AMOSTRAGEM:** Unidades, como por exemplo, indivíduos, residências dentre outras amostradas. Em geral, a unidade de amostragem é a mesma que a unidade de observação, mas não necessariamente. Por exemplo, residências podem ser amostradas mas, em seguida, cada indivíduo nas residências selecionadas é entrevistado ou observado.

**UNIDADES DE ANÁLISE:** O que ou quem está sendo estudado. Em pesquisa de ciência sócias as unidades mais típicas de análise são indivíduos.

**UNIDADES EXPERIMENTAIS:** Elementos ou indivíduos de um experimento.

**UNIDIMENSIONALIDADE:** Característica de um conjunto de indicadores que tem apenas um traço inerente ou conceito em comum. Da combinação entre os indicadores escolhidos e a definição teórica do construto unidimensional, o pesquisador deve estabelecer conceitual e empiricamente que os indicadores são medidas confiáveis e válidas somente do construto especificado, antes de estabelecer unidimensionalidade. Semelhante ao conceito de confiabilidade.

**UNIVERSALIDADE, ASPECTOS:** Termo usado para identificar três aspectos da universalidade de um fenômeno, os quais são: i) generalização temporal. Será o fenômeno, avaliação de docente, de interesse daqui a cinco anos?; ii) generalização geográfica. É o fenômeno interessante a outras pessoas na próxima

cidade, em outro estado, outro país?; iii) aplicabilidade a exemplos específicos do fenômeno geral. Há muitos exemplos específicos do fenômeno sendo estudados ou este é o único?.

**UNIVERSALISMO:** Um atributo tido como muito fundamental para caracterizar o sistema de valor científico é o uso dos critérios universais na avaliação do conhecimento científico. A validade ou valor de qualquer afirmativa científica é determinar apenas pela aplicação das normas técnicas das ciências; independentemente das características pessoais, sociais, políticas ou nacionais do autor.

**UNIVERSO:** Outro nome para população.

**UNIVERSO:** O mesmo que população ou a totalidade das observações estatísticas.

**UNIVERSO:** O mesmo que população. Em inglês universe.

**UNIVERSO:** É o conjunto de fenômenos a serem trabalhados, definido como critério global da pesquisa.

**UNIVERSO:** É a massa total de material dentro da área de interesse e a fonte total dos dados que poderiam ser de interesse para o projeto de amostragem. Na amostragem de depósitos minerais, o universo pode consistir de múltiplas características mensuráveis ou teores diversos de cada amostra do depósito. Os limites físicos do universo são geralmente estabelecidos antes da realização das medidas, podendo ser alteradas com a evolução dos conhecimentos. O universo pode ser designado como o depósito mineral inteiro, um corpo de minério dentro do depósito, ou um bloco de minério dentro do corpo, dependendo do objetivo do projeto de amostragem.

**UNIVERSO:** Massa geral e abstrata de pessoas sobre as quais se quer tirar conclusões. Exemplo: Brasileiros. Mas na seleção de amostras é necessário ser mais específico na identificação de quem temos em mente.

**UNÍVOCO:** Refere-se à definição cujo significado só admite uma interpretação. Nas ciências exatas, unívoco diz respeito à correspondência entre dois conjuntos, na qual um elemento do primeiro conduz a um, e somente a um, elemento do segundo. Na estatística diz-se que uma hipótese nula deve ser unívoca.

**ULTIMA RATIO:** Expressão latina que significa o argumento derradeiro, final.

**UNIDADE DE ANÁLISE:** Expressão mais ampla para sujeito de pesquisa. Objeto, organismo, conceito ou pessoa que está sendo estudado numa pesquisa.

**UNIDADE DE MEDIDA:** Referência arbitrária que atribui significado quantitativo a um valor bruto. Por exemplo, a unidade de medida de peso pode ser o quilograma (kg), a unidade de medida de volume pode ser o metro cúbico (m<sup>3</sup>), para potência elétrica, o watt (w), e assim por diante.

**UNIDADE EXPERIMENTAL:** Diz-se da unidade de análise que é utilizada em pesquisa experimentais. O mesmo que parcela experimental.

**UNIDADE SIGNIFICATIVA:** Expressão utilizada nas pesquisas qualitativas e hermenêuticas para designar o trecho de uma entrevista do sujeito da pesquisa que possui conteúdos relacionados às variáveis de contexto dos fatos que estão sendo objeto da análise hermenêutica. Normalmente, essas unidades significativas são organizadas em temas maiores, que serão posteriormente utilizados para uma compreensão ou verstehen do fenômeno.

**UNIVERSO:** Conjunto de tudo o que existe no tempo e no espaço.

**USO INDEVIDO DE GRÁFICOS:** Podem trazer uma ideia falsa dos dados que estão sendo analisados, chegando mesmo a confundir o leitor. Trata-se, na realidade, de um problema de construção de escalas.

**UNIDADE DE NÍVEL 2:** O indivíduo no segundo menor nível de uma estrutura com vários níveis, em uma estrutura de dois níveis, cada unidade de dois níveis como por exemplo, porcas origina certo número de unidades de nível 1 ou seja os leitões.

**UNIDADE EXPERIMENTAL:** É a unidade básica da experimentação agrícola, animal, industrial, etc. normalmente corresponde ao animal, ou uma planta, um grupo de plantas, um vaso com plantas, etc. individualmente, ou em grupo. O conjunto total das unidades experimentais é a população a partir da qual podemos obter uma amostra.

**UNIDADE EXPERIMENTAL OU PARCELA:** É a menor divisão do material experimental, tal que duas quaisquer parcelas possam receber diferentes tratamentos.

**UNIDADES AMOSTRAIS:** A população é fracionada em partes não-sobrepostas denominadas unidades como por exemplo, animais individuais, rebanhos diferentes, o procedimento de amostragem envolve a escolha de um subgrupo dessas unidades da população. As unidades podem, então, ser chamadas de unidades amostrais.

**UNIDADES NÍVEL 1:** O indivíduo no menor nível de uma estrutura de vários níveis ou um grupo de unidades de nível 1 como por exemplo, leitões está inserido em uma unidade de nível 2 como por exemplo a porca ou matriz.

**UNIMODAL:** Uma distribuição que tem uma única moda ou grupo modal.

**UTILITARISMO:** Conjunto de princípios, como regras, preceitos, normas, etc. que considera as teorias científicas somente como ferramenta para realizar previsões; norma que considera os conceitos e as teorias se estas tiverem praticidade.

**UTILIDADE:** Um julgamento subjetivo de preferência por um indivíduo que representa o valor ou utilidade holística de um objeto específico. Em análise conjunta, utilidade é assumida como sendo formada pela combinação de estimativas de utilidades parciais para qualquer conjunto especificado de níveis com o uso de um modelo aditivo, talvez em conjunção com efeitos de interação.

**UTILIDADE PARCIAL:** Estimativa da análise conjunta da preferência ou utilidade geral associada a cada nível de cada fator usado para definir o produto ou serviço.

**ÚLTIMO TEOREMA DE FERMAT:** É a afirmação de que se  $n > 2$ , então a equação  $x^n + y^n = z^n$  não apresenta soluções inteiras. Em inglês Fermat's last theorem.

**UNIVERSO:** É o conjunto de todos os elementos como por exemplo, pessoas, animais, células, objetos, dentre outros, que interessam a determinada pesquisa. Evidentemente, trata-se de um conjunto infinito, ou melhor, hipotético.

**UNIDADES AMOSTRAIS:** Correspondem às unidades selecionadas na amostragem para calcular as estatísticas.

$U_{(a,b)}$  : Distribuição de probabilidade contínua uniforme ou retangular entre os limites a e b.



**V DE CRAMER:** Uma medida de associação usada para dados categóricos, calculada diretamente a partir do qui-quadrado.

**V.A. :** Variável aleatória, casual, randômica ou estocástica.

**VACINA:** Preparação contendo microorganismos vivos ou mortos ou frações destes possuidora de propriedades antigênicas. As vacinas são empregadas para induzir em um indivíduo a imunidade ativa e específica contra um microrganismo.

**VALIDAÇÃO:** O processo de assegurar que as entrevistas foram de fato conduzidas de acordo com o que foi especificado.

**VALIDAÇÃO CRUZADA:** Procedimento de divisão da amostra em duas partes: a amostra de análise, usada na estimação da(s) função(ões) discriminante(s) ou do modelo de regressão logística, e a amostra de teste, usada para validar os resultados. A validação cruzada evita o superajuste da função discriminante ou da regressão logística, permitindo sua validação em uma amostra totalmente separada.

**VALIDAÇÃO EXTERNA:** Processo de testar a validade de uma medida, como um índice ou escala, examinando suas relações com outros supostos indicadores da mesma variável. Por exemplo, se o índice realmente mede preconceito, deve correlacionar com outros indicadores de preconceito.

**VALIDAÇÃO INTERNA:** Processo pelo qual os itens individuais de uma medida composta se correlacionam com a própria medida. Assim se testa a decisão de incluir todos os itens na medida composta. Ver ainda validação externa.

**VALIDAÇÃO POR PARTIÇÃO DE AMOSTRAS:** Ver validação cruzada.

**VALIDADE:** Grau em que uma medida ou um conjunto de medidas corretamente representa o conceito de estudo, ou seja, o grau em que se está livre de qualquer erro sistemático ou não aleatório. A validade se refere ao quão bem o conceito é definido pela(s) medida(s), ao passo que confiabilidade se refere à consistência da(s) medidas(s).

**VALIDADE:** Termo usado de forma generalizada para mensurar a qualidade experimental, como por exemplo, atender os pressupostos básicos da experimentação, como também apresentar um baixo valor do coeficiente de variação. Condição de algo que se encontra apto a produzir os efeitos dele esperados. Uso correto do equipamento de pesquisa para o fim que ele foi confeccionado como, por exemplo, um paquímetro para medir o diâmetro do caule de uma planta.

**VALIDADE:** Ver validade externa e validade interna.

**VALIDADE (DE UM INSTRUMENTO):** Diz-se que um instrumento é válido quando ele mede o que se espera que ele meça.



**VALIDADE:** Capacidade de refletir a verdadeira situação, A validade pode ser externa e interna. Em epidemiologia, a validade de um estudo é estimada, à parte os erros aleatórios, pela presença, em potencial, de vícios de seleção, de aferição e de confusão. Validade externa de um estudo: capacidade de generalização dos resultados. Validade interna de um estudo: grau em que as conclusões de um estudo são corretas, para a amostra de indivíduos investigados.

**VALIDADE:** Termo usado para descrever uma medida que reflete com exatidão o conceito que se pretende medir. Por exemplo, seu quociente de inteligência (QI) parece ser uma medida mais válida de inteligência que o número de horas que você passa numa biblioteca. É importante entender que nunca se pode provar a validade definitiva de uma medida. Podemos, porém, concordar quanto à sua validade relativa, baseados nas validades aparente, por critério, de conteúdo, por construção, interna e externa. Não deve ser confundido com confiabilidade.

**VALIDADE:** Extensão em que uma medida ou um conjunto de medidas representa corretamente o conceito do estudo, ou seja, o grau em que se está livre de qualquer erro sistemático ou não-aleatório. A validade se refere a quão bem o conceito é definido pela(s) medida(s), enquanto confiabilidade se refere à consistência da(s) medida(s).

**VALIDADE:** Habilidade dos indicadores de um construto em medir com precisão o conceito em estudo, como renda familiar ou inteligência. Validade é determinada em grande parte pelo pesquisador, pois a definição original do construto ou conceito é proposta pelo pesquisador e deve combinar com os indicadores ou medidas selecionados. Validade não garante confiabilidade e reciprocamente. Uma medida pode ser precisa ou válida, mas não consistente ou onfiável. Além disso, ela pode ser consistente sem ser precisa. Logo, validade e confiabilidade são duas condições separadas, mas inter-relacionadas.

**VALIDADE:** Símbolo ou conceito que pode assumir um de uma série de valores.

**VALIDADE (TIPOS):** Os psicólogos definem quatro tipos de validade (AMERICAN PSYCHOLOGICAL ASSOCIATION, 1974): i) validade preditiva: Os exames vestibulares teriam validade, se os escores deles fossem substancialmente correlacionados com sucesso posterior nas universidades; ii) validade concomitante ou empírica: Por meio deste tipo de validade se compara um desempenho com outro já existente, em vez de comparar com um futuro desempenho. Os escores de um vestibular teriam validade concomitante se correlacionassem com escores do segundo grau; iii) validade de conteúdo: Um teste tem este tipo de validade, quando ele contém uma amostra representativa das características que ele deve medir; iv) validade de constructo: Dizemos que um instrumento tem validade de constructo quando ele se correlaciona com características ou comportamento postulados no constructo.

**VALIDADE APARENTE:** Qualidade de um indicador que o faz parecer uma medida razoável de alguma variável. Parece fazer sentido, sem muita explicação, que ir frequentemente à igreja é indicador de religiosidade. Tem validade aparente.

**VALIDADE CONCORRENTE:** Capacidade da medida de refletir um padrão.

**VALIDADE CONVERGENTE:** Um alto grau de correlação entre diversos instrumentos de mensuração que visam medir o mesmo conceito.

**VALIDADE DE CONSTRUÇÃO:** O grau pelo qual um instrumento de mensuração representa e conecta de maneira lógica fenômenos observados em uma construção via teoria básica.

**VALIDADE DE CONSTRUÇÃO:** O grau em que uma medição se relaciona da forma esperada com outras variáveis num sistema de relações teóricas.

**VALIDADE DE CONTEÚDO:** O grau pelo qual os itens do instrumento representam o universo de um conceito sob estudo.

**VALIDADE DE CONTEÚDO:** O grau com que uma medida cobre a faixa de significados incluídos no conceito.

**VALIDADE DE CONTEÚDO:** Avaliação do grau de correspondência entre os itens selecionados para constituir uma escala múltipla e sua definição conceitual.

**VALIDADE DE CRITÉRIO:** Habilidade de agrupamentos em mostrar as diferenças esperadas sobre uma variável não usada para formar os agrupamentos. Por exemplo, se os agrupamentos foram formados sobre avaliações de desempenho, a mentalidade de marketing iria sugerir que os agrupamentos com avaliações mais altas de desempenho também teriam maiores escores de satisfação. Se isso ocorresse em teste empírico, então a validade de critério estaria suportada.

**VALIDADE DE EXPRESSÃO:** Ver validade de conteúdo.

**VALIDADE DE FACE:** Uma mensuração que parece medir o que deveria medir.

**VALIDADE DE PREVISÃO:** O grau pelo qual o nível futuro de um critério pode ser previsto por uma escala corrente de mensuração.

**VALIDADE (DE UM TESTE):** O mesmo que acuidade, acurácia e exatidão. Capacidade do teste para produzir resultados semelhantes ou próximos ao verdadeiro valor da característica que está sendo medida; a proporção em que o teste mede o que deve ser medido. Uma prova de função pulmonar é válida se separar, adequadamente, os indivíduos com função pulmonar normal e anormal. Tipos de validade: lógica ou consensual, de conteúdo, de constructo e relativa a um padrão. Validade lógica ou de consenso: grau em que os especialistas concordam, entre si, em suas opiniões, por exemplo, acerca da utilidade de um instrumento de mensuração. Validade de conteúdo: adequação do conteúdo de um teste para medir as facetas que devem ser medidas; por exemplo, um teste de qualidade de vida deve contemplar a saúde física, a autonomia de realizar as atividades do dia a dia, o estado mental e o ângulo social. Validade de constructo: trata-se de uma outra forma de validade de um teste. Se a característica a ser medida, pelo teste, correlaciona-se com uma outra, ou com o conhecimento disponível, os resultados do teste devem apresentar resultados compatíveis, ou seja, correlacionados, comportando-se como seria esperado; por exemplo, se a depressão aumenta com a idade, o resultado da aplicação de um novo teste, para medir depressão, deve mostrar tendência crescente com a idade. Validade relativa a um padrão: proporção ao grau em que um teste reflete um padrão; pode ser concorrente ou preditiva. É expressa por meio da sensibilidade, da especificidade e dos valores preditivos.

**VALIDADE DISCRIMINANTE:** Baixo grau de correlação entre construções que deveriam ser diferentes.

**VALIDADE ECOLÓGICA:** Refere-se ao quanto as situações comparadas no experimento são representativas da população de situações a qual se quer generalizar.

**VALIDADE EXPERIMENTAL:** Existem dois tipos dessa validade: i) interna: o mínimo necessário em termos de controle, mensuração, análise e procedimentos que torne possível interpretar os resultados do experimento; ii) externa: a propriedade e o quanto dos resultados do experimento pode ser generalizado.

**VALIDADE EXPERIMENTAL:** Tipo de validade que pode ser de duas categorias. Primeiro é a validade interna que trata das exigências necessárias no que diz respeito a mensuração, ao controle, aos procedimentos e as análises as quais possibilitam interpretar as respostas de um ensaio. Segundo é a validade externa que mostra quão geral as respostas do ensaio podem ser extrapoladas ou inferidas, de maneira mais real possível.

**VALIDADE EXTERNA (DE PESQUISA):** Refere-se ao quanto ou maneira os resultados de um experimento podem ser generalizados a diferentes sujeitos, situações, pesquisadores, e a testes.

**VALIDADE EXTERNA:** Está presente quando os resultados de um estudo são verdadeiros e significantes para uma grande população e não somente para os participantes do estudo.

**VALIDADE EXTERNA:** A extensão pela qual os relacionamentos casuais medidos em um experimento podem ser generalizados para épocas, pessoas ou ambientes externos.

**VALIDADE EXTERNA:** Refere-se à inferência estatística, ou seja, a generalização dos resultados para toda a população de interesse.

**VALIDADE EXTERNA:** A capacidade de generalização dos resultados de um estudo.

**VALIDADE EXTERNA:** Validade externa refere-se à propriedade que os resultados de uma pesquisa tem de serem aplicáveis em outra situações que não aquela em que a pesquisa foi executada.

**VALIDADE E FIDEDIGNIDADE (TESTE):** Validade refere-se ao atributo que um teste tem de medir o que ele deve medir, enquanto que fidedignidade refere-se à consistência do teste em medir o que quer que ele meça. Diz-se que um teste é fidedigno, mas não válido, se consistentemente mede algo que não devia e deixa de medir o que devia. Um teste não fidedigno não pode ser válido, uma vez que a falta de consistência elimina a possibilidade de medir o que devia medir.

**VALIDADE INTERNA:** A extensão pela qual explicações concorrentes para resultados experimentais podem ser evitadas.

**VALIDADE INTERNA:** Está presente quando os resultados de um estudo são verdadeiros e são significantes para os seus participantes.

**VALIDADE INTERNA:** A ausência de erros sistemáticos ou seja, vícios ou também de erros aleatórios em um estudo. Quando a validade interna existe, os resultados na população em estudo representam a verdade na população alvo.

**VALIDADE INTERNA:** Validade interna refere-se aos mínimos controles que uma pesquisa tem que ter, a fim de que os seus resultados sejam interpretáveis.

**VALIDADE INTERNA:** É a validação dos resultados apenas para a amostra considerada, ou seja, é a validade das inferências para os indivíduos que participaram do estudo. Os cálculos de tamanho de amostra feitos aqui são baseados principalmente nas técnicas estatísticas a serem utilizadas. Entretanto, é importante ressaltar que, o fato do número de observações ser suficiente não garante a utilização da

técnica estatística. Em geral, existem algumas suposições que devem ser satisfeitas como, por exemplo, a normalidade das observações em testes de comparação de médias.

**VALIDADE PREDITIVA:** Capacidade da medida de prever a ocorrência de um acontecimento futuro; é expressa pelo valor preditivo.

**VALIDADE PREDITIVA:** Ver validade de critério.

**VALIDADE RELACIONADA A CRITÉRIO:** O grau pelo qual um instrumento de mensuração pode prever uma variável designada como critério.

**VALIDADE SIMULTÂNEA:** O grau pelo qual uma variável, medida no mesmo espaço de tempo que uma variável de interesse, pode ser prevista pelo instrumento de mensuração.

**VALIDADE RELACIONADA A CRITÉRIO:** Em que grau uma medida se relaciona a algum critério externo. Por exemplo, a validade da junta escolar é demonstrada por sua capacidade de prever o sucesso dos alunos na faculdade.

**VALOR:** É o mesmo conceito de dado observado, dado numérico, dado ou estimativa. É o valor numérico, taxa ou resultado de uma variável particular, como por exemplo, o rendimento de algodão X é igual a 3,5 toneladas por hectare, ou seja,  $X = 3,5 \text{ t ha}^{-1}$ . Também serve para indicar categorias, tipos ou atributos de uma variável qualitativa, como a forma da folha de uma planta que pode ser  $y = \text{cordiforme}$ ,  $z = \text{linear}$ ,  $w = \text{espiralada}$ , etc.

**VALOR P:** Ver p e nível de significância.

**VARIÂNCIA:** Tipo de medida de dispersão de uma distribuição de dados. Representa o desvio-padrão elevado ao quadrado. Para calcular a variância indicada comumente pelo símbolo  $\sigma^2$  de um conjunto de dados, deve-se observar a seguinte fórmula:

$$\hat{\sigma}^2 = S^2 = \frac{\sum (x - \bar{x})^2}{n - 1}, \text{ onde: } x = \text{cada valor individual do conjunto de dados; } n = \text{número de valores}$$

do conjunto de dados;  $\bar{x} = \text{média dos valores do conjunto de dados}$ . A variância representa a média do quadrado dos desvios em relação à média. Ver dispersão.

**VARIÁVEIS:** São as medidas tomadas nas unidades experimentais, após terem sido submetidas aos tratamentos, constituem os valores das variáveis dependentes ou variáveis respostas. Estas são, no geral, predeterminadas pelo pesquisador, ou seja, ele sabe o que vai medir, contar, observar dentre outras. A grosso modo, podemos considerar as variáveis como sendo como sendo quantitativas contínuas e descontínuas e qualitativas atributivas. Muito frequentemente, são usadas as variáveis derivadas a partir de duas ou mais variáveis, como é o caso das percentagens, dos índices ou das relações. Por exemplo, de variáveis quantitativas tem-se altura de planta como contínuas e número de frutas por planta como discreta, e para variáveis qualitativas temos, como variável nominal cor de fruto branco, amarelo, verde ou a maturação de frutas como variável nominal verde, maduro, muito maduro e como ordinal índice de ataque de pragas numa planta severo, muito atacada, atacada, pouco atacada e sem ataque.

**VARIÁVEL CATEGÓRICA:** Variável que assume valores descontínuos ou discretos, ou seja, valores inteiros. Por exemplo: a variável número de filhos pode assumir os valores 1, 2, 3, 4 dentre outras. Não é possível o valor 4,5 filhos. Opõe-se à variável contínua; Alguns autores utilizam esta expressão também

para se referir às variáveis que possuem nível de mensuração nominal como por exemplo, sexo, cor dos olhos, dentre outras.

**VARIÁVEL CONTÍNUA:** Variável que pode assumir valores contínuos, ou seja, dados quantitativos para os quais qualquer valor dentro de um intervalo é possível. Exemplo: altura, peso, tempo, salário. Opõe-se à variável categórica.

**VARIÁVEL CONTROLE:** Variável que não é objeto de estudo numa pesquisa, porém é mantida sob controle. Utilizam-se as variáveis controle para neutralizar efeitos que não são o principal interesse do estudo, mas que podem afetar as outras variáveis estudadas. Por exemplo, num estudo sobre desempenho da atenção, o sono deve ser controlado, uma vez que se sabe que está relacionado com o desempenho da atenção de uma pessoa.

**VARIÁVEL DEMOGRÁFICA:** Qualquer variável que contenha dados demográficos. Exemplo: sexo, idade, nível de escolaridade, estado civil dentre outras.

**VARIÁVEL DEPENDENTE:** Variável cujo valor supõe-se que depende de outra variável chamada independente. É a variável que é observada e medida pelo pesquisador. Nos estudos experimentais, constituiu-se nos efeitos que estão sendo estudados. Por exemplo, numa pesquisa deseja-se estudar a ação das bebidas alcoólicas sobre o desempenho acadêmico dos alunos universitários. A variável desempenho acadêmico é a variável dependente ou efeito e a variável quantidade de bebida alcoólica ingerida constitui-se na variável independente ou causa do estudo.

**VARIÁVEL DICOTÔMICA:** Qualquer variável que somente possa assumir dois valores. Por exemplo: sexo masculino ou feminino, situação perante o fisco ativo ou inativo, exame final aprovado ou reprovado. Este tipo de variável possui obrigatoriamente nível de mensuração nominal.

**VARIÁVEL DICOTOMIZADA:** Variável não coletada diretamente numa pesquisa, mas que é proposta pelo pesquisador para fins de análise de dados, e que possui apenas dois valores. Por exemplo, após coletar a variável nota que apresenta valores contínuos, de zero a dez, o pesquisador decide criar uma nova variável denominada situação acadêmica, que possui dois valores: aprovado ou reprovado. A variável dicotômica que é categorizada, apresenta categorias que ocorrem naturalmente, como por exemplo, sexo: masculino ou feminino, estado de saúde; sadio ou doente, etc.

**VARIÁVEL DISCRETA:** Ver variável categórica.

**VARIÁVEL ENDÓGENA:** Qualquer variável interna ao sistema que está sendo estudado. Por exemplo, numa pesquisa sobre a produtividade, poderíamos ter as variáveis necessidade da empresa e necessidades de mercado como exógenas e as variáveis número de horas extras/mês, nível de estresse e produtividade como endógenas em relação ao estudo realizado. Ver variável exógena.

**VARIÁVEL EXÓGENA:** Variável externa ao sistema que está sendo estudado. Suas determinações estão fora do controle e do escopo do estudo: seus valores são dados a priori, e não analisados; Variável que afeta um estudo de forma indesejável, servindo apenas para distorcer e desorientar o pesquisador. Variável fora de controle direto do pesquisador, e que afeta o estudo de maneira não intencional. Opõe-se à variável endógena.

**VARIÁVEL INDEPENDENTE:** Variável que pode ser manipulada pelo pesquisador, a fim de avaliar os efeitos causados sobre outra variável chamada variável dependente. Hipoteticamente, é a influência causal nos estudos experimentais. Conhecida igualmente pela sigla VI.

**VARIÁVEL INTERMEDIÁRIA:** Ver variável interveniente.

**VARIÁVEL INTERVENIENTE:** Variável hipotética, não observada diretamente numa pesquisa, mas cuja existência é inferida na relação entre duas outras variáveis. Por exemplo, sabe-se que existe uma relação de causa e efeito factível entre a variável nível socioeconômico dos pais e mortalidade infantil, quanto maiores os valores da primeira, menores os valores da segunda. Ocorre que possivelmente existam outras variáveis intervenientes entre elas, como por exemplo, nível de acesso à informação, ou seja, pessoas com maior nível socioeconômico tendem a ter maior acesso a informações de saúde e higiene, o que acarretará um menor índice de mortalidade infantil.

**V DE CRAMER:** Ver coeficiente de correlação V de Cramer.

**VALOR ADAPTATIVO:** Proporção de progênie de um dado genótipo sobrevivente até a maturidade, relativa a aquela de outro genótipo. Expresso como um número puro entre 0 e 1. Assim, se apenas 80 % da progênie de um genótipo aa sobrevive, enquanto 100 % em um genótipo A\_\_ , o valor adaptativo de aa é 0,8.

**VALOR AJUSTADO:** Valor calculado por meio de um procedimento que usa um conjunto de valores, podendo ser comparado com o valor realmente observado. Em inglês Fitted value.

**VALOR AJUSTADO:** Para um conjunto de valores de variável independente, valor da variável dependente predito pela equação de regressão.

**VALOR CRÍTICO:** Valor que separa a região crítica e os valores da estatística de teste que não conduzem à rejeição da hipótese nula.

**VALOR CRÍTICO:** Valor da distribuição amostral que separa a região de rejeição da de não rejeição. Em inglês critical value.

**VALOR CRÍTICO:** Valor de um teste estatístico como por exemplo o teste t, teste F, dentre outros que denota um nível de significância especificado. Por exemplo, 1,96 denota um nível de significância de 0,05 para o teste t com grandes amostras.

**VALOR DE p:** Probabilidade de que a diferença observada possa ter sido obtida apenas por acaso, dada a variação aleatória e um único teste para a hipótese nula.

**VALOR DE PROVA P:** Chama-se valor de prova p à probabilidade, admitindo a veracidade da hipótese nula, de a estatística do teste cair na região crítica que tem um extremo igual ao valor observado.

**VALOR ESPERADO:** É o valor mais provável de um conjunto de valores, ou seja, o valor médio ou a média aritmética do mesmo.

**VALOR ESPERADO:** Ver esperança.

**VALOR GENÉTICO:** Valor de um animal como pai. Tecnicamente é a soma dos efeitos médios de todos os genes que um animal possui. Esta definição baseia-se no fato de que os progenitores transmitem seus



genes e não genótipos aos filhos. Na prática, representa o dobro dos desvios médios da progênie de um animal em relação a média da população, ou seja,  $2 \times$  *média dos filhos*-média da população. A diferença é dobrada porque somente uma amostra de metade dos genes ou seja, um gene de cada par é transmitida pelos pais à progênie. O valor genético existe para cada característica e é dependente da população na qual o animal é avaliado. Para um mesmo caráter, um animal pode estar acima da média em um rebanho e abaixo em outro.

**VALOR GENÉTICO PREDITO:** Predição do verdadeiro valor genético de um animal, para uma dada característica, baseado no desempenho individual e de parentes próximos. É um modo sistemático de combinar informação de desempenho disponível sobre o indivíduo, irmãos, irmãs, meios irmãos, meias irmãs e progênie do animal.

**VALOR INICIAL:** Estimativa de parâmetro inicial usada para processos de estimação incremental ou iterativa, como LISREL (Linear Structural Relations), tido como um dos mais robustos softwares para modelagem de equações estruturais com matrizes de covariância.

**VALOR DE PROJEÇÃO:** Ver matriz de projeção.

**VALOR DE UMA PROBABILIDADE:** Veja valor p, p valor ou p value. Conhecido também como nível descritivo de um teste de hipótese ou de significância.

**VALOR DE VEROSSIMILHANÇA:** Medida usada em regressão logística e análise logit para representar a falta de ajuste preditivo. Ainda que esses métodos não usem o procedimento dos mínimos quadrados na estimação do modelo, como se faz em regressão múltipla, o valor de verossimilhança é parecido com a soma de erros quadrados na análise de regressão.

**VALOR DISCREPANTE:** Em uma amostra de n observações é possível para um número limitado de dados estarem separados, em valores, do restante, surgindo a questão de se eles são de uma população diferente, ou a técnica de amostragem não foi adequada. Tais valores são chamados discrepantes. Em inglês Outlier.

**VALOR DO PARÂMETRO  $\theta_i$  NA INTERAÇÃO k:**  $\theta_i^k$ .

**VALOR ESPERADO:** O valor esperado de uma função de variáveis é seu valor médio em amostragens repetidas. Assim, se  $t(x_1, \dots, x_n)$  é uma estatística dependente das variáveis  $x_1, \dots, x_n$  com distribuição conjunta  $dF(x_1, \dots, x_n)$  o valor esperado de t, se existir, é:  $\int t dF(x_1, \dots, x_n)$ . Em inglês Expected value.

**VALOR ESPERADO:** Para uma variável aleatória discreta, ou contínua o valor médio dos resultados.

**VALOR EXTREMO:** Score muito afastado da maioria dos valores amostrais decorrente da inclusão no universo investigado de elemento estranho, de erro de observação, de falha instrumental, de equívoco de registro ou de introdução do dado na planilha de cálculos. Pode, entretanto, ser um valor verdadeiro, devendo-se examiná-lo cuidadosamente antes de sua exclusão da análise dos dados. Conhecido também como valor atípico ou aberrante ou discrepante. Na literatura inglesa denomina-se outlier.

**VALOR INTERPOLADO:** Em forma geral, um valor interpolado é uma previsão feita baseada nos dados para uma variável não observada. Em estatística espacial, um valor interpolado é um valor previsto para um local onde não existem medições da característica em estudo. Em inglês Interpolated value.

**VALOR-P:** A probabilidade de ter obtido um conjunto de dados se a hipótese nula for verdadeira. É o menor valor da probabilidade de risco de se rejeitar a hipótese de nulidade em um teste de hipótese ou de significância, quando na realidade ela for verdadeira. Também conhecido como nível descritivo de um teste ou p value.

**VALOR-P:** Probabilidade de que uma estatística de teste em um teste de hipóteses seja no mínimo tão extrema quanto a efetivamente obtida.

**VALOR-P:** A probabilidade exata de se obter um teste estatístico computadorizado devido em grande parte ao acaso. Quanto menor for o valor-p, menor é a probabilidade de que o resultado observado tenha ocorrido por acaso.

**VALOR PERDIDO:** Palavra para representar um valor que não foi observado ou, se observado, foi perdido por razões alheias à pesquisa. Existem técnicas para estimar estes valores. Em inglês Missing value.

**VALOR PREDITO:** Valor de uma variável dependente calculado a partir de uma regressão e de valores das variáveis independentes.

**VALOR PREDITIVO (OU DE PREDIÇÃO):** Acerto dos resultados de um teste; indica o número ou a proporção de doentes e sadios, em função do resultado de um teste diagnóstico. O valor preditivo pode ser positivo ou negativo. Valor preditivo negativo: indica, entre os resultados negativos de um exame, quantos pertencem a sadios. Valor preditivo positivo: indica, entre os resultados positivos de um exame, quantos pertencem a doentes.

**VALOR PREDITIVO DE UM TESTE NEGATIVO:** Probabilidade de que um indivíduo com teste negativo (-) seja uma pessoa sem o evento objeto da investigação, como por exemplo determinada doença.

**VALOR PREDITIVO DE UM TESTE POSITIVO:** Probabilidade de que um indivíduo com teste positivo (+) apresente o evento objeto da investigação como por exemplo determinada doença.

**VALOR PREDITIVO NEGATIVO:** Indica que proporção dos indivíduos com um resultado de teste negativo realmente não são portadores da doença.

**VALOR PREDITO NEGATIVO:** Refere-se à variabilidade pós-teste negativo, isto é, a proporção de verdadeiros testes negativos entre todos os testes negativos.

**VALOR PREDITO POSITIVO:** Refere-se a probabilidade pós-teste positivo, isto é, a proporção de verdadeiros testes positivos entre todos os testes.

**VALOR TÍPICO (VALOR REPRESENTATIVO):** Como o próprio nome indica, corresponde à escolha de um único valor para representar todo o conjunto de valores de uma série de observações, geralmente é um valor central dos dados, como por exemplo a média, a mediana ou a moda.

**VALORES DISCREPANTES:** São valores que se distanciam demais dos outros e pouco prováveis de ocorrerem novamente. É importante realizar um estudo para saber a razão da ocorrência desses valores. Pode-se citar três principais causas: i) erro de transcrição de dados; ii) algum fato importante ocorreu durante o trabalho e iii) o valor é verdadeiro e deve ser considerado como tal. Também conhecido como outliers.

**VALORES DISTANTES OU DISCREPANTES:** Valores extremos que se desviam amplamente da média.



**VALORES EXTREMOS:** Valores da variável menores ou maiores com relação a todos os outros membros de uma série. Em inglês Extreme values.

**VALORES PARCIAIS F (OU t):** O teste parcial F é simplesmente um teste estatístico da contribuição adicional de uma variável para precisão de previsão acima da contribuição das variáveis já na equação. Quando uma variável ( $X_u$ ) é acrescentada a uma equação de regressão depois que outras variáveis já estão na equação, sua contribuição pode ser muito pequena, ainda que tenha uma alta correlação com as variáveis na equação. O valor parcial de F é calculado para todas as variáveis simulando que cada uma seja a última a entrar na equação. Ele fornece a contribuição adicional de cada variável acima de todas as outras na equação. Um valor parcial F pequeno ou insignificante para uma variável que não está presente na equação indica sua contribuição pequena ou insignificante mediante ao modelo como já especificado. Um valor t pode ser calculado no lugar de valores de F em todos os casos, sendo o valor t aproximadamente a raiz quadrada do valor F.

**VALORES PREDITOS:** Valores de uma variável dependente obtidos com o uso de variáveis independentes em uma equação de regressão.

**VALIDADE:** Determina se a medida está realmente medindo o que deve estar medindo.

**VALOR ANÔMALO (OUTLIER):** Uma observação cujo valor é muito diferente do conjunto principal dos dados.

**VALOR DE P:** Em um teste de hipótese, é a probabilidade de obter os resultados observados (ou resultados mais extremos) quanto a hipótese nula é verdadeira.

**VALOR PREDITIVO:** Consultar valor preditivo positivo (VPP) e valor preditivo negativo (VPN).

**VALOR PREDITIVO NEGATIVO (VPN):** É a proporção de animais com teste negativo, ou seja, o teste indica que não apresentam a doença ou que não exibem a doença.

**VALOR PREDITIVO POSITIVO (VPP):** É a proporção de animais com teste positivo, ou seja, o teste indica que apresentam a doença ou que exibem a doença.

**VALORES:** Consultar leituras.

**VALORES CRÍTICOS (OU PONTOS PERCENTUAIS):** O valor de um teste estatístico determinado a partir de uma distribuição de probabilidade teórica, que corresponde à probabilidade de determinada área da cauda, por exemplo, 0,05.

**VARIAÇÃO AMOSTRAL:** Implica que os valores de uma estatística (por exemplo a média da amostra que estima a média populacional) não são idênticos em amostras diferentes. A variância e o desvio padrão (DP) de uma distribuição amostral da estatística indicam essa variação amostral. O DP de uma distribuição amostral é chamado de erro padrão da estatística ou estimador.

**VARIAÇÃO BIOLÓGICA:** Uma variabilidade inerente ao material biológico, de modo que mensurações obtidas de diferentes indivíduos dificilmente serão idênticas.

**VARIÂNCIA:** Uma medida de dispersão. É o quadrado do desvio padrão.

**VALOR CRÍTICO (CRITICAL VALUE):** O valor que um teste estatístico precisa ter para que sua probabilidade seja estatisticamente significativa. Cada teste estatístico é relacionado a uma distribuição de probabilidades que possui um valor crítico para um dado tamanho amostral e graus de liberdade.

**VALOR DE P (P-VALUE):** Ver valor de probabilidade.

**VALOR DE PROBABILIDADE (VALOR DE P) (PROBABILITY VALUE/P-VALUE):** A probabilidade de rejeitar uma hipótese nula estatística quando na realidade ela é verdadeira também é chamada de probabilidade de erro Tipo I.

**VALOR ESPERADO (EXPECTED VALUE):** O valor mais provável de uma variável aleatória também é chamado de expectativa de uma variável aleatória.

**VALORES SINGULARES (SINGULAR VALUES):** Os valores ao longo da diagonal de uma matriz quadrada  $W$  tal que uma matriz  $A$   $m \times n$  possa ser reescrita como o produto  $VWU$ , onde  $V$  é uma matriz  $m \times n$  com a mesma dimensão de  $A$ ,  $U$  e  $W$  são matrizes quadradas (dimensões  $n \times n$ ); e  $V$  e  $U$  são matrizes com colunas ortonormais.

**VARIAÇÃO (VARIATION):** Incerteza, diferença, dispersão ou heterogeneidade.

**VARIAÇÃO DENTRO DE GRUPOS (VARIATION WITHIN GROUPS):** Ver soma dos quadrados dos resíduos.

**VARIAÇÃO DO ERRO (ERROR VARIATION):** O que permanece não explicado por um modelo de regressão ou de análise de variância ANOVA (ANAVA). Também chamado de variação residual ou de soma dos quadrados dos resíduos, ele reflete erros de medidas e variação devido a causas que não são especificadas no modelo estatístico.

**VARIAÇÃO ENTRE GRUPOS (VARIATION AMONG GROUPS):** Em uma análise de variância, o desvio quadrado das médias dos grupos de tratamentos da média geral, somados para todos os grupos de tratamentos. Também é chamada de soma dos quadrados entre grupos.

**VARIÂNCIA RESIDUAL:** A parte da variância total de uma variável que permanece após a remoção dos efeitos de certos fatores; mede a variabilidade que não pode ser explicada pelo modelo. A variância residual é a média do quadrado residual em uma análise de variância ou ANOVA.

**VARIÁVEIS DUMMY:** É a variável que possui códigos, normalmente 0 e 1 para representar os resultados de uma variável binária nominal ou ordinal. Ao escolher uma categoria de uma variável categórica nominal com  $k$  categorias, como categoria de referência em uma análise de regressão, é possível criar uma série de  $(k-1)$  variáveis dummy, permitindo que uma das categorias seja comparada com a categoria de referência. Também conhecida como variável indicadora ou variável muda.

**VARIÁVEIS INDICADORAS:** Consultar variáveis dummy.

**VARIÁVEL:** Uma característica que pode assumir valores diferentes de elemento para elemento ou variáveis entre indivíduos ou grupos em uma população ou amostra, como por exemplo, altura, peso, diâmetro, número de sementes por fruto, número de insetos capturados, sexo que pode ser macho ou fêmea.

**VARIÁVEL:** Conceitos, características amostrais ou populacionais que podem assumir diferentes valores de elemento para elemento. Toda variável possui pelo menos dois componentes os quais são o nome

desta e sua escala de classificação nominal, ordinal, razão e intervalo junto com a unidade de medida que ela possa assumir. Exemplos, altura e diâmetro de uma árvore em centímetro, tipo de raça de caprino, canindé, moxotó, anglo nubiana, etc. tipo de raça bovina, tais como, gir, nelore, indubrasil, guzerá, sindi, etc. estado civil, solteiro, casado, divorciado, etc., tipo de capim como elefante, colômbio, etc.

**VARIÁVEL ALEATÓRIA:** Uma variável que pode assumir diversos valores em determinadas probabilidades. Todos os valores que a variável aleatória pode assumir, com suas probabilidades associadas, compreendem a distribuição de probabilidade da variável aleatória.

**VARIÁVEIS ENDÓGENAS:** São variáveis dependentes, num sistema de equações simultâneas. São as variáveis que são determinadas pelo sistema, ainda que elas também apareçam como variáveis explicativas em alguma outra equação do sistema.

**VARIÁVEIS EXÓGENAS:** São aquelas variáveis que são determinadas fora do modelo. Elas também incluem variáveis endógenas defasadas, desde que seus valores já sejam conhecidos em algum dado período. As variáveis exógenas e as variáveis endógenas defasadas são algumas vezes chamadas variáveis pré-determinadas.

**VARIÁVEL ALEATÓRIA:** Uma variável aleatória é uma variável cujos valores estão associados com alguma probabilidade de serem constatadas. Por exemplo, no lançamento de um dado honesto, temos 6 resultados mutuamente exclusivos que são 1, 2, 3, 4, 5 ou 6, cada um associado com uma probabilidade de ocorrer de  $\frac{1}{6}$ . Desta forma, o resultado do lançamento de um dado é uma variável aleatória.

**VARIÁVEL ALEATÓRIA DISCRETA:** Uma variável aleatória discreta é aquela que pode assumir somente valores finitos ou distintos. Por exemplo, os resultados do lançamento de um dado constituem variáveis aleatórias discretas porque elas se limitam aos valores 1, 2, 3, 4, 5 ou 6. Isto é, se diferenciam das variáveis contínuas, as quais podem assumir um número infinito de valores dentro de qualquer intervalo dado.

**VARIÁVEIS DUMMIES (OU MUDAS):** Variáveis explicativas qualitativas tais como tempo de guerra versus tempo de paz, períodos de greve versus período de não greve, homem versus mulher, etc. que podem ser introduzidas na análise de regressão adotando-se o valor de 1 para uma classificação por exemplo, tempo de guerra e 0 para outra por exemplo, tempo de paz. Essas são chamadas variáveis dummies, e são tratadas como qualquer outra variável. As variáveis dummies podem ser usadas para captar variações ou mudanças no intercepto, variações na declividade, e variações em ambos, intercepto e declividade:

$$Y = b_0 + b_1X + b_2D + u$$

$$Y = b_0 + b_1X + b_2XD + u$$

$$Y = b_0 + b_1X + b_2D + b_3XD + u$$

Onde D é 1 para uma classificação e 0 para outra, e X é a variável explicativa usual quantificável. As variáveis dummies também podem ser usadas para captar diferenças entre mais que duas classificações, tais como estações do ano e regiões:

$$Y = b_0 + b_1X + b_2D_1 + b_3D_2 + b_4D_3 + u$$

Onde  $b_0$  é o intercepto para a primeira estação ou região e  $D_1$ ,  $D_2$  e  $D_3$  referem-se, respectivamente, a estação ou região 2, 3, e 4. Observe que para qualquer número de classificações k, k-1 dummies são necessárias.

**VARIÁVEL DEPENDENTE QUALITATIVA:** Uma variável dependente qualitativa é chamada dicotômica, indicando a ocorrência ou não ocorrência de um evento ou a presença ou a ausência de uma condição. Por exemplo, uma pessoa ou está ou não está na força de trabalho, é empregada ou é desempregada. Uma pessoa possui um carro e uma casa, e vai ou não ao colégio. Nesses casos, costuma-se atribuir à ocorrência do evento ou presença da condição a um valor igual a 1, enquanto à não ocorrência e ausência é dado um valor igual a 0. Quando a variável dependente é dicotômica ou qualitativa, o método dos mínimos quadrados pode ainda ser usado para estimar uma equação de regressão, porém surgem diversos problemas. Primeiro, a hipótese da normalidade do termo erro,  $u$ , do modelo de regressão linear clássico, é violada. Esta hipótese é necessária para aplicar os testes de significância nos parâmetros. Contudo, a aproximação pode ser feita pelo teorema do limite central para amostras grandes ( $n \geq 30$ ). Segundo, a hipótese que o termo erro é não correlacionado com a variável explicativa,  $s$ , é também violada. Isto conduz a estimadores não viesados, porém ineficientes isto é, estimadores que não tem variância mínima. Vale salientar também que os valores estimados da variável dependente podem assumir resultados fora da variação de 0 a 1. Esta dificuldade pode ser superada pela condensação das probabilidades estimadas dentro da variação de 0 a 1, usando ou a função normal cumulativa denominada probit model ou a função logística conhecida por logit model. Essas estimações técnicas são apresentadas em livros e textos mais avançados de econometria.

**VARIÁVEL BINÁRIA:** Uma variável aleatória discreta com apenas dois valores possíveis. Também denominada variável dicotômica.

**VARIÁVEL CATEGÓRICA (QUALITATIVA):** Cada indivíduo pertence a um de dois ou mais categorias mutuamente exclusivas da variável.

**VARIÁVEL CONTÍNUA:** Pode assumir um conjunto infinito de valores possíveis em determinada amplitude.

**VARIÁVEL DEPENDENTE:** É a variável em um modelo de regressão que pode ser prevista pela(s) variável(eis) explicativa(s), independente(s), variável regressora, variável preditora, variável explanatória ou covariável. Também conhecida como resposta ou variável resposta.

**VARIÁVEL DISTINTA:** Envolve apenas um conjunto finito de valores possíveis.

**VARIÁVEL EXPLICATIVA:** Consultar variável independente.

**VARIÁVEL INDEPENDENTE:** É a variável em uma análise de regressão que é utilizada para prever o valor da variável dependente ou resposta. É também chamada de regressora, variável explicativa ou preditiva ou a covariável.

**VARIÁVEL ORDENADA:** Uma variável categórica que tem alguma base de ordem ou de postos nas várias categorias; por exemplo, escores da condição corporal, faixas etárias.

**VARIÁVEL PREDITIVA:** Consultar variável independente.

**VARIÁVEL QUALITATIVA:** É um indivíduo que pertence a qualquer uma das duas ou mais categorias distintas da variável. Também chamada de variável categórica.

**VARIÁVEL QUANTITATIVA:** Consiste em valores numéricos em uma escala bem definida. Também chamada variável numérica.

**VARIÁVEL REGRESSORA:** Consultar variável independente.

**VARIÁVEL RESPOSTA:** Consultar variável dependente.

**VARIÁVEL RESULTANTE:** Consultar variável dependente.

**VEROSSIMILHANÇA:** A probabilidade de obter os resultados observados em determinado modelo.

**VARIABILIDADE:** Uma característica que pode assumir diversos valores numéricos é chamada de variável, e a susceptibilidade com que muda de valor é chamada de variabilidade. Em inglês Variability.

**VARIABILIDADE:** Termo que indica o nível de dispersão dos dados de uma série ou distribuição de frequência em um levantamento ou experimento em torno de um valor central, tal como a média, mediana ou moda. A amplitude total, a variância ou quadrado médio, o desvio padrão e o desvio padrão da média são medidas de variabilidade absoluta, já o coeficiente de variação é uma medida de dispersão relativa, aplicadas de forma frequente em análises estatísticas.

**VARIABILIDADE:** O grau de heterogeneidade de um grupo, medido pela amplitude, pela amplitude semi-interquartil ou pelo desvio padrão. O mesmo que dispersão. Em inglês variability.

**VARIABILIDADE:** Muitas variáveis apresentam algum grau de variação em suas medidas, ou seja, seus valores não são constantes entre ou dentro das unidades de experimentação.

**VARIABILIDADE:** Capacidade de uma espécie, população, raça ou linhagem em expressar diferentes fenótipos.

**VARIABILIDADE INTEROBSERVADOR:** Medida da discordância entre diferentes observadores.

**VARIABILIDADE INTRA-OBSERVADOR:** Medida da inconsistência entre avaliações repetidas feitas por um mesmo observador.

**VARIABILIDADE METODOLÓGICA:** É o erro introduzido pelo processo de mensuração. Ela tem diversas causas tais como o aparelho, o operador, o meio ambiente, dentre outras, pela dificuldade em separá-las, a variação metodológica é, por vezes, globalmente designada como variação do observador ou erro do observador.

**VARIABILIDADE NATURAL:** É a variabilidade intrínseca do depósito mineral, medida pelo coeficiente de variação, veja classificação dos depósitos minerais.

**VARIAÇÃO ALEATÓRIA:** Tipo de variação em um processo devida à chance; o tipo de variação, inerente a qualquer processo, que não seja capaz de produzir bens ou serviços exatamente da mesma maneira cada vez.

**VARIAÇÃO BIOLÓGICA:** espectro de valores apresentados pelas características biológicas; por exemplo, da pressão arterial. Nas mensurações de uma característica, em um grupo de pessoas, há que se atentar para a variabilidade biológica e para a variação metodológica introduzida no processo de mensuração.

**VARIAÇÃO CONCOMITANTE:** O grau pelo qual uma causa presumível e um efeito presumível ocorrem ou variam simultaneamente.

**VARIAÇÃO CONCOMITANTE:** É o relacionamento de uma série de valores numéricos pareados. Ela mede o grau de associação entre esses valores. Se as variáveis mudam na mesma direção, o relacionamento é positivo.

**VARIAÇÃO CONCOMITANTE, PRINCÍPIO DE:** Se a variação da intensidade de um fator resulta numa variação paralela do efeito, então, este fator é uma causa.

**VARIAÇÃO DENTRO DE AMOSTRAS:** Em análise da variância, a variação que é devida à chance.

**VARIAÇÃO DETERMINÁVEL:** Tipo de variação em um processo que resulta de causas que podem ser identificadas.

**VARIAÇÃO DEVIDA AO ERRO:** Veja variação dentro de amostras.

**VARIAÇÃO DEVIDA AO TRATAMENTO:** Veja variância entre amostras.

**VARIAÇÃO DO OBSERVADOR:** Erro metodológico, introduzido pelo observador no processo de mensuração de um evento. Pode ser de dois ou mais observadores ao mesmo tempo que é a variação entre observadores ou de um mesmo observador em diferentes ocasiões que é a variação intra-observador. Ver variação biológica.

**VARIAÇÃO EXPLICADA:** Soma dos quadrados dos desvios explicados para todos os pares de dados bivariados em uma amostra. Em inglês explained variation.

**VARIAÇÃO MARGINAL:** Para variáveis relacionadas por uma equação de regressão, a variação na variável dependente quando uma das variáveis independentes sofre uma variação de uma unidade e as outras variáveis independentes permanecem constantes.

**VARIAÇÃO NÃO-EXPLICADA:** Soma dos quadrados dos desvios não-explicados para todos os pares de dados bivariados em uma amostra. Em inglês unexplained variation.

**VARIAÇÃO SAZONAL DE UMA DOENÇA:** Distribuição cronológica de casos em que os valores máximos e mínimos de frequência ocorrem sempre no mesmo período, ou seja, do ano, do mês, da semana ou do dia, caracterizando uma sazonalidade.

**VARIAÇÃO TOTAL (VT):** É a soma das variações explicadas mais a da variação não explicada. Em inglês total variation.

**VARIAÇÃO TOTAL:** Igual à soma dos desvios em relação à média elevados ao quadrado. É geralmente chamada de soma total dos quadrados (STQ), mas algumas vezes é referida simplesmente como a soma dos quadrados (SQ).

**VARIAÇÕES ESTACIONAIS:** São as flutuações dos termos de uma série de tempo em redor de sua tendência secular, que se repetem nas mesmas épocas e que podem ser atribuídas ao ritmo próprio às causas naturais do fenômeno em estudo. Como exemplo podemos citar o barateamento de um gênero alimentício, na época da colheita e venda, seguido do seu encarecimento na época do plantio; tal variação não denota nem prosperidade nem crise. Diz-se, também, variações sazonais.

**VARIÂNCIA:** É uma medida de variabilidade mais comumente usada. São expressos em unidades ao quadrado na escala de mensuração, e indicam a dispersão na distribuição.



**VARIÂNCIA:** A variância cujo conceito se deve ao matemático Inglês Sir Ronald Aylmer Fisher (FISHER, 1918), é o segundo momento de uma distribuição de frequência considerada em torno da média aritmética. Em inglês Variance.

**VARIÂNCIA:** Medida de dispersão para uma variável aleatória ou para uma lista de números; simboliza-se por  $\sigma^2$ , onde  $\sigma$  é a letra grega minúscula sigma; o desvio-padrão é a raiz quadrada da variância.

**VARIÂNCIA:** É a medida de dispersão associada à média:  $Var(X) = \sum_{i=1}^n P(X_i) (X_i - \bar{X})^2$ . Assumindo a

probabilidade de ocorrência das n variáveis aleatórias igual a  $\frac{1}{n}$ , tem-se:

$$Var[X] = \sum_{i=1}^n \frac{1}{n} (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \text{ ou desenvolvendo-a: } Var[X] = E[X^2] - \bar{X}^2.$$

**VARIÂNCIA:** A variância é o quadrado do desvio padrão. Também conhecida como quadrado médio, ou soma de quadrados de desvios ou dos resíduos.

**VARIÂNCIA:** Desvio padrão elevado ao quadrado. É, como o desvio padrão, uma medida de dispersão. Em inglês variance.

**VARIÂNCIA:** É a medida que permite avaliar o grau de dispersão dos valores da variável em relação à média.

**VARIÂNCIA:** A média do quadrado do desvio de uma observação, em relação à média aritmética; uma medida de variação. Assume valores nulos ou maiores que zero, nunca valores negativos.

**VARIÂNCIA:** Soma dos quadrados dos desvios em relação à média, dividida pelo número de observações da amostra menos um ( $n-1$ ), que é o número de graus de liberdade ou correção de Bessel.

**VARIÂNCIA:** Estatística que descreve a variação observada em uma característica. Sem variação, nenhum progresso genético é possível, uma vez que animais geneticamente superiores não seriam distinguíveis dos geneticamente inferiores.

**VARIÂNCIA:** Soma dos desvios em relação à média elevados ao quadrado, dividida pelo número de observações menos 1 (número de graus de liberdade que é n-1).

**VARIÂNCIA:** É o parâmetro  $\sigma^2$  de uma distribuição normal, em que  $\sigma$  é o desvio padrão.

**VARIÂNCIA:** É o desvio padrão elevado ao quadrado e é simbolizado por  $S^2$ . A variância é uma medida que tem pouca utilidade como estatística descritiva, porém é extremamente importante na inferência estatística e em combinações de amostras.

**VARIÂNCIA:** Também conhecida como média quadrática é a média dos quadrados das diferenças entre as observações e a média geral. Ela nos dá uma estimativa da quantidade de variação em um conjunto de dados.

**VARIÂNCIA:** As somas dos desvios ao quadrado da média dividido pelo número de observações menos um.

**VARIÂNCIA:** É a média do quadrado das distâncias euclidianas que cada ponto do conjunto está da média aritmética. Em inglês variance.

**VARIÂNCIA AMOSTRAL ( $S^2$ ):** Serve para estimar a variância populacional, através de observações obtidas em uma amostra retirada desta população.

**VARIÂNCIA AMOSTRAL:** Variância calculada com base nos valores de uma amostra; usada para estimar o valor da variância populacional.

**VARIÂNCIA DA VARIÁVEL ALEATÓRIA Y:** Símbolo estatístico representado ou dado por  $Var(Y) = V(X) = \sigma^2 = \sigma_X^2 = \sigma_{(X)}^2$ .

**VARIÂNCIA AGRUPADA:** Uma estimativa da variância de uma população baseada na combinação de duas ou mais variâncias de amostra. A estimativa de variância agrupada é apropriada toda vez que as variâncias de duas ou mais populações são consideradas iguais.

**VARIÂNCIA AMOSTRAL:** Variância de uma amostra. Em inglês Sample variance.

**VARIÂNCIA A PRIORI:** Mesmo que variância amostral ou simplesmente variância, é conhecida em notação geoestatística como:  $C_0 = E[Z(X) - m^2]$ , ou seja, a covariância  $C(h)$ , para distância nula.

**VARIÂNCIA COMUM:** Variância compartilhada com outras variáveis na análise fatorial.

**VARIÂNCIA DE CO-KRIGAGEM:** A variância de estimativa por co-krigagem é calculada

$$\text{como: } \sigma_{CK}^2 = \bar{C}(V, V) + \mu_i - \sum_{\alpha=1}^n \eta_{\alpha} \bar{C}(V_{\alpha}, V) - \sum_{\beta=1}^m \lambda_{\beta} \bar{C}(V_{\beta}, V).$$

**VARIÂNCIA DE DISPERSÃO:** Com definição semelhante àquela dada na estatística clássica, refere-se a medida de dispersão de uma variável aleatória de suporte  $v$ , dentro do volume maior  $V$ , que pode ser calculada como:  $D^2(v, V) = E\{S^2(X)\} = E\left\{\frac{1}{N} \sum_{i=1}^N [Z_v(X) - Z_v(X_i)]^2\right\}$ . Onde  $Z_v(x_i)$  são os teores unitários de

suporte  $v_i$ ;  $Z_v(x)$  é o teor médio no volume  $V$ , que pode ser determinado como:  $Z_v(X) = \frac{1}{N} \sum_{i=1}^N Z_v(X_i)$ .

**VARIÂNCIA DE ESTIMATIVA:** É o erro cometido pelas múltiplas extensões de amostras de teores  $Z(x)$ , para estimar o teor do bloco.

**VARIÂNCIA DE EXTENSÃO:** É o erro cometido pela extensão de uma pequena amostra de teor  $Z(x)$  para estimar o teor do bloco, de volume muito maior.

**VARIÂNCIA DE INTERPOLAÇÃO:** É uma medida local utilizada em geoestatística da variância de estimativa baseada nos ponderadores da krigagem ordinária:  $\sigma^2 = \sum_{i=1}^n \lambda_i [Z(X_i) - Z_v^*]^2$ , onde  $Z(x_i)$  é o teor do  $i$ -ésimo ponto de amostragem;  $Z_v^*$  é o valor estimado do teor.

**VARIÂNCIA DE KRIGAGEM DA MÉDIA:** Ou variância de estimativa da média, pode ser calculada, como:  $\sigma_{KM}^2 = \sum_{\alpha=1}^n \sum_{\beta=1}^m \lambda_{\alpha} \lambda_{\beta} C(V_{\alpha}, V_{\beta}) = \mu$ . Neste caso, a variância de estimativa da média é igual ao próprio multiplicador de Lagrange ( $m$ ).

**VARIÂNCIA DE KRIGAGEM ORDINÁRIA:** É a quantificação da variância de estimativa, a partir do variograma de semivariâncias ou covariâncias e da configuração espacial dos dados ponderadores:  $\sigma_E^2 = \bar{C}(V, V) + \mu - \sum_{\alpha=1}^n \lambda_{\alpha} \bar{C}(V_{\alpha}, V)$ , ou em termos de semivariâncias:  $\sigma_E^2 = \sum_{\alpha=1}^n \lambda_{\alpha} \bar{\gamma}(V_{\alpha}, V) + \mu - \bar{\gamma}(V, V)$ .

**VARIÂNCIA DE KRIGAGEM TRANSITIVA:** É a quantificação da variância de estimativa, quando se utilizar desta técnica:  $\sigma_{TK}^2 = \rho(0) - \sum_{\alpha=1}^n \lambda_{\alpha} \rho(X_{\alpha})$ .



**VARIÂNCIA DE KRIGAGEM UNIVERSAL:** Permite determinar a variância de estimativa usando a técnica da krigagem universal:  $\sigma_{UK}^2 = \bar{C}(V, V) + \sum_{i=1}^k \mu_i b_i - \sum_{\alpha=1}^n \lambda_{\alpha} \bar{C}(V_{\alpha}, V)$ .

**VARIÂNCIA DENTRO DOS GRUPOS:** Medida baseada na variação dentro de cada grupo isto é, variação em torno da média de um único grupo.

**VARIÂNCIA DO ERRO:** Variância de uma variável devido a erros na coleta de dados ou na medida.

**VARIÂNCIA ENTRE AMOSTRAS:** Em análise da variância é a variação entre as diferentes amostras.

**VARIÂNCIA ENTRE OS GRUPOS:** Medida da variação entre as médias de mais de um grupo, com base nas variáveis independentes em estudo.

**VARIÂNCIA ESPACIAL:** É variância observada entre a variância a priori e o efeito pepita.

**VARIÂNCIA ESPECÍFICA:** Variância de cada variável, única àquela variável e que não é explicada ou associada com outras variáveis na análise fatorial.

**VARIÂNCIA POPULACIONAL ( $\sigma^2$ ):** É um valor que mede o grau de dispersão dos valores da variável, na população, em relação à média populacional. Definida como a soma dos quadrados dos desvios dos valores da variável em relação a sua média, dividido pelo número de observações:  $\frac{\text{Soma}(X_i - \text{média})^2}{N}$ .

**VARIÂNCIA POPULACIONAL:** Variância calculada quando se conhecem todos os elementos da população. Em contraste, ver variância amostral.

**VARIÂNCIA RELATIVA:** É o quociente entre a variância e o quadrado da média. Em inglês relative variance.

**VARIÂNCIA DO QUADRADO MÉDIO:** Ver quadrado médio.

**VARIÁVEL:** Condição ou característica observada em cada paciente, como por exemplo, idade, história de infarto do miocárdio, nível de glicose no sangue, que pode assumir valores diferentes e é observada e registrada uma ou mais vezes ao longo da pesquisa.

**VARIÁVEL:** Determinada característica dos indivíduos ou elementos objetos da investigação, como são exemplos o sexos dos animais, o estado civil, a etnia, peso de recém-nascidos, dentre outras.

**VARIÁVEL:** Termo de uma função ou relação sujeita á alterações de valor; quantidade que pode assumir qualquer valor de um conjunto específico de valores. Propriedade real medida por observações individuais.

**VARIÁVEL:** Geralmente é qualquer quantidade que varia de elemento para elemento. Mais precisamente, uma variável no sentido matemático, por exemplo, uma quantidade que pode tomar qualquer uma de uma série específica de valores. É conveniente usar a mesma palavra para se referir a características não mensuráveis. Por exemplo, sexo é uma variável neste sentido, desde que um indivíduo humano pode tomar um de dois valores, masculino ou feminino. Em inglês Variable.

**VARIÁVEIS EXPLANATÓRIAS OU COVARIÁVEIS:** Em análise de sobrevivida ou de sobrevivência, interessa estudar o tempo de vida dos indivíduos, tempo esse que pode ser influenciado por inúmeros fatores, tais como as características do próprio indivíduo como sexo, idade, etc. e fatores exteriores ao

indivíduo tratamentos a que foi sujeito, fatores ambientais, etc.. A esses fatores denominam-se variáveis explanatórias, concomitantes ou covariáveis.

**VARIÁVEL:** É qualquer característica populacional mensurável de acordo com alguma escala que varia de elemento para elemento ou de indivíduo para indivíduo, se contrapõe ao conceito de constante K que é qualquer característica populacional mensurável de acordo com alguma escala de medida e que não varia de indivíduo para indivíduo.

**VARIABILIDADE:** O grau de heterogeneidade de um grupo, medido pela amplitude, pela amplitude semi-interquartil ou pelo desvio-padrão. O mesmo que dispersão.

**VARIÂNCIA:** Desvio padrão elevado ao quadrado. É, como o desvio padrão, uma medida de dispersão.

**VARÁVEL ALEATÓRIA:** A uma grandeza numérica, que assume diferentes valores, estando cada um destes valores associado a uma certa probabilidade, dá-se o nome de variável aleatória.

**VARÁVEL ALEATÓRIA CONTÍNUA:** É aquela que pode assumir valores, mesmo dentro de um intervalo finito, resultando que a probabilidade de um dado valor ou ponto é nula.

**VARÁVEL ALEATÓRIA DISCRETA:** É aquela que pode assumir, com probabilidade diferente de zero, um número finito de valores dentro de um intervalo finito. Um caso típico é quando se efetua contagens.

**VARÁVEL CONTÍNUA:** Diz-se da variável que pode assumir, teoricamente, qualquer valor em um certo intervalo real. Exemplo: a altura dos alunos constitui uma variável contínua, pois, teoricamente, um aluno poderá ter altura igual a 1,800 m, 1,810 m, 1,811 m , 1,812 m,...

**VARÁVEL DISCRETA:** É aquela que assume valores em pontos da reta real. Exemplo: número de erros em um livro: 0, 1, 2, 3,...

**VARÁVEL DEPENDENTE:** É a variável não controlada num experimento, sendo, por definição, aleatórios seus valores.

**VARÁVEL INDEPENDENTE:** É a variável que pode ser controlada. Em outras palavras, seus valores são exatos.

**VARIÁVEL:** É uma característica da população. Toda questão de pesquisa define um número de construções teóricas que o pesquisador quer associar. O grau de operacionalização destas construções não faz parte de um consenso. Por essa razão, a seção que trata das definições das variáveis deve permitir ao leitor avaliar a adequação dos instrumentos utilizados, as variáveis escolhidas e as construções teóricas descritas no quadro conceitual.

**VARIÁVEL:** Símbolo utilizado para representar um objeto matemático que pode assumir mais de um resultado. Em inglês variable.

**VARIÁVEL:** Corresponde, por exemplo, em genética, aos diferentes fenótipos de um mesmo caráter.

**VARIÁVEL:** Propriedade que determina a maneira pela qual os elementos de qualquer conjunto são diferentes entre si.

**VARIÁVEL:** Característica ou atributo que pode tomar diferentes valores. Exemplos: sexo como masculino e feminino, idade, profissão e número de filhos. Em termos práticos, variável é a característica em que se coletam dados em uma investigação.

**VARIÁVEL:** Medida de uma única característica mensurável ou contável.

**VARIÁVEL:** Uma característica comum a todos os dados, uma característica populacional mensurável de acordo com alguma escala de medida ou de mensuração ou também de contagem. Por exemplo o peso ao nascer de bezerros machos bovinos em quilogramas da raça guzerá. Em inglês variable.

**VARIÁVEL (TIPOS):** As variáveis podem ser dos tipos qualitativa tais como sexo e cor, quantitativa discreta ou descontínua, como número de batimentos cardíacos e número de filhos e quantitativa contínua como temperatura, peso e comprimento.

**VARIÁVEL, CLASSIFICAÇÃO:** Sujeitos experimentais podem ser classificados numa característica que estava presente antes e destacadamente do experimento. Tal característica não é o resultado das manipulações do pesquisador. Exemplos podem ser: sexo, idade, doença, quociente de inteligência (QI), status econômico, dentre outras.

**VARIÁVEL ALEATÓRIA:** É uma variável que recebe um certo número de valores de acordo com uma certa distribuição de probabilidades. As variáveis aleatórias, podem ser variáveis discretas tendo um número limitado de valores tal como o dado, ou variáveis contínuas tal como uma amostra de ouro, que pode assumir qualquer valor dentro do intervalo total.

**VARIÁVEL ALEATÓRIA:** É aquela que pode assumir qualquer conjunto de valores, positivos ou negativos, com uma dada probabilidade. É sinônimo de variável estocástica.

**VARIÁVEL ALEATÓRIA:** Variável cujo valor depende do resultado de um experimento.

**VARIÁVEL ALEATÓRIA:** Para um determinado espaço amostral  $\Omega$ , uma variável aleatória, denotada por  $X$ , é uma função com domínio  $\Omega$  e contradomínio a reta real. Em inglês Random variable.

**VARIÁVEL ALEATÓRIA:** É uma função que associa a cada elemento  $s$  ( $s$  minúsculo) pertencente a um espaço amostra  $S$  ( $S$  maiúsculo) de um experimento ou ensaio aleatório, um número real  $x = X(s)$ . Em inglês random variable.

**VARIÁVEL ALEATÓRIA:** Variável que pode assumir qualquer um de um conjunto de valores diferentes, associados, cada um, a determinada probabilidade.

**VARIÁVEL ALEATÓRIA:** Variável tipicamente representada por uma letra do alfabeto latino como  $x$ ,  $y$  ou  $w$ , que tem um valor numérico único determinado pela chance para cada resultado de um experimento. Toda variável associada a uma probabilidade de ocorrência em um experimento ou ensaio aleatório, casual, estocástico ou não determinístico.

**VARIÁVEL ALEATÓRIA:** É uma variável cujo valor numérico atual é determinado por probabilidades. Por exemplo,  $X$ : pontuação na escala de atitudes em relação à estatística,  $Y$ : número de disciplinas reprovadas em Estatística, dentre outras. Observe que o resultado depende do aluno selecionado. A variável aleatória tem uma distribuição de probabilidades associada, o que nos permite calcular a probabilidade de ocorrência de certos valores.

**VARIÁVEL ALEATÓRIA BERNOULLI:** Variável aleatória com distribuição de Bernoulli. Em inglês Bernoulli random variable.

**VARIÁVEL ALEATÓRIA BINOMIAL:** Variável aleatória discreta que tem distribuição binomial. Em inglês Binomial random variable.

**VARIÁVEL ALEATÓRIA BINOMIAL NEGATIVA:** Variável aleatória discreta com distribuição binomial negativa. Em inglês Negative binomial random variable.

**VARIÁVEL ALEATÓRIA:** Variável descritora de populações infinitas, a cujos valores são associadas probabilidades da ocorrência de seus possíveis valores, os quais representam eventos aleatórios resultantes de um experimento não determinístico.

**VARIÁVEL ALEATÓRIA CONTÍNUA (V.A.C.):** Variável aleatória com um número infinito de valores que podem ser associados a pontos de um intervalo contínuo da reta.

**VARIÁVEL ALEATÓRIA CONTÍNUA:** Variável aleatória que pode assumir valores nos números reais. Em inglês Continuous random variable.

**VARIÁVEL ALEATÓRIA DE POISSON:** Variável aleatória discreta com distribuição de Poisson. Em inglês Poisson random variable.

**VARIÁVEL ALEATÓRIA DISCRETA (V.A.D.):** Variável aleatória que tem um número finito, ou uma quantidade numerável de valores.

**VARIÁVEL ALEATÓRIA EXPONENCIAL:** Variável aleatória que possui distribuição exponencial. Em inglês Exponential random variable.

**VARIÁVEL ALEATÓRIA GAMA:** Variável aleatória com distribuição gama. Em inglês Gamma random variable.

**VARIÁVEL ALEATÓRIA GEOMÉTRICA:** Variável aleatória discreta que tem distribuição geométrica. Em inglês Geometric random variable.

**VARIÁVEL ALEATÓRIA HIPERGEOMÉTRICA:** Variável aleatória discreta que tem distribuição hipergeométrica. Em inglês Hypergeometric random variable.

**VARIÁVEL ALEATÓRIA LOGNORMAL:** Variável aleatória com distribuição lognormal. Em inglês Lognormal random variable.

**VARIÁVEL ALEATÓRIA MULTINOMIAL:** Variável aleatória discreta que tem distribuição multinomial. Em inglês Multinomial random variable.

**VARIÁVEL ALEATÓRIA NORMAL PADRONIZADA:** Variável aleatória com distribuição normal de média zero (0) e variância ou desvio padrão igual a um (1). Em inglês Standard normal random variable.

**VARIÁVEL ALEATÓRIA NORMAL:** Variável aleatória contínua que tem distribuição Normal. Em inglês Normal random variable.

**VARIÁVEL ALEATÓRIA PASCAL:** Variável aleatória discreta com distribuição Pascal, também chamada distribuição binomial negativa. Em inglês Pascal random variable.

**VARIÁVEL ALEATÓRIA POSITIVA:** Variável aleatória que assume valores positivos. Em inglês Positive random variable.

**VARIÁVEL ALEATÓRIA UNIFORME:** Variável aleatória, discreta ou contínua, que tem distribuição teórica de probabilidade uniforme ou retangular. Em inglês Uniform random variable.

**VARIÁVEL ALEATÓRIA CONTÍNUA:** Uma variável aleatória é dita contínua se o seu conjunto de valores  $(X(S))$  for contínuo, isto é, for o conjunto dos reais ou um intervalo real. Em inglês continuous random variable.

**VARIÁVEL ALEATÓRIA CONTÍNUA:** Variável aleatória que pode tomar qualquer valor real em um dado intervalo; é caracterizado por uma função de densidade tal que a área sob curva respectiva entre dois números representa a probabilidade de a variável aleatória estar entre esses números. Em contraste, ver variável aleatória discreta.

**VARIÁVEL ALEATÓRIA DISCRETA:** Uma variável aleatória é dita discreta se o seu conjunto de valores  $[X(S)]$  for discreto, isto é, finito ou infinito enumerável. Em inglês discrete random variable.

**VARIÁVEL ALEATÓRIA DISCRETA:** Variável aleatória cujos valores podem ser todos relacionados. Em contraste, ver variável aleatória contínua.

**VARIÁVEL ALEATÓRIA UNIDIMENSIONAL:** É uma variável aleatória definida sobre um espaço amostral unidimensional. Em inglês One-dimensional random variable.

**VARIÁVEL ANTECEDENTE (Z):** Variável de natureza relacional, sendo aquela variável de prova com a finalidade de explicar a relação  $X \rightarrow Y$ .

**VARIÁVEL BINÁRIA:** Variável que só assume um de dois valores possíveis, zero ou . veja variável dicotômica.

**VARIÁVEL BINOMIAL:** Uma variável com apenas dois atributos é binomial. Exemplo é sexo, com os atributos masculino e feminino.

**VARIÁVEL CASUAL (OU HIPOTETICAMENTE CASUAL):** É o fator casual em estudo. É uma das duas variáveis principais de uma investigação epidemiológica: a outra é a variável efeito. Poder ser um fator de risco, como por exmplo, hábito de fumar, nível elevado de colesterol ou uma intervenção como aplicação de uma vacina, de um medicamento e de um programa. A variável casual é sempre uma variável independente. Por vezes, usa-se exposição principal ou simplesmente a palavra exposição, para designá-la.

**VARIÁVEL CATEGÓRICA:** É a variável que usa valores que servem meramente como um rótulo, atributo, qualidade própria de um elemento ou indivíduo de uma população ou amostra ou meio de identificação. Também é chamada de variável não-métrica, nominal, binária, qualitativa ou taxonômica. O sexo, o estado civil, tipo de religião, tipo de raça, a cor da pele, o tipo sanguíneo, o número de um uniforme de futebol são exemplos deste tipo de variável.

**VARIÁVEL COEXISTÊNCIA:** Variável que expressa relação de causa e efeito, onde se X ocorre, então também ocorrerá Y.

**VARIÁVEL CONCOMITANTE:** Fenômeno mensurável que está estatisticamente associado com as respostas de principal interesse. É um método usado para reduzir variabilidade num experimento. Entretanto, as variáveis concomitantes não devem ser afetadas pelos tratamentos.

**VARIÁVEL CONTÍNUA:** Uma variável que pode assumir qualquer valor dentro de um intervalo, na escala de mensuração. Uma variável contínua pode receber qualquer valor dentro de uma abrangência definida de valores. Os possíveis valores da variável pertencem a uma série contínua. Entre quaisquer dois valores da variável pode ocorrer um grande número de valores intermediários. Altura, peso e tempo são exemplos de variáveis contínuas.

**VARIÁVEL CONTÍNUA:** Variável cujos possíveis valores formam um intervalo de número reais  $\mathfrak{R}$  ou cujo campo de definição é um subintervalo sob o conjunto dos números reais e que resultam, normalmente, de uma mensuração, como por exemplo, peso em quilogramas, pressão arterial, temperatura, dentre outras.

**VARIÁVEL CONTÍNUA:** Resulta normalmente de uma mensuração, e a escala numérica de seus possíveis valores corresponde ao conjunto  $\mathfrak{R}$  dos números Reais, ou seja, podem assumir, teoricamente, qualquer valor entre dois limites.

**VARIÁVEL CONTINGENCIAL:** Variável que expressa ligação de relação, onde se X ocorre, então ocorrerá Y somente se Z estiver presente.

**VARIAÇÃO RESIDUAL (RESIDUAL VARIATION):** O que é deixado ou não é explicado por um modelo de regressão ou de análise de variância conhecida como ANOVA ou ANAVA. Também é chamada de variação do erro ou de soma dos quadrados dos resíduos.

**VARIÂNCIA (VARIANCE):** Uma medida de quão longe os valores observados diferem do valor esperado.

**VARIÂNCIA AMOSTRAL (SAMPLE VARIANCE):** Uma estimativa sem tendência da variância de uma amostragem. Igual à soma dos quadrados dos desvios dividida pelo tamanho amostral menos um.

**VARIÁVEL ALEATÓRIA (RANDOM VARIABLE):** A função matemática que atribui um valor numérico a cada resultado experimental.

**VARIÁVEL ALEATÓRIA BINOMIAL (BINOMIAL RANDOM VARIABLE):** O resultado de um experimento que consiste de múltiplas tentativas de Bernoulli.

**VARIÁVEL ALEATÓRIA (ESTOCÁSTICA) CONTÍNUA (CONTINUOUS RANDOM VARIABLE):** O resultado de um experimento que pode assumir qualquer valor quantitativo. Comparar com variável aleatória discreta.

**VARIÁVEL ALEATÓRIA (ESTOCÁSTICA) DE BERNOULLI (BERNOULLI RANDOM VARIABLE):** O resultado de um experimento em que há somente dois resultados possíveis, tais como a presença ou ausência; defeituoso ou não defeituoso; vivo ou morto; cara ou coroa; sadio ou doente. Ver também variável aleatória.

**VARIÁVEL ALEATÓRIA DE POISSON (POISSON RANDOM VARIABLE):** O resultado discreto de um dado experimento conduzido em uma área finita ou em uma quantidade de tempo finita. Diferente da variável aleatória binomial, a de Poisson pode assumir qualquer valor inteiro.



**VARIÁVEL CONTROLE:** É qualquer variável que é mantida constante numa pesquisa; observando-se apenas um dos seus níveis. Usam-se variáveis controle em pesquisa para neutralizar os efeitos de variáveis que não são de principal interesse ao estudo, mas que podem afetar o comportamento observado. Por exemplo, num estudo sobre leitura, sexo deve ser controlado, uma vez que se sabe que é relacionado com desempenho de leitura. Uma variável, que não a variável independente de interesse principal, cujos efeitos são determinados pelo pesquisador. As variáveis controles são incluídas nas pesquisas como variáveis independentes com a finalidade de explicar ou controlar variações. Também chamadas simplesmente de controles. São espécimes similares, submetidos tanto quanto possível ao mesmo tratamento aplicado aos objetos do experimento, exceto com relação à mudança da variável em estudo.

**VARIÁVEL CORRELAÇÃO:** Variável que expressa relação de correlação, onde pode significar dependência para alguma espécie de associação.

**VARIÁVEL CRITÉRIO ( $\gamma$ ):** Ver variável dependente. Também é conhecida como variável resposta.

**VARIÁVEL DEPENDENTE ( $\gamma$ ):** Variável que está sendo prevista ou explicada pelo conjunto de variáveis independentes.

**VARIÁVEL DEPENDENTE:** Pode ser definida como aquela variável que é observada e medida em resposta a uma variável independente. Espera-se que a variável dependente aumente, diminua ou varie de alguma forma sistemática à proporção que mudem os níveis da variável independente. Com qualquer outra variável dependente pode ser medida em diferentes escalas. A variável que é afetada ou tida como afetada pela variável independente.

**VARIÁVEL DEPENDENTE:** Variável que é modelada como função de variáveis independentes em análise de regressão.

**VARIÁVEL DEPENDENTE:** Variável que, em uma regressão, supomos ser causada pela variável independente.

**VARIÁVEL DEPENDENTE:** A variável que está sendo explicada ou depende de outra variável denominada de explicativa ou independente, também conhecida como variável resposta.

**VARIÁVEL ALEATÓRIA DISCRETA (DISCRETE RANDOM VARIABLE):** O resultado de um experimento que assume apenas valores inteiros. Comparar com variável aleatória contínua.

**VARIÁVEL ALEATÓRIA GAUSSIANA (GAUSSIAN RANDOM VARIABLE):** Ver variável aleatória normal.

**VARIÁVEL ALEATÓRIA LOG-NORMAL:** Uma variável aleatória cujo logaritmo natural é uma variável aleatória normal. Em inglês Lognormal random variable.

**VARIÁVEL ALEATÓRIA MULTINOMIAL:** É uma variável aleatória ou casual que pode assumir múltiplos valores. Ela é a extensão multivariada de uma variável aleatória binomial. Em inglês Multinomial random variable.

**VARIÁVEL ALEATÓRIA NORMAL (NORMAL RANDOM VARIABLE):** Uma variável aleatória cuja função de distribuição de probabilidades é simétrica em torno da média e que pode ser descrita por dois

parâmetros, sua média ( $\mu$ ) e sua variância ( $\sigma^2$ ). Também chamada de variável aleatória gaussiana, em forma de sino, em forma de campânula ou em forma de chapéu de Napoleão.

**VARIÁVEL ALEATÓRIA NORMAL MULTIVARIADA (MULTIVARIATE NORMAL RANDOM VARIABLE):** O análogo de uma variável aleatória normal ou gaussiana para dados multivariados. Sua distribuição de probabilidades é simétrica ao redor de seu vetor médio, e é caracterizada por um vetor médio n-dimensional ou multidimensional e por uma matriz de variância-covariância nxn.

**VARIÁVEL DEPENDENTE:** O mesmo que variável-efeito, variável-resposta, variável explicada ou variável predita; por exemplo, uma doença, ou o estado nutricional. A variável dependente é influenciada pelas variáveis independentes.

**VARIÁVEL DEPENDENTE:** Representada no eixo dos Y. Constitui a variável efeito, ou seja, efeito presumido da variável independente.

**VARIÁVEL DEPENDENTE:** Um símbolo ou conceito que se espera seja explicado ou causado pela variável independente ou cujo valor, acredita-se, muda em resposta à variável independente.

**VARIÁVEL DEPENDENTE (Y):** Variável de natureza relacional, sendo aquelas categorias ou valores a serem explicados ou descobertos, dado que são influenciados, determinados ou afetados pela variável independente X.

**VARIÁVEL DETERMINÍSTICA:** Variável que expressa relação de causa e efeito, onde X causa sempre Y, ou se X ocorre sempre ocorrerá Y.

**VARIÁVEL DE CONFUNDIMENTO (DE CONFUSÃO, CONFUNDIDORA OU CONFUNDÍVEL):** É a que confunde a interpretação casual. Trata-se de uma terceira variável que explica a associação ilusória entre duas outras variáveis ou seja, a exposição principal, hipoteticamente casual, e o efeito, em estudo. Por exemplo, na investigação entre contraceptivos orais e coronariopatias, o fumo é variável de confundimento. Toda variável que se suspeita ser de confundimento é candidata, em potencial, a ser controlada, isto é, ter seus efeitos neutralizados, nas investigações causais, quer na fase de planejamento, quer na fase de análise de um estado. Conhecido também como fator de confusão. Ver confundimento.

**VARIÁVEL DE CONFUSÃO:** O mesmo que variável de confundimento.

**VARIÁVEL DE CONTROLE:** A variável que é mantida constante numa tabulação cruzada tripla para apresentar três variáveis pelo uso de tabelas contingenciais.

**VARIÁVEL DE CONTROLE:** Variável mantida constante, para esclarecer a relação entre duas outras variáveis. Exemplo: tendo descoberto uma relação entre educação e preconceito, a variável sexo pode ser mantida constante, examinando a relação entre educação e preconceito somente entre homens e depois somente entre mulheres. Neste exemplo, sexo é a variável de controle.

**VARIÁVEL DE DISTORÇÃO:** No modelo de elaboração, uma variável de controle que interage com as variáveis independente e dependente, fazendo a relação das duas parecer o contrario do que de fato é.

**VARIÁVEL DICOTÔMICA:** Variável independente usada para explicar o efeito que diferentes níveis de uma variável não-métrica têm na previsão da variável dependente.



Para listar L níveis de uma variável independente não-métrica, L-1 variáveis dicotômicas são necessárias. Por exemplo, o sexo é medido como masculino ou feminino e poderia ser representado por duas variáveis dicotômicas,  $X_1$  e  $X_2$ . Quando o respondente é do sexo masculino,  $X_1 = 1$  e  $X_2 = 0$ . Do mesmo modo, quando o respondente é do sexo feminino,  $X_1 = 0$  e  $X_2 = 1$ . No entanto, quando  $X_1 = 1$ , sabemos que  $X_2$  deve ser igual a 0. Assim, precisamos de apenas uma variável  $X_1$  ou  $X_2$ , para representar o sexo. Não precisamos incluir ambas, pois uma é perfeitamente prevista pela outra e os coeficientes de regressão não podem ser estimados. Se uma variável tem três níveis, apenas duas variáveis dicotômicas são necessárias. Assim, o número de variáveis dicotômicas é uma a menos do que o número de níveis da variável não-métrica.

**VARIÁVEL DICOTÔMICA:** Variável em que só existem duas respostas possíveis, como por exemplo, sim/não, doente/não doente.

**VARIÁVEL DISCRETA:** Variável quantitativa cujos possíveis valores formam um conjunto finito ou infinito enumerável de números e que geralmente resultam de uma contagem, como por exemplo, o número de filhos em uma família, o número de grãos por espiga de milho em um experimento, dentre outras.

**VARIÁVEL DISCRETA:** Uma variável descontínua ou discreta pode receber apenas valores específicos. Tamanho da família é uma variável discreta. Uma família pode ser composta de 1, 2, 3 ou mais membros, mas não é possível possuir membros entre esses valores. Ver variável contínua.

**VARIÁVEL ECOLÓGICA:** A que descreve o que ocorre em grupos de indivíduos; por exemplo, porcentagem de obesos ou de fumantes.

**VARIÁVEL EFEITO:** É a variável dependente. Por exemplo: um agravamento à saúde ou à qualidade de vida.

**VARIÁVEL ESTATÍSTICA:** Combinação linear que representa a soma ponderada de duas ou mais variáveis independentes que formam a função discriminante. Também chamada de combinação linear ou composta linear.

**VARIÁVEL ESTATÍSTICA:** Combinação linear de variáveis. Em uma análise de variância multivariada (MANOVA), as variáveis dependentes são formadas em variáveis estatísticas na(s) função(ões) discriminante(s).

**VARIÁVEIS ESTATÍSTICAS CANÔNICAS:** Combinações lineares que representam a soma ponderada de duas ou mais variáveis e podem ser definidas para variáveis dependentes ou independentes. Também chamadas de composições lineares, compostos lineares e combinações lineares.

**VARIÁVEL ESTATÍSTICA CONJUNTA:** Combinação de variáveis conhecidas como fatores especificadas pelo pesquisador que constituem o valor ou a utilidade total dos estímulos. O pesquisador também especifica todos os possíveis valores para cada fator, sendo esses valores conhecidos como níveis.

**VARIÁVEL ESTATÍSTICA DE AGRUPAMENTO:** Conjunto de variáveis ou características que representam os objetos a serem agrupados e usado para calcular a similaridade entre objetos.

**VARIÁVEL ESTATÍSTICA DE REGRESSÃO:** Combinação linear de variáveis independentes ponderadas usadas coletivamente para prever a variável dependente.

**VARIÁVEL EXÓGENA:** Variável que afeta um estudo de forma indesejável, sem o nosso interesse e somente serve para distorcer e desorientar o pesquisador. Variáveis fora do controle direto do pesquisador, que afetam um estudo de maneira não intencional.

**VARIÁVEL EXPERIMENTAL, EFEITO DA:** O efeito da variável experimental ou variável independente é determinado pela revisão da variável dependente nos grupos experimental e controle.

Grupos	Antes	Depois
Grupo experimental	$E_a$	$E_d$
Grupo controle	$C_a$	$C_d$

Efeitos da variável experimental (V. E.) =  $(E_d - C_d) - (E_a - C_a)$ .

**VARIÁVEL EXTERNA (OU EXTRÍNSECA):** É toda e qualquer variável, outra que não seja a exposição principal e a variável-efeito, em tela. É uma variável independente e confundidora, em potencial, da associação exposição-doença, em investigação: por exemplo, na associação entre consumo de álcool e pressão arterial, são diversas as variáveis extrínsecas, entre as quais a idade, o sexo, a obesidade e o fumo. Em qualquer investigação, a variável externa é candidata a ter seus efeitos neutralizados. A neutralização se faz no planejamento ou na análise dos dados. Na fase de análise dos dados, pode-se diferenciar, entre as variáveis externas, as que não são confundidoras das que o são, estas últimas, as variáveis de confusão, devem ser neutralizadas por estratificação ou análise multivariada.

**VARIÁVEL INDEPENDENTE:** O mesmo que variável antecedente, explicativa ou preditora; por exemplo, um tratamento, um fator de risco ou uma característica pessoal, que podem influenciar o efeito, onde este é chamado de variável dependente.

**VARIÁVEL INDEPENDENTE:** Uma variável que explica mudanças em outra variável.

**VARIÁVEL INDEPENDENTE:** Variável (eis) selecionada(s) como previsora e potencial variável de explicação da variável dependente.

**VARIÁVEL INDEPENDENTE:** Um símbolo ou conceito sobre o qual o pesquisador tem algum controle ou pode manipular até certo ponto e que, hipoteticamente, pode causar ou influenciar a variável dependente ou a variável que se acredita afetar o valor da variável dependente.

**VARIÁVEL INDEPENDENTE:** É uma variável que é manipulada, medida, ou selecionada pelo pesquisador, a fim de observar seu relacionamento à resposta do sujeito ou alguma outra variável observada, isto é, variável dependente. Portanto, uma variável independente é uma variável que se usa para influenciar alguma outra variável; é uma condição antecedente ao comportamento observado. Condições ambientais manipuláveis que são conhecidas com base em evidências anteriores ou tidas hipotetizadas como associadas casualmente com alguns eventos observáveis.

**VARIÁVEL INDEPENDENTE:** Representada no eixo dos x. Em estudos experimentais, é a variável manipulável cujos valores são colhidos e determinados pelo pesquisador. Constitui a variável causal, ou seja, a causa presumida da variável dependente.

**VARIÁVEL INDEPENDENTE:** Uma variável na regressão que afeta a variável dependente, mas que se supõe não seja afetada pelas outras variáveis na regressão.

**VARIÁVEL INDEPENDENTE (X):** Variável de natureza relacional, sendo aquela que influencia, determina ou afeta a denominação da variável dependente (Y).

**VARIÁVEL INDEPENDENTE (IMPORTÂNCIA DE DESCRIÇÃO):** Para se estimar o quanto os resultados experimentais podem ser generalizados a outras situações, faz-se necessária uma completa e específica descrição da variável independente.

**VARIÁVEL ALEATÓRIA NORMAL PADRÃO (STANDARD NORMAL RANDOM VARIABLE):** Uma função de distribuição de probabilidades que produz uma variável aleatória normal com média=0 e variância=1. Se uma variável aleatória contínua X é uma normal padrão, ela é escrita como  $X \sim N(0,1)$ ,  $X \cap N(0,1)$  ou ainda  $X:N(0,1)$ . Em geral indicada por Z. Calculada através da equação  $Z = \frac{(X - \mu)}{\sigma}$ . Conhecida também por score reduzido, padrão ou padronizado.

**VARIÁVEL ALEATÓRIA UNIFORME (UNIFORM RANDOM VARIABLE):** Uma variável aleatória em que qualquer resultado é igualmente provável.

**VARIÁVEL CATEGÓRICA (CATEGORIC VARIABLE):** Características, tipos, atributos que são classificadas em dois ou mais grupos distintos. Comparar com variável contínua.

**VARIÁVEL CONTÍNUA (CONTINUOUS VARIABLE):** Características que são medidas usando números inteiros ou números reais fracionários. Comparar com variável categórica.

**VARIÁVEL DEPENDENTE (DEPENDENT VARIABLE):** Em uma declaração de causa e efeito, a variável resposta, ou o objeto afetado a partir do qual se deseja determinar a causa. Comparar com variável independente e variável preditora.

**VARIÁVEL INDEPENDENTE (INDEPENDENT VARIABLE):** Em uma declaração de causa e efeito, é a variável preditora, ou o objeto que é postulado como o causador do efeito observado. Comparar com variável dependente e variável resposta.

**VARIÁVEL PREDITORA (PREDICTOR VARIABLE):** O fator hipotético de causa. Ver também variável independente e comparar com variável resposta.

**VARIÁVEL RESPOSTA (RESPONSE VARIABLE):** O efeito hipotético de um fator causal. Ver também variável dependente e comparar com variável preditora.

**VARIÁVEL INDICADORA:** É a transformação não linear da variável original, por meio da função indicadora.

$$i(X; Z) = \begin{cases} 1, & \text{se } Z(X) \leq z \\ 0, & \text{se } Z(X) > z \end{cases}, \text{ Onde: } i(x; z) \text{ é a função indicadora de } z; Z(x) \text{ é o teor no ponto } x, x \text{ e } D;$$

z é o teor de corte.

**VARIÁVEL IRREVERSÍVEL:** Variável que expressa ligação de relação, onde se X ocorre, então Y ocorrerá, mas se Y ocorre, então nenhuma ocorrência se dará.

**VARIÁVEL INTERVALAR:** Variável em escala de intervalo, com as características das escalas nominais e ordinais. Além disso, apresentam distâncias iguais entre os intervalos que se estabelecem sobre o atributo medido.

**VARIÁVEL INTERVENIENTE:** É uma variável hipotética que não é observada diretamente na pesquisa, mas inferida dos relacionamentos entre a variável independente e a variável dependente. Sua importância está em sua habilidade de explicar o relacionamento entre as variáveis independente e dependente sob as condições do estudo e sob outras condições diferentes. Uma variável cuja existência é inferida, mas que não pode ser manipulada ou medida.

**VARIÁVEL INTERVENIENTE (OU INTERMEDIÁRIA):** É uma variável independente situada na seqüência casual, entre uma outra variável independente e a dependente.

**VARIÁVEL INTERVENIENTE (W):** Variável de natureza relacional, sendo aquela que é colocada entre as variáveis X e Y tendo como finalidade anular, ampliar ou diminuir a influência de X sobre Y.

**VARIÁVEL MANIFESTA:** Valor observado de um item ou questão específica, obtido de respondentes em questões como em um questionário ou a partir de observações feitas pelo pesquisador. Variáveis manifestas são usadas como indicadores de construtos ou variáveis latentes.

**VARIÁVEL MÉTRICA:** Variável com uma unidade constante de medida. Se uma variável tem intervalo de 1 a 9, a diferença entre 1 e 2 é a mesma que aquela entre 8 e 9.

**VARIÁVEL MUDA:** Variável independente ou explicativa, utilizada na análise de regressão, que toma o valor 1 quando determinada condição é satisfeita, e o valor 0 em caso contrário. Também conhecida como variável Dummy.

**VARIÁVEL NECESSÁRIA:** Variável que expressa relação de causa e efeito, onde se ocorrer X, somente X, então ocorrerá Y.

**VARIÁVEL NOMINAL:** Variável de escala nominal, sendo o tipo de variável mais simples e com menor informação. Os elementos são agrupados em classes, obedecendo determinado critério de classificação, sem indicar hierarquia.

**VARIÁVEL OBJETIVA:** Uma variável que o estatístico procura maximizar ou minimizar na teoria da decisão.

**VARIÁVEL ORDINAL:** Variável de escala ordinal. A variável resulta da operação de ordenar por posto. Assim, além de classificar os elementos de um conjunto, se estabelece uma ordem hierárquica, conforme determinado critério.

**VARIÁVEL ORGANÍSMICA:** Uma variável que é uma característica natural ou existente dos indivíduos em estudo. Por exemplo, sexo do indivíduo.

**VARIÁVEL PREDITORA ( $X_n$ ):** Ver variável independente.

**VARIÁVEL PROBABILÍSTICA:** Variável que expressa relação de causa e efeito, onde se X ocorre, provavelmente, a % das vezes ocorrerá Y.

**VARIÁVEL PRODUTO-PRODUTO:** Variável que expressa ligação de relação produto-produto, onde se  $X_1$  ocorre Y pode ou não se produzir, dependendo de  $X_2$  ocorrer. A probabilidade de que Y ocorrerá quando se dá  $X_1$ , em  $O_1$ , que é o ambiente de  $X_1$ , dependendo da probabilidade de  $X_2$  vir a ocorrer em  $O_2$ . Deste modo  $X_1$  e  $X_2$  que são os produtores de Y são condições probabilísticas de Y.

**VARIÁVEL QUALITATIVA:** Variável de natureza analítica definida em termos de seus atributos, portanto não são mensuráveis, nem numéricas ou não contáveis, apenas são descritas, sendo a condição qualitativa definida por escala. conhecida também como variável categórica.

**VARIÁVEL QUALITATIVA:** Variável que inclui diferenças radicais. Exemplo de variáveis de interesse epidemiológico: sexo, idade, local de residência, local de trabalho, ocupação. Também conhecida como variável categórica.

**VARIÁVEL QUANTITATIVA:** Variável de natureza analítica, aplicada aos fenômenos, objetos, dentre outras, que podem ser medida, contados, pesados, dentre outras, exigindo um sistema lógico que permita atribuir números para contagem ou mensuração formando uma escala. É obtida por mensuração quando as unidades podem variar por suas propriedades formando escalas.

**VARIÁVEL QUANTITATIVA:** Variável que envolve diferenças não-substanciais, diferenças apenas de grau. Referem-se a propriedades que mantém a mesma natureza em toda a sua extensão e que se mostram ora com maior, ora com menor intensidade, podendo ser expressas em termos numéricos. Exemplos: peso e estatura.

**VARIÁVEL QUANTITATIVA CONTÍNUA:** Variável que admite valores fracionários entre quaisquer valores consecutivos. Exemplo: temperatura corporal.

**VARIÁVEL QUANTITATIVA DESCONTÍNUA:** Variável que não admite a inclusão de valores fracionários entre dois valores consecutivos expressos por números inteiros. Exemplo: número de batimentos cardíacos. Conhecida também como variável discreta.

**VARIÁVEL RAZÃO:** Variável em escala de razão, esse tipo reúne todas as propriedades dos números naturais: classificação, ordem, distância e origem.

**VARIÁVEL REGIONALIZADA:** É qualquer função numérica com uma distribuição espacial, que varia de um lugar a outro com continuidade aparente, mas cujas variações não podem ser representadas por uma função determinística, para enfatizar as feições particulares dessas variáveis. Em geologia, todas as observações quantitativas feitas em duas ou três dimensões, ou seja, área ou volume, respectivamente, considerando-se obviamente o campo geométrico, sejam elas geoquímicas, geofísicas, sedimentológicas, estruturais, dentre outros exemplos, podem ser consideradas como exemplos de variáveis regionalizadas. A maioria das variáveis regionalizadas apresenta um aspecto aleatório consistindo de variações altamente irregulares e imprevisíveis, e um aspecto estruturado refletindo as características estruturais do fenômeno regionalizado. Uma formulação apropriada para solução de problemas de estimativa, deve levar em consideração essas duas características aparentemente contraditórias, por meio de uma representação simples da variabilidade espacial.

**VARIÁVEL REVERSÍVEL:** Variável que expressa relação de causa e efeito, onde se X ocorre, então Y ocorrerá, e se Y ocorre, então X ocorrerá.

**VARIÁVEL SEQUENCIAL:** Variável que expressa ligação de relação, onde se X ocorre, então mais tarde ocorrerá Y.

**VARIÁVEL SUBSTITUÍVEL:** Variável que expressa ligação de relação, onde se X ocorre, então Y ocorre, mas se Z ocorre, então também Y ocorre.

**VARIÁVEL SUFICIÊNCIA:** Variável que expressa relação de causa e efeito, onde X causa Y. A ocorrência de X é suficiente para a subsequente ocorrência de Y.

**VARIÁVEL DEPENDENTE:** Variável que é modelada como função de variáveis independentes em análise de regressão. Em inglês *Dependent variable*.

**VARIÁVEL DEPENDENTE:** É a variável decorrente de uma ou mais variáveis, estas denominadas independentes ou preditivas.

**VARIÁVEL DEPENDENTE:** A variável que se supõe depender de outra variável ou ser causada por outra chamada variável independente ou explicativa. Se for descoberto que a renda é em parte função da quantidade de educação formal, a renda está sendo tratada como variável dependente. Também conhecida como variável resposta.

**VARIÁVEL DEPENDENTE:** Variável  $y$  em uma equação de regressão ou de regressão múltipla.

**VARIÁVEL DEPENDENTE:** Efeito presumido, ou resposta, a uma mudança na(s) variável (eis) independente(s).

**VARIÁVEL DEPENDENTE (VD):** Mede o fenômeno que se estuda e que se quer explicar. São aquelas cujos efeitos são esperados de acordo com as causas. Elas se situam, habitualmente, no fim do processo causal e são sempre definidas na hipótese ou na questão de pesquisa. No nosso exemplo: desempenho em estatística e atitudes em relação à estatística.

**VARIÁVEL DICOTÔMICA:** O mesmo que variável binária. Por exemplo a variável aleatória discreta de Bernoulli.

**VARIÁVEL DICOTÔMICA:** Variável métrica especial usada para representar uma única categoria ou atributo de uma variável não métrica. Para dar conta de  $L$  níveis de uma variável não métrica  $L-1$  são necessárias variáveis dicotômicas. Por exemplo, sexo é medido como masculino ou feminino e poderia ser representado por duas variáveis dicotômicas ( $X_1$  e  $X_2$ ). Quando o respondente é do sexo masculino,  $X_1=1$  e  $X_2=0$ . Do mesmo modo quando o respondente é do sexo feminino,  $X_1=0$  e  $X_2=1$ . No entanto, quando  $X_1=1$ , sabemos que  $X_2$  deve ser igual a 0. Logo, precisamos de apenas uma variável,  $X_1$  ou  $X_2$  para representar a variável sexo. Se uma variável não-métrica tem três níveis, apenas duas variáveis dicotômicas são necessárias. Sempre temos uma variável dicotômica a menos do que o número de níveis para a variável não métrica.

**VARIÁVEL DICOTÔMICA:** Variável métrica binária usada para representar uma única categoria de uma variável não métrica.

**VARIÁVEL DICOTÔMICA:** Variável não-métrica transformada em uma variável métrica designando-se 1 ou 0 a um objeto, dependendo se este possui ou não uma característica particular.

**VARIÁVEL DISCRETA:** Variável que só assume determinados valores em um intervalo enumerável ou que possui como escala de definição o conjunto dos números naturais. Veja variável contínua. É aquela em que a menor diferença não nula entre dois possíveis valores é um número finito.

**VARIÁVEL DISCRETA OU DESCONTÍNUA:** Seus valores são expressos geralmente por meio de números inteiros não negativos. Resulta normalmente de contagens ou enumeração.



**VARIÁVEL DUMMY (VARIÁVEL LATENTE):** Uma quantidade escrita em uma expressão matemática em forma de uma variável, embora represente uma constante. O termo é usado também de forma mais flexível para denotar uma variável artificial expressando características qualitativas; por exemplo, a presença ou ausência de uma característica pode ser indicada pelos valores 0 ou 1 associados aos indivíduos. Neste sentido a palavra latente/dummy deve ser evitada. Em inglês *Dummy variable*.

**VARIÁVEL ESTATÍSTICA:** Combinação linear de variáveis formada na técnica multivariada determinando-se pesos empíricos aplicados a um conjunto de variáveis especificado pelo pesquisador.

**VARIÁVEL ESTATÍSTICA CONJUNTA:** Combinação de variáveis conhecidas como fatores especificadas pelo pesquisador que constituem o valor ou a utilidade total dos estímulos. O pesquisador também especifica todos os possíveis valores para cada fator, sendo esses valores conhecidos como níveis.

**VEROSSIMILHANÇA (LIKELIHOOD):** É uma distribuição empírica que permite quantificar a preferência entre hipóteses. Ela é proporcional à probabilidade de um conjunto específico de dados ou observações, dada uma determinada hipótese estatística ou um conjunto de parâmetros:  $L(\text{hipótese} | \text{dados observados}) = cP(\text{dados observados} | \text{hipótese})$ , mas diferente de uma distribuição de probabilidades, a verossimilhança não é restringida ao intervalo entre 0,0 e 1,0. Ela é um dos dois termos no numerador do Teorema de Bayes. Conceito criado por Sir Ronald Aylmer Fisher (FISHER, 1922). FRS foi um estatístico, biólogo evolutivo e geneticista inglês (1890-1962). Ver também máxima verossimilhança.

**VETOR COLUNA (COLUMN VECTOR):** Uma matriz com muitas linhas, mas apenas uma coluna. Comparar com vetor linha.

**VETOR LINHA (ROW VECTOR):** Uma matriz com muitas colunas, mas apenas uma linha. Comparar com vetor coluna.

**VETOR MÉDIO (MEAN VECTOR):** O vetor com valores mais prováveis das variáveis aleatórias multivariadas, ou dados multivariados.

**VARIÁVEL F:** Nome dado ao quociente da divisão de duas variáveis aleatórias independentes que apresentam a distribuição de qui quadrado ( $\chi^2$ ).

**VARIÁVEL INDEPENDENTE:** Causa presumida de qualquer mudança na variável dependente.

**VARIÁVEL INDEPENDENTE:** Variável cujos valores não são problemáticos numa análise, mas considerados como dados. Supõe-se que uma variável independente cause ou determine o valor de uma variável dependente. Se descobrirmos que religiosidade é em parte função do sexo, mulheres são mais religiosas que homens, sexo é a variável independente e religiosidade pode se tornar uma variável independente na explicação de crime.

**VARIÁVEL INDEPENDENTE:** Este termo é comumente usado em análise de regressão. Quando uma variável  $y$  é expressa como função de variáveis  $x_1, \dots, x_k$ , mais um termo estocástico, as variáveis  $x$ 's são conhecidas como variáveis independentes. Em inglês *Independent variable*.

**VARIÁVEL INDEPENDENTE:** A variável  $x$  em uma equação de regressão ou uma das variáveis  $x$  em uma equação de regressão múltipla.

**VARIÁVEL INDEPENDENTE (VI):** São aquelas variáveis candidatas a explicar a(s) variável(eis) dependente(s), cujos efeitos queremos medir. Aqui devemos ter cuidado, pois mesmo encontrando relação entre as variáveis isto, não necessariamente, significa relação causal.

**VARIÁVEL LATENTE:** Ver variável Dummy.

**VARIÁVEL OBSERVÁVEL:** Uma variável matemática ou estocástica, cujos valores podem ser diretamente observados, a diferença de variáveis não observáveis que são incluídas em equações estruturais, mas, que não são diretamente observadas. Em inglês *Observable variable*.

**VARIÁVEL OCULTA:** Variável que afeta as variáveis consideradas, sem estar, entretanto, incluída no estudo.

**VARIÁVEL ORDINAL:** Variável cujos atributos diferem entre si em alguma medida de grandeza, como por exemplo variar de alto a baixo ou de fraco a forte.

**VARIÁVEL QUALITATIVA:** Quando seus valores são expressos por atributos.

**VARIÁVEL QUALITATIVA (OU CATEGÓRICA) NOMINAL:** São aquelas cujas respostas podem ser encaixadas em categorias, sendo que cada categoria é independente, sem nenhuma relação com as outras: sexo masculino, feminino, raça branco, preto, outro, dentre outras.

**VARIÁVEL QUANTITATIVA:** Quando os dados são de caráter nitidamente quantitativo, e o conjunto dos resultados possui uma estrutura numérica, trata-se, portanto da estatística de variável.

**VARIÁVEL QUANTITATIVA DISCRETA:** São aquelas resultantes de contagens, constituem um conjunto finito de valores: número de filhos, número de reprovações em matemática, idade em anos completos, dentre outras.

**VARIÁVEL QUANTITATIVA CONTÍNUA:** Resultados de mensurações, podem tomar infinitos valores: pontuação na escala de atitude, nota na prova de matemática, pontuação no vestibular, dentre outros.

**VARIÁVEL SUBSTITUTA:** Seleção de uma única variável com a maior carga fatorial para representar um fator no estágio de redução de dados, em vez de usar uma escala múltipla ou um escore fatorial.

**VARIÁVEL SUPRESSORA:** No modelo de elaboração, uma variável de controle que interage com as variáveis independente e dependente, camuflando a relação entre elas.

**VARIÁVEIS:** Variáveis são características que são medidas, controladas ou manipuladas em uma pesquisa. Diferem em muitos aspectos, principalmente no papel que a elas é dado em uma pesquisa e na forma como podem ser medidas.

**VARIÁVEIS:** Agrupamentos lógicos de atributos. A variável sexo se compõem dos atributos masculinos e feminino.

**VARIÁVEIS:** Condições suficientes e necessárias para que um dado evento aconteça. Características que assumem diferentes valores para diferentes indivíduos do estudo. Condições essenciais cuja determinação garante a ocorrência de um dado evento.

**VARIÁVEIS ALEATÓRIAS DISTRIBUÍDAS CONJUNTAMENTE:** Variáveis aleatórias que possuem uma distribuição conjunta. Em inglês *Jointly distributed random variables*.



**VARIÁVEIS ALEATÓRIAS INDEPENDENTES (EXPLICATIVAS):** Várias variáveis aleatórias são independentes se a distribuição conjunta é igual ao produto das distribuições marginais. Em inglês Independent random variables.

**VARIÁVEIS ALEATÓRIAS INDEPENDENTES:** Duas variáveis aleatórias que não se afetam mutuamente; o conhecimento do valor de uma delas não dá qualquer informação sobre o valor da outra.

**VARIÁVEIS BINÁRIAS:** O mesmo que variáveis dicotômicas.

**VARIÁVEIS CATEGÓRICAS:** As variáveis categóricas representam dados qualitativos. Essa representação pode ser nominal ou ordinal e são do tipo alfanuméricas. Estas variáveis só podem assumir determinados valores num intervalo de valores sendo cada um representativo de uma categoria. Exemplos: As categorias do sexo de um indivíduo vão ser representadas por dois valores numéricos inteiros, por exemplo, o masculino com o valor 1 e o feminino com o valor 2, não sendo possível existir por exemplo o valor 1,54. Um outro exemplo de uma variável categórica pode ser o número de filhos de um casal. Neste caso o valor 1 representa a categoria casal com um filho.

**VARIÁVEIS CONFUNDIDAS:** Variáveis operantes numa situação específica de tal forma que seus efeitos não podem ser separados.

**VARIÁVEIS CONTÍNUAS:** São as variáveis que podem tomar qualquer valor de um determinado intervalo.

**VARIÁVEIS CONTÍNUAS:** Variável que expressa continuidade, as quais podem assumir um conjunto ordenado de valores dentro de determinados limites. Refletem pelo menos uma ordem hierárquica.

**VARIÁVEIS CONTÍNUAS:** As variáveis contínuas representam dados quantitativos, são do tipo numérico e podem assumir qualquer valor de um dado intervalo de valores. É possível qualquer variável contínua ser categorizada. Um bom exemplo disso é a idade que é uma variável contínua que geralmente é categorizada em grupos etários que é variável categórica. Exemplos: Um indivíduo pesar 74,2458 Kg e medir 1,70158 m.

**VARIÁVEL QUALITATIVA (OU CATEGÓRICA) ORDINAL:** São aquelas cujas categorias mantêm uma relação de ordem com as outras, que podem ser regulares ou não pois existe uma ordem natural nas categorias: classe social alta, média, baixa, auto-percepção de desempenho em matemática péssimo, ruim, regular, bom, ótimo, dentre outras. A rigor, no tratamento estatístico das variáveis categóricas, não existe diferença se ela for nominal ou ordinal, a única observação é que quando você está lidando com uma variável ordinal, é aconselhável manter a ordem natural das categorias, de menor para maior, na hora da apresentação, seja em tabela ou em gráficos.

**VARIÁVEL QUALITATIVA (OU CATEGÓRICA) ORDINAL:** São aquelas cujas categorias mantêm uma relação de ordem com as outras, que podem ser regulares ou não pois existe uma ordem natural nas categorias: classe social alta, média, baixa, auto-percepção de desempenho em matemática péssimo, ruim, regular, bom, ótimo, dentre outras. A rigor, no tratamento estatístico das variáveis categóricas, não existe diferença se ela for nominal ou ordinal, a única observação é que quando você está lidando com uma variável ordinal, é aconselhável manter a ordem natural das categorias, de menor para maior, na hora da apresentação, seja em tabela ou em gráficos.

**VARIÂNCIA POPULACIONAL ( $\sigma^2$ ):** É um valor que mede o grau de dispersão dos valores da variável, na população, em relação à média populacional. Definida como a soma dos quadrados dos desvios dos valores da variável em relação a sua média, dividido pelo número de observações:  $\frac{\text{Soma}[(X_i - \text{média})^2]}{N}$ .

**VARIÁVEIS CONTÍNUAS:** Variáveis em que os dados são medidos por meio de uma escala numérica ininterrupta. São exemplos a altura, o peso e a idade.

**VARIÁVEIS DE RAZÃO:** Variáveis derivadas de uma escala contínua que contém um ponto zero verdadeiro.

**VARIÁVEIS DICOTÔMICAS:** Variáveis com apenas dois níveis.

**VARIÁVEIS DIMENSIONAIS:** O mesmo que variáveis contínuas.

**VARIÁVEIS DISCRETAS:** São as variáveis que podem tomar um número finito ou uma infinidade numerável de valores.

**VARIÁVEIS DISCRETAS:** Variável que expressa continuidade, as quais são constituídas de partes ou categorias separadas e distintas.

**VARIÁVEIS DISCRETAS:** Variáveis dicotômicas e variáveis nominais são às vezes chamadas de variáveis discretas porque as diferentes categorias estão completamente separadas uma das outras.

**VARIÁVEIS EXPLANATÓRIAS:** São variáveis experimentais. Objetos do desenho de pesquisa. Busca-se relacionamentos entre elas: variáveis efeitos ou dependentes e variáveis causas, causais ou independentes.

**VARIÁVEIS EXÓGENAS (TIPOS):** Existem ou melhor, se definem na literatura sete tipos de variáveis exógenas envolvidas na pesquisa científica que devem ser controladas para que um experimento tenha validade interna: i) história; ii) maturação; iii) teste; iv) degenerescência do instrumento; v) regressão estatística; vi) seleção tendenciosa; vii) mortalidade.

**VARIÁVEIS FICTÍCIAS:** Quando trabalhamos com modelos de regressão, sejam eles simples ou múltiplos, podemos lidar com situações em que uma ou mais variáveis nominais (categóricas), devam ser incorporadas no modelo, como, por exemplo, sexo, filiação partidária, grupo social dentre outras. Para que alguma variável categórica possa ser introduzida no modelo, é necessário que sejam criadas uma ou mais variáveis assumindo valores numéricos, e que possam representar as categorias da variável nominal considerada. Essas variáveis que deverão ser criadas são chamadas de variáveis fictícias.

**VARIÁVEIS QUANTITATIVAS:** As variáveis quantitativas possuem uma subclassificação, elas podem ser discretas ou contínuas. O primeiro caso ocorre quando os possíveis valores da variável podem ser enumerados. Esta situação é típica de dados oriundos de contagens, como por exemplo o número diário de assaltos em um quarteirão que pode assumir valores no conjunto  $\{0, 1, 2, 3, \dots\}$ . A segunda subclassificação ocorre nos casos em que a variável pode assumir valores em um intervalo contínuo, por consequência os possíveis valores são infinitos e não-enumeráveis. A variável idade, por exemplo, é uma variável contínua pois se for medida com bastante precisão pode gerar como resultado 32,1023 anos de idade e, quanto maior for a precisão, dificilmente dois indivíduos apresentarão idades iguais. Apresentamos a seguir alguns exemplos de variáveis quantitativas.

**VARIÁVEL QUALITATIVA:** Toda variável que não é quantitativa, será classificada como qualitativa. A variável qualitativa é não numérica, portanto é expressa em classes, níveis ou categorias. Caso estes níveis

sejam ordenáveis, a variável é dita ser ordinal, caso contrário ela é classificada como nominal. É importante ressaltar que esta ordenação nos níveis ou categorias da variável é natural tal como ocorre com a variável classe social. Nesta situação, classe A > classe B > classe C > classe D. Como já foi comentado, o tipo de variável determina o tipo de análise e, para variáveis qualitativas ordinais, um resumo numérico, uma técnica gráfica ou até uma tabela de frequência deve incorporar a ideia de ordenação. Conhecida também como categórica.

**VARIÁVEIS INDEPENDENTES:** São variáveis consideradas preditivas de uma outra variável denominada dependente.

**VARIÁVEIS INTERVALARES:** Permitem não apenas ordenar em postos os itens que estão sendo medidos, mas também quantificar e comparar o tamanho das diferenças entre eles.

**VARIÁVEIS INTERVENIENTES:** Fatores intermediários envolvidos na casualidade indireta.

**VARIÁVEIS NOMINAIS:** Permitem apenas classificação qualitativa. Ou seja, elas podem ser medidas apenas em termos de quais itens pertencem a diferentes categorias, mas não se pode quantificar nem mesmo ordenar tais categorias.

**VARIÁVEIS NOMINAIS:** Variáveis nomeadas ou categóricas que não têm uma escala de medidas.

**VARIÁVEIS ORDENADAS OU RANQUEADAS:** O mesmo que variáveis ordinais.

**VARIÁVEIS ORDINAIS:** Permitem ordenar os itens medidos em termos de qual tem menos e qual tem mais da qualidade representada pela variável, mas ainda não permitem que se diga o quanto mais.

**VARIÁVEIS ORDINAIS:** Dados médicos que podem ser caracterizados em termos de mais do que dois valores e em que há a implicação clara de uma direção, do melhor para o pior, porém não são mensurados por uma escala de medidas contínua.

**VARIÁVEIS PERTUBADORAS:** São variáveis exógenas não controladas que podem ser confundidas com as variáveis explanatórias.

**VARIÁVEIS PREDITORAS:** Variáveis independentes em uma equação de regressão.

**VARIÁVEIS QUALITATIVAS:** O mesmo que atributos qualitativos.

**VARIÁVEIS QUANTITATIVAS:** O mesmo que atributos quantitativos.

**VARIÁVEIS RANDOMIZADAS:** São variáveis exógenas não-controladas, mas que podem ser tratadas como erros aleatórios.

**VARIÁVEIS NOMINAIS:** São variáveis categóricas em que não existe uma ordem entre as suas categorias. Exemplo: Quando falamos da variável categórica sexo e das suas categorias masculino e feminino não estabelecemos uma ordem entre elas. Tanto podemos definir o valor 1 para o sexo masculino e o 2 para o feminino como o inverso.

**VARIÁVEIS OCULTAS:** Uma variável é oculta se ela tem um efeito relevante sobre as relações entre variáveis em estudo, como por exemplo nos estudos de regressão e correlação mas não está incluída nelas.

**VARIÁVEIS ORDINAIS:** São variáveis categóricas em que existe uma ordem ou ranking, isto é, uma escala qualitativa entre as várias categorias. Exemplo: As notas de um aluno classificadas com as categorias medíocre, suficiente, bom e muito bom. Entre estas categorias existe um escala implícita em que sabemos, por exemplo, que o bom significa que obteve uma nota mais elevada que o suficiente.

**VARIEDADE:** Conjunto ou série de pontos para os quais existe uma bijeção sobre  $n$ -uplas  $(x_1, x_2, \dots, x_n)$  de números reais que são as coordenadas de cada ponto. Por exemplo, uma reta é uma variedade unidimensional enquanto que a superfície da esfera é uma variedade bidimensional. Uma variedade  $n$ -dimensional é um espaço topológico que é localmente Euclidiano. Isto é, cada ponto está numa região que se assemelha ao espaço  $R_n$ . Em inglês manifold.

**VARIMAX:** É um dos métodos de rotação fatorial ortogonal mais popular, na análise estatística multivariada.

**VARIOGRAMA:** É a representação gráfica da função variograma. É a ferramenta básica da geoestatística que permite descrever quantitativamente a variação no espaço de um fenômeno regionalizado. Neste gráfico cartesiano analisa-se a variabilidade espacial em função da distância numa determinada direção e desta em relação a outras para inferir uma possível anisotropia no depósito. Ver propriedades do variograma.

**VARIOGRAMA:** Para uma variável  $Y$ , observável no local  $x$  do espaço, o variograma é definido como:  $\gamma(h) = \frac{E[Y(x+h) - Y(x)]^2}{2}$ , sendo  $h$  uma distância entre dois locais quaisquer. Em inglês Variogram.

**VARIOGRAMA ABSOLUTO:** É o variograma obtido a partir do cálculo da função variograma absoluto. Na realidade, utiliza-se comumente da função semi-variograma, mas denomina-se simplesmente variograma.

**VARIOGRAMA CRUZADO:** É o gráfico cartesiano da função variograma cruzado.

**VARIOGRAMA EXPERIMENTAL:** É o variograma calculado a partir das observações disponíveis, passível de modelagem ou não, dependendo do comportamento espacial da variável regionalizada.

**VEIL:** Método de estimação de componentes de variância por verossimilhança integrada.

**VERDADE (COMO CHEGAR À):** Os meios como o pesquisador pode chegar a verdade científica são os seguintes: i) experiência: própria ou de outros é um pré-requisito para se chegar à verdade; ii) raciocínio: instrumento indispensável na derivação da verdade, mas não a gera ou a identifica; iii) pesquisa: abordagem disciplinada ao estudo dos fenômenos naturais. A sua aplicação e desenvolvimento dependem de vários instrumentos, inclusive de verdade empírica.

**VERIFICAÇÃO COMPUTADORIZADA DE DADOS:** Checagem computadorizada final de erros nos dados.

**VALOR(ES) CRÍTICO(S):** O valor de uma estatística correspondente a um nível de significância estabelecido, determinado a partir da distribuição amostral. Por exemplo, se  $P(Z \geq z_{0,05}) = P(Z \geq 1,96) = 0,05$ , então  $Z_{0,05} = 1,96$  é o valor crítico de  $Z$  no nível de significância igual a 0,05.

**VALOR DE REFERÊNCIA:** Um parâmetro colocado em um algoritmo tabular de CUSUM, determinado por meio de magnitude da mudança do processo, que deve ser detectado.

**VALOR ESPERADO:** O valor esperado de uma variável aleatória  $X$  é sua média a longo prazo ou um valor médio. No caso contínuo, o valor esperado de  $X$  é

$$E(x) = \int_{-\infty}^{+\infty} xf(x) dx, \text{ em que } f(x) \text{ é a função densidade da variável aleatória } X.$$

**VALOR P:** O nível exato de significância de um teste estatístico; isto é, a probabilidade de obter um valor da estatística de teste que seja no mínimo tão extremo quanto aquele observado quando a hipótese nula é verdadeira. Conhecido também como nível descritivo, p valor ou p value.

**VARIÂNCIA:** Uma medida de variabilidade definida como o valor esperado do quadrado da variável aleatória em torno da média.

**VARIÂNCIA CONDICIONAL:** A variância da distribuição de probabilidades condicionais de uma variável aleatória.

**VARIÂNCIA DA AMOSTRA:** Uma medida de variabilidade de dados amostrais, definida como

$$s^2 = \left[ \frac{1}{(n-1)} \right] \sum_{i=1}^n (x_i - \bar{x})^2, \text{ em que } \bar{x} \text{ é a média da amostra.}$$

**VARIÂNCIA DA POPULAÇÃO:** Veja variância.

**VARIÂNCIA DO ERRO:** A variância de um termo ou componente do erro em um modelo.

**VARIÁVEIS ALEATÓRIAS INDEPENDENTES:** Variáveis aleatórias para as quais  $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$  para qualquer conjunto A e B na faixa de X e Y, respectivamente. Há várias descrições equivalentes de variáveis aleatórias independentes.

**VARIÁVEL ALEATÓRIA:** Uma função que atribui um número real para cada resultado no espaço amostral de um experimento aleatório.

**VARIÁVEL ALEATÓRIA BINOMIAL:** Uma variável aleatória discreta que é igual ao número de sucessos em um número fixo de tentativas de Bernoulli.

**VARIÁVEL ALEATÓRIA BINOMIAL NEGATIVA:** Uma variável aleatória discreta que é o número de tentativas até que um número especificado de sucessos ocorra em tentativas de Bernoulli.

**VARIÁVEL ALEATÓRIA CONTÍNUA:** Uma variável aleatória com um intervalo tanto finito como infinito de números reais para sua faixa.

**VARIÁVEL ALEATÓRIA CONTÍNUA UNIFORME:** Uma variável aleatória contínua com faixa de um intervalo finito e uma função densidade de probabilidade.

**VARIÁVEL ALEATÓRIA DE ERLANG:** Uma variável aleatória contínua que é a soma de um número fixo de variáveis aleatórias independentes exponenciais.

**VARIÁVEL ALEATÓRIA DE POISSON:** Uma variável aleatória discreta que é o número de eventos que ocorrem em um processo de Poisson.

**VARIÁVEL ALEATÓRIA DE WEIBULL:** Uma variável aleatória contínua que é frequentemente utilizada para modelar o tempo até a falha de um sistema físico. Os parâmetros da distribuição são flexíveis o suficiente para que a função densidade de probabilidade possa assumir diferentes formas.

**VARIÁVEL ALEATÓRIA DISCRETA:** Uma variável aleatória com uma faixa finita ou infinita contável.

**VARIÁVEL ALEATÓRIA EXPONENCIAL:** Uma variável aleatória contínua que representa o tempo entre eventos em um processo de Poisson.

**VARIÁVEL ALEATÓRIA GAMA:** Uma variável aleatória que generaliza uma variável aleatória de Erlang para valores não-inteiros do parâmetro r.

**VARIÁVEL ALEATÓRIA GEOMÉTRICA:** Uma variável aleatória discreta que é o número de tentativas de Bernoulli até que um sucesso ocorra.

**VARIÁVEL ALEATÓRIA HIPERGEOMÉTRICA:** Uma variável aleatória discreta que é o número de sucessos obtidos a partir de uma amostra retirada sem reposição de populações finitas.

**VARIÁVEL ALEATÓRIA LOGNORMAL:** Uma variável aleatória contínua com distribuição de probabilidades igual àquela de  $\exp(W)$  para uma variável aleatória normal W.

**VARIÁVEL ALEATÓRIA NORMAL:** Uma variável aleatória contínua que é a mais importante em estatística porque ela resulta do teorema do limite central. Veja teorema do limite central.

**VARIÁVEL ALEATÓRIA NORMAL PADRÃO:** Uma variável aleatória normal ou gaussiana com média zero e variância um, que tem sua função de distribuição cumulativa tabelada.

**VARIÁVEL ALEATÓRIA QUI-QUADRADO:** Uma variável aleatória contínua que resulta da soma dos quadrados das variáveis aleatórias normais padrões independentes. Ela é um caso especial de uma variável aleatória gama.

**VARIÁVEL ALEATÓRIA UNIFORME:** Refere-se tanto a uma variável aleatória uniforme discreta como contínua.

**VARIÁVEL ALEATÓRIA UNIFORME DISCRETA:** Uma variável aleatória com uma faixa finita e função de probabilidade constante.

**VARIÁVEL CASUAL:** Quando  $y = f(x)$  e y é considerado causado por x, x é algumas vezes chamado de variável causal.

**VARIÁVEL DEPENDENTE:** A variável de resposta em uma regressão ou em um planejamento de experimentos.

**VARIÁVEL INDEPENDENTE:** As variáveis preditivas ou regressoras em um modelo de regressão.

**VARIÁVEL(IS) PREDITORA(S):** As variáveis independentes ou regressoras em um modelo de regressão.

**VARIÁVEL(IS) INDICADORA(S):** Variáveis a que são atribuídos valores numéricos para identificar os níveis de uma resposta qualitativa ou categórica. Por exemplo, uma resposta com dois níveis categóricos como sim e não poderia ser representada com uma variável indicadora assumindo valores 0 e 1.

**VARIÁVEL REGRESSORA:** A variável independente ou preditora em um modelo de regressão.

**VERIFICAÇÃO DO MODELO:** A verificação do modelo consiste em avaliar o grau de ajuste deste aos dados disponíveis. Existem diversas ferramentas com esta finalidade. Em inglês Model checking.



**VEROSSIMILHANÇA:** Segundo Bussab e Morettin (2002), o novo dicionário Aurélio da Língua Portuguesa (2ª Edição, 1986) define verossímil ou verossimilhante aquilo que é semelhante à verdade, provável, e verossimilhança ou verossimilitude, ou ainda verossimilitude, à qualidade ou caráter de verossímil. Segundo ainda os autores uma amostra verossímil é aquela que fornece a melhor informação possível sobre um parâmetro de interesse da população, desconhecido, e que pretendemos estimar. o princípio da verossimilhança afirma que devemos escolher aquele valor do parâmetro desconhecido que maximiza a probabilidade de obter a amostra particular observada, ou seja, o valor que torna aquela amostra a mais provável. O uso desse princípio conduz a um método de estimação pelo qual se obtêm os denominados estimadores de máxima verossimilhança que, em geral, têm propriedades muito boas. Esse princípio foi enunciado por Sir Ronald Aylmer Fisher pela primeira vez em 1912 (FISHER, 1912) e, em 1922 (FISHER, 1922), deu-lhe forma mais completa, introduzindo a expressão Likelihood (verossimilhança).

**VEROSSIMILHANÇA PERFILHADA OU CONCENTRADA:** É a abordagem de verossimilhança para eliminar parâmetros de nuisance e refere-se a substituir tais parâmetros por suas estimativas de máxima verossimilhança para cada valor fixo do parâmetro de interesse. Isso porque na definição de verossimilhança se contempla modelos multiparamétricos, no entanto, muitas vezes o interesse reside apenas em um subconjunto de parâmetros, sendo os demais denominados parâmetros de perturbação (nuisance) e participam do modelo apenas para ajudar a descrever melhor a variabilidade, e um caso típico é quando o interesse reside nos componentes de variância e os efeitos fixos são considerados nuisance, e neste caso, é necessário um método para concentrar a verossimilhança em um só parâmetro ou grupo de parâmetros por meio da eliminação do parâmetro de nuisance.

**VEROSSIMILHANÇA CONDICIONAL:** Supondo que o vetor de dados  $X$  pode ser particionado como  $(T,U)$ , com função de densidade dada por  $l(\theta, \phi; t, u) = l(\theta, \phi; t)l(\theta, \phi; u|t)$ . O primeiro termo do lado direito é chamado de verossimilhança marginal e o segundo é a verossimilhança condicional. Em inglês Conditional likelihood.

**VEROSSIMILHANÇA CONDICIONAL:** Algumas vezes é possível encontrar estatísticas suficientes para parâmetros de nuisance e condicionar a função de verossimilhanças nesses valores das estatísticas de tal forma que ela não irá depender deles. Suponha  $U$  uma estatística tal que a fatorização é dada por  $p_Y(y; \psi, \lambda) = p_U(u; \psi, \lambda)p_{Y|U=u}(y; \mu, \psi)$ . Negligenciando  $p_U(\cdot)$  temos a função de verossimilhança condicional baseada em condicionamento sobre  $U=u$ ,  $L_c(\psi) = L_c(\psi; y|u) = p_{Y|U=u}(y; \mu, \psi)$ . Quando se trabalha com distribuições que pertence a família exponencial as fatorizações

$p_{U,V}(u, v; \psi, \lambda) = p_V(v; \psi)p_{U|V=v}(u, v; \psi, \lambda)$  e  $p_Y(y; \psi, \lambda) = p_U(u; \psi, \lambda)p_{Y|U=u}(y; \mu, \psi)$  quase sempre acontecem. Pode-se observar que em ambos métodos as inferências não são baseadas na função de verossimilhança original, mas sim, numa pseudo-verossimilhança que é no entanto, uma função de verossimilhança genuína. É importante ressaltar que o uso dessas verossimilhanças é raramente obtido e quando existem, os cálculos destas funções de verossimilhanças é frequentemente difícil.

**VEROSSIMILHANÇA GENUÍNA:** Termo que é caracterizada por satisfazer: que o valor esperado da função score é zero, o estimador de máxima verossimilhança (EMV) segue uma distribuição assintoticamente normal, o teste da razão de verossimilhança (RV) segue uma distribuição assintoticamente de qui quadrado  $(\chi^2)$ , e.t.c.

**VEROSSIMILHANÇA MARGINAL:** Seja  $t$  uma estatística usada para estimar um parâmetro  $\theta$ , a função de verossimilhança  $L(x_1, \dots, x_n; \theta) = L(t; \theta)$  que depende dos dados por meio da estatística  $t$  é chamada de verossimilhança marginal. Em inglês Marginal likelihood.

**VEROSSIMILHANÇA MARGINAL:** É a função de verossimilhança para os quais alguns parâmetros foram marginalizados, ou seja, houve integração da função completa para um conjunto de parâmetros de menor interesse. Muitas vezes é difícil de ser obtida. Suponha que  $(U, V)$  seja uma estatística suficiente para  $\theta = (\psi, \lambda)$  o qual a fatorização é dada por,  $p_{U,V}(u, v; \psi, \lambda) = p_V(v; \psi)p_{U|V=v}(u, v; \psi, \lambda)$ . Negligenciando  $p_{U|V}(\cdot)$ , tem-se a função de verossimilhança marginal baseada em  $V=v$ :  $L_M(\psi) = L_M(\psi, v) = p_V(v; \psi)$ .

**VEROSSIMILHANÇA PERFILADA:** É o tipo de verossimilhança quando se consegue escrever alguns parâmetros com funções de outros, reduzindo a dimensionalidade total do espaço. A função de probabilidade conjunta dos dados, expressa pelas densidades simplificadas por essas funções é verossimilhança perfilada. A idéia neste caso é que em geral pode-se obter uma função de pseudo-verossimilhança num modelo paramétrico substituindo o parâmetro de perturbação  $\lambda$  por seu estimador de máxima verossimilhança  $\hat{\lambda}_\psi$  para  $\psi$  fixado. Função de verossimilhança perfilada para  $\psi$  é denotada por  $L_P(\psi) = L(\psi; \hat{\lambda}_\psi)$

A função score é  $l_* = (l_\psi, l_\lambda)$ , onde,  $l_\psi = \frac{\partial l(\psi, \lambda)}{\partial \psi}$  vetor de dimensão  $k$  e  $l_\lambda = \frac{\partial l(\psi, \lambda)}{\partial \lambda}$  vetor de dimensão  $p-k$ . Sob condições de regularidade,  $\hat{\lambda}_\psi$  é a solução em  $\lambda$  de  $l_\lambda(\psi, \lambda) = 0$ .  $L_P(\psi)$  não é uma verossimilhança genuína, isto é,  $E\left(\frac{\partial l_P(\psi)}{\partial \psi}\right) \neq 0$   $E\left(\frac{\partial l_P(\psi)}{\partial \psi} \frac{\partial l_P(\psi)}{\partial \psi^T}\right) \neq -E\left(\frac{\partial^2}{\partial \psi \partial \psi^T} l_P(\psi)\right)$ . No entanto, tem algumas

propriedades interessantes que fazem parecer uma verossimilhança verdadeira: i) O EMV perfilada de  $\psi$  é o mesmo da EMV baseado em  $L(\psi, \lambda)$ ; ii) A estatística da razão de log-verossimilhança perfilada é o mesmo à estatística da razão de verossimilhança verdadeira

$$RV_P = RV_P(\psi) = 2\{l_P(\hat{\psi}) - l_P(\psi)\} = 2\{l(\hat{\psi}, \hat{\lambda}) - l(\psi, \hat{\lambda}_\psi)\}.$$

Sobre condições de regularidade, a distribuição assintótica da  $RV_P$  é  $\chi_k^2$ : i) Regiões de confiança baseada no teste da razão de verossimilhança, perfilada, é equivalente àquelas baseadas no teste da razão de verossimilhança usual; ii) Informações perfiladas observadas

$$j_P(\psi) = -\frac{\partial^2}{\partial \psi \partial \psi^T} l_P(\psi).$$

Inversa da matriz de informação perfilada observada  $(j_P(\psi))^{-1} = j^{\psi\psi}(\psi, \hat{\lambda}_\psi)$ . Usar  $L_P(\psi)$  assemelha-se a tratar o parâmetro de  $\lambda$  como se fosse conhecido  $\hat{\lambda}_\psi$ , isso não é adequado quando a dimensão de  $\lambda$  é grande. Um fraco desempenho da função de verossimilhança perfilada em alguns casos pode ser parcialmente explicado pelo fato desta função não estabelecer uma tentativa de aproximação para funções de verossimilhança condicional e marginal genuínas.

**VEROSSIMILHANÇA PERFILADA MODIFICADA:** Tipo de verossimilhança descrita como mostrado a seguir. Inicialmente é importante enfatizar que a função de verossimilhança perfilada modificada proposta por Cox e Reid (1987), é o conceito fundamental para o início da definição desse tipo de verossimilhança. Senão vejamos. A verossimilhança perfilada modificada proposta por Cox e Reid (1987) pode



ser obtido como uma aproximação de uma verossimilhança condicional. Se a função de log-verossimilhança tem parâmetros separáveis, isto é,  $l(\theta) = l(\psi, \lambda) = l_1(\psi) + l_2(\lambda)$  e, portanto,  $\hat{\lambda}_\psi = \hat{\lambda}$ ,  $l_p(\psi) = l_1(\psi)$ ,  $j_p(\psi) = -\frac{\partial^2 l_1(\psi)}{\partial \psi \partial \psi^T}$ . Uma forma de tentar aproximar-se a esta situação é por meio de uma parametrização

ortogonal. Os componentes  $\psi$  e  $\lambda$  de  $\theta$  são chamados parâmetros ortogonais se os componentes  $l_\psi$  e  $l_\lambda$  do vetor escore são não correlacionados, ou seja, se  $I_{\psi\lambda} = 0$ . Se  $\psi$  e  $\lambda$  são parâmetros ortogonais então,

$l(\psi, \lambda) = l_1(\psi) + l_2(\lambda) + O_p\left(n^{-1/2}\right)$ . A obtenção de uma nova parametrização  $(\psi, \phi)$  ortogonal só é

garantido se  $\psi$  for unidimensional. Para inferências sobre  $\psi$  o modelo condicional dado  $\hat{\lambda}$  pode fornecer uma pseudo-verossimilhança mais exato do que a verossimilhança perfilada. A log-verossimilhança perfilada modificada de Cox e Reid (1987), é obtida por  $l_M(\psi) = l_p(\psi) - \frac{1}{2} \log |j_\lambda(\psi, \hat{\lambda}_\psi)|$ . A equação de

estimação para verossimilhança perfilada modificada é obtida como segue. Seja Y um vetor aleatório cujos componentes são independentes e Y tem função densidade:  $f_Y(y; \psi, \lambda)$  onde  $\psi$ : parâmetro interesse

(escalar)  $\lambda$ : parâmetro de perturbação (dimensão p). Função de log-verossimilhança e a informação observada  $l(\psi, \lambda) = \log f_Y(y; \psi, \lambda)$  e  $nj_{\lambda\lambda}(\psi, \lambda) = -\frac{\partial^2 l(\psi, \lambda)}{\partial \lambda \partial \lambda^T}$ . Função de log-verossimilhança perfilada

para  $\psi$   $l_p(\psi) = l(\psi, \hat{\lambda}_\psi)$ . Função de verossimilhança perfilada modificada para  $\psi$ , ou seja,

$l_M(\psi) = l(\psi, \hat{\lambda}_\psi) - \frac{1}{2} \log |nj_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|$ . Cox e Reid (1987) sugerem que quando os parâmetros  $\lambda$  e  $\psi$  são

ortogonais, isto é, em relação a informação esperada de Fisher definida por:  $I_{\psi\lambda} = E\left(-\frac{\partial^2 l}{\partial \psi \partial \lambda}\right)/n$ , obtendo

$I_{\psi\lambda} = 0$ . A princípio é possível expressar o parâmetro de perturbação  $\phi$  como  $\phi = \phi(\psi, \lambda)$  com  $\lambda$  ortogonal  $\psi$ . Considerando propriedades da equação de estimação  $l'_M(\psi) = U_M(\psi) = 0$ . A equação é desenvolvida usando uma expansão assintótico estocásticos para  $U_M(\psi)$ . Expansão da estatística Escore. Considere-se Y identicamente distribuídos com  $l_i = \log f(y_i; \psi, \lambda)$  define-se

$$Z_{\psi\lambda} = \frac{1}{\sqrt{n}} \frac{\partial^2 l(\psi, \lambda)}{\partial \psi \partial \lambda} = \frac{1}{\sqrt{n}} \sum_i \frac{\partial^2 l_i(\psi, \lambda)}{\partial \psi \partial \lambda} \text{ e } Z_{\psi\psi\lambda} = \frac{1}{\sqrt{n}} \frac{\partial^3 l(\psi, \lambda)}{\partial \psi^2 \partial \lambda} - \sqrt{n} I_{\psi\psi\lambda}.$$

um escalar. As variáveis aleatórias Z são  $O_p(1)$  e tem média zero. Os I's são de  $O(1)$  representam os valores esperados das derivadas da log-verossimilhança, por exemplo,

$$I_{\psi\psi} = -\frac{1}{n} E\left(\frac{\partial^2}{\partial \psi^2} l(\psi, \lambda)\right) \text{ e } I_{\psi\lambda} = \frac{1}{n} E\left(\frac{\partial^3}{\partial \psi \partial \lambda^2} l(\psi, \lambda)\right), I_{\lambda\lambda} = \frac{1}{n} E\left(\frac{\partial l}{\partial \lambda}\right)^2 \text{ e } I_{\lambda\psi\lambda} = \frac{1}{n} E\left(\frac{\partial l(\psi, \lambda)}{\partial \lambda}\right)\left(\frac{\partial^2 l(\psi, \lambda)}{\partial \psi \partial \lambda}\right).$$

Derivada da função de verossimilhança perfilada modificada é expressa por:

$$l'_M(\psi) = l'_p(\psi) - \frac{1}{2} \frac{\partial \left\{ \log |j_\lambda(\psi, \hat{\lambda}_\psi)| \right\}}{\partial \psi}$$

Expandindo  $l(\psi, \hat{\lambda}_\psi)$  em séries de Taylor em relação a  $\lambda$  e diferenciando o resultado em relação a  $\psi$ , obtemos

$l'_p(\psi) = \sqrt{n}a_{11} + a_{12} + \frac{1}{\sqrt{n}}a_{13} + \frac{1}{n}a_{14} + O_p(n^{-3/2})$ , onde os  $a_{ij}$  são polinômios de ordem j em Z, com

coeficientes que são funções de I. Consideremos.  $a_{11} = Z_\psi$ ,  $a_{12} = \frac{Z_{\psi\lambda}Z_\lambda}{I_{\lambda\lambda}} + \frac{1}{2} \frac{Z_\lambda^2 I_{\psi\lambda\lambda}}{I_{\lambda\lambda}^2}$ . Combinando as

expressões acima com os resultados das expansões, obtemos.  $El'_p(\psi) = -\frac{I_{\psi\lambda\lambda}}{2I_{\lambda\lambda}} + \frac{1}{n}A_{14} + O(n^{-2})$ , onde

$A_{14} = E(\sqrt{n}a_{13} + a_{14})$ . Note que, de  $El'_p(\psi) = -\frac{I_{\psi\lambda\lambda}}{2I_{\lambda\lambda}} + \frac{1}{n}A_{14} + O(n^{-2})$  a escore perfilada é  $O_p(\sqrt{n})$  e

um vício de ordem  $O(1)$ . Diferenciando  $\log |nj_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|$  em relação  $\psi$  e expandindo em série de Taylor

tem-se que,  $\frac{\partial}{\partial \psi} \log |nj_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)| = a_{20} + \frac{1}{\sqrt{n}}a_{21} + \frac{1}{n}a_{22} + O_p(n^{-3/2})$ . Combinando estes resultados obtemos,

$$El'_M(\psi) = \frac{1}{n} \left\{ \frac{1}{8I_{\lambda\lambda}^2} (3I_{\psi\lambda\lambda\lambda} + 8I_{\lambda,\psi\lambda\lambda} + 8I_{\psi\lambda,\lambda\lambda} + 12I_{\lambda\lambda,\psi\lambda} + 8I_{\lambda,\psi\lambda,\lambda} + 4I_{\lambda,\lambda,\psi\lambda}) \right.$$

$$+ \frac{1}{12I_{\lambda\lambda}^3} [I_{\lambda\lambda\lambda} + (4I_{\psi\lambda\lambda\lambda} + 15I_{\psi\lambda,\lambda\lambda} - 9I_{\lambda,\lambda,\psi\lambda}) - 24I_{\lambda,\lambda\lambda} (I_{\lambda,\psi\lambda\lambda} + 4I_{\lambda,\lambda,\psi\lambda})$$

$$+ I_{\psi\lambda\lambda} (-I_{\lambda\lambda\lambda} + 2I_{\lambda,\lambda\lambda} - 12I_{\lambda,\lambda,\lambda} - 4I_{\lambda,\lambda,\lambda,\lambda})] + \frac{I_{\psi\lambda\lambda}}{8I_{\lambda\lambda}^4} (7I_{\lambda\lambda\lambda}^2 + 36I_{\lambda\lambda\lambda} I_{\lambda,\lambda\lambda}$$

$$+ 24I_{\lambda,\lambda\lambda}^2) \left. \right\} + O(n^{-2}).$$
 Derivadas similares como as obtidas acima, mostram que

$$-El''_p(\psi) - Var l'_p(\psi) = \frac{1}{2I_{\lambda\lambda}} (I_{\psi\psi\lambda\lambda} - 2I_{\psi\lambda,\psi\lambda} - 2I_{\psi,\lambda,\psi\lambda}) - \frac{I_{\psi\psi\lambda}}{2I_{\lambda\lambda}^2} (I_{\lambda,\lambda\lambda} + I_{\lambda,\lambda,\lambda})$$

$$+ \frac{I_{\psi\lambda\lambda}}{2I_{\lambda\lambda}^2} (2I_{\psi,\lambda} + I_{\psi\lambda\lambda}) + O(n^{-1}) \text{ e}$$

$$-El''_M(\psi) - Var l'_M(\psi) = \frac{1}{I_{\lambda\lambda}} (I_{\psi\psi\lambda\lambda} + I_{\lambda,\psi\psi\lambda}) + \frac{I_{\psi\psi\lambda}}{I_{\lambda\lambda}^2} (I_{\lambda\lambda\lambda} + I_{\lambda,\lambda\lambda}) + O(n^{-1}).$$
 Se considerarmos o erro

quadrático médio da função escore modificada  $U_M$ , num sentido condicional, então  $U_M$  é melhor do que  $U_p$ . Outro aspecto das propriedades de segunda ordem de  $U_M$  que pode ser relevante para a teoria de equação de estimação é a quantidade

$$\frac{E\{U_M^2(\psi)\}}{[E\{-U'_M(\psi)\}]^2}$$

Godambe (1976) mostrou que esta quantidade é minimizada pela equação de estimação ótima. A expressão para esta quantidade em termos de esperança da derivada da log-verossimilhança é dada por:

$$\frac{1}{nI_{\psi\psi}} \left[ 1 + \frac{1}{n} \left\{ \frac{I_{\psi\lambda,\psi\lambda}}{I_{\lambda\lambda}} - \frac{1}{2} \left( \frac{I_{\psi\lambda\lambda}^2}{I_{\lambda\lambda}^2} \right) \right\} + O(n^{-2}) \right]$$

Como exemplo considere o caso da distribuição gaussiana inversa. Seja a função de densidade,

$$f(y; \mu, \sigma) = \frac{1}{\sqrt{(2\pi\sigma)}} y^{-3/2} \exp\left\{-\frac{1}{2\sigma\mu^2 y} (y - \mu)\right\}$$

$y, \mu, \sigma > 0$ ,  $\mu$  o parâmetro de interesse e  $\sigma$  o: parâmetro de perturbação, Os parâmetros canônicos são  $\left(-\frac{1}{(2\sigma\mu^2)}, -\frac{1}{(2\sigma)}\right)$ . As estatísticas suficientes são  $\left(y, \frac{1}{y}\right)$ . A função de log-verossimilhança é dada

por:  $l(\mu, \sigma) = -\frac{n}{2} \log \sigma - \frac{1}{2\sigma\mu^2} \sum \frac{(y_i - \mu)}{y_i}$  com  $\hat{\mu} = \bar{y}$  e  $\hat{\sigma}_\mu = \sum \frac{(y_i - \mu)^2}{y_i n \mu^2}$ . A função de log-veros-

similhança perfilada é:  $l_p(\mu) = -\left(\frac{n}{2}\right) \log \hat{\sigma}_\mu$ . E a função de log-verossimilhança perfilada modificada é

dada por:  $l_M(\mu) = -\left(\frac{n-2}{2}\right) \log \hat{\sigma}_\mu$ , mostrando que  $l_p$  e  $l_M$  contém o mesmo ponto de estimação  $\hat{\mu}$ , a

partir da equação de estimação, temos,  $U_p(\mu) = -\frac{n(\bar{y} - \mu)}{\hat{\sigma}_\mu \mu^3} = 0$  e  $U_M(\mu) = \left(\frac{n-2}{n}\right) U_p(\mu) = 0$ . Embora a

média e a variância da equação de estimação podem ser calculadas diretamente, usando as fórmulas

$$-EI_p''(\psi) - VarI_p'(\psi) = \frac{1}{2I_{\lambda\lambda}} (I_{\psi\psi\lambda\lambda} - 2I_{\psi\lambda,\psi\lambda} - 2I_{\psi,\lambda,\psi\lambda}) - \frac{I_{\psi\psi\lambda}}{2I_{\lambda\lambda}^2} (I_{\lambda,\lambda\lambda} + I_{\lambda,\lambda,\lambda}),$$

$$+ \frac{I_{\psi\lambda\lambda}}{2I_{\lambda\lambda}^2} (2I_{\psi,\lambda\lambda} + I_{\psi\lambda\lambda}) + O(n^{-1}) \mathbf{e}$$

$$-EI_M''(\psi) - VarI_M'(\psi) = \frac{1}{I_{\lambda\lambda}} (I_{\psi\psi\lambda\lambda} + I_{\lambda,\psi\psi\lambda}) + \frac{I_{\psi\psi\lambda}}{I_{\lambda\lambda}^2} (I_{\lambda\lambda\lambda} + I_{\lambda,\lambda\lambda}) + O(n^{-1})$$

$$-EU_p' - VarU_p = -\frac{2}{(\sigma\mu^3)} + O(n^{-1}) \mathbf{e} \quad -EU_M' - VarU_M = O(n^{-1})$$

$l(\theta) = l(\psi, \lambda) = l_1(\psi) + l_2(\lambda)$  e, portanto,  $\hat{\lambda}_\psi = \hat{\lambda}$ ,  $l_p(\psi) = l_1(\psi)$ ,  $j_p(\psi) = -\frac{\partial^2 l_1(\psi)}{\partial \psi \partial \psi^T}$ . Uma forma

de tentar aproximar-se a esta situação é por meio de uma parametrização ortogonal.

Os componentes  $\psi$  e  $\lambda$  de  $\theta$  são chamados parâmetros ortogonais se os componentes  $l_\psi$  e  $l_\lambda$  do vetor escore são não correlacionados, ou seja, se  $I_{\psi\lambda} = 0$ . Se  $\psi$  e  $\lambda$  são parâmetros ortogonais

$l(\psi, \lambda) = l_1(\psi) + l_2(\lambda) + O_p\left(n^{-1/2}\right)$ . A obtenção de uma nova parametrização  $(\psi, \phi)$  ortogonal

só é garantido se  $\psi$  for unidimensional. Para inferências sobre  $\psi$  o modelo condicional dado  $\hat{\lambda}$  pode fornecer uma pseudo-verossimilhança mais exato do que a verossimilhança perfilada. A log-verossimilhança perfilada modificada de Cox e Reid (1987) é obtida por

$l_M(\psi) = l_p(\psi) - \frac{1}{2} \log |j_\lambda(\psi, \hat{\lambda}_\psi)|$ . A equação de estimação para verossimilhança perfilada modificada é

obtida como segue. Seja  $Y$  um vetor aleatório cujos componentes são independentes e  $Y$  tem densidade:  $f_Y(y; \psi, \lambda)$ , onde  $\psi$ : parâmetro interesse (escalar) e  $\lambda$ : parâmetro de perturbação (dimensão  $p$ ). Função

de log-verossimilhança e a informação observada  $l(\psi, \lambda) = \log f_Y(y; \psi, \lambda)$  e  $nj_{\lambda\lambda}(\psi, \lambda) = \frac{-\partial^2 l(\psi, \lambda)}{\partial \lambda \partial \lambda^T}$ .

Função de log-verossimilhança perfilada para  $\psi$   $l_p(\psi) = l(\psi, \hat{\lambda}_\psi)$ . Função de verossimilhança perfilada modificada para  $\psi$   $l_M(\psi) = l(\psi, \hat{\lambda}_\psi) - \frac{1}{2} \log |nj_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|$ . Cox e Reid (1987) sugerem que quando os parâ-

metros  $\lambda$  e  $\psi$  são ortogonais, isto é, em relação a informação esperada de Fisher definida por:

$I_{\psi\lambda} = E\left(-\frac{\partial^2 l}{\partial \psi \partial \lambda}\right) / n$ , obtendo  $I_{\psi\lambda} = 0$ . A princípio é possível expressar o parâmetro de perturbação

$\phi$  como  $\phi = \phi(\psi, \lambda)$  com  $\lambda$  ortogonal  $\psi$ . Considerando propriedades da equação de estimação

$l_M'(\psi) = U_M(\psi) = 0$ . A equação é desenvolvida usando uma expansão assintótico estocásticos para

$U_M(\psi)$  que é a expansão da estatística escore. Considere-se  $Y$  identicamente distribuídos com

$l_i = \log f(y_i; \psi, \lambda)$  define-se  $Z_{\psi\lambda} = \frac{1}{\sqrt{n}} \frac{\partial^2 l(\psi, \lambda)}{\partial \psi \partial \lambda} = \frac{1}{\sqrt{n}} \sum_i \frac{\partial^2 l_i(\psi, \lambda)}{\partial \psi \partial \lambda}$  e  $Z_{\psi\psi\lambda} = \frac{1}{\sqrt{n}} \frac{\partial^3 l(\psi, \lambda)}{\partial \psi^2 \partial \lambda} - \sqrt{n} I_{\psi\psi\lambda}$ , neste

caso considere-se  $\lambda$  um escalar. As variáveis aleatórias  $Z$  são  $O_p(1)$  e tem média zero. Os  $I$ 's são de  $O(1)$  representam os valores esperados das derivadas da log-verossimilhança, por exemplo,

$I_{\psi\psi} = -\frac{1}{n} E\left(\frac{\partial^2}{\partial \psi^2} l(\psi, \lambda)\right)$  e  $I_{\psi\lambda} = \frac{1}{n} E\left(\frac{\partial^3}{\partial \psi \partial \lambda^2} l(\psi, \lambda)\right)$ ,  $I_{\lambda\lambda} = \frac{1}{n} E\left(\frac{\partial l}{\partial \lambda}\right)^2$  e  $I_{\lambda\psi} = \frac{1}{n} E\left(\frac{\partial l(\psi, \lambda)}{\partial \lambda}\right) \left(\frac{\partial^2 l(\psi, \lambda)}{\partial \psi \partial \lambda}\right)$ . Derivada

da função de verossimilhança perfilada modificada é expressa por  $l_M(\psi) = l_p(\psi) - \frac{1}{2} \frac{\partial \left\{ \log |j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)| \right\}}{\partial \psi}$ .

Expandindo  $l(\psi, \hat{\lambda}_\psi)$  em séries de Taylor em relação a  $\lambda$  e diferenciando o resultado em relação a  $\psi$ , obtemos  $l_p'(\psi) = \sqrt{n} a_{11} + a_{12} + \frac{1}{\sqrt{n}} a_{13} + \frac{1}{n} a_{14} + O_p(n^{-3/2})$ , onde os  $a_{ij}$  são polinômios de ordem

$j$  em  $Z$ , com coeficientes que são funções de  $I$ . Consideremos  $a_{11} = Z_\psi$ ,  $a_{12} = \frac{Z_{\psi\lambda} Z_\lambda}{I_{\lambda\lambda}} + \frac{1}{2} \frac{Z_\lambda^2 I_{\psi\lambda\lambda}}{I_{\lambda\lambda}^2}$ .

Combinando as expressões acima com os resultados das expansões, obtém-se

$$El_p'(\psi) = -\frac{I_{\psi\lambda\lambda}}{2I_{\lambda\lambda}} + \frac{1}{n} A_{14} + O(n^{-2}) \text{ onde } A_{14} = E(\sqrt{n} a_{13} + a_{14}). \text{ Note que, de}$$

$$El_p'(\psi) = -\frac{I_{\psi\lambda\lambda}}{2I_{\lambda\lambda}} + \frac{1}{n} A_{14} + O(n^{-2}) \text{ a escore perfilada é } O_p(\sqrt{n}) \text{ e um vicio de ordem } O(1).$$

Diferenciando  $\log |nj_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|$  em relação  $\psi$  e expandindo em série de Taylor tem-se

$$\frac{\partial}{\partial \psi} \log |nj_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)| = a_{20} + \frac{1}{\sqrt{n}} a_{21} + \frac{1}{n} a_{22} + O_p(n^{-3/2}). \text{ Combinando estes resultados obtém-se}$$

$$El_M'(\psi) = \frac{1}{n} \left\{ \frac{1}{8I_{\lambda\lambda}^2} (3I_{\psi\lambda\lambda\lambda} + 8I_{\lambda,\psi\lambda\lambda} + 8I_{\psi\lambda,\lambda\lambda} + 12I_{\lambda\lambda,\psi\lambda\lambda} + 8I_{\lambda,\psi\lambda,\lambda} + 4I_{\lambda,\lambda,\psi\lambda\lambda}) \right.$$

$$+ \frac{1}{12I_{\lambda\lambda}^3} [I_{\lambda\lambda\lambda} + (4I_{\psi\lambda\lambda\lambda} + 15I_{\psi\lambda,\lambda\lambda} - 9I_{\lambda,\lambda,\psi\lambda}) - 24I_{\lambda,\lambda\lambda} (I_{\lambda,\psi\lambda\lambda} + 4I_{\lambda,\lambda,\psi\lambda})$$

$$+ I_{\psi\lambda\lambda} (-I_{\lambda\lambda\lambda} + 2I_{\lambda,\lambda\lambda\lambda} - 12I_{\lambda,\lambda,\lambda\lambda} - 4I_{\lambda,\lambda,\lambda,\lambda})] + \frac{I_{\psi\lambda\lambda}}{8I_{\lambda\lambda}^4} (7I_{\lambda\lambda\lambda}^2 + 36I_{\lambda\lambda\lambda} I_{\lambda,\lambda\lambda}$$

$$+ 24I_{\lambda,\lambda\lambda}^2) \left. \right\} + O(n^{-2}). \text{ Derivadas similares como as obtidas acima, mostram que}$$

$$-El'_p(\psi) - VarI'_p(\psi) = \frac{1}{2I_{\lambda\lambda}}(I_{\psi\psi\lambda\lambda} - 2I_{\psi\lambda,\psi\lambda} - 2I_{\psi,\lambda,\psi\lambda}) - \frac{I_{\psi\psi\lambda}}{2I_{\lambda\lambda}^2}(I_{\lambda,\lambda\lambda} + I_{\lambda,\lambda,\lambda})$$

$$+ \frac{I_{\psi\lambda\lambda}}{2I_{\lambda\lambda}^2}(2I_{\psi,\lambda\lambda} + I_{\psi\lambda\lambda}) + O(n^{-1}) \quad e \quad -El'_M(\psi) - VarI'_M(\psi) = \frac{1}{I_{\lambda\lambda}}(I_{\psi\psi\lambda\lambda} + I_{\lambda,\psi\psi\lambda}) + \frac{I_{\psi\psi\lambda}}{I_{\lambda\lambda}^2}(I_{\lambda\lambda\lambda} + I_{\lambda,\lambda\lambda}) + O(n^{-1}).$$

**VERTEX:** Instrumento de medição de distância e altura das árvores que funciona por meio de tecnologia digital e ultrassom. Para o seu uso faz-se necessário que a uma unidade emissora de pulsos ultrassônicos (transponder) fique estacionada no tronco da árvore à altura de 1,30 m. A unidade receptora dos pulsos (vertex) calcula a distância ou altura por meio de velocidade de deslocamento do som no meio. Para obter exatidão nas medidas é necessária a calibração do instrumento.

**VETOR:** Segmento de reta com direção e sentido. Uma matriz que apresenta apenas uma coluna. Em inglês vector.

**VETOR:** Representação da direção e magnitude do papel de uma variável como retratada em uma interpretação gráfica de resultados da análise discriminante.

**VETOR:** Método para representar um ponto ideal ou atributo em um mapa perceptual. Envolve o uso de projeções para determinar a ordem de um objeto no vetor.

**VETOR:** Um vetor é uma n-upla de números  $u_i$  dados por  $u = (u_1, u_2, \dots, u_n)$ . Os  $u_i$  são chamados componentes de u. Se todos os  $u_i$  forem iguais a zero isto é  $u_i = 0$ , então u é chamado vetor nulo. Por um múltiplo escalar de u, ku (onde k é um número real), entendemos o vetor que é obtido de u, multiplicando-se seus componentes por k, ou seja:  $ku = (ku_1, ku_2, \dots, ku_n)$ . Vale notar ou lembrar que dois vetores são iguais se e somente se suas componentes correspondentes são iguais.

**VETOR:** Conjunto de números reais (por exemplo,  $X_1 = X_n$ ) que podem ser escritos em colunas ou linhas. Os vetores coluna são considerados convencionais e os vetores linha são tidos como transpostos. Os vetores coluna e vetores linha são mostrados como se segue:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \quad X^T = [X_1 X_2 \dots X_n]$$

**Vetor coluna**

**Vetor linha**

O T no vetor linha indica que é transposta do vetor coluna.

**VETOR DE PROBABILIDADE:** Um vetor  $u = (u_1, u_2, \dots, u_n)$  é chamado vetor de probabilidade, se suas componentes são não negativas e somam 1. Por exemplo considere os seguintes vetores:

$$u = \left(\frac{3}{4}, 0, -\frac{1}{4}, \frac{1}{2}\right), v = \left(\frac{3}{4}, \frac{1}{2}, 0, \frac{1}{4}\right) e w = \left(\frac{1}{4}, \frac{1}{4}, 0, \frac{1}{2}\right)$$

Então o vetor u é um vetor de probabilidade, pois sua terceira componente é negativa, já o vetor v não é um vetor de probabilidade, pois a soma de suas componentes é maior do que 1, e w é um vetor de probabilidade.

**VÍCIO:** Conceito referente a uma estatística ou a um estimador. Um estimador é dito sem vício, não viciado, não tendencioso, não viesado ou imparcial se a média de sua distribuição amostral  $\{E[\hat{\theta}]\}$  coincide com o parâmetro  $\theta$  a ser estimado. Em inglês bias.

**VÍCIO:** Ver erro sistemático.

**VÍCIO:** Para um estimador t do parâmetro  $\theta$ , o vício é definido como a diferença entre o valor esperado de t e o próprio parâmetro  $\theta$ . Em inglês Bias.

**VÍCIO DE AFERIÇÃO:** Origina-se quando as variáveis de um estudo são medidas erroneamente, de modo a distorcer os resultados. Para acontecer um vício de aferição, em estudos analíticos, o processo de medida deve ser sistematicamente diferente nos dois grupos em comparação.

**VÍCIO DE CONFUSÃO:** Acontece quando dois fatores ou processos estão inter-relacionados, e é erroneamente concluído que um dos fatores é o responsável pelos resultados. Existe a distorção nos resultados, se o fator de confusão está presente desigualmente nos dois grupos que estão sendo comparados.

**VÍCIO DE SELEÇÃO:** Distorção dos resultados de uma pesquisa, decorrente do modo dos participantes serem recrutados ou perdidos no curso do estudo. Isto é, o vício de seleção ocorre quando observações são feitas em um grupo de pacientes selecionados incorretamente.

**VIÉS:** Qualidade de um instrumento de medida que tende a resultar em má interpretação do que está sendo medido numa determinada direção. Por exemplo, se o item do questionário de uma pesquisa de opinião pública perguntar o seguinte: você não concorda que o governador está fazendo uma boa administração? seria tendencioso, pois, em geral, incentivaria respostas mais favoráveis.

**VIÉS:** Distorção, tendenciosidade ou erro sistemático, introduzido em qualquer fase de um estudo; erro que não é devido ao acaso.

**VIÉS:** Ver erro sistemático.

**VIÉS DE AFERIÇÃO (DA INFORMAÇÃO OU DA OBSERVAÇÃO):** Erro sistemático de diagnóstico de um evento, quando os resultados podem ser imputados à maneira como as variáveis são conceituadas ou medidas.

**VIÉS DE CONFUNDIMENTO (OU DE CONFUSÃO DE VARIÁVEIS):** Erro sistemático que ocorre quando os resultados de uma associação entre dois fatores podem ser imputados, total ou parcialmente, a um outro fator não levado em consideração; este outro fator é a variável de confusão.

**VIÉS DE DURAÇÃO:** Viés que ocorre quando os casos mais moderados e mais indolentes de uma doença são detectados desproporcionalmente em programas populacionais de rastreamento. Os casos mais agressivos já terão resultado em morte ou em sintomas que exigiram intervenção médica.

**VIÉS DE MENSURAÇÃO:** Viés que resulta da qualificação distorcida das exposições ou dos desfechos por causa do uso de técnicas inadequadas ou da subjetividade da escala de medidas.

**VIÉS DE NEYMAN:** O mesmo que viés do exame tardio.

**VIÉS DE PUBLICAÇÃO:** Tendência dos investigadores, revisores e editores de revistas em propor ou aceitar a publicação de manuscritos em função da direção ou da intensidade dos resultados do estudo.

**VIÉS DE RECORDAÇÃO:** Viés que resulta da diferente capacidade de recordar de uma exposição a fatores causais entre aqueles que têm uma doença, quando comparados com aqueles que não têm.

**VIÉS DE SELEÇÃO:** Viés que ocorre quando a alocação dos indivíduos em um estudo ou em um grupo específico do estudo do estudo está influenciada por características individuais que também têm influência na probabilidade do desfecho.

**VIÉS DE SELEÇÃO (DA POPULAÇÃO OU DA AMOSTRA):** É o erro na escolha da população ou dos grupos para estudo, devido a diferenças sistemáticas entre características daqueles indivíduos incluídos no estudo e daqueles que não o são. Distorção sistemática introduzida pela maneira como os participantes são selecionados para o estudo, por perdas ou por não resposta dos incluídos na amostra original, diferenças nas características dos grupos que são, entre si, comparados.

**VIÉS DO OBSERVADOR:** Interferência do pesquisador sobre os dados obtidos em ensaios, levantamentos ou observação de fenômenos. Este fato pode comprometer as hipóteses em estudo.

**VIÉS OU BIAS:** Diferença entre o valor de uma estatística e a do parâmetro.

**VIÉS OU BIAS OCASIONAIS:** São aqueles decorrentes de erros de cobertura, tais como superposição, inclusão ou exclusão de unidades, da escolha não-aleatória das amostras, de erros de observação, de defeitos instrumentais e outros.

**VIÉS OU TENDENCIODADE:** Introdução de um erro que produz desvios ou distorções que ocorrem predominantemente em uma direção, de modo oposto ao que ocorre no erro aleatório. O mesmo que Erro diferencial.

**VISUALIZAÇÃO DE DADOS:** Técnicas para retratar dados em formato gráfico, tipicamente usadas de um modo exploratório para identificar relações básicas.

**VIZINHOS:** Em estatística espacial, definem-se vizinhos como locais que apresentam fronteiras geográficas comuns. Este conceito é claramente definido para o caso em que as observações são de contagens e feitas em regiões fechadas de mapas, como municípios, condados, estados, dentre outras, precisando ser adaptada para o caso em que a quantidade de interesse é observada de forma contínua ao longo de toda a região de estudo. Em inglês Neighbors.

**VOLUME:** Medida usada tradicionalmente para expressar a produção de madeira tanto de árvores ou toras individuais, quanto de florestas ou povoamentos florestais. O contexto, escopo ou forma de medir a produção de madeira resulta em diferentes tipos de volume, por exemplo, utiliza-se os termos volume com casca e volume sem casca quando o interesse da produção florestal inclui ou não a casca. O volume de árvores individuais resulta de três atributos básicos da árvore: diâmetro à altura do peito (DAP), altura e forma.

**VOLUME CILÍNDRICO:** Volume de uma árvore considerando um cilindro hipotético cuja altura é igual à altura total ou comercial da árvore, e o diâmetro igual ao seu diâmetro à altura do peito (DAP). Volume que exclui o atributo de forma das árvores, representando, geralmente, apenas um cálculo intermediário para se obter o volume sólido. O volume cilíndrico de povoamentos é definido como a soma dos volumes cilíndricos das árvores individuais.

**VOLUME COMERCIAL:** Produção de madeira de árvores e povoamentos florestais que é apropriada a uma ou várias formas de utilizações da madeira. Geralmente, as toras das árvores são consideradas apropriadas para uma dada utilização quando o menor diâmetro da tora é maior que o diâmetro mínimo de utilização. Por exemplo, é comum serrarias não serrarem toras cujo menor diâmetro seja inferior a 20 cm. Por isso, a produção florestal expressa na forma de volume comercial deve informar também o diâmetro mínimo de utilização.

**VOLUME EMPILHADO:** Representação da produção de madeira de árvores ou povoamentos florestais por meio de volume das pilhas de madeira. Inclui não só o volume efetivo das toras, como também o espaço vazio entre elas.

**VOLUME SÓLIDO:** Designa a produção de madeira em termos do volume efetivo das toras, em contraposição ao termo volume empilhado.

**VOLUME TOTAL:** Conceito que se refere a produção de madeira de árvores e povoamentos florestais independentemente das formas de utilizações da madeira, ou seja, sem considerar um diâmetro mínimo de utilização.

**VAR (X):** Variância ou quadrado médio da variável aleatória X. Também conhecida por  $V(X)$ ,  $\sigma^2$ ,  $\sigma_x^2$  ou ainda  $\sigma_{(X)}^2$ .







**WALD-HERRING, FÓRMULA DE:** Ver fórmula de Wald-Herring.

**WALKER, PROVA DE:** Ver prova de Walker.

**WAPPÄUS, FÓRMULA DE:** Ver fórmula de Wappäus.

**WASH-OUT:** Suspensão temporária de medicação para remover os efeitos residuais da droga em uso pelo paciente.

**WASH OUT (OU FASE DE WASH OUT):** Etapa de limpeza; designa o período, em geral, em um ensaio terapêutico, no qual o tratamento é retirado para que seus efeitos desapareçam e não influenciem a etapa subsequente da investigação. A fase de wash out, ou de limpeza, tem o sentido de remover o efeito do tratamento e fazer com que os indivíduos retornem às suas características habituais. É usado em pesquisas nas quais os participantes mudam de um tratamento para outro. Ver estudo cruzado.

**WATERS, FÓRMULA DE:** Ver fórmula de Waters.

**WAUGH, FÓRMULA DE:** Ver fórmula de Waugh.

**WAVELET:** É uma função matemática útil para processamento de sinais e compressão de imagens, embora seu uso com estes propósitos seja recente a teoria não é nova. Os princípios são semelhantes ao da Análise de Fourier. Na WWW (Web ou Internet) wavelets tem sido utilizadas para a compressão de imagens com mais sucesso do que o método familiar JPEG o qual é um acrônimo de Joint Photographic Experts Group, sendo um método de compressão de imagens fotográficas e também considerado como um formato de arquivo. Em inglês Wavelet.

**WEISE, HIPSÔMETRO:** Hipsômetro baseado em semelhança de triângulos. Trata-se de um tubo metálico com 15 a 20 cm de comprimento, com uma mira em uma extremidade e um orifício de visada ou ocular em outra. A escala de altura é marcada sobre uma peça serrilhada fixada ao lado do tubo. Uma segunda escala, conhecida como escala de distância observador árvore, é inserida por meio de escala de alturas na marca zero formando com esta um ângulo reto. Na extremidade superior desta escala esta acoplado um pêndulo que se desloca na escala serrilhada do tubo para a medição das alturas.

**WHIPPLE, CRITÉRIO DE:** Ver critério de Whipple.

**WISHART, DISTRIBUIÇÃO DE:** Ver distribuição de Wishart.

**WISHART, TEOREMA DE:** Ver teorema de Wishart.

**WIRSTEIN, FÓRMULA (DE PEREQUAÇÃO MECÂNICA) DE:** Ver fórmula de perequação mecânica de Wittstein.

**WOOLHOUSE, MÉTODO DE:** É um método usado na ciência atuarial para calcular períodos subanuais.





**X:** Símbolo de uma incógnita. Exemplo: descobrir o x do problema: achar a solução do problema.

**X:** Dez em algarismo romano.

**X:** Variável aleatória contínua com distribuição de probabilidade normal ou Gaussiana, a qual é reduzida no escore padrão Z para obter áreas ou probabilidades sob a curva da distribuição normal padrão.

**X:** Variável aleatória discreta com distribuição de Bernoulli a qual assume apenas dois valores: 0 ou 1, sendo denominada portanto de variável binária.

**X:** Variável aleatória discreta com distribuição binomial, cujo campo de definição é o conjunto dos números 0, 1, 2, 3, ..., n.

**X:** Variável discreta com distribuição de Poisson, cujo campo de definição é o conjunto infinito enumerável dos números 0, 1, 2, 3, 4, 5, 6, ... 20, ...,  $\infty$ .

**X:** Denominação que se dá ao ponto médio de uma classe em uma tabela de distribuição de frequências, o qual é obtido somando os limites inferior e superior dessa classe e dividindo o resultado por 2.

**X:** Letra usada para simbolizar uma variável independente, uma resposta em um experimento, um valor obtido em levantamentos de dados, uma variável estudada num processo observacional, etc.

**X:** Símbolo para representar um conjunto, ou um elemento de um conjunto quando considerada maiúscula ou minúscula respectivamente.

**X:** Símbolo usado na estatística descritiva para representar o valor de uma variável quantitativa discreta ou quantitativa contínua.

**X:** Símbolo usado para representar uma matriz qualquer.

$\underline{X}$  : Símbolo usado para representar um vetor qualquer.

$X \cap F_X$  : X tem distribuição  $F_X$ , ou também  $X \sim F_X$  ou ainda  $X : F_X$ .

$x_i$  : Valor da observação i, centro ou ponto médio da classe i, dependendo do contexto, mas sendo claro em cada situação.

$x_{(1)}$  : Valor da menor das observações, é o valor  $x_{\min}$ . Usado como estatística de ordem ou ordinal.

$x_{(i)}$  : i-ésimo valor observado.

$x_{(n)}$  : Valor da maior das observações é o valor  $x_{\max}$ . Usado como estatística de ordem ou ordinal.

$\bar{x}$  : Valor da média aritmética amostral.



$\bar{x}_g$  : Valor da média geométrica amostral.

$\bar{x}_h$  : Valor da média harmônica amostral.

$\bar{x}_q$  : Valor da média quadrática amostral.

$\dot{x}$  : Valor da mediana amostral.

$\ddot{x}$  : Valor da moda amostral.

$x_i - x_0$  : Desvio, afastamento, erro ou resíduo de  $x_i$  em relação a  $x_0$ .

$|x_i - x_0|$  : Desvio absoluto de  $x_i$  em relação a  $x_0$ .

$[x]$  : Característica de  $x$ .

**XILÔMETRO:** Instrumento utilizado para medição do volume de toras por meio do princípio de Arquimedes. Consiste em um recipiente com água que permite observar o deslocamento do nível da água quando as toras são submergidas.

**X-BARRA:** Simbologia indicada por  $\bar{X}$  e utilizada para representar a média aritmética de uma amostra de dados, de uma série, de uma distribuição, de um conjunto de valores, de um rol, de dados brutos, etc. Na língua inglesa, a palavra Média é representada tanto pelo termo MEAN como por AVERAGE. A origem de average é a palavra árabe AWAR, que significa mercadorias danificadas. Na lei marítima, average (média) significa a divisão da perda entre as partes interessadas (WONNACOTT; WONNACOTT, 1985).



**Y:** Letra do alfabeto latino que é usada para representar uma variável dependente, uma variável resposta na análise de regressão, uma variável conjunta na análise de correlação e também utilizada em resultados oriundos de experimentos ou ensaios, etc.

**YATES:** Estatístico inglês que criou no século vinte os chamados blocos incompletos balanceados (BIB).

**YATES, CORREÇÃO DE:** Ver correção da continuidade para  $\chi^2$ , que é o mesmo.

**YOU DEN, QUADRADO:** Ver quadrado Youden.

**YULE, COEFICIENTE DE ASSOCIAÇÃO DE:** Ver associação, coeficiente de, que é o mesmo.

**Y-BARRA:** Simbologia indicada por  $\bar{Y}$  e utilizada para representar a média aritmética, valor médio, média dos y's em uma amostra de dados quantitativos, de uma série, de uma distribuição, de um conjunto de valores, de um rol, de dados brutos, etc.



# Z

**Z**: Conjunto dos números inteiros relativos.

**Z**: Utilizado para representar o conjunto dos números inteiros. Em inglês Z.

**Z**: Eixo diagonal utilizada nas representações gráficas tridimensionais e muito usado em análise estatística multidimensional. Nos gráficos bidimensionais, tem-se o eixo horizontal ou dos X's, também denominado de eixo das abcissas e o eixo vertical ou dos Y's, também chamado de eixo das ordenadas.

**Z DE FISHER, TRANSFORMAÇÃO**: Ver transformação z de Fisher.

**ZERO DE UMA FUNÇÃO**: Nome especial que recebe o valor x de uma raiz de uma função matemática, tal que  $y = f(x) = 0$ , isto é, ponto onde a curva ou função corta ou intercepta o eixo dos x ou das abcissas. Em inglês function zero.

**ZEUNER, DIAGRAMA DE**: Ver diagrama de Zeuner.

**Z NORMAL PADRÃO**: É a variável ou escore reduzido que transforma a variável aleatória normal X, dada por

$$Z = \frac{X - \mu}{\sigma}$$

a qual serve para calcular a área, densidade ou probabilidade sob a curva da distribuição normal ou Gaussiana, e representa a distância em que a variável X está da média, em número de desvios padrões.

**ZONA**: Relativa a um atributo, é a faixa geográfico tal que os valores do atributo em causa, apresentados pelas unidades amostrais nela contidas formam um estrato.

**ZONA DE ACEITAÇÃO**: Se o valor calculado da estatística de teste está dentro da zona de aceitação, a hipótese nula é aceita. Em contraste, ver região crítica.

**ZONA DE REJEIÇÃO**: Se o valor calculado da estatística de teste está dentro da zona de rejeição, a hipótese nula é rejeitada. Em contraste, ver região de aceitação.

**ZONA SOCIAL**: O mesmo que trato estatístico.

$z_{\alpha}$  : Quantil da distribuição normal estandarizada, padrão ou reduzida. O Modelo normal também é conhecido como distribuição Gaussiana a qual tem forma de sino ou de chapéu de napoleão, campanular sendo simétrica em relação a média aritmética.





# APÊNDICE

## GLOSSÁRIO DE ANOTAÇÕES

- $\infty$  Infinito
- $\pm$  Isso significa que adicionamos e subtraímos o valor à direita do sinal àquele à esquerda deste sinal, para obter dois diferentes valores. Assim,  $x \pm y$  nos fornece  $(x + y)$  e  $(x - y)$ .
- $\geq$  O valor à esquerda do sinal é maior ou igual ao valor à direita do sinal.
- $\leq$  O valor à esquerda do sinal é menor ou igual ao valor à direita do sinal.

$\log x$  Este é o logaritmo ( $\log$ ) do número  $x$ . As vezes, coloca-se  $x$  entre parênteses, quando isso auxilia na compreensão. Consultar Transformações: Para verificar os usos da transformação logarítmica. Os logaritmos podem apresentar diferentes bases, sendo as mais comuns 10 e  $e$  (ver a seguir). Em termos gerais, se  $\log_y (x) = z$ , então  $x = y^z$ , em que a base do logaritmo é o número  $y$ . Contando que esteja-se coerente com o uso de uma base particular de um logaritmo e especifique quais bases está se usando, pode-se empregar um logaritmo de qualquer base. Lembre-se de que o  $\log$  de zero é infinito, e que só pode-se utilizar  $\log$  de números positivos. Pode-se obter os valores de logaritmos a partir de tabelas especiais ou, mais frequentemente, com o auxílio de calculadoras ou computador. Tendo-se dois números,  $a$  e  $b$ , por exemplo, então o  $\log$  do seu produto é igual à soma dos logaritmos, ou seja,  $\log(ab) = \log(a) + \log(b)$ , o  $\log$  de seu quociente é igual à diferença de seus logaritmos, isto é,  $\log\left(\frac{a}{b}\right) = \log(a) - \log(b)$ .

$\log_e x$  este é o logaritmo neperiano ou natural de  $x$  com base  $e$ , normalmente escrito como  $\ln x$ , em que  $e$  é a constante 2,71828. O logaritmo de base  $e$  de uma quantidade  $x$  é o valor  $z$ , de tal modo que  $x = e^z$ . Esses logs são utilizados com mais frequência em estatística e geralmente são empregados em programas de computador.

$\log_{10} x$  Este é o conhecido logaritmo de base 10 de  $x$ . Se  $\log_{10}(x) = z$ , então  $x = 10^z$ . Por exemplo,  $\log_{10}(10) = 1$  e  $\log_{10}(100) = 2$ . Os logs com base 10 são utilizados para simplificar a multiplicação e a divisão, ao invés de contar com a adição e subtração, entretanto, atualmente são pouco utilizados com essa finalidade, em razão do uso de calculadora e computador.

Logotipo ( $p$ ) Transformação logística (logotipo) de uma proporção,  $p$ , cuja equação é obtida conforme o modelo logotipo

$$(p) = \log_e \left\{ \frac{P}{(1-P)} \right\}.$$

$e^x$  É a função exponencial, às vezes expressa como  $\exp(x)$ . Se  $x = \log_e(y)$ , então  $y = e^x$  é o antilogaritmo de  $x$ . Desse modo, quando empregamos transformação logarítmica de uma variável, é possível transformá-la de volta à escala original utilizando o seu antilog, por meio da função exponencial de uma calculadora. Note que, mesmo quando utilizamos  $\log$  com base 10 para transformação, podemos transformá-lo de volta à escala original utilizando a função antilog  $10^x$ .

$|x|$  As linhas verticais à esquerda e à direita de  $x$  indicam que devemos ignorar o sinal de  $x$ , isto é, devemos considerar o seu valor absoluto. Por exemplo,  $|-15,5| = |15,5| = 15,5$ .

$\sum_{i=1}^n x_i$  Normalmente é abreviado como  $\Sigma x$ . A letra grega maiúscula sigma,  $\Sigma$ , indica a soma dos valores

da variável quantitativa,  $x$ , para todos os  $n$  indivíduos ou elementos da amostra; isto é, de  $i = 1$  até  $i = n$ .

A abreviação ( $\Sigma x$ ) que omite os valores de  $i$ , acima e abaixo do sinal de soma, como mencionado para  $x$ , é utilizada quando não resulta em dúvida. Assim,  $\sum_{i=1}^n x_i = x_1 + x_2 + x_3 + \dots + x_n$ .

$\sum_{i=1}^n (x_i - \bar{x})^2$  Normalmente é abreviado como  $\Sigma(x - \bar{x})^2$ . É denominada soma dos resíduos ou desvios

quadrados, soma de quadrados ajustada ou simplesmente soma de quadrados. Subtraímos a média da amostra,  $\bar{x}$ , de cada valor de  $x$  na amostra e, então, elevamos essa diferença ao quadrado; repetindo esse processo para todos os  $n$  valores de  $x$  obtemos um conjunto de  $n$  diferenças quadradas que, então, são somadas.

ANCOVA Análise de covariância.

ANOVA Análise de variância.

RRA Redução de risco absoluto; a diferença entre dois riscos.

IC Intervalo de confiança ou fiducial; contém o parâmetro de interesse com uma probabilidade prescrita.

CONSORT Um protocolo que fornece, em forma de checklist ou fluxograma, uma diretriz aos autores acerca de como devem relatar um teste clínico aleatório.

GL Graus de liberdade de uma estatística; isto é, o número de observações independentes que contribuem para o valor da estatística.

MBE Medicina baseada em evidência.

MVBE Medicina veterinária baseada em evidência.

EEG Equações de estimativa generalizada.

MLG Modelo linear generalizado.

CCI Coeficiente de correlação intraclasse; se as unidades, individualmente, estão contidas em um conglomerado, ele representa as variações entre os conglomerados, como uma proporção da variação total.

PTI Proporção da taxa de incidência; a proporção de duas taxas de incidência, normalmente a taxa de incidência da doença no grupo tratado dividida pela incidência da enfermidade no grupo controle.

RV Razão de verossimilhança; a razão de duas verossimilhanças. Por exemplo, no contexto de um teste diagnóstico, pode ser a razão das verossimilhanças de um resultado positivo de um teste em animais com e sem a doença.

ERV Estatística da razão de verossimilhança que emprega a razão de duas verossimilhanças para comparar dois modelos de regressão.

MANOVA Análise de variância multivariada.

EVM Estimação de verossimilhança máxima; um procedimento que propicia estimativas dos parâmetros em um modelo por meio da maximização de sua verossimilhança.

MME Modelo multinível.

NNT Número de animais necessários para tratar em um teste com um novo tratamento, ao invés do tratamento controle, para evitar uma reação adversa.

VPN Valor preditivo negativo de um teste diagnóstico ou de triagem, por exemplo, a proporção de animais negativos que realmente não apresentam a doença.

PP Proporção de probabilidade; a proporção de duas probabilidades, normalmente as probabilidades de ocorrer a doença quando exposto ao fator divididas pela probabilidade da doença naqueles são expostos ao fator.

VPP Valor preditivo positivo de um teste diagnóstico ou de triagem, como por exemplo, a proporção daqueles com teste positivo que, na verdade, apresentam a doença.

TAC Teste aleatório controlado.

ROC Curva receiver operating.

RR Risco relativo; a proporção de dois riscos, normalmente o risco de doença daqueles expostos ao fator dividido pelo risco daqueles não expostos ao fator.

RRR Redução de risco relativo igual a 1-RR.

DP Desvio-padrão de um conjunto de observações.

EP (b) Erro padrão de uma estatística  $b$ .

EPM Erro padrão da média da amostra. Às vezes é abreviada como EP.

DNP Desvio Normal Padrão obtido por:  $z = \frac{x - \mu}{\sigma}$ , em que a variável,  $x$ , apresenta distribuição Normal, com média =  $\mu$  e DP =  $\sigma$ . O DNP tem uma distribuição Normal, com média=0 e DP=1.

## SÍMBOLOS UTILIZADOS NA ESTATÍSTICA

=: Igual.

≠: Diferente.

<: Menor que.

>: Maior que.

$\leq$ : Menor ou igual a.  
 $\geq$ : Maior ou igual a.  
 $\equiv$ : Idêntico a.  
 $\sim$ : Não.  
 $\approx$ : Aproximadamente igual a.  
 $\cong$ : Congruente a.  
 $\sim$ : Similar a.  
 $:$ : Razão.  
 $||$ : Valor absoluto.  
 $!$ : Fatorial.  
 $\sqrt{\quad}$ : Raiz.  
 $\wedge$ : E.  
 $\therefore$ : Portanto.  
 $\cap$ : Interseção (e).  
 $\cup$ : União (ou).  
 $\subset$ : Está contido.  
 $\supset$ : Contém.  
 $\in$ : Pertence a.  
 $\notin$ : Não pertence a.  
 $\vee$ : Ou.  
 $\exists$ : Existe.  
 $\forall$ : Para todo.  
 $\infty$ : Infinito.  
 $\forall$ : Qualquer que seja.  
 $\Sigma$ : Somatório.  
 $\int$ : Integral.  
 $\partial$ : Derivada.  
 $\lim_{x \rightarrow \infty} f(x)$ : Limite.  
 $\Leftrightarrow$ : Se e somente se.

$\Rightarrow$ : Implica em.  
 $\emptyset$ : Vazio.  
 $\nexists$ : não existe.  
 $\not\subset$ : não está contido.  
 $\not\supset$ : não contém.  
 $/$ : tal que.  
 $\mathbf{N}$ : conjunto dos números naturais.  
 $\mathbf{Q}$ : conjunto dos números racionais.  
 $\mathbf{Q}' = \mathbf{I}$ : conjunto dos números irracionais.  
 $\mathbf{R}$ : conjunto dos números reais.

### ÍNDICE DOS SÍMBOLOS

$\mathbf{a}$ : Ordenada na origem de uma reta de regressão, de uma reta dos mínimos retângulos ou de uma plano de regressão dos valores observados.  
 $\mathbf{a}_k$ : Momentos em relação á origem dos valores observados.  
 $\mathbf{AH}_0$ : Aceitação de uma hipótese nula.  
 $\mathbf{b}, \mathbf{b}_{yx}, \mathbf{b}_{xy}$ : Coeficientes de regressão dos valores observados.  
 $\mathbf{b}_{zx,y}, \mathbf{b}_{zy,x}$ : Coeficientes de regressão parcial dos valores observados.  
 $\mathbf{b}_1, \mathbf{b}_2$ : Coeficientes de regressão parcial, coeficientes de PEARSON dos valores observados.  
 $\mathbf{C}_x^n$ : Número de combinações.  
 $\mathbf{cov}, \mathbf{cov}(\mathbf{x}, \mathbf{y}), \mathbf{cov}(\mathbf{X}, \mathbf{Y})$ : Covariâncias dos valores observados e teóricos.  
 $\mathbf{e}$ : Base de logaritmos neperianos.  
 $\mathbf{e}_m$ : Desvio médio absoluto dos valores observados.  
 $\mathbf{E}(\mathbf{X})$ : Esperança matemática.  
 $\mathbf{exp}$ : Exponencial.  
 $\mathbf{F}$ : Variável de SNEDECOR.  
 $\mathbf{f}(\mathbf{x}), \mathbf{f}_1(\mathbf{x}), \mathbf{f}(\mathbf{x}, \mathbf{y})$ : Funções densidade de probabilidade.  
 $\mathbf{F}(\mathbf{x}), \mathbf{F}_1(\mathbf{x}), \mathbf{F}(\mathbf{x}, \mathbf{y})$ : Funções de distribuição.  
 $\mathbf{g}, \mathbf{g}_i$ : Parâmetros quaisquer dos valores observados.



**G**: Variável aleatória associada ao parâmetro  $g$ .

**$g_1, g_2$** : Coeficientes de FISHER dos valores observados.

**H**: Hipótese alternativa.

**$H_0$** : Hipótese nula.

**$k, k_1, k_2$** : Parâmetros, números de graus de liberdade.

**L,  $L(\gamma)$** : Funções de verossimilhança.

**log**: Logaritmo.

**$m, m_x, m_i$** : Médias aritméticas dos valores teóricos.

**$\tilde{m}, \tilde{m}_x$** : Medianas dos valores teóricos.

**$m_{x|y}$** : Médias condicionadas dos valores teóricos.

**$m_k, m_k(x), m_{kl}$** : Momentos centrados dos valores observados.

**$m_{11}$** : Covariância dos valores observados.

**$m(t), m_x(t)$** : Funções geradoras dos momentos.

**$n, n_i, n_{ij}, n_{ijk}$** : Efetivos, frequências absolutas.

**N**: Efetivo de uma população finita.

**$n'_i, n'_{ij}$** : Frequências relativas.

**$n''_i$** : Frequências unitárias.

**$n_{i.}, n_{.j}, n_{ij}, n_{i.k}, n_{.jk}, n_{i..}, n_{.j.}, n_{..k}$** : Frequências marginais absolutas.

**$n'_i, n'_j$** : Frequências condicionadas relativas.

**$N^*(x)$** : Função cumulativa de frequências.

**p**: Proporção, parâmetro de uma distribuição binomial.

**$P(A), P(X=x), P(x), P_x, P_1(x), P(x, y)$** : Probabilidades.

**$P(A|B), P(x|y)$** : Probabilidades condicionadas.

**q**: Número de classes ou probabilidade de fracasso ou insucesso.

**p**: Probabilidade de sucesso.

**1-p**: Probabilidade de fracasso ou insucesso.

**$q_1, q_3$** : Quartis dos valores observados.

**$r, r_{xy}, r(x, y)$** : Coeficientes de correlação dos valores observados.

**$r^2, r^2_{xy}, r^2(x, y)$** : Coeficientes de determinação dos valores observados.

**$r_s$** : Coeficiente de correlação ordinal de SPEARMAN dos valores observados.

**$r_{yz.x}$** : Coeficiente de correlação parcial dos valores observados.  $r$

**$R_{z(xy)}$** : Coeficiente de correlação múltipla dos valores observados.

**$R^2_{z(xy)}$** : Coeficiente de determinação múltipla dos valores observados.

**$RH_0$** : Rejeição de uma hipótese nula.

**$s, s_x$** : Desvios padrões dos valores observados.

**$s^2, s^2_x$** : Variâncias dos valores observados.

**$S^2$** : Variável aleatória correspondente à variância de uma amostra.

**$s_{corr}$** : Desvio padrão corrigido ou correção de SHEPPARD.

**$s^2_{corr}$** : Variância corrigida ou correção de SHEPPARD.

**$s_{y.x}, s^2_{z.xy}$** : Desvios padrões residuais dos valores observados.

**$s^2_{y.x}, s^2_{z.xy}$** : Variâncias residuais dos valores observados.

**$s_{y|x}, s_{y|i}$** : Desvios padrões condicionados dos valores observados.

**$SQD, SQD_x$** : Somas dos quadrados dos desvios.

**$SQD_{y.x}$** : Soma dos quadrados dos desvios residual.

**$SPD, SPD_{xy}$** : Somas dos produtos dos desvios.

**t**: Variável de Student.

**$u_i, u_{obs}$** : Valor observado de U.

**$u_{1-\alpha}, u_{1-\alpha/2}$** : Valores teóricos de U.

**U,  $U_i$** : Variáveis aleatórias reduzidas ou variáveis normais reduzidas.

**$CV, CV_x, CV_x$** : Coeficientes de variação dos valores observados e teóricos.

**$var(x), var(X)$** : Variância dos valores observados e teóricos.

**W, At, A, R**: Amplitude dos valores observados.

**W**: Variável aleatória correspondente à amplitude de uma amostra.

**$w_i$** : Coeficientes de ponderação ou peso.

**$x, x_i$** : Valores observados.

**$X, X_i$** : Variáveis aleatórias.

**$\bar{x}, \bar{x}_i$** : Médias aritméticas.

**$\bar{X}$** : Variável aleatória correspondente à média aritmética de uma amostra.

$\bar{x}_g$ : Média geométrica.

$\bar{x}_h$ : Média harmônica.

$\bar{x}_p$ : Média ponderada.

$\bar{x}_q$ : Média quadrática.

$\bar{x}_y, \bar{x}_j, \bar{x}_{yz}, \bar{x}_{jk}$ : Médias condicionadas dos valores observados.

$\tilde{x}$ : Mediana dos valores observados.

$\tilde{X}$ : Variável aleatória correspondente à mediana de uma amostra.

$\alpha$ : Nível de significância, risco de primeira espécie.

$1 - \alpha$ : Grau de confiança.

$\alpha_k$ : Momentos em relação à origem dos valores teóricos.

$\beta$ : Risco de segunda espécie.

$\beta_{yx}, \beta_{xy}$ : Coeficientes de regressão dos valores teóricos.

$\beta_1, \beta_2$ : Coeficientes de PEARSON dos valores teóricos.

$\gamma, \gamma_i$ : Parâmetros quaisquer dos valores teóricos.

$\gamma_1, \gamma_2$ : Coeficientes de FISHER dos Valores teóricos.

$\delta$ : Diferença de médias.

$\Delta x, \Delta x_i$ : Intervalos de classe, acréscimos.

$\epsilon_m$ : Desvio médio absoluto dos valores teóricos.

$\eta_{yx}$ : Razão de correlação dos valores observados.

$\theta_{yx}$ : Razão de correlação dos valores teóricos.

$\lambda_{obs}$ : Valor observado de  $\Lambda$ .

$\lambda_a$ : Valor teórico de  $\Lambda$ .

$\Lambda$ : Razão de verossimilhança.

$\mu_k, \mu_{kl}$ : Momentos centrados dos valores teóricos.

$\mu_{11}, \mu_{11}(X, Y)$ : Covariâncias dos valores teóricos.

$\prod_{i=a}^b$ : Símbolo de multiplicação ou produto.

$\rho, \rho_{xy}, \rho(X, Y)$ : Coeficientes de correlação dos valores teóricos.

$\sigma, \sigma_x, \sigma(X)$ : Desvios padrões dos valores teóricos.

$\tilde{\sigma}$ : Desvio padrão estimado.

$\sigma^2, \sigma_x^2, \sigma^2(X)$ : Variâncias dos valores teóricos.

$\tilde{\sigma}^2$ : Variância estimada.

$\sigma_{y,x}^2$ : Variância residual dos valores teóricos.

$\sigma_{y|x}^2$ : Variâncias condicionadas.

$\sum_{i=a}^b$ : Símbolo de somatório.

$\phi$ : Coeficiente de correlação fi.

$\phi(\mathbf{u}), \phi(\mathbf{u}, \mathbf{v})$ : Funções densidade de probabilidade das distribuições normais reduzidas.

$\Phi(\mathbf{u})$ : Função de distribuição da distribuição normal reduzida a uma dimensão.

$\chi^2$ : Variável de PEARSON.

$\chi_{obs}^2$ : Valor esperado de  $\chi^2$ .

$\chi_{1-\alpha}^2$ : Valor teórico de  $\chi^2$ .

## ABREVIACÕES UTILIZADAS NA ESTATÍSTICA

**Conforme:** cf.

**Coordenador:** coord.

**Figura:** fig.

**Folha:** f.

**Isto é:** i.e.

**Não se aplica:** n.a.

**Nota do autor:** N. do A.

**Nota do editor:** N. do E.

**Nota do tradutor:** N. do T.

**Organizador:** org.

**Página:** p.

**Páginas:** pp.

**Plural:** pl.

**Por exemplo:** p. ex.

**Século:** séc.

**Segundo informações coletadas:** sic

**Sem data:** s.d.

**Separata:** sep.

**Tabela:** tab.

**Tomo:** t.

**Tradutor:** trad.

**Veja:** v.

**Ver também:** v. tb.

**Versus:** vs.

**Volume:** v.

#### SIMBOLOGIA OU NOMENCLATURA USADA NOS TESTES DE HIPÓTESES OU DE SIGNIFICÂNCIA:

Nível de significância alfa ( $\alpha$ )	Conclusão	Simbologia
Valor de alfa menor ou igual a 0,1% ( $\alpha \leq 0,1\%$ )	Diferença muito altamente significativa ou significativa	Três asteriscos (***)
Valor de alfa menor ou igual a 1% ( $\alpha \leq 1\%$ )	Diferença altamente significativa ou significativa	Dois asteriscos (**)
Valor de alfa entre 1% e 5% ( $1\% < \alpha < 5\%$ )	Diferença significativa ou significativa	Um asterisco (*)
Valor de alfa entre 5 e 10% ( $5\% \leq \alpha \leq 10\%$ )	Diferença provavelmente significativa ou significativa	Um asterisco (*)
Valor de alfa maior que 10% ( $\alpha > 10\%$ )	Diferença não significativa ou não significativa	n.s.

#### ABREVIACÕES LATINAS USADAS NA ESTATÍSTICA

**apud:** apud, Citado por.

**circa:** ci., Cerca, aproximadamente.

**et alii:** et al., e outros.

**et sequentia:** et. seq., e seguintes(s), que se segue(m).

**exempli gratia:** e.g., exemplo.

**ibidem:** ibid., no mesmo lugar, na mesma obra.

**idem:** idem, o mesmo, o mesmo autor.

**in:** In, dentro, em.

**loco citato:** loc. cit., no local citado.

**nota bene:** n. b., observe-se bem.

**opere citato:** op. cit., obra já citada.

**prima facie:** p. fac., à primeira vista, presumivelmente.

**quantum satis:** q.s., quanto basta.

**sine loco:** s.l., sem local de publicação.

**sine nomine:** s.n., sem nome do editor.

**status quo:** s.q., estado atual, situação atual.

**sui generis:** s.g., único, peculiar.

**ultima ratio:** u.rat., argumento definitivo, último.

**verbi gratia:** v.g., exemplo.

#### GLOSSÁRIO MATEMÁTICO

### 1. ARITMÉTICA, ÁLGEBRA ELEMENTAR E CÁLCULO DIFERENCIAL E INTEGRAL.

#### 1.1. MÓDULO OU VALOR ABSOLUTO

Módulo ou valor absoluto de um número real  $x$  é o valor de  $x$ , independentemente do sinal e é denotado por  $|X|$ .

Exemplo:  $|-5| = |5| = 5$ .

#### 1.2. FATORIAL

Fatorial de um número inteiro  $n$  é o produto de todos os inteiros de  $1, 2, \dots, n$  e é indicado por  $n!$ .

Exemplo:  $5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$ .  $n! = nx(n-1)x(n-2)x(n-3)x \dots x1$

$0! = 1$  (por definição).

#### 1.3. REGRAS PARA OS ÍNDICES (POTENCIAÇÃO)

$$X^a \cdot X^b = X^{a+b}$$

$$\frac{X^a}{X^b} = X^{a-b}$$

$$(X^a)^b = X^{ab}$$

$X^0 = 1, X \neq 0$  (Toda potência de expoente 0 é igual a 1).

$$\text{Exemplo: } \frac{6^2}{6^2} = \frac{36}{36} = 1; \frac{6^2}{6^2} = 6^{2-2} = 6^0 = 1$$

$X^1 = X$  (Toda potência de expoente 1 é igual a base).

$$\text{Exemplo: } \frac{2^5}{2^4} = \frac{32}{16} = 2; \frac{2^5}{2^4} = 2^{5-4} = 2^1$$

$$X^{-a} = \frac{1}{X^a}$$

$-X^a = X^a$ , sendo "a" par

**Exemplos:**  $2^2 \cdot 2^3 = 4 \cdot 8 = 32 = 2^5 = 2^{2+3}$ ;  $\frac{3^3}{3^2} = \frac{27}{9} = 3 = 3^{3-2}$ ;

$$(2^3)^2 = 8^2 = 64 = 2^6 = 2^{2 \cdot 3}; -3^2 = 9; -3^5 = -243$$

#### 1.4. FRAÇÃO

i) Soma ou adição de frações homogêneas ou de mesmo denominador.

Como fazer: Somam-se os numeradores e conserva-se o denominador comum.

$$\text{Exemplo: } \frac{2}{8} + \frac{3}{8} = \frac{2+3}{8} = \frac{5}{8}.$$

ii) Adição de frações que não têm o mesmo denominador comum, ou seja, frações heterogêneas.

Como fazer: Para somar frações que não tenham o mesmo denominador, é preciso reduzi-las ao mesmo denominador e aplicar, então, a regra do primeiro caso i.

$$\text{Exemplo: Somar as frações } \frac{2}{3}, \frac{5}{8} \text{ e } \frac{1}{6}.$$

Utilizando-se da regra 2 de redução de frações ao mesmo denominador comum, temos que o mínimo múltiplo comum dos números 3, 6, e 8 mmc (3,6,8) = 24 e:

$$\frac{2}{3} + \frac{5}{8} + \frac{1}{6} = \frac{16}{24} + \frac{15}{24} + \frac{4}{24} = \frac{16+15+4}{24} = \frac{35}{24}.$$

iii) Subtração de duas frações com o mesmo denominador.

**Como fazer:** Subtrai-se o numerador da menor do numerador da maior e conserva-se o denominador comum.

$$\text{Exemplo: } \frac{7}{12} - \frac{5}{12} = \frac{7-5}{12} = \frac{2}{12} = \frac{1}{6}.$$

iv) Subtração de duas frações que não têm o mesmo denominador.

Vale lembrar que, como colocado para a adição, que somente podemos subtrair quantidades de mesma natureza.

**Como fazer:** Da mesma forma que na adição, para se obter a subtração de frações heterogêneas, é preciso, primeiro, reduzi-las ao mesmo denominador, e, então, aplicar o caso i.

$$\text{Exemplo: } \frac{6}{7} - \frac{2}{5} = \frac{30}{35} - \frac{14}{35} = \frac{30-14}{35} = \frac{16}{35}.$$

v) Multiplicar uma fração por outra.

Como fazer: Para se multiplicar uma fração por outra, multiplicam-se seus numeradores para obter o numerador da fração produto e seus denominadores para obter o denominador da fração produto.

$$\text{Exemplo: } \frac{3}{5} \cdot \frac{5}{9} = \frac{3 \cdot 5}{8 \cdot 9} = \frac{15}{72} = \frac{5}{24}.$$

vi) Dividir uma fração por outra.

Como fazer: Multiplica-se a fração do dividendo pela fração do divisor invertida. Em outras palavras conserva-se a primeira que é o dividendo e multiplica-se pelo inverso da segunda que é o divisor.

$$\frac{\frac{a}{b}}{\frac{c}{d}} = \frac{a}{b} \cdot \frac{d}{c}$$

$$\text{Exemplo: } \frac{\frac{3}{5}}{\frac{4}{7}} = \frac{3}{5} \cdot \frac{7}{4} = \frac{3 \cdot 7}{5 \cdot 4} = \frac{21}{20}.$$

#### 1.5. NÚMERO PAR, NÚMERO ÍMPAR, NÚMERO MISTO, NÚMERO PRIMO E NÚMERO COMPOSTO.

i) Número Par: É todo número divisível por 2.

ii) Número Ímpar: É todo número que não é divisível por 2.

iii) Número Misto: Número constituído de um valor inteiro e uma fração.

Transformar um número misto em uma fração

$$a \frac{b}{c} = \frac{a \cdot c + b}{c}$$

Exemplo:

$$4 \frac{3}{5} = \frac{4 \cdot 5 + 3}{5} = \frac{23}{5}.$$

iv) Número Primo: Número natural maior que 1 e que só é divisível por 1 e por ele mesmo.

v) Número Composto: Número natural que é divisível por mais de dois números naturais.

#### 1.6. LOGARÍTMOS

Se  $X = a^Y$  então  $Y = \log_a X$ , isto é Y é o logaritmo de base a do número X.

#### 1.7. BASE 10

Se  $X = 10^Y$  então  $Y = \log_{10} X$

**Exemplo:**  $X = 100 = 10^2$ , isto é, o logaritmo de base 10 do número 100 é 2 ou  $\log_{10} 100 = 2$ ; e  $\log_{10} e = 0,4343$

#### 1.8. BASE e

Logaritmo de base e = 2,71828 são chamados natural ou logaritmo neperiano, o qual é escrito assim:  $\log_e x$  ou  $\ln x$ . Exemplo:  $\log_e 10 = 2,3026$ .

#### 1.9. REGRAS OU PROPRIEDADES DOS LOGARÍTMOS

$\log_a (xy) = \log_a x + \log_a y$ . Exemplo:  $\log_a (2 \cdot 3) = \log_{10} 2 + \log_{10} 3$

$\log_a \left(\frac{x}{y}\right) = \log_a x - \log_a y$ . Exemplo:  $\log_e \left(\frac{3}{4}\right) = \log_e 3 - \log_e 4$



$\log_a x^k = k \log_a x$ . Exemplo:  $\log_{10} 2^3 = 3 \log_{10} 2$

Essas regras são oriundas das regras para os índices da potenciação, mostrados anteriormente.

**1.10. MUDANÇA DE BASE NO LOGARÍTMO**

$\log_b x = \log_b a \times \log_a x$ . Exemplo:  $\log_e 100 = \log_e 10 \times \log_{10} 100 = 2,3025 \times 2 = 4,605$

**1.11. 1.11. PROGRESSÃO E SÉRIES**

**1.11.1. PROGRESSÃO ARITMÉTICA (PA)**

$a, a + d, a + 2d, \dots, a + (n - 1)d$ . Exemplo: seja o conjunto  $\{1, 3, 5, 7\}$

Soma dos n primeiros termos:

$S_n = \frac{(a_1 + a_n)}{2} \cdot n$ , onde  $S_n = a + (a + d) + (a + 2d) + \dots + [a + (n - 1)d]$   $S_n = \frac{n}{2} \{2a + (n - 1)d\}$

Exemplo:  $S_4 = \frac{4}{2} \{2 \times 1 + 3 \times 2\} = 2 \times 8 = 16$ . Soma dos n primeiros números naturais:

$S_n = 1 + 2 + 3 + 4 + \dots + n = \frac{n}{2} \{2 + (n - 1)\} = \frac{n(n + 1)}{2}$

**1.11.2. PROGRESSÃO GEOMÉTRICA (PG)**

$a, ar, ar^2, \dots, ar^{n-1}$ . Exemplo:  $(a = 1, r = \frac{1}{2}, n = 4)$

Soma dos n primeiros termos:  $S_n = a + ar + ar^2 + \dots + ar^{n-1} = \frac{a(1 - r^n)}{(1 - r)}$ .

Exemplo:  $S_4 = 1 + \frac{1}{2} + \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^3 = 1,875$ ;  $S_4 = \frac{1 - \left(\frac{1}{2}\right)^4}{\left(1 - \frac{1}{2}\right)} = 1,875$ .

Para  $|r| < 1$ , é possível determinar a soma desse tipo de PG infinita, por meio da expressão dada por:

$S_\infty = \frac{a}{(1 - r)}$ . Exemplo:  $S_\infty = \frac{1}{(1 - 0,5)} = 2$ .

**1.12. RAIZES QUADRADAS, CÚBICAS, QUARTAS, QUINTAS, SEXTAS, ...**

**Raiz quadrada:** Denominamos raiz quadrada de um número real Y, a todo número que elevado ao quadrado dá o valor Y. X é a raiz quadrada de Y, implica  $X^2 = Y$ . Denominaremos raiz n-ésima de um número real Y a todo número que elevado ao expoente n dá o valor Y. X é a raiz n-ésima de Y, implica que  $X^n = Y$ . Para todo número real Y podemos definir  $\sqrt[n]{Y} = Y$ .

**1.12.1. PROPRIEDADES FUNDAMENTAIS**

i)  $\sqrt[n]{Y^m} = \sqrt[n \cdot p]{Y^{m \cdot p}}$

ii)  $\sqrt[n]{Y^m} = \sqrt[\frac{n}{p}]{Y^{\frac{m}{p}}}$

iii)  $\sqrt[n]{X \cdot Y} = \sqrt[n]{X} \cdot \sqrt[n]{Y}$

iv)  $\sqrt[n]{\frac{X}{Y}} = \frac{\sqrt[n]{X}}{\sqrt[n]{Y}}$

v)  $(\sqrt[n]{Y})^m = \sqrt[n]{Y^m}$

vi)  $(\sqrt[n]{Y})^n = \sqrt[n]{Y^n} = Y$

vii)  $Y^{\frac{m}{n}} = \sqrt[n]{Y^m}$

viii)  $\sqrt[n]{Y^m} = Y^{\frac{m}{n}} = Y^{\frac{m \cdot p}{n \cdot p}} = \sqrt[n \cdot p]{Y^{m \cdot p}}$

ix)  $\sqrt[n]{X} \cdot \sqrt[n]{Y} = X^{\frac{1}{n}} \cdot Y^{\frac{1}{n}} = (X \cdot Y)^{\frac{1}{n}} = \sqrt[n]{X \cdot Y}$

x)  $\frac{\sqrt[n]{X}}{\sqrt[n]{Y}} = \frac{X^{\frac{1}{n}}}{Y^{\frac{1}{n}}} = \left(\frac{X}{Y}\right)^{\frac{1}{n}} = \sqrt[n]{\frac{X}{Y}}$

xi)  $(\sqrt[n]{X})^m = \left(X^{\frac{1}{n}}\right)^m = X^{\frac{1}{n} \cdot m} = X^{\frac{m}{n}} = \sqrt[n]{X^m}$

xii)  $\sqrt[n]{\sqrt[m]{X}} = \sqrt[n]{X^{\frac{1}{m}}} = \left(X^{\frac{1}{m}}\right)^{\frac{1}{n}} = X^{\frac{1}{m} \cdot \frac{1}{n}} = X^{\frac{1}{nm}} = \sqrt[nm]{X}$

**1.13. CÁLCULO DE EXPRESSÕES**

i)  $X(Y + Z) = XY + XZ$

ii)  $(X + Y)(Z + W) = XZ + XW + YZ + YW$

iii)  $(X + Y)^2 = X^2 + 2XY + Y^2$

iv)  $(X - Y)^2 = X^2 - 2XY + Y^2$

v)  $(X + Y)(X - Y) = X^2 - Y^2$

**1.14. FÓRMULAS TRIGONOMÉTRICAS**

**1.14.1. PARA QUALQUER TRIÂNGULO COM ÂNGULO RETO**

$Sen \theta = \frac{Cateto\ oposto}{Hipotenusa}$  ;

$Cos \theta = \frac{Cateto\ adjacente}{Hipotenusa}$  ;

$Tan \theta = \frac{Cateto\ oposto}{Cateto\ adjacente} = \frac{Sen \theta}{Cos \theta}$  ;

$Co\ sec \theta = \frac{1}{Sen \theta}$  ;

$Sec \theta = \frac{1}{Cos \theta}$  ;

$Cot \theta = \frac{1}{Tan}$

### 1.14.2. ÂNGULOS

**ÂNGULO:** Ângulo é uma figura geométrica formada por duas linhas que possuem um ponto como intersecção. Este ponto é o vértice do ângulo. A abertura do ângulo é uma propriedade invariante deste e é medida, no sistema internacional (SI), em radianos.

**RADIANO:** A medida em radiano de um ângulo é o comprimento do arco cortado pelo ângulo, dividido pelo raio do círculo. O sistema internacional (SI) de medidas utiliza o radiano como unidade derivada para ângulos. Devido ao seu relacionamento com o comprimento do arco, radianos são uma unidade especial. Senos e cossenos cujos argumentos estão em radianos possuem propriedades analíticas particulares, tal como criar funções exponenciais em base e.

**GRAUS:** A medida em graus de um ângulo é o comprimento de um arco, dividido pela circunferência de um círculo e multiplicada por 360. O símbolo de graus é um pequeno círculo sobrescrito °.  $2\pi$  radianos é igual a  $360^\circ$  (um círculo completo), então um radiano é aproximadamente  $57^\circ$  e um grau é  $\frac{\pi}{180}$  radianos.

**GRADIANO:** O gradiano, também chamado de grado, é uma medida angular onde o arco é dividido pela circunferência e multiplicado por 400. Essa forma é usado mais em triangulação. Sinônimo de Grado que é o nome de unidade de medida de arcos, correspondendo é centésima parte de um arco formado por um ângulo reto ( $90^\circ$ ). Como uma circunferência completa é formada por 4 arcos, o grado é  $\frac{1}{400}$  de toda a circunferência, valendo, portanto,  $0,9^\circ$  ( $0,9^\circ$  vezes  $400=360^\circ$ ).

Ângulo					
Degrado	$0^\circ$	$30^\circ$	$45^\circ$	$60^\circ$	$90^\circ$
Radiano	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$
Sen	0	$\frac{1}{2}$	$\frac{1}{\sqrt{2}}$	$\frac{\sqrt{3}}{2}$	1
Cos	1	$\frac{\sqrt{3}}{2}$	$\frac{1}{\sqrt{2}}$	$\frac{1}{2}$	0
Tan	0	$\frac{1}{\sqrt{3}}$	1	$\sqrt{3}$	$\infty$

Considere uma circunferência de centro O e um arco AB nessa circunferência. Se o arco AB tem comprimento igual ao raio, dizemos que ele mede 1 radiano. Portanto, radiano é a medida de um arco cujo comprimento é igual ao raio da circunferência que contém o referido arco. Como ao arco está associado um ângulo central, também pode-se dizer que radiano é a medida do ângulo central que determina na circunferência um arco cujo comprimento é igual ao raio. Uma vez que uma circunferência qualquer tem comprimento  $C = (2\pi.r)$  umc (unidades de medida de comprimento), o arco de uma volta tem medida igual a  $2\pi$  radianos. De modo geral, dado um arco cujo comprimento é L umc, dizemos que sua medida, em radianos, é igual a  $\frac{L}{r}$ . Assim, se a circunferência do arco considerado tem raio unitário, a medida do

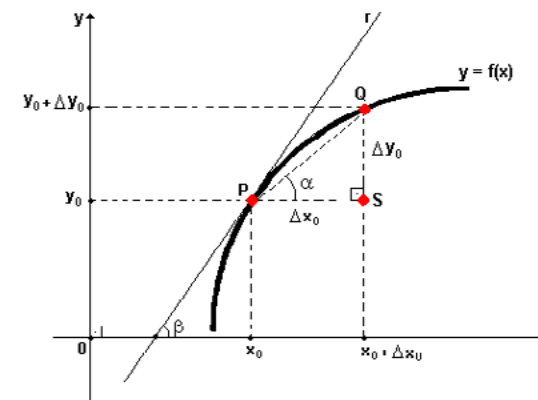
arco, em radianos, é numericamente igual ao comprimento do arco. Já a medida expressa em degree, é mensurada em graus.

### 1.15. TABELA DE DERIVADAS PADRÕES

#### 1.15.1. DEFINIÇÃO DE DERIVADA

Derivada de uma função  $y = f(x)$  num ponto  $x = x_0$ .

Considere a figura abaixo, que representa o gráfico de uma função  $y = f(x)$ , definida num intervalo de números reais.



Observando a figura, pode-se definir o seguinte quociente, denominado razão incremental da função  $y = f(x)$ , quando  $x$  varia de  $x_0$  para  $x_0 + \Delta x_0$ :

Define-se a derivada da função  $y = f(x)$  no ponto  $x = x_0$ , como sendo o limite da razão ou quociente incremental acima, quando  $\Delta x_0$  tende a zero, e é representada por  $f'(x_0)$ , ou seja:

$$\lim_{\Delta x_0 \rightarrow 0} \frac{\Delta y_0}{\Delta x_0} = f'(x_0)$$

**Nota:** a derivada de uma função  $y = f(x)$ , pode ser representada também pelos símbolos  $y'$  ou  $\frac{dy}{dx}$ .

Observe que quando  $\Delta x_0 \rightarrow 0$ , o ponto Q no gráfico acima, tende a coincidir com o ponto P da mesma figura., definindo a reta  $r$ , que forma um ângulo  $\beta$  com o eixo horizontal (eixo das abcissas), e, neste caso, o ângulo  $SPQ = \alpha$  tende ao valor do ângulo  $\beta$ . Quando  $\Delta x_0 \rightarrow 0$ , já vimos que o quociente  $\frac{\Delta y_0}{\Delta x_0}$

representa a derivada da função  $y = f(x)$  no ponto  $x_0$ . Mas, o quociente  $\frac{\Delta y_0}{\Delta x_0}$  representa, como se sabe da

trigonometria, a tangente do ângulo  $SPQ = \alpha$ , onde P é o vértice do ângulo. Quando  $\Delta x_0 \rightarrow 0$ , o ângulo  $SPQ = \alpha$ , tende ao ângulo  $\beta$ .

Assim, não é difícil concluir que a derivada da função  $y = f(x)$  no ponto  $x = x_0$ , é igual numericamente à tangente do ângulo  $\beta$ . Pode-se escrever então:  $f'(x_0) = tg \beta$ .

Neste caso pode-se afirmar que a derivada de uma função  $y = f(x)$  num ponto  $x = x_0$ , coincide numericamente com o valor da tangente trigonométrica do ângulo formado pela tangente geométrica à curva representativa de  $y = f(x)$ , no ponto  $x = x_0$ . Existem fórmulas para o cálculo das derivadas das funções,

como será mostrado a seguir. Veja por exemplo o cálculo da derivada da função  $y = x^2$ , no ponto  $x = 10$ , neste caso tem-se:

$$y = f(x) = x^2$$

$$f(x + \Delta x) = (x + \Delta x)^2 = x^2 + 2x \cdot \Delta x + (\Delta x)^2$$

$$f(x + \Delta x) - f(x) = x^2 + 2x \cdot \Delta x + (\Delta x)^2 - x^2 = 2x \cdot \Delta x + (\Delta x)^2$$

$$\Delta y = f(x + \Delta x) - f(x) = x^2 + 2x \cdot \Delta x + (\Delta x)^2 - x^2 = 2x \cdot \Delta x + (\Delta x)^2$$

Portanto a derivada é definida conforme a expressão a seguir:

$$y' = \frac{dy}{dx} = \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{2x \cdot \Delta x + (\Delta x)^2}{\Delta x} = \lim_{\Delta x \rightarrow 0} (2x + \Delta x) = 2x$$

Observe que se for colocado na expressão acima,  $\Delta x$  em evidência, simplifica-se o resultado obtido. Portanto a derivada da função  $y = x^2$  é igual a  $y' = 2x$ . Logo, a derivada da função  $y = x^2$ , no ponto  $x = 10$ , será igual a  $y'(10) = 2 \cdot 10 = 20$ .

$y = f(x)$	$\frac{dy}{dx}$
$x^n$	$nx^{n-1}$
$\sqrt{x}$	$\frac{1}{2\sqrt{x}}$
$\frac{1}{x}$	$-\frac{1}{x^2}$
$e^{ax}$	$ae^{ax}$
$\log_e x$	$\frac{1}{x}$
$a^x$	$a^x \log_e a$
$Sen x$	$Cos x$
$Cos x$	$-Sen x$
$\tan x$	$Sec^2 x$
$Sen^{-1}\left(\frac{x}{a}\right)$	$\frac{1}{\sqrt{a^2 - x^2}}$
$Cos^{-1}\left(\frac{x}{a}\right)$	$-\frac{1}{\sqrt{a^2 - x^2}}$
$\tan^{-1}\left(\frac{x}{a}\right)$	$\frac{a}{(a^2 + x^2)}$

**Exemplo:**  $3X^2 + 7 = 2 \cdot (3X^{2-1}) + 0 = 6X$

### 1.16. TABELA DE INTEGRAIS PADRÕES

#### 1.16.1. DEFINIÇÃO DE INTEGRAL

Seja  $f(x)$  uma função limitada definida no intervalo fechado  $[a, b]$  e seja P uma partição qualquer de  $[a, b]$ . A integral de  $f(x)$  no intervalo  $[a, b]$ , denotada por  $\int_a^b f(x) dx$  é dada por  $\int_a^b f(x) dx = \lim_{n \rightarrow \infty} \sum_{i=1}^n f(c_i) \cdot \Delta x_i$ ,

desde que o limite do segundo membro exista. Na notação  $\int_a^b f(x) dx$ ,  $f(x)$  é chamada função integrando,

$\int$  é o símbolo da integral, e os números  $a$  e  $b$  são chamados limites de integração, onde  $a$  é o limite inferior e  $b$  é o limite superior de integração. Se  $\int_a^b f(x) dx$  existe, diz-se que  $f$  é integrável em  $[a, b]$  e geometri-

camente, a integral representa a área da região limitada pela função  $f(x)$ , as retas  $x = a$  e  $x = b$  e o eixo  $x$ , desde que  $f(x) \geq 0 \forall x \in [a, b]$ .

Deve ser salientado que a integral não significa necessariamente uma área. Dependendo do problema, ela pode representar grandezas, como, volume, quantidade de bactérias presentes em certo instante, trabalho realizado por uma força, momentos e centro de massa ou ponto de equilíbrio. A definição de integral pode ser ampliada, de modo a incluir o caso em que o limite inferior seja maior do que o limite superior, e o caso em que os limites inferior e superior são iguais, senão vejamos.

Se  $a > b$ , então  $\int_a^b f(x) dx = -\int_b^a f(x) dx$ , se a integral à direita existir. Se  $a = b$  e  $f(a)$  existe, então  $\int_a^a f(x) dx = 0$ . Se  $f(x)$  é uma função contínua no intervalo fechado  $[a, b]$  então  $f(x)$  é integrável em  $[a, b]$ . A seguir são mostradas algumas propriedades fundamentais da integral definida.

i) Se a função  $f(x)$  é integrável no intervalo fechado  $[a, b]$  e se  $K$  é uma constante real qualquer, então

$$\int_a^b K f(x) dx = K \int_a^b f(x) dx;$$

ii) Se as funções  $f(x)$  e  $g(x)$  são integráveis em  $[a, b]$ , então  $f(x) \pm g(x)$  é integrável em  $[a, b]$

$$\text{e } \int_a^b f(x) \pm g(x) dx = \int_a^b f(x) dx \pm \int_a^b g(x) dx;$$

iii) Se  $a < c < b$  e a função  $f(x)$  é integrável em  $[a, c]$  e em  $[c, b]$ , então,  $f(x)$  é integrável em

$$[a, b] \text{ e } \int_a^b f(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx;$$

iv) Se a função  $f(x)$  é integrável e se  $f(x) \geq 0$  para todo  $x$  em  $[a, b]$ , então,  $\int_a^b f(x) dx \geq 0$ ;

v) Se as funções  $f(x)$  e  $g(x)$  são integráveis em  $[a, b]$  e  $f(x) \geq g(x)$  para todo  $x$  em  $[a, b]$ , então,

$$\int_a^b f(x) dx \geq \int_a^b g(x) dx;$$

vi) Se  $f(x)$  é uma função integrável em  $[a, b]$ , então  $|f(x)|$  é integrável em  $[a, b]$  e

$$\left| \int_a^b f(x) dx \right| \leq \int_a^b |f(x)| dx$$

Exemplo: Seja a função  $f(x) = 2x^3 + 7$  para  $1 \leq x \leq 4$ , calcule a integral dessa função para os limites especificados.

$$\begin{aligned} \int_1^4 f(x) dx &= \int_1^4 2x^3 + 7 dx = \int_1^4 2x^3 dx + \int_1^4 7 dx = 2 \int_1^4 x^3 dx + 7 \int_1^4 dx = \\ &= 2 \left[ \frac{x^4}{4} \right]_1^4 + 7[x]_1^4 = 2 \left[ \frac{4^4}{4} - \frac{1^4}{4} \right] + 7[4-1] = 2 \left[ \frac{256}{4} - \frac{1}{4} \right] + 7[3] = \\ &= 2[64-1] + 21 = 126 + 21 = 147 \end{aligned}$$

$y = f(x)$	$\int f(x) dx$
$a$	$ax + C$
$x^a$ ( $a \neq -1$ )	$\frac{x^{a+1}}{(a+1)} + C$
$\sqrt{x}$	$\frac{2x^{\frac{3}{2}}}{3} + C$
$\frac{1}{x}$	$\log_e  x  + C$
$\frac{1}{(a+x)}$	$\log_e  (a+x)  + C$
$e^{ax}$	$\frac{e^{ax}}{a} + C$

$y = f(x)$	$\int f(x) dx$
$\text{Sen } x$	$-\text{Cos } x + C$
$\text{Cos } x$	$\text{Sen } x + C$
$\tan x$	$-\log_e  \cos x  + C$
$\frac{1}{x^2 + a^2}$	$\frac{1}{a} \tan^{-1} \left( \frac{x}{a} \right) + C$
$\frac{1}{\sqrt{a^2 + x^2}}$	$\text{sen}^{-1} \left( \frac{x}{a} \right) + C$
$\log_e x$	$x \log_e  x  - x + C$

**Exemplo de aplicação:** Seja a função  $f(x) = 7Y^3 + 12$ . A integral dessa função, para x variando entre os limites reais de 1 a 2 é dada por:

$$I = \int_1^2 (7y^3 + 12) dy = \int_1^2 7y^3 dy + \int_1^2 12 dy = 7 \int_1^2 y^3 dy + 12 \int_1^2 dy = 7 \left[ \frac{y^{3+1}}{3+1} \right]_1^2 + 12[y]_1^2$$

$$I = 7 \left[ \frac{y^4}{4} \right]_1^2 + 12[y]_1^2 = 7 \left[ \frac{2^4 - 1^4}{4} \right] + 12[2-1] = 7 \left[ \frac{16-1}{4} \right] + 12 = 7 \left[ \frac{15}{4} \right] + 12$$

$$I = \frac{105}{4} + 12 = 38,25.$$

### 1.17. LIMITES

O conceito de limite de uma função é muito importante em toda teoria matemática envolvida com o cálculo diferencial e integral. Há uma cadeia ordenada muito bem estabelecida no cálculo. A teoria de limites é fundamental. O motivo para isto é que nem tudo o que queremos realizar, ocorre no meio físico e quase sempre é necessário introduzir um modelo que procura algo que está fora das coisas comuns e esta procura ocorre com os limites nos estudos de sequências, séries, cálculos de raízes de funções. Por exemplo, obter uma raiz de uma função polinomial de grau maior do que 4 somente é possível por meio de métodos numéricos que utilizam fortemente as idéias de limites e continuidade. Na verdade, este cálculo depende do teorema do valor Intermediário que é uma consequência do estudo de continuidade de funções. Idéia intuitiva de limite matemático. Estudaremos o comportamento de uma função f nas proximidades de um ponto. Para fixar idéias, consideremos a função  $f: \mathbb{R} - \{1\} \rightarrow \mathbb{R}$  definida por:

$$f(x) = \begin{cases} X^2 - 1 \\ X - 1 \end{cases}$$

Para x diferente de 1, f pode ser simplificada e reescrita na forma mais simples:

$$f(x) = x + 1$$

Ao analisar o comportamento desta função nas vizinhanças do ponto  $x = 1$ , ponto este que não pertence ao domínio de f, constatamos que esta função se aproxima rapidamente do valor  $L = 2$ , quando os valores de x se aproximam de  $x = 1$ , tanto por valores de  $x < 1$  (à esquerda de 1) como por valores  $x > 1$  (à direita de 1).

Do ponto de vista numérico, as tabelas abaixo mostram o comportamento da função f, para valores x à esquerda e à direita de  $x = 1$ .

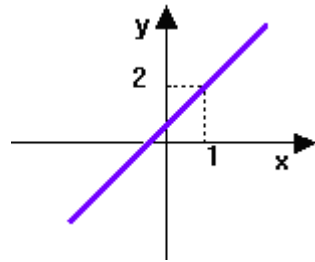
Pela esquerda de $x = 1$							Pela direita de $x = 1$								
x	0	0,5	0,8	0,9	0,99	0,999	1	x	2	1,5	1,2	1,1	1,01	1,001	1
f(x)	1	1,5	1,8	1,9	1,99	1,999	2	f(x)	3	2,5	2,2	2,1	2,01	2,001	2

Neste caso, dizemos  $L = 2$  é o limite da função f quando x se aproxima de 1, o que é simbolizado por:

$$\lim_{x \rightarrow 1} f(x) = 2$$

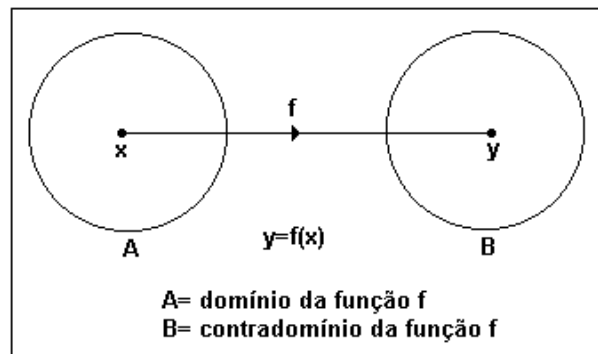
Este resultado pode ser visto por meio de análise gráfica de f, cujo esboço vemos na figura a seguir:





### 1.18. FUNÇÃO

Dados dois conjuntos A e B não vazios, chama-se função ou aplicação de A em B, representada por  $f: A \rightarrow B; y = f(x)$ , a qualquer relação binária que associa a cada elemento de A, um único elemento de B. Portanto, para que uma **relação de A em B** seja uma função, exige-se que a cada  $x \in A$  esteja associado um único  $y \in B$ , podendo entretanto existir  $y \in B$  que não esteja associado a nenhum elemento pertencente ao conjunto A.



Na notação  $y = f(x)$ , entende-se que **y é imagem ou contradomínio ou campo de valores da função de x pela função f, ou seja:** y está associado a x por meio da função f. Sendo x o domínio da função  $f(x)$  ou campo de definição da função. Exemplos:  $f(x) = 4x + 3$ ; então  $f(2) = 4 \cdot 2 + 3 = 11$  e portanto, 11 é imagem de 2 pela função f;  $f(5) = 4 \cdot 5 + 3 = 23$ , portanto 23 é imagem de 5 pela função f,  $f(0) = 4 \cdot 0 + 3 = 3$ , dentre outros valores.

#### ALFABETO GREGO

NOME DA LETRA	SÍMBOLOS	
	MAIÚSCULAS	MINÚSCULAS
Alfa	A	$\alpha$
Beta	B	$\beta$
Gama	$\Gamma$	$\gamma$
Delta	D	$\delta$
Épsilon	E	$\epsilon$
Zeta	Z	$z$
Eta	H	$h$
Téta	$\Theta$	$\theta$
Iota	I	$i$
Capa	K	$k$
Lambda	$\Lambda$	$\lambda$

Mu(mi)	M	$\mu$
Nu(ni)	N	$\nu$
Csi	X	$\chi$
Omicron	O	$\omicron$
Pi	$\Pi$	$\pi$
Ró	P	$\rho$
Sigma	S	$\sigma$
Tau	T	$\tau$
Upsilon(ipsilon)	Y	$\upsilon$
Fi	F	$\phi$
Chi(qui)	C	$\psi$
Psi	Y	$\psi$
Omega	$\Omega$	$\omega$

#### TEORIA DOS CONJUNTOS (ÁLGEBRA DE CONJUNTOS) E ANÁLISE COMBINATÓRIA (TÉCNICAS DE CONTAGEM).

### 1. Introdução

A teoria dos conjuntos foi desenvolvida por volta do ano de 1872 pelo matemático alemão de origem russa que nasceu na cidade de São Petersburgo na Rússia, Georg Ferdinand Ludwig Philipp Cantor (1845–1918) e aperfeiçoada no início do século XX por outros matemáticos, entre eles, Ernst Zermelo (alemão, 1871–1956), Adolf Fraenkel (alemão, 1891–1965), Kurt Gödel (austríaco, 1906–1978), Janos von Newman (húngaro, 1903–1957), dentre outros.

O estudo sobre a teoria de conjuntos se faz necessário em estatística devido a moderna teoria de probabilidades ter como base teórica a álgebra de conjuntos ou de eventos, a qual é baseada em axiomas matemáticos, sendo necessário conhecermos essa teoria para compreendermos os axiomas, teoremas, leis ou propriedades referentes ao estudo dos experimentos aleatórios, de seus eventos e possibilidades de ocorrência desses acontecimentos.

### 2. Conceitos básicos

#### 2.1. Conjuntos

Em matemática, um conjunto é qualquer lista ou coleção bem definida de objetos os quais são os elementos do conjunto. Não interessa a ordem e quantas vezes os elementos estão listados na coleção. Ou seja, a noção de conjunto em Matemática é praticamente a mesma utilizada na linguagem cotidiana: agrupamento, classe, coleção. Por exemplo: Conjunto das letras maiúsculas do alfabeto; Conjunto dos números inteiros pares; Conjunto dos dias da semana; Conjunto dos Presidentes da República do Brasil, etc.

Designamos os conjuntos por letras maiúsculas latinas, tais como A, B, C, X, Y, U, V, W,..., e os elementos que os compõe por letras minúsculas latinas dispostas entre chaves e separados por vírgula, como por exemplo,  $\{a, b, c, x, y, u, v, w, \dots\}$

##### 2.1.1. Exemplos de conjuntos

i) Os números inteiros entre 1 e 50, inclusive;

ii) Os pontos de uma reta;

iii) Os números reais entre 0 e 12, exclusive

### 2.1.2. Elemento

É cada membro ou objeto que entra na formação do conjunto. Por exemplo: V, I, C, H, E são elementos do primeiro conjunto acima; 2, 4, 6 são elementos do segundo; e assim por diante.

### 2.1.3. Relação de pertinência ( $\in$ )

**Pertinência:** é a característica associada a um elemento que faz parte de um conjunto.

**Símbolo de pertinência:** Se um elemento pertence a um conjunto utilizamos o símbolo  $\in$  que se lê: “pertence”. Para afirmar que 1 é um número natural ou que 1 pertence ao conjunto dos números naturais, escrevemos:  $1 \in \mathbb{N}$ .

Para afirmar que 0 não é um número natural ou que 0 não pertence ao conjunto dos números naturais, escrevemos:  $0 \notin \mathbb{N}$ .

Um símbolo matemático muito usado para a negação é a barra / traçada sobre o símbolo normal.

### 2.2. Diagrama de EULER-VENN (Lê-se “Ôiler-ven”)

Pode-se expressar as operações entre conjuntos através dos DIAGRAMAS DE EULER-VENN (John Venn, 1834-1923) que são úteis na verificação de propriedades de operações entre conjuntos, mas não devem ser considerados instrumentos de prova matemática rigorosa. Sendo assim é possível usar essas figuras para representar conjuntos e, como em muitos casos não interessa saber quais são os elementos de um conjunto, podemos representá-los por uma região do plano delimitada por uma curva fechada, essas figuras que são conhecidas como Diagramas de Euler-Venn, são muitos úteis no estudo de conjuntos.

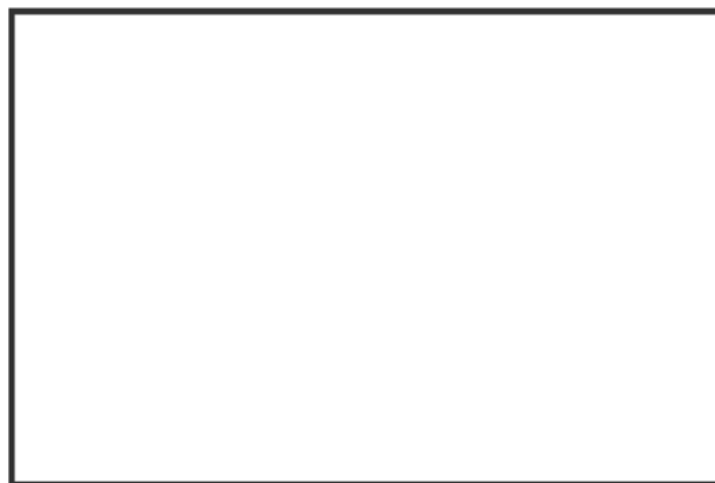


Figura 1: Diagrama de Euler – Venn.

### 2.3. Formas de designação de conjuntos e dos elementos de um conjunto

Existem duas maneiras de designar os elementos e os conjuntos simbolicamente, as quais são descritas a seguir.

#### 2.3.1. Método da enumeração ou listagem completa ou método tabular

É aquele onde ocorre a designação dos nomes de todos os elementos do conjunto.

Este método é usado quando o número de elementos não é grande; escrevem-se os nomes dos elementos entre chaves. Por exemplo, o conjunto A dos números pares positivos, menores do que 10, pode ser representado por:

$$A = \{2, 4, 6, 8,\}$$

Essa mesma notação poderá ser usada se o número de elementos é grande, mas de tal maneira que, escrevendo-se os primeiros elementos do conjunto, se possa inferir quais são os demais elementos omitidos. Por exemplo, o conjunto dos números pares positivos, menores do que 40 podem ser simbolizados por  $\{2, 4, 6, 8, \dots, 38\}$ , onde escrevemos os primeiros elementos e o último, separados por reticência. Outro exemplo é o do conjunto P de todos os números pares positivos, o qual poderá ser representado por  $P = \{2, 4, 6, 8, 10, \dots\}$ . Outros exemplos usados para ilustrar esse método podem ser:

O conjunto A dos números primos, menores do que 10 isto é:

$$A = \{2, 3, 5, 7\};$$

O conjunto B dos números pares positivos, menores do que 8:

$$B = \{2, 4, 6\}.$$

#### 2.3.2. Método da propriedade ou regra

É aquele onde ocorre a designação de uma propriedade satisfeita por todos os elementos do conjunto e não é válida para elementos de outros conjuntos. Por exemplo, o conjunto dos números fracionários menores do que 3, o conjuntos dos números primos, o conjunto das cidades do estado do Rio Grande do Norte.

Seja, então, o conjunto F dos números fracionários entre 0 e 2. Cada elemento do conjunto F deve ser um número fracionário compreendido entre 0 e 2 (propriedade definidora do conjunto).

Para representar um elemento qualquer do conjunto F, usamos um símbolo, chamado variável, que pode ser uma letra qualquer, x, por exemplo. Dizemos, então, que uma variável é um símbolo que pode ser substituído por qualquer elemento de um conjunto, denominado domínio da variável. No exemplo anterior, o domínio da variável x é o conjunto dos números fracionários compreendidos entre 0 e 2.

Sendo assim, o conjunto F pode ser denotado como mostrado a seguir.

$$F = \{x \text{ tal que } x \text{ é fracionário e } 0 < x < 2\}.$$

Se denominarmos de D o conjunto dos números fracionários e usarmos um símbolo para tal que, usualmente " $|$ " ou então ":", tem-se o seguinte:

$$D = \{x \mid x \in D \text{ e } 0 < x < 2\} \text{ ou}$$

$$D = \{x: x \in D \text{ e } 0 < x < 2\}.$$

Em geral, o conjunto dos elementos  $x$  que satisfazem a uma propriedade “ $p$ ” pode ser escrito assim:

$$\{x \mid x \text{ satisfaz } p\} \text{ ou } \{x \mid p(x)\}, \text{ ou ainda,}$$

$$\{x : p(x)\}$$

Mostramos a seguir alguns exemplos de aplicação desse método.

O conjunto  $A$  dos números inteiros não-negativos menores do que 2000:

$$A = \{x \in \mathbb{N} \mid x < 2000\}$$

O conjunto  $B$  dos números fracionários cujos quadrados são maiores ou iguais a 25: sendo  $D =$  conjunto dos números fracionários.

$$B = \{x \in D \mid x^2 \geq 25\}$$

Às vezes, essa representação pode dar margem à representação anterior da listagem completa. Por exemplo, o conjunto  $C$  dos números naturais que satisfazem à equação  $x^2 - 81 = 0$  é o seguinte:

$C = \{x \in \mathbb{N}^* \mid x^2 - 81 = 0\}$  ou então  $C = \{9\}$ , onde  $\mathbb{N}^*$  é o conjunto dos números naturais, que é o conjunto  $\mathbb{N}$  dos números inteiros não-negativos [ $\mathbb{N} = \{0, 1, 2, 3, 4, 5, 6, 7, \dots\}$ ] sem o zero, ou seja,  $\mathbb{N}^* = \{1, 2, 3, 4, 5, 6, 7, \dots\}$ .

### 3. Simbologias usadas na álgebra de conjuntos

#### 3.1. Quadro representativo dos principais símbolos utilizados na teoria dos conjuntos

Na tabela 1 são mostrados diversos símbolos utilizados na teoria de conjuntos, os quais são de extrema importância para se entender a álgebra de conjuntos e eventos.

Vale lembrar que nem todos esses símbolos são usados ao mesmo tempo, e que nem todos serão aplicados nos próximos itens que serão descritos adiante, no entanto para se efetuar operações de álgebra de conjuntos ou de eventos o leitor pode se valer desses símbolos contidos na tabela 1.

Tabela 1: Símbolos representativos utilizados na teoria de conjuntos

$\in$ : pertence	$\exists$ : existe
$\notin$ : não pertence	$\nexists$ : não existe
$\subset$ : está contido	$\forall$ : para todo (ou qualquer que seja)
$\not\subset$ : não está contido	$\emptyset$ : conjunto vazio
$\supset$ : contém	$\mathbb{N}$ : conjunto dos números naturais
$\not\supset$ : não contém	$\mathbb{Z}$ : conjunto dos números inteiros

$/$ : tal que	$\mathbb{Q}$ : conjunto dos números racionais
$\Rightarrow$ : implica que	$\mathbb{Q}' = \mathbb{I}$ : conjunto dos números irracionais
$\Leftrightarrow$ : se, e somente se	$\mathbb{R}$ : conjunto dos números reais

#### 3.2. Simbologia utilizada entre conjuntos

##### 3.2.1. Pertence $(\in)$

Relação usada entre elemento e um conjunto.

Se  $a$  é um elemento de  $A$ , nós podemos dizer que o elemento  $a$  pertence ao conjunto  $A$  e podemos escrever  $a \in A$ .

Por exemplo, seja o conjunto  $A = \{1, 2, 3, 4, 5\}$ , então  $1 \in A$ .

##### 3.2.2. Não pertence $(\notin)$

Relação usada entre elemento e um conjunto.

Se  $a$  não é um elemento de  $A$ , nós podemos dizer que o elemento  $a$  não pertence ao conjunto  $A$  e podemos escrever  $a \notin A$ .

Por exemplo, seja o conjunto  $B = \{10, 20, 30, 40, 50\}$ , então  $5 \notin B$ .

##### 3.2.3. Está contido $(\subset)$

Relação usada entre conjuntos.

Diz-se que o conjunto  $A$  está contido no conjunto  $B$ , ou que  $A$  é subconjunto de  $B$ , se, e somente se, todo elemento ou componente de  $A$  também pertence ao conjunto  $B$ , e indica-se por  $A \subset B$ . Por exemplo:

$$\{0, 2\} \subset \{0, 2\}$$

$$\{0, 3\} \subset \{0, 1, 3\}$$

Vale salientar a diferença existente entre os símbolos  $\in$  e  $\subset$ , pois o primeiro é usado para relacionar elemento e conjunto, enquanto o segundo relaciona dois conjuntos, por exemplo, dizemos que  $1 \in \{1, 4\}$  e não  $1 \subset \{1, 4\}$ . Todavia,  $\{1\} \subset \{1, 4\}$ , pois  $\{1\}$  é um conjunto (unitário).

##### 3.2.4. Não está contido $(\not\subset)$

Relação usada entre conjuntos.

A negação para está contido  $(\subset)$  indica-se com o sinal de está contido com uma barra cortando o símbolo, na diagonal. Assim  $A \not\subset B$ . Por exemplo, Considere os seguintes conjuntos finitos  $A = \{4, 5, 6\}$  e  $B = \{1, 10, 20\}$ , nesse caso  $A \not\subset B$ .

Se  $A \subset B$  ou  $B \subset A$ , dizemos que  $A$  e  $B$  são comparáveis. Se  $A \not\subset B$  e  $B \not\subset A$ , então  $A$  e  $B$  não são comparáveis. Sendo assim, os conjuntos  $A = \{3, 6\}$  e  $B = \{0, 2, 5, 9\}$  não são comparáveis.

### 3.2.5. Contém ( $\supset$ )

Relação usada entre conjuntos.

Se  $A \subset B$ , diz-se também que B contém A e indica-se assim ( $B \supset A$ ). Por exemplo, Sejam os conjuntos  $A = \{1, 5, 10\}$  e  $B = \{1, 5\}$  então  $A \supset B$ .

### 3.2.6. Não contém ( $\not\supset$ )

Relação usada entre conjuntos.

A negação para contém é dada pelo símbolo de contém com uma barra cortando o símbolo, na diagonal indicado por  $\not\supset$ . Por exemplo, Sejam os seguintes conjuntos  $A = \{1, 5, 10\}$  e  $B = \{2, 6\}$  então  $A \not\supset B$ .

### 3.2.7. Igual (=)

Relação usada entre conjuntos.

Diz-se que dois conjuntos A e B são iguais se, e somente, todo elemento de A pertence a B e todo elemento de B pertence a A. Sendo assim, a maneira de se constatar se dois conjuntos A e B são iguais ou semelhantes é verificar a dupla inclusão.  $A \subset B$  e  $B \subset A$ . Por exemplo, os conjuntos  $A = \{0, 5\}$  e  $B = \{5, 0\}$  são iguais isto é  $A = B$  ou  $\{0, 5\} = \{5, 0\}$ . Este exemplo ilustra o fato de que a ordem dos elementos no conjunto é indiferente, segundo a definição anterior, ou seja, a afirmação de que  $A \subset B$  não exclui a possibilidade de B estar contido em A, e quando isso ocorre, dizemos que A e B são iguais e denotamos por  $A = B$ .

### 3.2.8. Diferente ( $\neq$ )

Relação usada entre conjuntos.

Se não for verificada a definição de igualdade entre conjuntos, dizemos então que A é diferente de B e indica-se assim  $A \neq B$ . Por exemplo, sejam os conjuntos  $A = \{1, 3, 5, 6, 8, 12\}$ ;  $B = \{2, 4, 10\}$ , então  $A \neq B$ .

## 4. Tipos de conjuntos

### 4.1. Conjunto universo (U)

É o conjunto ao qual pertencem todos os elementos envolvidos em um determinado assunto ou estudo, e é simbolizado pela letra U. Assim, se procuramos determinar as soluções reais de uma equação do segundo grau, nosso conjunto Universo U é R (conjunto dos números reais);



Figura 2: Diagrama ilustrativo do conjunto universo.

#### 4.1.1. Exemplo

$U = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$

### 4.2. Conjunto vazio ( $\emptyset$ )

**Conjunto vazio:** é um conjunto que não possui elementos. O conjunto vazio é representado por  $\{ \}$  ou  $\emptyset$ . Convém notar a diferença entre o conjunto vazio e o conjunto cujo único elemento é o zero, isto é, o conjunto  $\{0\}$ , que não é vazio. O conjunto vazio é finito, pois tem zero elemento.

Pode-se admitir sem demonstração, a afirmação de que o conjunto vazio é único.

O conjunto vazio é um subconjunto de qualquer conjunto A. Dessa forma, para qualquer conjunto A, temos que  $\emptyset \subset A \subset U$ .

#### 4.2.1. Exemplo

Seja A o conjunto dos números primos pares entre 4 e 9. Então  $A = \emptyset$ .

### 4.3. Conjunto unitário

É aquele que é constituído de apenas um elemento.

#### 4.3.1. Exemplo

$A = \{x \mid x \text{ é par e } 4 < x < 8\}$  ou  $A = \{6\}$

### 4.4. Conjunto binário

É aquele que possui apenas dois elementos

#### 4.4.1. Exemplo

$B = \{7, 9\}$

### 4.5. Conjunto finito

É aquele que é vazio ou constituído de exatamente “n” elementos, onde “n” é um número inteiro positivo.

#### 4.5.1. Exemplo

Seja D o conjunto das letras que são vogais do alfabeto latino, isto é:

$D = \{a, e, i, o, u\}$ , então D é um conjunto finito.

### 4.6. Conjunto infinito

É aquele contrário a finito, ou seja, o conjunto que não é vazio ou não é constituído de um número “n” de elementos, onde “n” não é um número inteiro positivo ou zero.

#### 4.6.1. Exemplo

Seja I o conjunto dos números inteiros pares positivos, isto é:

$Y = \{2, 4, 6, 8, \dots\}$ ,

Então I é um conjunto infinito.



#### 4.7. Conjunto enumerável

É aquele conjunto que é finito ou se seus elementos podem ser arranjados em forma de seqüência, e neste caso é chamado de infinito enumerável.

##### 4.7.1. Exemplo

Seja  $N$  o conjunto dos números inteiros não negativos.

$$N = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, \dots\}$$

O conjunto  $n$  é infinito enumerável

#### 4.8. Conjunto não enumerável

É aquele conjunto que é infinito, e que seus elementos não podem ser arranjados em forma de seqüência.

##### 4.8.1. Exemplo

Seja  $F$  o intervalo unitário de números reais, isto é:

$$F = \{x \mid 0 \leq x \leq 1\}$$

### 5. Subconjuntos

**Subconjunto pode ser definido assim:** quando todos os elementos de um conjunto  $A$  qualquer pertencem a um outro conjunto  $B$ , diz-se, então, que  $A$  é um subconjunto de  $B$ , ou seja,  $A \subset B$ . isto é um conjunto  $A$  é um subconjunto de (está contido em)  $B$  se, e somente se, todo elemento  $x$  pertencente a  $A$  também pertence a  $B$ ; por exemplo: Todo o conjunto  $A$  é subconjunto dele próprio, ou seja  $A \subset B$ . O conjunto vazio, por convenção, é subconjunto de qualquer conjunto, ou seja,  $\emptyset \subset A$ .

A menos que afirme-se o contrário, pode-se dizer que todos os conjuntos em estudo são supostos subconjuntos de um conjunto fixo chamado conjunto universo, o qual é denotado pela letra “ $U$ ”.

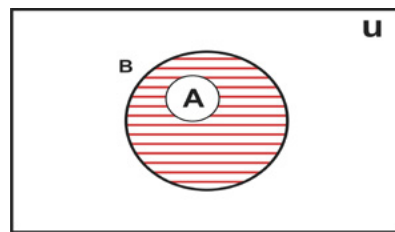


Figura 3: Diagrama ilustrativo de um subconjunto  $A$  de um conjunto  $B$ .

##### 5.1. Exemplo

Sejam os conjuntos  $A = \{6, 7\}$  e  $B = \{5, 6, 7, 8, 9, 10\}$ , então  $A$  é subconjunto próprio de  $B$ .

##### 5.2. Tipos

###### 5.2.1. Subconjunto próprio

Vale lembrar que se  $A \subset B$  não exclui a possibilidade de  $A = B$ . Contudo, se  $A \subset B$ , mas  $A \neq B$ , então dizemos que  $A$  é subconjunto próprio de  $B$ .

Alguns autores usam o símbolo  $\subseteq$  para designar qualquer subconjunto e o símbolo  $\subset$  somente para um subconjunto próprio.

###### 5.2.2. Subconjunto impróprio

Todo conjunto é subconjunto dele próprio, chamado de subconjunto impróprio. Quando  $A$  é um subconjunto de  $B$ , diz-se que  $B$  é um superconjunto de  $A$ .

###### 5.2.3. Cardinalidade

Cardinalidade, em teoria dos conjuntos, é uma forma de medir a quantidade de elementos de um conjunto. Existem duas formas de tratar a cardinalidade: Usando números cardinais, ou seja, associando a cada conjunto um outro conjunto, de forma que conjuntos com o mesmo número de elementos estão associados ao mesmo número cardinal. Comparando diretamente os conjuntos, através de funções.

O cardinal indica o número ou quantidade dos elementos constituintes de um conjunto. É interessante destacar que se diferencia do ordinal, porque o ordinal introduz ordem e dá idéia de hierarquia: Primeiro, segundo, terceiro, etc. O cardinal, por sua vez, nomeia o número de elementos constituintes e esse é o nome do conjunto correspondente. Para a nomenclatura destes números ver nomes dos números.

Dado um conjunto  $A$ , o cardinal deste conjunto é simbolizado por  $|A|$ . Por exemplo: Se  $A$  tem 3 elementos o cardinal indica-se  $|A| = 3$

### 6. Classes de conjuntos ou conjuntos das partes (P)

Frequentemente, os elementos de um conjunto são eles mesmos conjuntos. Por exemplo, cada linha num conjunto de linhas é um conjunto de pontos. Para ajudar a esclarecer essas situações, usamos muitas vezes a palavra classe ou família para tais conjuntos. As palavras subclasse e subfamília têm significado análogo ao de subconjunto.

O conjunto de todos os subconjuntos de um conjunto dado  $A$  é chamado de conjunto de partes (ou conjunto potência) de  $A$ , denotado por  $P(A)$  ou  $2^A$ .

Dado um conjunto  $A$  pode-se considerar o conjunto  $P(A)$ , chamado conjunto das partes de  $A$  tal que  $P(A) = \{X \mid X \subset A\}$ .

Em geral, se  $A$  é finito e tem  $n$  elementos, então  $P(A)$  terá  $2^n$  elementos, ou seja:

$$\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \binom{n}{3} + \dots + \binom{n}{n} = 2^n$$

Existe uma relação entre o cardinal de um conjunto e o conjunto de partes ou conjunto potência:  $|A| = n \Rightarrow |P(A)| = 2^n$ . Onde  $|P(A)|$  é o cardinal do conjunto de partes.

##### 6.1. Exemplo 1.

Se  $S$  é o conjunto de três elementos  $\{x, y, z\}$  a lista completa de subconjuntos de  $S$  é:

$\{\}$  (conjunto vazio);

$\{x\}$ ;

$\{y\}$ ; $\{z\}$ ; $\{x, y\}$ ; $\{x, z\}$ ; $\{y, z\}$ ; $\{x, y, z\}$ ;

e portanto o conjunto de partes de S é o conjunto de 8 elementos:

$$P(S) = \{\{\}, \{x\}, \{y\}, \{z\}, \{x, y\}, \{x, z\}, \{y, z\}, \{x, y, z\}\}.$$

Notemos que o número de elementos de A é 3 e o de P(A) é 8, ou  $2^3$ , ou melhor:

$$\binom{3}{0} + \binom{3}{1} + \binom{3}{2} + \binom{3}{3} = 2^3 = 8$$

## 6.2. Exemplo 2

Se  $A = \phi$ , então  $P(A) = \{\phi\}$ , que não é vazio, e P(A) tem  $2^0 = 1$  elemento.

## 7. Propriedades ou leis da álgebra de conjuntos (ou eventos)

Os conjuntos sob as operações de união, interseção, diferença e complemento absoluto satisfazem como já mencionado anteriormente às várias leis ou identidades que estão enumeradas a seguir.

1:  $A \cup B = B \cup A \Rightarrow$  LEI COMUTATIVA DA UNIÃO

2:  $A \cup (B \cap C) = (A \cup B) \cap C = A \cup B \cap C \Rightarrow$  LEI ASSOCIATIVA DA UNIÃO

3:  $A \cap (B \cup C) = (A \cap B) \cup C \Rightarrow$  LEI ASSOCIATIVA DA INTERSEÇÃO

4:  $A \cap B = B \cap A \Rightarrow$  LEI COMUTATIVA DA INTERSEÇÃO

5:  $A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \Rightarrow$  PRIMEIRA LEI DISTRIBUTIVA

6:  $A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \Rightarrow$  SEGUNDA LEI DISTRIBUTIVA

7:  $A - B = A \cap B^c$

8: SE  $A \subset B$ , ENTÃO  $A^c \supset B^c$  OU  $B^c \subset A^c$ ,  $(A \cup B) = B$ ,  $A \cap B = A \Rightarrow$  LEI DAS ABSORÇÕES

9:  $A \cup \phi = A$ ,  $A \cap \phi = \phi \Rightarrow$  LEI DE IDENTIDADE

10:  $A \cup U = U$ ,  $A \cap U = A \Rightarrow$  LEI DE IDENTIDADE

11:  $(A \cup B)^c = A^c \cap B^c \Rightarrow$  PRIMEIRA LEI DA DUALIDADE OU “DE MORGAN”<sup>1</sup>

<sup>1</sup> Augustus De Morgan ou Augustus De Morganquinto, Matemático e lógico indiano radicado na Inglaterra e tornou-se cidadão Britânico, nasceu na cidade de Madura na Índia em 1806 e faleceu na cidade de Londres em 1871, foi membro da real sociedade, introduziu em seu país a lógica matemática.

12:  $(A \cap B)^c = A^c \cup B^c \Rightarrow$  SEGUNDA LEI DA DUALIDADE OU “DE MORGAN”

13:  $A = (A \cap B) \cup (A \cap B^c) \Rightarrow$  PARA QUAISQUER CONJUNTOS A e B

14:  $A \cup A = A$ ;  $A \cap A = A \Rightarrow$  LEIS IDEMPOTENTES

15:  $A \cup (A \cap B) = A$

16:  $A \cup (A^c \cap B) = A \cup B$

17:  $A \cup B = (A^c \cap B) \cup (A \cap B^c) \cup A \cap B$

18:  $A \cup A^c = U \Rightarrow$  LEI DO COMPLEMENTO OU COMPLEMENTARES

19:  $A \cap A^c = \phi \Rightarrow$  LEI DO COMPLEMENTO

20:  $U^c = \phi$ ,  $\phi^c = U \Rightarrow$  LEI DO COMPLEMENTO

21:  $(A^c)^c = A \Rightarrow$  LEI DO COMPLEMENTO (DUPLA NEGAÇÃO)

Onde:  $\left\{ \begin{array}{l} U = OU = SOMA (+) \\ \cap = e = MULTIPLICAÇÃO (X) \end{array} \right.$

## 8. Operações entre conjuntos

Neste item estuda-se determinadas operações que podem ser efetuadas com os conjuntos em geral, sendo neste caso de fundamental importância e necessário distinguir dois conectivos, os quais são o “ou” e o “e”.

O conectivo “e” é usado quando liga duas afirmações que são válidas simultaneamente. Por exemplo, se dissermos Vou ao circo e ao museu, significa que irei ao circo e também ao museu.

Geralmente quando se usa o conectivo “ou”, este é inclusivo, isto é, quando dizemos A ou B, o significado é A ou B ou ambos.

As definições dadas para dois conjuntos podem ser estendidas para três ou mais conjuntos, reduzindo-se sempre a uma operação com dois conjuntos.

Sejam os seguintes conjuntos: Universo (U) e sejam os conjuntos A, B, C, D, E, F, G, H e L, contidos nesse conjunto universo, e sejam ainda dois conjuntos H e L adicionais. Esses conjuntos serão usados para demonstrar e ilustrar as operações entre conjuntos.

$U = \{1, 2, 3, 4, 5, 6, 10, 20\}$

$A = \{1, 3, 5, 6\}$

$B = \{2, 4, 6\}$

$C = \{1, 10, 20\}$

$D = \{20, 1, 10\}$

$E = \{1, 20\}$

$F = \{1, 2, 3, 4, 5\}$

$$G = \{1, 2, 3\}$$

$$H = \{0, 2, 4, 6, \dots\}$$

$$L = \{1, 3, 5, 7, \dots\}$$

### 8.1. União ( $\cup$ )

#### 8.1.1. Introdução

Chama-se reunião de A e B ao conjunto S dos elementos de U que pertencem a A ou a B (ou inclusivo). Indica-se assim:  $S = A \cup B$  e lê-se A **U** B que é a parte hachurada da figura abaixo. Simbolicamente, temos que:

$$A \cup B = \{x \in U \mid x \in A \text{ ou } x \in B\}$$

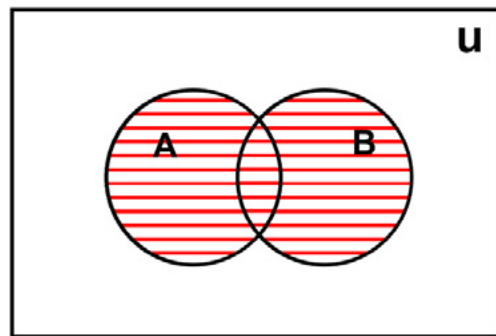


Figura 4: Diagrama da operação reunião entre dois conjuntos A e B.

#### 8.1.2. Exemplo 1

Sejam os eventos A e B dados por:  $A = \{1, 3, 5, 6\}$  e  $B = \{2, 4, 6\}$ .

Temos que:

$$A \cup B = \{1, 2, 3, 4, 5, 6\}$$

#### 8.1.3. Exemplo 2

Sejam os conjuntos D e E dados abaixo, então temos:  $D = \{20, 1, 10\}$  e

$$E = \{1, 20\}$$

$D \cup E = \{1, 10, 20\}$ , isto é, se  $E \subset D$ ,  $D \cup E = D$  (veja a figura abaixo, ilustrando essa operação).

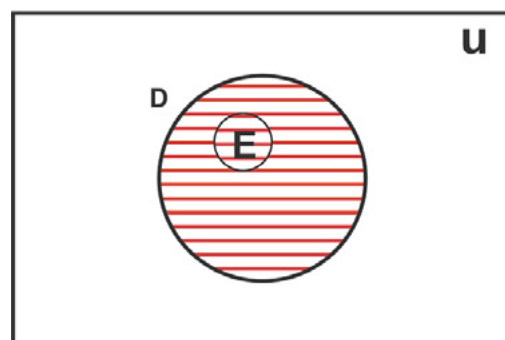


Figura 5: Diagrama da operação reunião entre um conjunto D e um subconjunto E

### 8.2. Interseção ( $\cap$ )

#### 8.2.1. Introdução

Chama-se interseção de A e B ao conjunto I dos elementos de U que pertencem a A e B. Indica-se por  $I = A \cap B$  e lê-se A **I** B. A região hachurada na figura abaixo representa a interseção de A e B.

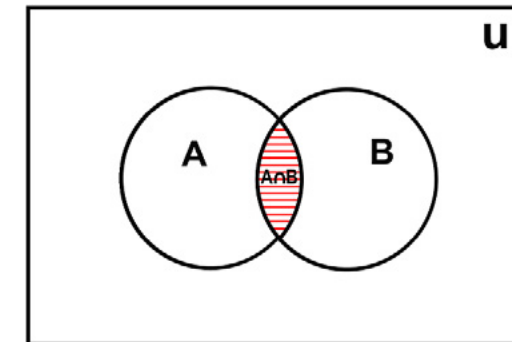


Figura 6: Diagrama da operação interseção entre dois conjuntos A e B.

A operação realizada também tem o nome de interseção e o símbolo é do próprio resultado. Em símbolos temos que:

$$A \cap B = \{x \in U \mid x \in A \text{ e } x \in B\}$$

Para três conjuntos A, B e C mutuamente exclusivos, temos o diagrama ilustrado na figura a seguir:

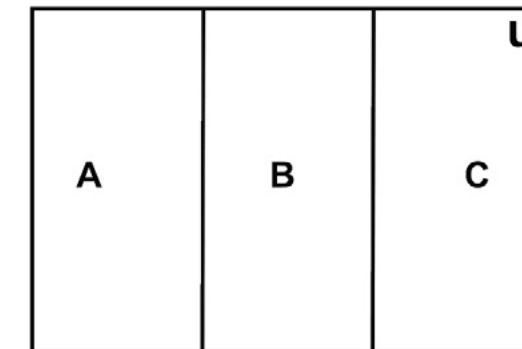


Figura 7: Diagrama de três conjuntos A, B e C mutuamente exclusivos.

#### 8.2.2. Exemplo 1

Sejam os eventos A e B dados por:  $A = \{1, 3, 5, 6\}$  e  $B = \{2, 4, 6\}$ .

Temos que:  $A \cap B = \{6\}$

#### 8.2.3. Exemplo 2

Sejam os conjuntos B e C, então temos que:  $B = \{2, 4, 6\}$  e  $C = \{1, 10, 20\}$ .

**$B \cap C = \emptyset$** , nesse caso B e C são denominados de disjuntos ou incompatíveis. A figura abaixo ilustra a operação entre os conjuntos B e C.

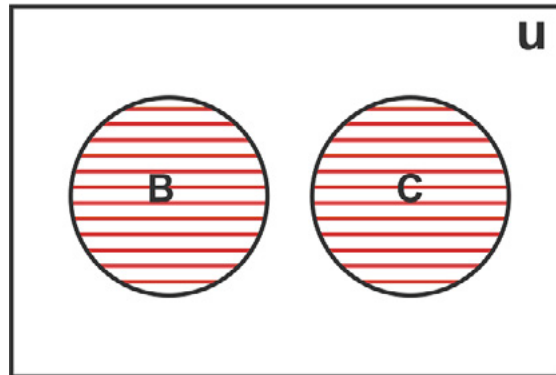


Figura 8: Diagrama de dois conjuntos B e C disjuntos, incompatíveis ou mutuamente exclusivos.

### 8.2.4. Exemplo 3

Sejam os conjuntos D e E, então temos que:  $D = \{20, 1, 10\}$  e  $E = \{1, 20\}$

Então  $D \cap E = E$ , ou seja, se  $E \subset D$ , segue-se que  $D \cap E = E$ , veja esse caso na figura a seguir:

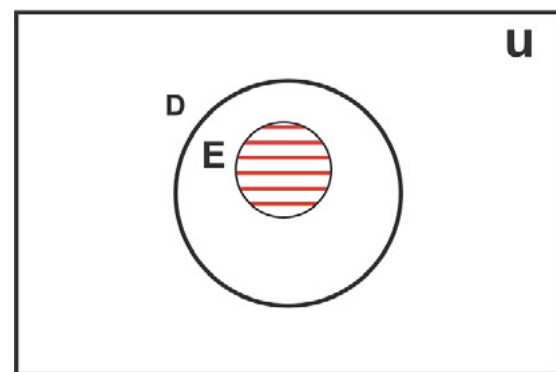


Figura 9: Diagrama entre um conjunto D e um subconjunto E.

### 8.2.5. Exemplo 4

Sejam os conjuntos B e o conjunto vazio  $\emptyset$  então temos que:

Se B e dado por  $B = \{2, 4, 6\}$  e  $\emptyset$ .

Então temos que:

**$\emptyset \cap B = \emptyset$** , isso vale para quaisquer conjuntos, pois se o conjunto vazio  $\emptyset$  não possui elementos então não existe, portanto elementos comuns entre ele e qualquer outro conjunto considerado.

## 8.3. Complementar [ $A^c = \bar{A}$ ] ou complemento absoluto

### 8.3.1. INTRODUÇÃO

Dado o conjunto A contido no conjunto universo U, chama-se complementar de A ao conjunto dos elementos de U que não pertencem a A. Indica-se por

$$A^c = \bar{A} = \tilde{A} = A' = \sim A = CA.$$

A operação realizada denomina-se complementação. A região hachurada na figura abaixo representa o complementar de A.

Em símbolos essa operação é dada por:

$$\bar{A} = A^c = \{x | x \in U \text{ e } x \notin A\}$$

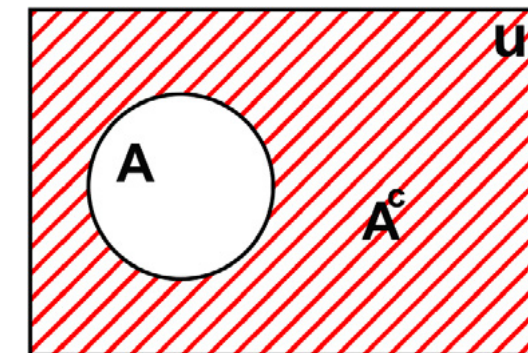


Figura 10: Diagrama da operação complementar de um conjunto A.

### 8.3.2. Exemplo 1

Sejam os conjuntos universo U e A, dados por:

$$U = \{1, 2, 3, 4, 5, 6, 10, 20\}$$

$$A = \{1, 3, 5, 6\}$$

$$\text{Então tem-se que } A^c = \bar{A} = \{2, 4, 10, 20\}$$

### 8.3.3. Exemplo 2

Sejam os seguintes conjuntos universo:

$$U = N = \{0, 1, 2, 3, 4, 5, 6, \dots\} \text{ e } A = N^* = \{1, 2, 3, 4, 5, 6, \dots\}, \text{ então } A^c = \bar{A} = \{0\}$$

### 8.3.4. Exemplo 3

Sejam os conjuntos universo U e os conjuntos A e complementar de  $A^c = \bar{A}$ , então temos que:

$$i) (A \cap A^c) = \emptyset \text{ e,}$$

$$ii) (A \cup A^c) = U$$

## 8.4. Diferença ( $A \setminus B$ ) ou complemento relativo

### 8.4.1. Introdução

Chama-se diferença A – B ao conjunto dos elementos do conjunto Universo U que pertencem a A e não pertencem a B, veja a ilustração dessa operação nas duas figuras a seguir:  $A \setminus B$  e  $B \setminus A$  sendo iguais é denominado diferença simétrica.

Simbolicamente tem-se que:

$$A - B = \{x \in U | x \in A \text{ e } x \notin B\}$$



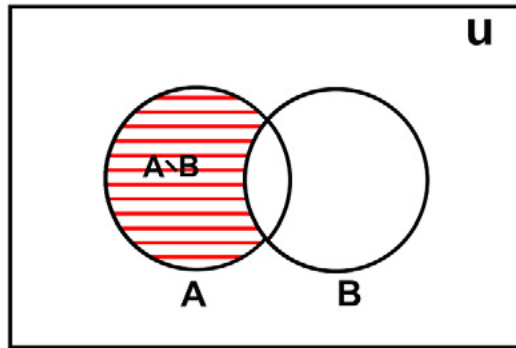


Figura 11: Diagrama da operação diferença entre dois conjuntos A e B.

Por outro lado se for considerado o conjunto  $(B - A)$  tem-se que:

$$B - A = \{x \in U \mid x \in B \text{ e } x \notin A\}, \text{ como ilustra a figura a seguir:}$$

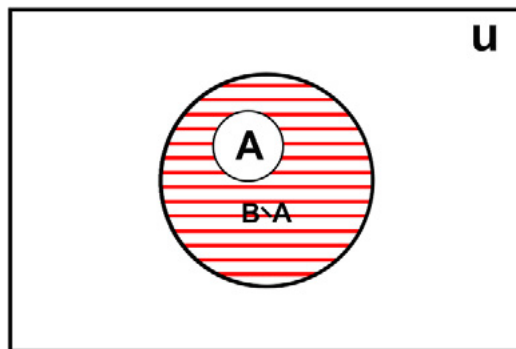


Figura 12: Diagrama da operação diferença entre um conjunto B e um subconjunto A.

Vale salientar que:  $(A \setminus B) \cap B = \emptyset$  OU  $(A \cap B^c) \cap B = \emptyset$ .

### 8.4.2. Exemplo 1

Sejam os conjuntos A e B, então tem-se que:

$$A = \{1, 3, 5, 6\}$$

$$B = \{2, 4, 6\}$$

$$A - B = A \setminus B = \{1, 3, 5\}$$

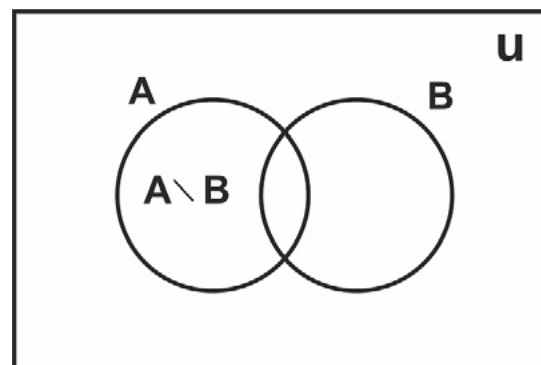


Figura 13: Diagrama da operação diferença entre dois conjuntos A e B.

### 8.4.3. Exemplo 2

Sejam os conjuntos H e L, então tem-se que:

$$U = \{1, 2, 3, 4, 5, 6, 10, 20\};$$

$$H = \{0, 2, 4, 6, \dots\}; L = \{1, 3, 5, 7, \dots\} \text{ e } H - L = H$$

### 8.4.4. Exemplo 3

Sejam os conjuntos F e G, então tem-se que:

$$U = \{1, 2, 3, 4, 5, 6, 10, 20\}$$

$$F = \{1, 2, 3, 4, 5\}$$

$$G = \{1, 2, 3\}$$

$$G - F = \emptyset$$

### 8.4.5. Exemplo 4

Considere os conjuntos Universo U e dois conjuntos quaisquer A e B, então tem-se que:

$$i) A^c = \bar{A} = (U - A)$$

$$ii) A - B = (A \cap B^c)$$

## 9. Produto cartesiano [AXB]: (conjuntos produto)

Dados dois conjuntos A e B, chama-se produto cartesiano de A por B, simbolizado por AXB, ao conjunto dos pares ordenados cujos primeiros elementos pertencem a A e cujos segundos elementos pertencem a B, isto é,

$$AXB = \{(x, y) \mid x \in A \text{ e } y \in B\},$$

O produto cartesiano de um conjunto por ele mesmo, por exemplo, AXB, é representado por  $A^2$ .

Verifique que, se  $n(A)$  indica o número de elementos de A, e  $n(B)$  indica o número de elementos de B e  $n(A \times B)$  é o número de elementos de AXB. Então, tem-se que:

$$n(a \times b) = n(A) \cdot n(B)$$

O conceito de produto cartesiano é estendido para um número finito de conjuntos de forma natural. O produto cartesiano de  $A_1, A_2, \dots, A_m$ , escrito por  $A_1 \times A_2 \times \dots \times A_m$ , é o conjunto de todas as m-úplas ordenadas  $\{a_1, a_2, \dots, a_m\}$ , onde  $a_i \in A_i$  para todo i.

Deve-se salientar que  $AXB \neq BXA$ .

### 9.1. Exemplo

Por exemplo, seja os conjuntos  $A = \{1, 2\}$  e  $B = \{3, 4, 5\}$

Podemos formar um novo conjunto de pares ordenados, cujos primeiros elementos pertencem ao conjunto A e cujos segundos elementos pertencem a B, isto é,  $\{(1, 3), (1, 4), (1, 5), (2, 3), (2, 4), (2, 5)\}$

Este conjunto é chamado o produto cartesiano de A por B e indica-se  $A \times B$ .

Deve-se verificar neste exemplo que  $A \times B \neq B \times A$ , pois  $B \times A = \{(3, 1), (3, 2), (4, 1), (4, 2), (5, 1), (5, 2)\}$ , e nenhum elemento de  $B \times A$  pertence a  $A \times B$ , já que, por exemplo, o par  $(1,3) \neq (3, 1)$ .

### 10. Exercícios de aplicação

#### 10.1. Exemplo 1

Sejam os conjuntos A, B e C não mutuamente exclusivos e tais que existem  $A \cap B$ ,  $A \cap C$  e  $B \cap C$ , definidas no conjunto universo “U”. Estabeleça as expressões das operações entre conjuntos e indique o diagrama de Venn para o conjunto em que.

##### 10.1.1. “A” E “B” OCORREM, MAS “C” NÃO

$$A \cap B \cap C^c$$

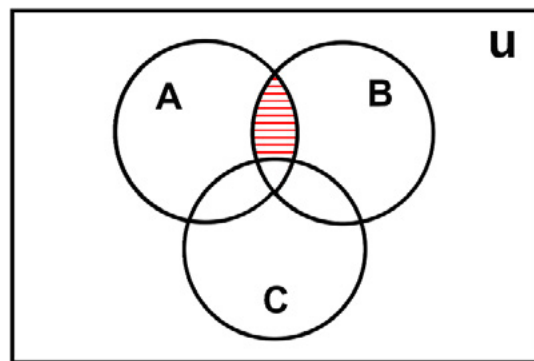


Figura 14: Diagrama mostrando a operação  $A \cap B \cap C^c$  entre três conjuntos A, B e C.

Ou seja, A e B ocorrem, mas C não.

##### 10.1.2. Somente “A” ocorre

$$A \cap B^c \cap C^c$$

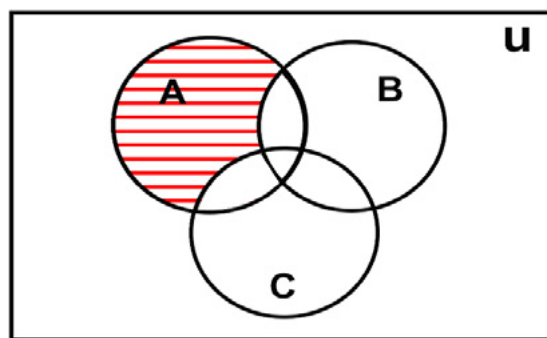


Figura 15: Diagrama da operação  $A \cap B^c \cap C^c$  entre três conjuntos A, B e C.

Isto é somente A ocorre.

#### 10.2. Exemplo 2

Sejam três conjuntos mutuamente exclusivos, Estabeleça as expressões das operações entre conjuntos e indique o diagrama de Venn para o conjunto em que.

A ocorre ou B ocorre ou C ocorre.

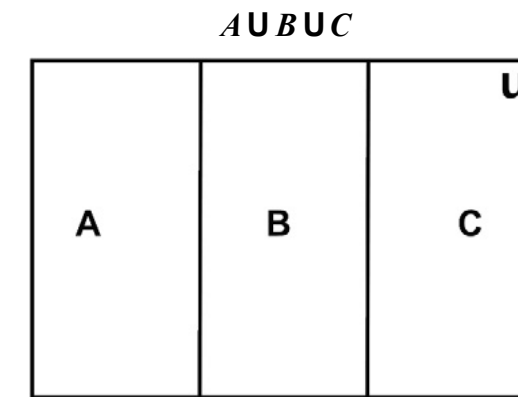


Figura 16: Diagrama da operação  $A \cup B \cup C$  entre três conjuntos A, B e C.

10.3. Sejam os conjuntos A e B. Encontre as expressões das operações entre conjuntos e indique o diagrama de Venn para o conjunto em que:

10.3.1. Apenas A ou apenas B ocorre, isto é, ocorre exatamente um dos dois conjuntos

$$(A \cap B^c) \cup (B \cap A^c)$$

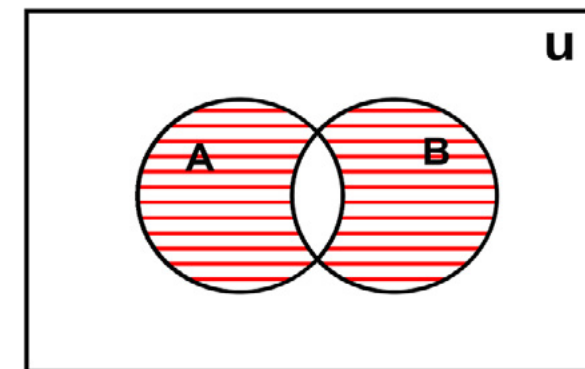


Figura 17: Diagrama da operação  $(A \cap B^c) \cup (B \cap A^c)$  entre dois conjuntos A e B.

Ou seja, Apenas A ou apenas B ocorre, mas não ambos. Ou ainda A ocorrer, mas B não, ou B ocorrer, mas A não.

**10.3.2. A ocorre, mas B não, isto é, somente A ocorre**

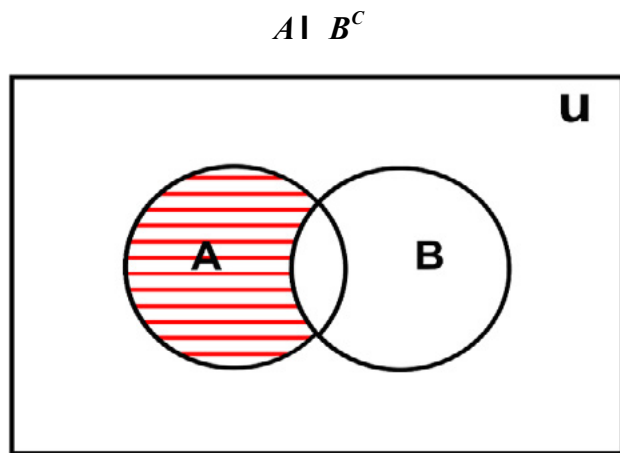


Figura 18: Diagrama da operação  $A \cap B^c$  entre dois conjuntos A e B.

**Técnicas de contagem ou análise combinatória**

**1. Notação fatorial**

O produto dos inteiros positivos de 1 a n, inclusive, aparece freqüentemente em matemática e, por isso, é representado pelo símbolo especial n! (lê-se “n” fatorial):

$n! = 1.2.3.....(n-2) (n-1) n$ . Exemplo:  $4! = 4.3.2.1 = 24$

**1.1. Permutação (arranjo):  $n P_r = \frac{n!}{(n-r)!}$**

Um arranjo de um conjunto de n objetos, em dada ordem, é chamado de permutação dos objetos (tomados todos ao mesmo tempo). Um arranjo de quaisquer  $r \leq n$  destes objetos, em dada ordem, é chamado de r-permutação ou permutação dos n objetos tomados r a r.

**Exemplo:** Encontre o número de permutações de 6 objetos, a, b, c, d, e e f, tomadas 3 a 3. Em outras palavras, encontre o número de “palavras de 3 letras”, com letras distintas, que podem ser formadas com as 6 letras acima. Pelo princípio fundamental de contagem, existem  $6.5.4 = 120$  palavras de três letras distintas, formadas com as 6 letras dadas; ou existem 120 permutações de objetos tomadas 3 a 3. isto é,  $P(6,3) = 120$ .

**1.2. Combinação:  $C_r^n = \frac{n!}{r!(n-r)!}$**

Suponha que temos uma coleção de n objetos. Uma combinação destes n objetos tomados r a r, ou uma r-combinação, é qualquer subconjunto de r elementos. Em outras palavras, uma r-combinação é qualquer seleção de r dos n objetos, sem considerar a ordem.

**Exemplo:** de um grupo de 8 caprinos da raça Caniné, quantos grupos de animais podem ser formados com 3 deles? Cada grupo é essencialmente uma combinação dos 8 animais, tomados 3 a 3. Assim, podem ser formados

$C(8,3) = \binom{8}{3} = \frac{8.7.6}{1.2.3} = 56$  grupos diferentes.

**1.3. Noções gerais sobre análise combinatória (ou técnicas de contagem)**

A análise combinatória surge então, como um método auxiliar na contagem do número de casos favoráveis e do número de casos possíveis de eventos aleatórios em determinado espaço amostral associado a um particular experimento ou ensaio aleatório.

**1.3.1. Permutações**

$P_n = n!$  dá o número de permutações de n elementos.

**1.3.1.1. Exemplo**

Seja um grupo de 4 elementos, então tem-se as seguintes permutações que são:

$P_4 = 4! (= 24)$ , será, por exemplo, o cardinal do conjunto (permutações de 4 elementos: 1, 2, 3, 4):

- {1, 2, 3, 4}, {1, 2, 4, 3}, {1, 3, 2, 4}, {1, 3, 4, 2}, {1, 4, 2, 3}, {1, 4, 3, 2},
- {2, 3, 1, 4}, {2, 3, 4, 1}, {2, 4, 1, 3}, {2, 4, 3, 1}, {2, 1, 3, 4}, {2, 1, 4, 3},
- {3, 4, 1, 2}, {3, 4, 2, 1}, {3, 1, 2, 4}, {3, 1, 4, 2}, {3, 2, 1, 4}, {3, 2, 4, 1},
- {4, 1, 2, 3}, {4, 1, 3, 2}, {4, 2, 1, 3}, {4, 2, 3, 1}, {4, 3, 1, 2}, {4, 3, 2, 1}

**1.3.2. Arranjos**

$A_k^n = \frac{n!}{(n-k)!}, (0 \leq k \leq n)$

O qual fornece o número de n elementos, k a k.

**1.3.2.1. Algumas propriedades**

- i)  $A_0^n = 1$
- ii)  $A_1^n = n$
- iii)  $A_n^n = A_{n-1}^n = P_n = n!$
- iv)  $A_k^n = 0$  se  $k > n$

**1.3.2.2. Exemplos**

Para os exemplos a seguir considerem-se os 4 elementos: 1, 2, 3, 4.

- i)  $A_1^4 = \frac{4!}{3!} = 4$ , é por exemplo, o cardinal do conjunto  $\{\{1\}, \{2\}, \{3\}, \{4\}\}$
- ii)  $A_2^4 = \frac{4!}{2!} = 4 \times 3 = 12$ , é por exemplo, o cardinal do conjunto

{1, 2}, {1, 3}, {1, 4},

{2, 1}, {2, 3}, {2, 4},

{3, 1}, {3, 2}, {3, 4},

{4, 1}, {4, 2}, {4, 3}

iii)  $A_3^4 = \frac{4!}{1!} = 4 \times 3 \times 2 = 24$ , é por exemplo, o cardinal do conjunto

{1, 2, 3}, {1, 3, 2}, {1, 2, 4}, {1, 4, 2}, {1, 3, 4}, {1, 4, 3},

{2, 3, 4}, {2, 4, 3}, {2, 1, 3}, {2, 3, 1}, {2, 1, 4}, {2, 4, 1},

{3, 1, 2}, {3, 2, 1}, {3, 1, 4}, {3, 4, 1}, {3, 2, 4}, {3, 4, 2},

{4, 1, 2}, {4, 2, 1}, {4, 1, 3}, {4, 3, 1}, {4, 2, 3}, {4, 3, 2}

iv)  $A_4^4 = P_4 = 4! = 24$

Verifique anteriormente no item referente a permutações.

### 1.3.3. Combinações

$$C_k^n = \binom{n}{k} = \frac{n!}{k!(n-k)!}, (0 \leq k \leq n)$$

Fornece o número de combinações de n elementos, k a k.

Numa definição sem grande rigor diremos que são como arranjos saem repetições, considerando-se que, por exemplo, uma vez incluída no conjunto das combinações a combinação {1, 2, 3}, então as combinações {1, 3, 2}, {3, 2, 1}, {3, 1, 2}, {2, 1, 3} e {2, 3, 1} são repetições da combinação inicial e portanto não entram para o cálculo do número de combinações.  $C_k^n$  dá-nos, por exemplo, o número de diferentes conjuntos de k pessoas que se podem formar a partir de um conjunto de n ( $n \geq k$ ) pessoas; ou o número de diferentes formas em que um conjunto de n pessoas podem segurar  $k \leq n$  bolas (dado que cada pessoa segura ou não uma bola e cada pessoa não pode segurar mais de uma bola).

#### 1.3.3.1. Algumas propriedades

i)  $C_k^n = C_{n-k}^n$

ii)  $C_0^n = C_n^n = 1$

iii)  $C_1^n = A_1^n = n$

iv)  $C_k^n = 0$  se  $k > n$

#### 1.3.3.2. Exemplos

Para os exemplos a seguir considerem-se os 4 elementos: 1, 2, 3, 4

i)  $C_1^4 = \frac{4!}{1!3!} = 4 (= A_1^4)$

é, por exemplo, o cardinal do conjunto {{1}, {2}, {3}, {4}}.

ii)  $C_2^4 = \frac{4!}{2!2!} = \frac{4 \times 3}{2} = 6$

é, por exemplo, o cardinal do conjunto.

{{1, 2}, {1, 3}, {1, 4},

{2, 1}, {2, 3}, {2, 4},

{3, 1}, {3, 2}, {3, 4},

{4, 1}, {4, 2}, {4, 3}}

iii)  $C_3^4 = \frac{4!}{3!1!} = 4$

é, por exemplo, o cardinal do conjunto

{{1, 2, 3}, {1, 3, 2}, {1, 2, 4}, {1, 4, 2}, {1, 3, 4}, {1, 4, 3},

{2, 3, 4}, {2, 4, 3}, {2, 1, 3}, {2, 3, 1}, {2, 1, 4}, {2, 4, 1},

{3, 1, 2}, {3, 2, 1}, {3, 1, 4}, {3, 4, 1}, {3, 2, 4}, {3, 4, 2},

{4, 1, 2}, {4, 2, 1}, {4, 1, 3}, {4, 3, 1}, {4, 2, 3}, {4, 3, 2}}

iv)  $C_4^4 = \frac{4!}{4!0!} = 1$

é, por exemplo, o cardinal do conjunto {{1, 2, 3, 4}} cujo único elemento é o conjunto dado por {1, 2, 3, 4}.

v)  $C_0^4 = \frac{4!}{4!0!} = 1$

é o cardinal do conjunto cujo único elemento é o conjunto vazio.

vi) Note que  $C_0^4 + C_1^4 + C_2^4 + C_3^4 + C_4^4 = 16 = 2^4$ , o que não é uma coincidência; veja os resultados a seguir:

#### 1.3.3.3. Alguns resultados envolvendo combinações

i)  $\sum_{k=0}^{\min(n,N)} \binom{N}{k} \binom{M}{n-k} = \binom{N+M}{n}$

ii)  $(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}, \forall n \in N_0$  (Teorema binomial)

iii)  $\sum_{k=0}^n \binom{n}{k} = 2^n$  (basta fazer na igualdade acima, a = b = 1).



iv) Dicotomias:  $2^n = \sum_{k=0}^n \binom{n}{k}$  é o número de dicotomias de n elementos, i.e. o número (total) de com-

binhações de n elementos k e k, para k = 0, 1,..., n. Note-se que cada elemento pode estar ou não estar presente em cada dicotomia, independentemente de qualquer outro elemento estar ou não presente, o que gera de fato um total de 2<sup>n</sup> diferentes possibilidades de combinações.

**1.3.3.4. Exemplos**

Porque 2<sup>n</sup> é o número de subconjuntos de um conjunto com n elementos?

Porque 2<sup>k</sup> é o número de maneiras diferentes de, partindo de um estado k + 1, chegar ao estado 0, passando ou não por cada um dos k estados intermédios?

Porque 2<sup>k</sup> é o número de diferentes números (coleções de dígitos) com k dígitos que se podem construir com os dígitos de 0 a k, cuja soma dos dígitos é igual a k e onde cada dígito d maior que zero é sempre antecedido de d – 1 zeros?

**Propriedades da Esperança Matemática [E(X)], da Variância [Var(X)], da Covariância [Cov(X,Y)] e do Coeficiente de Correlação [r] de variáveis aleatórias discretas e contínuas.**

1. Propriedades da esperança matemática

- i)  $E(K) = K$ , sendo K uma constante
- ii)  $E(K \cdot X) = K \cdot E(X)$ , sendo K uma constante
- iii)  $E(X \pm K) = E(X) \pm K$
- iv)  $E(X \pm Y) = E(X) \pm E(Y)$
- v)  $E(XY) = E(X) E(Y)$  Se X e Y são independentes
- vi)  $E[(X - \mu_{(x)})] = E(X) - E[\mu_{(x)}] = \mu_{(x)} - \mu_{(x)} = 0$ , isto é a média ou esperança de uma variável aleatória centrada na média é sempre nula.

a) Dedução das propriedades da esperança matemática ou média de uma variável aleatória discreta ou contínua:  $\{E(X) = \mu_{(x)} = \mu_x = \mu = m\}$

i) A média de uma constante é igual a própria constante.

Sendo K uma constante, então:  $E(K) = K$ , pois temos que:

$$E(K) = [(K \cdot P_1) + (K \cdot P_2) + \dots + (K \cdot P_n)] = K \cdot (P_1 + P_2 + \dots + P_n) = (K) \cdot (1), \text{ pois,}$$

$$(P_1 + P_2 + \dots + P_n) = \sum_{i=1}^n P(X_i) = 1,$$

ii) Se for multiplicado os valores de uma variável aleatória por um valor constante e arbitrário, a média fica multiplicada por essa constante

$$E(K \cdot X) = K \cdot E(X)$$

ii.1) Para o caso de variáveis aleatórias discretas tem-se que:

$$E(K \cdot X) = \sum_{i=1}^n K \cdot X_i \cdot P_i = K \cdot \sum_{i=1}^n X_i \cdot P_i = K \cdot E(X)$$

ii.2) Para o caso de variáveis aleatórias contínuas tem-se que:

$$E(K \cdot X) = \int_{-\infty}^{+\infty} K \cdot X \cdot f(X) \cdot dX = K \cdot \int_{-\infty}^{+\infty} X \cdot f(X) \cdot dX = K \cdot E(X)$$

iii) Se for somado ou subtraído um valor constante e arbitrário aos valores de uma variável aleatória, a média fica somada ou diminuída dessa mesma constante

$$E(X \pm K) = E(X) \pm K$$

Demonstraremos o caso de E(X+K), mas ressaltando que vale para E(X-K)

$$E(X+K) = \sum_{i=1}^n (X_i + K) P_i = \sum_{i=1}^n X_i \cdot P_i + K \sum_{i=1}^n P_i = E(X) + K, \text{ pois } \sum_{i=1}^n P_i = 1$$

iv) A média de uma soma ou diferença de variáveis aleatórias é igual à soma ou diferença das médias dessas variáveis

$$E(X \pm Y) = E(X) \pm E(Y)$$

iv.1) Para o caso de variáveis aleatórias discretas tem-se que:

Suponha que as variáveis X e Y são definidas para o conjunto de valores X<sub>1</sub>, X<sub>2</sub>,...,X<sub>m</sub>, e Y<sub>1</sub>, Y<sub>2</sub>, ...,Y<sub>n</sub>. Indicaremos por P(X<sub>i</sub>,Y<sub>j</sub>) A probabilidade de ocorrerem simultaneamente X<sub>i</sub> e Y<sub>j</sub> (i=1,2,...,m ; j=1,2,...,n). Assim P(X<sub>1</sub>Y<sub>2</sub>) indica a probabilidade de ocorrer X<sub>1</sub> com Y<sub>2</sub>, P(X<sub>2</sub>Y<sub>1</sub>) a probabilidade de ocorrer X<sub>2</sub> com Y<sub>1</sub>, etc.

Pela definição de esperança matemática, pode-se escrever:

$$E(X+Y) = (X_1 + Y_1) \cdot P(X_1, Y_1) + (X_1 + Y_2) \cdot P(X_1, Y_2) + \dots + (X_1 + Y_n) \cdot P(X_1, Y_n) + (X_2 + Y_1) \cdot P(X_2, Y_1) + (X_2 + Y_2) \cdot P(X_2, Y_2) + \dots + (X_2 + Y_n) \cdot P(X_2, Y_n) + \dots + (X_m + Y_1) \cdot P(X_m, Y_1) + (X_m + Y_2) \cdot P(X_m, Y_2) + \dots + (X_m + Y_n) \cdot P(X_m, Y_n)$$

Esta soma pode ser colocada sob a seguinte forma:

$$E(X+Y) = X_1 [P(X_1, Y_1) + P(X_1, Y_2) + \dots + P(X_1, Y_n)] + X_2 [P(X_2, Y_1) + P(X_2, Y_2) + \dots + P(X_2, Y_n)] + \dots + X_m [P(X_m, Y_1) + P(X_m, Y_2) + \dots + P(X_m, Y_n)] + Y_1 [P(X_1, Y_1) + P(X_2, Y_1) + \dots + P(X_m, Y_1)] + Y_2 [P(X_1, Y_2) + P(X_2, Y_2) + \dots + P(X_m, Y_2)] + \dots + Y_n [P(X_1, Y_n) + P(X_2, Y_n)] + \dots + P(X_m, Y_n]$$

Mas pelo teorema da soma de probabilidades tem-se:

$$P_1 = P(X_1) = P(X_1, Y_1) + P(X_1, Y_2) + \dots + P(X_1, Y_n)$$

$$P_2 = P(X_2) = P(X_2, Y_1) + P(X_2, Y_2) + \dots + P(X_2, Y_n)$$

.....

$$P_m = P(X_m) = P(X_m, Y_1) + P(X_m, Y_2) + \dots + P(X_m, Y_n)$$

$$q_1 = P(Y_1) = P(X_1, Y_1) + P(X_2, Y_1) + \dots + P(X_m, Y_1)$$

$$q_2 = P(Y_2) = P(X_1, Y_2) + P(X_2, Y_2) + \dots + P(X_m, Y_2)$$

.....

$$q_n = P(Y_n) = P(X_1, Y_n) + P(X_2, Y_n) + \dots + P(X_m, Y_n)$$

Sendo assim tem-se que:

$$E(X+Y) = X_1 P(X_1) + X_2 P(X_2) + \dots + X_m P(X_m) + Y_1 P(Y_1) + Y_2 P(Y_2) + \dots + Y_n P(Y_n)$$

$$E(X+Y) = E(X) + E(Y)$$

iv.2) Para o caso de variáveis aleatórias contínuas essas propriedades são válidas de forma semelhante.

v) A média do produto de duas variáveis aleatórias independentes é igual ao produto das médias dessas variáveis.

Sendo assim tem-se

$$E(X.Y) = E(X) . E(Y)$$

v.1) Para o caso de variáveis aleatórias discretas tem-se que:

Para o caso do produto de duas variáveis aleatórias, considere as variáveis X e Y, definidas para os conjuntos de valores a seguir:

$$X_1, X_2, \dots, X_m, \text{ e } Y_1, Y_2, \dots, Y_n.$$

Sendo assim a esperança matemática do produto de duas variáveis aleatórias é definida por.

$$E(X.Y) = X_1 Y_1 P(X_1, Y_1) + X_1 Y_2 P(X_1, Y_2) + \dots + X_1 Y_n P(X_1, Y_n) + X_2 Y_1 P(X_2, Y_1) + X_2 Y_2 P(X_2, Y_2) + \dots + X_2 Y_n P(X_2, Y_n) + \dots + X_m Y_1 P(X_m, Y_1) + X_m Y_2 P(X_m, Y_2) + \dots + X_m Y_n P(X_m, Y_n)$$

Onde  $P(X_i, Y_j)$  é a probabilidade de ocorrerem as duas variáveis simultaneamente  $X_i$  e  $Y_j$  ( $i=1,2,\dots,m$ ;  $j=1,2,\dots,n$ ).

Nesse caso temos que impor a condição de que os eventos sejam independentes. Para isto deve-se ter  $P(X_i, Y_j) = P(X_i).P(Y_j)$ , onde  $P(X_i)$  e  $P(Y_j)$  são as probabilidades individuais ou marginais de ocorrerem um particular valor de X ou de Y.

Respectivamente, obtém-se assim:

$$E(XY) = X_1 Y_1 P(X_1)P(Y_1) + X_1 Y_2 P(X_1)P(Y_2) + \dots + X_1 Y_n P(X_1)P(Y_n) + X_2 Y_1 P(X_2)P(Y_1) + X_2 Y_2 P(X_2)P(Y_2) + \dots + X_2 Y_n P(X_2)P(Y_n) + \dots + \dots + X_m Y_1 P(X_m)P(Y_1) + X_m Y_2 P(X_m)P(Y_2) + \dots + X_m Y_n P(X_m)P(Y_n)$$

Esta expressão pode ser colocada na seguinte forma.

$$E(XY) = X_1 P(X_1)[(Y_1)P(Y_1) + Y_2 P(Y_2) + \dots + Y_n P(Y_n)] + X_2 P(X_2)[P(Y_1) + Y_2 P(Y_2) + \dots + Y_n P(Y_n)] + \dots + \dots + X_m P(X_m)[Y_1 P(Y_1) + Y_2 P(Y_2) + \dots + Y_n P(Y_n)]$$

$$E(XY) = [X_1 P(X_1) + X_2 P(X_2) + \dots + X_m P(X_m)][Y_1 P(Y_1) + Y_2 P(Y_2) + \dots + Y_n P(Y_n)]$$

$$E(XY) = E(X).E(Y)$$

v.2) Para o caso de variáveis aleatórias contínuas essas propriedades são válidas de forma semelhante:

vi) A média de uma variável aleatória centrada ou do desvio afastamento ou erro é zero ou nula, isto é,  $E(\text{erro}) = E[(X - \mu)] = 0$ ,

vi.1) Para o caso de variáveis aleatórias discretas tem-se que:

$$E(e) = E[(X - \mu)] = (X_1 - \mu) . P_1 + (X_2 - \mu) . P_2 + \dots + (X_n - \mu) . P_n$$

$$E(e) = E[(X - \mu)] = [(X_1 . P_1) + (X_2 . P_2) + \dots + (X_n . P_n)] - \mu . (P_1 + P_2 + \dots + P_n), \text{ e como,}$$

$$[(X_1 . P_1) + (X_2 . P_2) + \dots + (X_n . P_n)] = E(X) = \mu, \text{ e } (P_1 + P_2 + \dots + P_n) = 1,$$

Tem-se finalmente que:  $E(e) = \mu - \mu = 0$

vi.2) Para o caso das variáveis aleatórias contínuas o procedimento é semelhante:

Então fica assim:

$$E(e) = \int_{-\infty}^{+\infty} (X - \mu) . dP = \int_{-\infty}^{+\infty} X . dP - \mu \int_{-\infty}^{+\infty} dP$$

$$\text{E como } \int_{-\infty}^{+\infty} X . dP = \mu, \int_{-\infty}^{+\infty} dP = 1,$$

Obtem-se finalmente que:

$$E(e) = \mu - \mu = 0, \text{ como no caso anterior.}$$

vii)  $E(e_1) . E(e_2) = 0$

Sendo  $X_1$  e  $X_2$  duas variáveis casuais independentes, de médias  $\mu_1$  e  $\mu_2$ , os desvios correspondentes são:  $e_1 = X_1 - \mu_1$ , e  $e_2 = X_2 - \mu_2$ , e já sabe-se que

$$E(e_1) = E(X_1) - E(\mu_1) = \mu_1 - \mu_1 = 0, \text{ e que } e_2 = E(X_2) - E(\mu_2) = \mu_2 - \mu_2 = 0.$$

Demonstra-se que  $E(e_1 . e_2) = 0$ ,

Com efeito, pelo teorema da esperança matemática do produto, tem-se:

$$E(e_1 . e_2) = E(e_1) . E(e_2) = (0) . (0) = 0$$

i.5) Funções de variáveis aleatórias

Seja x uma variável aleatória. Então  $Y = g(X)$  é também uma variável aleatória, ou seja o caso agora é de uma variável aleatória que depende de X, por exemplo  $g(X) = X^2$  ou ainda a seguinte função  $g(X) = (X - \mu)^2$  segue-se que:

$$P(Y = y) = \sum_{[x: g(x)=y]} P(X = x)$$

O valor esperado ou esperança matemática de  $g(X)$  também pode ser obtido a partir da função de probabilidade ou de densidade de probabilidade de x.

i.5.1.) Esperança matemática ou média para a função de variável aleatória  $g(X)$  de uma variável aleatória discreta  $X$ .  $\{E[g(X)]\}$ .

Seja  $X$  uma variável aleatória discreta com função de probabilidade  $f(X) = P(X_i)$ , então

$$E[g(X)] = g(x_1)f(x_1) + g(x_2)f(x_2) + \dots + g(x_n)f(x_n)$$

$$E[g(X)] = \sum_{i=1}^n g(x_i)f(x_i)$$

i.5.2.) Esperança matemática ou média para a função de variável aleatória  $g(X)$  de uma variável aleatória contínua  $X$ .  $\{E[g(X)]\}$ .

Se  $x$  é uma variável aleatória contínua com função densidade de probabilidade  $f(X)$ , então.

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x)f(x)dx$$

**2. Propriedades da variância**  $[Var(X), V(X), \sigma_{(X)}^2, \sigma_X^2, \sigma^2]$

- i)  $V(K) = 0$ , sendo  $K$  uma constante
- ii)  $V(KX) = K^2V(X)$ , sendo  $K$  uma constante
- iii)  $V(X \pm K) = V(X)$ , sendo  $K$  uma constante
- iv)  $V(X \pm Y) = V(X) + V(Y)$ , se  $X, Y$  são variáveis aleatórias independentes.
- v)  $V(X \pm Y) = V(X) + V(Y) \pm 2.COV. (X, Y)$ , se  $X, Y$  são variáveis aleatórias dependentes discretas ou contínuas.

**2.1. Dedução das propriedades da variância:**  $[V(X)] = Var(X) = \sigma_{(X)}^2 = \sigma_X^2 = \sigma^2$

i) A variância de uma constante é nula

$$Var(K) = 0, Var(K) = E\{[K-E(K)]^2\} = E\{[K-K]^2\} = E(K^2) - [E(K)]^2$$

$$A E(K^2) = \sum_{i=1}^n K^2 \cdot P(K) = K^2 \cdot \sum_{i=1}^n P(K) = K^2 \cdot 1 = K^2, e$$

$$[E(K)] = \sum_{i=1}^n K \cdot P(K) = K \cdot \sum_{i=1}^n P(K) = K \cdot 1 = K, \text{ então a } Var(K) = K^2 - [K]^2 = K^2 - K^2 = 0$$

ii) Se for multiplicado todos os valores de uma variável aleatória por uma constante arbitrária  $K$ , sua variância fica multiplicada pelo quadrado dessa constante

$$Var(K \cdot X) = K^2 \cdot Var(X), \text{ com efeito tem-se } Var(K \cdot X) = E\{[(KX-E(X))]^2\}, \text{ onde:}$$

$$E(KX) = KE(X) = K \cdot \mu, \text{ portanto:}$$

$$Var(K \cdot X) = E\{[K \cdot X - E(K \cdot X)]^2\} = E\{[K \cdot X - K \cdot E(X)]^2\} = E\{K[X - E(X)]^2\} =$$

$$Var(K \cdot X) = E\{K^2[X - E(X)]^2\} = K^2 \frac{E[X - E(X)]^2}{Var(X)}$$

$$Var(K \cdot X) = K^2 \cdot Var(X)$$

iii) Se for somado ou subtraído uma constante arbitrária aos valores de uma variável aleatória, sua variância permanece inalterada

$$Var(X_i \pm K) = Var(X)$$

$$Var(X_i + K) = Var(X_i) + Var(K)$$

$Var(X_i + K) = Var(X_i) + 0$ , como a  $Var(X_i) = Var(X)$ , temos que:

$$Var(X_i + K) = Var(X)$$

iv) A variância de uma soma ou diferença de variáveis aleatórias independentes é igual à soma das variâncias dessas variáveis

$Var(X + Y) = Var(X) + Var(Y)$ , se  $X$  e  $Y$  são variáveis casuais independentes mostraremos que  $Var(X+Y) = Var(X) + Var(Y)$ , de fato:

$Var(X+Y) = E[(X+Y - \mu)]^2$ , onde  $\mu = E[X + Y] = E(X) + E(Y) = \mu_1 + \mu_2$ , onde:

$\mu_1 = E(X)$  e  $\mu_2 = E(Y)$ . Temos, pois, com  $e_1 = X - \mu_1$  e  $e_2 = Y - \mu_2$

$$Var(X + Y) = E[(X+Y-\mu_1-\mu_2)^2]$$

$$Var(X + Y) = E[(e_1 + e_2)^2]$$

$$Var(X + Y) = E[(e_1^2 + 2e_1e_2 + e_2^2)] =$$

$$Var(X + Y) = E(e_1^2) + 2 \cdot E(e_1)E(e_2) + E(e_2^2) =$$

$Var(X + Y) = Var(X) + Var(Y)$ , pois, como se sabe a média ou esperança dos erros é sempre nulas ou zero, ou seja, tem-se sempre que,  $E(e_1) = E(e_2) = 0$

Esta propriedade pode ser facilmente generalizada para o caso da soma de  $n$  variáveis aleatórias independentes. Obte-se então o seguinte:

$$Var(X_1 + X_2 + \dots + X_n) = Var(X_1) + Var(X_2) + \dots + Var(X_n)$$

Já este teorema, por sua vez, pode ser combinado com o anterior, para fornecer a variância de uma função linear de variáveis aleatórias independentes, da seguinte forma:

$$Var(K_1X_1 + K_2X_2 + \dots + K_nX_n) = Var(K_1X_1) + Var(K_2X_2) + \dots + Var(K_nX_n) =$$

$$= K_1^2 Var(X_1) + K_2^2 Var(X_2) + \dots + K_n^2 Var(X_n)$$

Se essas variâncias forem todas iguais, isto é, se tivermos:

$$Var(X_1) = Var(X_2) = \dots = Var(X_n) = \sigma^2, \text{ obteremos o seguinte:}$$

$$Var(K_1X_1) + Var(K_2X_2) + \dots + Var(K_nX_n) = (K_1^2 + K_2^2 + \dots + K_n^2) \cdot \sigma^2 = \sigma^2 \cdot \sum_{i=1}^n K_i^2$$

Como caso particular temos o da variância da média aritmética.

$$Var(\bar{X}) = Var\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) = \sum_{i=1}^n \frac{1}{n^2} \cdot Var(X_i) = \frac{n}{n^2} \cdot \sigma^2 = \frac{\sigma^2}{n}$$

v)  $\text{Var}(X-Y) = \text{Var}(X) + \text{Var}(Y)$ , se X e Y são independentes.

Prova:

$$\text{Var}(X-Y) = \text{Var}(X) + (-1)^2 \text{Var}(Y) = \text{Var}(X) + \text{Var}(Y)$$

vi) Se X e Y não são variáveis aleatórias independentes, então,

$$\text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y) \pm 2 \cdot \frac{E(X \cdot Y) \cdot E(X) \cdot E(Y)}{\text{COVARIÂNCIA}}$$

$$\text{Var}(X+Y) = E\{[(X+Y)-E(X+Y)]^2\} =$$

$$\text{Var}(X+Y) = E\{[(X-E(X))+(Y-E(Y))]^2\} =$$

$$\text{Var}(X+Y) = E\{[X-E(X)]^2 + [Y-E(Y)]^2 + 2 \cdot [X-E(X)] \cdot [Y-E(Y)]\} =$$

$$\text{Var}(X+Y) = E\{[X-E(X)]^2\} + E\{[Y-E(Y)]^2\} + 2 \cdot E\{[X-E(X)] \cdot [Y-E(Y)]\} =$$

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2 \cdot \text{Cov}(X, Y)$$

vii)  $\text{Var}(X_i - \mu) = \frac{(n-1)}{n} \cdot \sigma^2$

Consideremos o desvio, erro, afastamento, discrepância ou resíduo:  $e_1 = (X_1 - \bar{X}_1)$ , em relação à média aritmética. Qual será a sua variância? Então veja.

Tem-se:

$$e_1 = X_1 - \bar{X} = X_1 - \frac{(X_1 + X_2 + \dots + X_n)}{n} = \frac{(n-1)}{n} X_1 - \frac{1}{n} X_2 - \dots - \frac{1}{n} X_n, \text{ logo:}$$

$$\text{Var}(e_1) = \left(\frac{n-1}{n}\right)^2 \cdot \text{Var}(X_1) + \left(\frac{1}{n}\right)^2 \cdot \text{Var}(X_2) + \dots + \left(\frac{1}{n}\right)^2 \cdot \text{Var}(X_n)$$

Se tiver  $\text{Var}(X_i) = \sigma^2$ , para  $i = 1, 2, \dots, n$ , obtem-se:

$$\text{Var}(e_1) = \text{Var}(X_1 - \bar{X}) = \left(\frac{n-1}{n}\right)^2 \cdot \sigma^2 + \left(\frac{n-1}{n^2}\right) \cdot \sigma^2 = \left(\frac{n-1}{n}\right) \cdot \sigma^2$$

Por outro lado, pela definição de variância tem-se que:

$$\begin{aligned} \text{Var}(X_1 - \bar{X}) &= E\left\{\left[(X_1 - \bar{X}) - (E(X_1 - \bar{X}))\right]^2\right\} = E\left(\frac{n-1}{n} X_1 - \frac{1}{n} X_2 - \dots - \frac{1}{n} X_n\right)^2 = \\ &= \left(\frac{n-1}{n}\right) \cdot E(X_1) - \left(\frac{1}{n}\right) \cdot E(X_2) - \dots - \left(\frac{1}{n}\right) \cdot E(X_n) \end{aligned}$$

Suponha que para todas as variáveis aleatórias  $X_i$ , com  $i = 1, 2, \dots, n$ , tem-se  $E(X_i) = \mu$ ,  $\text{Var}(X_i) = \sigma^2$ .

Então fica assim:

$$E(X_1 - \bar{X}) = \left(\frac{n-1}{n}\right) \cdot \mu - \left(\frac{n-1}{n}\right) \cdot \mu = 0,$$

$$\text{Var}(X_1 - \bar{X}) = E\left[(X_1 - \bar{X})^2\right],$$

Logo temos:

$$E\left[(X_1 - \bar{X})^2\right] = \left(\frac{n-1}{n}\right) \cdot \sigma^2$$

Como se sabe a fórmula da variância de um conjunto de valores é dada pela equação a seguir, sendo assim tem-se que:

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1},$$

Pela demonstração acima verifica-se que:

$$E(S^2) = \left(\frac{1}{n-1}\right) \cdot E\left[\sum_{i=1}^n (X_i - \bar{X})^2\right] = \left(\frac{1}{n-1}\right) \cdot \sum_{i=1}^n E[X_i - \bar{X}]^2 = \left(\frac{1}{n-1}\right) \cdot n \cdot \left(\frac{n-1}{n}\right) \cdot \sigma^2 = \sigma^2$$

viii)  $\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) + 2 \cdot \sum_{i < j} \text{Cov}(X_i, X_j)$ , para  $E(X_i^2) < \infty$ , sendo  $1 \leq i \leq n$

ix)  $\text{Var}(a \cdot X \pm b) = a^2 \cdot \text{Var}(X)$ , se a e b são constantes.

Demonstração:

$$\text{Var}(a \cdot X \pm b) = \text{Var}(a \cdot X) + \text{Var}(b) \pm 2 \cdot \text{Cov}(a \cdot X, b)$$

Como  $\text{Cov}(a \cdot X, b) = E\{[a \cdot X - E(a \cdot X)][b - E(b)]\} = 0$ , temos enfim que,

$$\text{Var}(a \cdot X \pm b) = a^2 \cdot \text{Var}(X)$$

### 3. Propriedades da covariância

i)  $\text{COV}(X, Y) = E(XY) - E(X)E(Y)$

ii)  $\text{COV}(X, Y) = \text{COV}(Y, X)$ , logo a covariância é simétrica.

iii) Se  $V(X) = 0$  (ou  $V(Y) = 0$ ) então  $\text{COV}(X, Y) = 0$ .

iv)  $\text{COV}(aX, Y) = a \text{COV}(X, Y)$  para  $a \in \Re$  então a  $\text{COV}(X, Y)$  é bilinear

v)  $\text{COV}(X + Z, Y) = \text{COV}(X, Y) + \text{COV}(Z, Y)$ .

vi)  $\text{COV}(KX, Y) = \text{COV}(X, KY) = K \text{COV}(X, Y)$ , se K for uma constante

vii)  $\text{COV}(K, X) = \text{COV}(X, K) = 0$  se K for uma constante

#### 3.1. Demonstração das Propriedades da Covariância [COV(X)]

i)  $\text{Cov}(X, Y) = E(X \cdot Y) - E(X) \cdot E(Y)$

Demonstração:

$$\text{Cov}(X, Y) = E\{[X - E(X)][Y - E(Y)]\} = E\{X \cdot Y - Y \cdot E(X) - X \cdot E(Y) + E(X) \cdot E(Y)\} =$$

$$\text{Cov}(X, Y) = E(X \cdot Y) - E[Y \cdot E(X)] - E[X \cdot E(Y)] + E[E(X) \cdot E(Y)] =$$

$$\text{Cov}(X, Y) = E(X \cdot Y) - E(X) \cdot E(Y) - E(Y) \cdot E(X) + E(X) \cdot E(Y) =$$

$$\text{Cov}(X, Y) = E(X \cdot Y) - E(X) \cdot E(Y)$$

ii)  $\text{Cov}(X, Y) = \text{Cov}(Y, X)$ , logo a covariância é simétrica.

Demonstração:



$$\text{Cov}(X,Y) = E(X,Y) - E(X).E(Y)$$

$$\text{Cov}(Y,X) = E(Y,X) - E(Y).E(X)$$

$$\text{Logo a Cov}(X,Y) = \text{Cov}(Y,X)$$

$$\text{iii) Se } V(X) = 0 \text{ [ou } V(Y) = 0] \text{ então } \text{Cov}(X, Y) = 0$$

Demonstração:

$$\text{Cov}(X,Y) = E\{[X-E(X)][Y-E(Y)]\}$$

$$\text{Var}(X) = E\{[X-E(X)]^2\}$$

Se a  $\text{Var}(X) = 0$ , implica que  $X = E(X)$ , logo,

$$\text{Cov}(X,Y) = E\{[X-X][Y-E(Y)]\} = E\{[0][Y-E(Y)]\}$$

$$\text{Cov}(X, Y) = 0$$

$$\text{iv) Cov}(X+Z, Y) = \text{Cov}(X, Y) + \text{Cov}(Z, Y)$$

$$\text{Cov}(X+Z, Y) = E\{[(X+Z) - E(X+Z)][Y-E(Y)]\} =$$

$$\text{Cov}(X+Z, Y) = E\{[(X - E(X)) - (Z - E(Z))][Y-E(Y)]\} =$$

$$\text{Cov}(X+Z, Y) = E\{[(X-E(X))][Y-E(Y)] + [Z-E(Z)][Y-E(Y)]\} =$$

$$\text{Cov}(X+Z, Y) = E\{[(X+E(X))][Y-E(Y)]\} + E\{[Z-E(Z)][Y-E(Y)]\} =$$

$$\text{Cov}(X+Z, Y) = \text{Cov}(X, Y) + \text{Cov}(Z, Y)$$

$$\text{v) Cov}(KX, Y) = \text{Cov}(X, KY) = K.\text{Cov}(X, Y), \text{ para } K \in \mathfrak{R}, \text{ isto é a covariância é bilinear}$$

$$\text{Cov}(KX, Y) = E\{[KX-E(KX)][Y-E(Y)]\}$$

$$\text{Cov}(KX, Y) = K.E\{[X-E(X)][Y-E(Y)]\}$$

$$\text{Cov}(KX, Y) = K.\text{Cov}(X, Y)$$

Ou ainda,

$$\text{Cov}(K.X, Y) = E(KX, Y) - K.E(X).E(Y)$$

$$\text{Cov}(K.X, Y) = K.E(X, Y) - K.E(X).E(Y)$$

$$\text{Cov}(K.X, Y) = K\{E(X, Y) - E(X).E(Y)\} = K.\text{Cov}(X, Y)$$

$$\text{vi) Cov}(K, X) = \text{Cov}(X, K) = 0$$

$$\text{Cov}(K, X) = E\{[K-E(K)][X-E(X)]\} = E\{[K-E(K)][X-E(X)]\} = E\{[K-K][X-E(X)]\} =$$

$$\text{Cov}(K, X) = E\{[0][X-E(X)]\} = 0.[X-E(X)] = 0$$

4. Propriedades do coeficiente de correlação.

$$\text{i) } r(X, Y) = r(Y, X)$$

$$\text{ii) } -1 \leq r \leq 1$$

$$\text{iii) } r(X, Y) = r(ax + b, cY + d), \text{ se } a, c \neq 0$$

$$\text{iv) } r(X, Y) = 1, r(X, -X) = -1$$

É importante destacar que pares de variáveis aleatórias com distribuições marginais (idênticas) podem ter covariâncias e correlações distintas, assim  $\text{COV}(X, Y)$  e  $r(X, Y)$  são medidas do inter-relacionamento entre  $X$  e  $Y$ .

Devemos frisar que, embora  $\text{COV}(X, Y) = 0$  sempre que  $X$  e  $Y$  são variáveis aleatórias independentes, o inverso não é verdadeiro, isto é, se  $\text{COV}(X, Y) = 0$  não podemos concluir que  $X$  e  $Y$  são variáveis aleatórias independentes.

Portanto: Se  $X$  e  $Y$  são duas variáveis aleatórias independentes, então  $\text{COV}(X, Y) = 0$ .

De outro modo, se  $X$  e  $Y$  são independentes, isto implica  $X$  e  $Y$  não correlacionadas. A recíproca não é verdadeira, isto é  $\text{COV}(X, Y) = 0$  não implica  $X$  e  $Y$  independentes. Entretanto, é possível demonstrar que se as variáveis tem distribuição normal, o fato de a covariância ser igual a zero é condição suficiente para podermos afirmar que são variáveis independentes.

TÉCNICAS DE SOMATÓRIO  $\left(\sum_{i=1}^n X_i\right)$

### ÍNDICES OU NOTAÇÃO POR ÍNDICES

O símbolo  $X_i$  (leia-se  $X$  índice  $i$ ) representa um dos “ $n$ ” valores,  $X_1, X_2, \dots, X_n$ , assumidos pela variável “ $X$ ”, na amostra ou no conjunto de dados. A letra “ $i$ ”, usada como índice, pode representar qualquer um dos valores: 1, 2, 3, ...,  $n$ . Evidentemente pode ser usada para qualquer outra letra além de “ $i$ ” (MARTINS; DONAIRE, 1991; TOLEDO; OVALLE, 1987).

### SOMATÓRIO

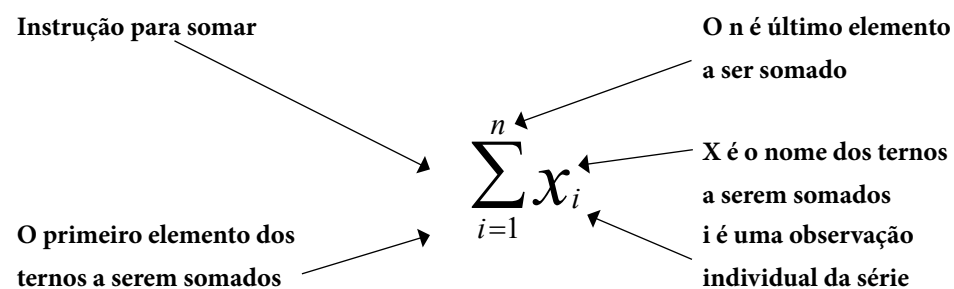
Muitas vezes precisamos escrever expressões que envolvem somas com muitos termos, ou cujos termos obedecem a certa lei de formação. Por exemplo, a soma dos 100 primeiros números naturais: o. Por exemplo, a soma dos 100 primeiros números naturais:  $1 + 2 + 3 + \dots + 99 + 100$

Simbolizaremos por  $x_i$  o  $i$ -ésimo termo da soma. Assim,  $x_1$  representa o primeiro termo,  $x_2$  representa o segundo,  $x_3$ , o terceiro,  $x_{100}$  representa o centésimo elemento. Também chamaremos  $n$  o número de termos da soma. Assim, na ilustração,  $n = 100$ .

A soma de  $n$  termos pode ser simbolicamente representada por  $\sum_{i=1}^n x_i$ .

No caso anterior, temos 100 termos; então  $n = 100$  e a soma desses cem números será representada por  $\sum_{i=1}^{100} X_i$ .

**PARTES DO SÍMBOLO DO SOMATÓRIO (NOTAÇÃO DE SOMATÓRIO)**



Dessa forma, a leitura acima se faz assim: “somatório de  $x_i$ , para  $i$  variando de 1 a “ $n$ ” ou “soma de  $x_i$ , para  $i$  variando de 1 a “ $n$ ”. O símbolo  $\sum$  é a letra grega sigma maiúsculo. Se estamos interessados na soma dos segundo, terceiro,..., centésimo elemento, devemos escrever:  $\sum_{i=2}^{100} x_i$ .

**PROPRIEDADES DOS SOMATÓRIOS**

i) O quadrado da soma é diferente da soma dos quadrados;

$$\left(\sum_{i=1}^n X_i\right)^2 \neq \sum_{i=1}^n X_i^2$$

ii) O produto de duas somas é diferente da soma dos produtos;

$$\left(\sum_{i=1}^n X_i\right) \cdot \left(\sum_{i=1}^n Y_i\right) \neq \left(\sum_{i=1}^n X_i Y_i\right)$$

iii) O número de parcelas de um somatório:

O número “P” de parcelas ou termos do somatório  $\sum_{i=a}^n$  é dado pela seguinte expressão.

$$P = n - a + 1$$

iv) Somatórios Múltiplos (Duplos);

O somatório múltiplo (duplo) de um produto é igual ao produto dos somatórios tomados separadamente;

$$\sum_{i=1}^n \sum_{j=1}^m X_i Y_j = \left(\sum_{i=1}^n X_i\right) \cdot \left(\sum_{j=1}^m Y_j\right)$$

v) O somatório do produto de uma constante por uma variável que depende do somatório é igual ao produto da constante pelo somatório da variável. Ou é o mesmo que se cada elemento da série é multiplicado por uma constante, os elementos podem ser somados, e a soma multiplicada pela constante.

$$\sum_{i=1}^n K X_i = K \sum_{i=1}^n X_i$$

vi) A soma de uma constante sobre  $n$  termos é igual a  $n$  vezes a constante.  $\sum_{i=1}^n K = n \cdot K$

vii) O somatório da soma (ou diferença) é igual à soma (ou diferença) de somatórios.

$$\sum_{i=1}^n (X_i \pm Y_i) = \sum_{i=1}^n X_i \pm \sum_{i=1}^n Y_i$$

viii) Quando não houver possibilidade de dúvidas, poderemos eliminar os índices. Assim:

$$\sum X^2 \text{ serão usados, ao invés de } \sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2$$

ix) Sempre que houver operações indicadas em frente do somatório, deveremos desenvolvê-las para em seguida aplicarmos as propriedades. Por exemplo.

$$\begin{aligned} &\sum (3X + 2)^2 \\ &\sum (3X + 2)^2 = \sum (9X^2 + 12X + 4) \\ &\sum (3X + 2)^2 = \sum 9X^2 + \sum 12X + \sum 4 \\ &\sum (3X + 2)^2 = 9\sum X^2 + 12\sum X + n4 \end{aligned}$$

**SOMATÓRIO DUPLO**

É frequentemente, na representação dos dados estatísticos, o uso de tabelas de dupla entrada, onde os valores são expressos em função de duas variáveis. Uma variável linha e uma variável coluna. Assim, podemos representar: estado civil (solteiro, casado, outros) x sexo (masculino e feminino), faixa etária x faixas de rendas; escolaridade x departamentos etc.

A indicação da soma dos elementos de tabelas de dupla entrada pode ser feita pelo somatório duplo já mostrado anteriormente.

**EXEMPLO DE APLICAÇÃO 1**

Seja  $X_{ij}$  um elemento genérico, sujeito a  $i$ -ésima linha e a  $j$ -ésima coluna da tabela a seguir:

**Tabela:** Valores genéricos de uma variável quantitativa  $x_{ij}$ .

i	j				
	1	2	3	...	k
1	$X_{11}$	$X_{12}$	$X_{13}$	...	$X_{1k}$
2	$X_{21}$	$X_{22}$	$X_{23}$	...	$X_{2k}$
3	$X_{31}$	$X_{32}$	$X_{33}$	...	$X_{3k}$
⋮	⋮	⋮	⋮	⋮	⋮
L	$X_{L1}$	$X_{L2}$	$X_{L3}$	...	$X_{Lk}$

Para provar algumas propriedades dadas anteriormente, usaremos as seguintes séries de números:

Variáveis X:  $X_1 = 5 \quad X_2 = 3 \quad X_3 = -2 \quad X_4 = 0$

Variáveis Y:  $Y_1 = 2 \quad Y_2 = 3 \quad Y_3 = -3 \quad Y_4 = 1$

Para a v propriedade. Vamos supor que a constante seja igual a 3, então:

$$\begin{aligned} &\sum_{i=1}^4 K X_i = K \sum_{i=1}^4 X_i \\ &KX_1 + KX_2 + KX_3 + KX_4 = K(X_1 + X_2 + X_3 + X_4) \end{aligned}$$

$$3(5) + 3(3) + 3(-2) + 3(0) = 3[5 + 3 + (-2) + 0]$$

$$15 + 9 - 6 + 0 = 3(6)$$

$$18 = 18$$

Para a vi propriedade. Vamos supor que a constante seja igual a 3 e n = 4, então:

$$\sum_{i=1}^4 K = n.K$$

$$K + K + K + K = n.K$$

$$3 + 3 + 3 + 3 = 4.(3)$$

$$12 = 12$$

Para a vii propriedade, teremos:

$$\sum_{i=1}^4 (x_i + y_i) = \sum_{i=1}^4 x_i + \sum_{i=1}^4 y_i$$

$$(X_1 + Y_1) + (X_2 + Y_2) + (X_3 + Y_3) + (X_4 + Y_4) =$$

$$(X_1 + X_2 + X_3 + X_4) + (Y_1 + Y_2 + Y_3 + Y_4)$$

$$(5 + 2) + (3 + 3) + (-2 - 3) + (0 + 1) =$$

$$[5 + 3 + (-2) + 0] + [2 + 3 + (-3) + 1]$$

$$7 + 6 - 5 + 1 = 6 + 3$$

$$9 = 9$$

Sendo assim temos que:

$$i) \quad X_{11} + X_{12} + \dots + X_{1K} + X_{21} + X_{22} + \dots + X_{LK} = \sum_{i=1}^L \sum_{j=1}^k X_{ij}$$

$$ii) \quad \sum_{i=3}^L \sum_{j=3}^k X_{ij}^2 = X_{33}^2 + X_{34}^2 + \dots + X_{3k}^2 + X_{43}^2 + \dots + X_{Lk}^2$$

$$iii) \quad \sum_{i=1}^k X_{2i} - \sum X_{2i} = X_{21} + X_{22} + X_{23} + \dots + X_{2K}$$

$$iv) \quad \sum_{i=1}^L X_{i3} - \sum X_{i3} = X_{31} + X_{23} + X_{33} + \dots + X_{L3}$$

### EXEMPLO DE APLICAÇÃO 2

Se  $X_{ij}$  representar o elemento sujeito a i-ésima linha e a j-ésima coluna da tabela então temos o seguinte. Note também que as propriedades utilizadas anteriormente também são válidas nesses casos.

**Tabela:** Valores fictícios de uma variável quantitativa  $X_{ij}$ .

		j			
i		1	2	3	4
1		5	-2	0	1
2		2	1	0	-2
3		1	2	4	3

Calcular:

$$i) \quad \sum_{i=1}^3 \sum_{j=1}^4 X_{ij} = \sum \sum X_{ij} = 5 + (-2) + 0 + 1 + 2 + \dots + 3 = 15$$

$$ii) \quad \sum_{i=3}^4 X_{3j} = \sum X_{3j} = 1 + 2 + 4 + 3 = 10$$

$$iii) \quad \sum_{i=1}^3 X_{i4} = \sum X_{i4} = 1 + (-2) + 3 = 2$$

$$iv) \quad \sum_{i=2}^3 \sum_{j=3}^4 X_{ij}^3 = 0^3 + (-2)^3 + 4^3 + 3^3 = 83$$

$$v) \quad \sum \sum (X_{ij} - 1)^2 = \sum \sum (X_{ij}^2 - 2X_{ij} + 1)$$

$$\sum \sum X_{ij}^2 - 2 \sum \sum X_{ij} + \sum \sum 1 =$$

$$5^2 + (-2)^2 + 0^2 + 1^2 + 2^2 + \dots + 3^2 - 2(15) + 12(1) = -51$$

### EXERCÍCIOS SOBRE SOMATÓRIO

O QUE SIGNIFICA A NOTAÇÃO SOMATÓRIO E PARA QUE É UTILIZADA?

EXEMPLO

Sendo,  $X = \{7, 3, 9, 5, 6\}$  e  $Y = \{3, 2, 8, 1, 1\}$ , calcule as seguintes somas:

$$i) \sum X \quad ii) \sum Y \quad iii) \sum X^2 \quad iv) \sum Y^2$$

$$v) \sum XY \quad vi) \sum (X + 2) \quad vii) \sum (X + Y) \quad viii) \sum (X - 2)$$

$$ix) \sum (X + 1)(Y - 3) \quad x) \sum (X + 2)^2 \quad xi) \frac{\sum (X + 4)}{\sum (Y + 4)} \quad xii) \sum \left( \frac{X + 4}{Y + 4} \right)$$

$$xiii) \sum (X - Y) \quad xiv) \sum \left( \frac{X}{Y} \right) \quad xv) \sum \left( \frac{Y}{X} \right)$$

Com os dados do exercício anterior, mostre que.

$$i) \sum (X + Y) = \sum X + \sum Y;$$

$$ii) \sum (X - Y) = \sum X - \sum Y;$$

$$iii) (\sum X)^2 \neq \sum X^2;$$

$$iv) (\sum Y)^2 \neq \sum Y^2;$$

$$v) \sum (X.Y) \neq \sum X . \sum Y$$

$$vi) \sum \left( \frac{X}{Y} \right) \neq \frac{\sum X}{\sum Y}$$

$$vii) \sum \left( \frac{X}{Y} \right) \neq \frac{\sum X}{\sum Y}$$

$$viii) \sum_{i=1}^5 \sum_{j=1}^5 X_i Y_j = \sum_{i=1}^5 X_i . \sum_{j=1}^5 Y_j$$

Sendo,  $X = \{7, 2, 9\}$  e  $K = 10$ , Mostre que  $\sum_{i=1}^3 KX_i = k \sum_{i=1}^3 X_i$

Desenvolver os somatórios:

i)  $\sum_{i=1}^5 9$                       iv)  $\sum_{i=2}^8 6i$

ii)  $\sum_{i=3}^8 200$                       v)  $\sum_{i=-2}^4 i^3$

iii)  $\sum_{i=0}^6 X^{8-i}$                       vi)  $\sum_{j=1}^5 5X_j Y_j$

Expressar cada uma das seguintes somas, mediante o emprego da notação de somatório:

i)  $(X_1 + 2)^2 + (X_2 + 2) + (X_3 + 2)^2 + (X_4 + 2)^2$ ;

ii)  $m_1 X(Y_1 - b)^2 + m_2 X(Y_2 - b)^2 + \dots + m_{10} X(Y_{10} - b)^2$ ;

iii)  $\frac{X_1 Y_1 + X_2 Y_2 + \dots + X_7 Y_7}{X_1 + X_2 + \dots + X_7}$

iv)  $\left(\frac{X_1 - 4}{Y_1}\right)^3 + \left(\frac{X_2 - 4}{Y_2}\right)^3 + \dots + \left(\frac{X_{15} - 4}{Y_{15}}\right)^3$

v)  $(4x_1 - 2y_1) + (4x_2 - 2y_2) + \dots + (4x_n - 2y_n)$

De acordo com a tabela abaixo, calcule:

a)  $\left(\sum_{i=1}^5 X_i\right) \cdot \left(\sum_{j=1}^4 Y_j\right)$       b)  $\sum_{j=1}^4 \sum_{i=1}^5 X_i Y_j$

c)  $\sum_{i=1}^3 X_i^2$                       d)  $\sum_{j=1}^4 Y_j^2$

Tabela: Valores fictícios de duas variáveis quantitativas  $X_i, Y_j$ .

$X_i$	$Y_j$
$X_1 = 2$	$Y_1 = 1$
$X_2 = 4$	$Y_2 = 3$
$X_3 = 6$	$Y_3 = 5$
$X_4 = 8$	$Y_4 = 7$
$X_5 = 10$	-----
$\left(\sum_{i=1}^5 X_i\right) =$	$\left(\sum_{j=1}^4 Y_j\right) =$

Calcular os seguintes somatórios:

i)  $\left(\sum_{i=1}^4 (X_i + 2)\right)$  para  $X = \{2, 3, 4, 5\}$

ii)  $\left(\sum_{i=1}^4 (X_i + 2)^2\right)$  para  $X = \{1, 2, 3, 4\}$

Provar através de exemplos numérico que o número P de parcelas ou termos do somatório  $\sum_{i=a}^n$  é dado pela seguinte expressão  $P = n - a + 1$ .

Prove que o somatório de uma constante “K” é igual ao produto do número de termos pela constante.

Desenvolver os seguintes somatórios:

i)  $\sum Y_i$  ( $i = 1, 2, \dots, 8$ );

ii)  $\sum V_i$  ( $i = 1, 2, \dots, n$ );

iii)  $\sum Z_i$  ( $i = 3, 4, \dots, 9$ );

iv)  $\sum Y_i$  ( $i = 1, 2, \dots, 100$ );

v)  $\sum (V_i W_i - K)$  ( $i = 1, 2, \dots, n$ );

### TÉCNICAS DE PRODUTÓRIO $\left(\prod_{i=1}^n X_i\right)$

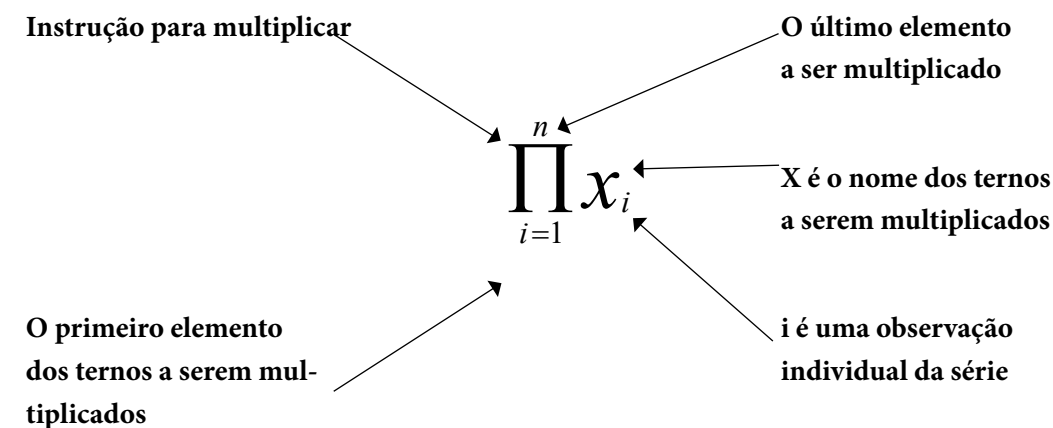
#### INTRODUÇÃO

Da mesma forma que foi mostrada anteriormente a técnica de somatório a qual representa a soma de termos de uma série numérica, pode-se necessitar da representação do produtório de termos de uma sucessão (MARTINS; DONAIRE, 1991; TOLEDO; OVALLE, 1987).

O produto de n fatores pode ser simbolicamente representado por:

$$\prod_{i=1}^n x_i = X_1 \cdot X_2 \cdot X_3 \cdot \dots \cdot X_n$$

Partes do símbolo do produtório



O símbolo  $\prod$  é a letra do alfabeto grego “Pi” maiúscula. Esta letra corresponde ao nosso “P” no alfabeto latino e, naturalmente, nos faz lembrar a palavra produto. Lê-se: “Produtório de  $x_i$ , para  $i$  variando de 1 a  $n$ , ou produto de  $x_i$  para  $i$  variando de 1 a  $n$ ”.



**EXEMPLOS**

- i) o produto  $2 \cdot 4 \cdot 6 \cdot 8 \cdot 10 \cdot 12 \cdot \dots \cdot 36$  pode ser representado por  $\prod_{j=1}^{18} 2j$
- ii) o produto  $1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot \dots \cdot 20$ , pode ser indicado por  $\prod_{i=2}^{21} (i-1)$ .

**PROPRIEDADES DOS PRODUTÓRIOS**

i) Se cada elemento da série é multiplicado por uma constante, o produto dos termos ficará multiplicado pela constante elevada ao número de termos do produtório. Para a essa propriedade, vamos considerar que a constante seja igual a 3, então:

$$\prod_{i=1}^4 K \cdot X_i = K^4 \prod_{i=1}^4 X_i$$

$$(KX_1)(KX_2)(KX_3)(KX_4) = K^4(X_1 \cdot X_2 \cdot X_3 \cdot X_4)$$

$$3(5) \cdot 3(3) \cdot 3(-2) = 3^4(5 \cdot 3 \cdot -2 \cdot 1)$$

$$15 \cdot 9 \cdot -6 \cdot 3 = 81(-30)$$

$$-2 \cdot 430 = -2 \cdot 430$$

ii) O produtório do produto é igual ao produto dos produtórios.

$$\prod_{i=1}^n (X_i + Y_i) = \left( \prod_{i=1}^n X_i \right) \left( \prod_{i=1}^n Y_i \right)$$

Para conduzir tais propriedades, usaremos os seguintes números:

Variáveis X:  $X_1=5, X_2=3, X_3=-2, X_4=1$

Variáveis Y:  $Y_1=2, Y_2=3, Y_3=-3, Y_4=1$

Para esta propriedade, teremos:

$$\prod_{i=1}^4 (X_i Y_i) = \left( \prod_{i=1}^4 X_i \right) \left( \prod_{i=1}^4 Y_i \right)$$

$$(X_1 Y_1)(X_2 Y_2)(X_3 Y_3)(X_4 Y_4) = (X_1 \cdot X_2 \cdot X_3 \cdot X_4)(Y_1)(Y_2)(Y_3)(Y_4)$$

$$(5 \cdot 2)(3 \cdot 3)(-2 \cdot -3)(1 \cdot 1) = (5 \cdot 3 \cdot -2 \cdot 1)(2 \cdot 3 \cdot -3 \cdot 1)$$

$$10 \cdot 9 \cdot 6 \cdot 1 = (-30)(-18)$$

$$540 = 540$$

iii) De uma maneira geral, o produto dos “n” primeiros números inteiros é dado por:

$$\prod_{i=1}^n i = 1 \cdot 2 \cdot 3 \cdot \dots \cdot n = n!$$

**EXERCÍCIOS SOBRE PRODUTÓRIO**

Qual o significado da notação produtório e para que é utilizada?

Escrever sob a forma de produtório os seguintes produtos.

- i)  $2 \cdot 4 \cdot 8 \cdot 16 \cdot 32 \cdot \dots \cdot 1024$
- ii)  $3 \cdot 6 \cdot 9 \cdot 12 \cdot 15 \cdot \dots \cdot 60$
- iii)  $1 \cdot 3 \cdot 5 \cdot 7 \cdot 9 \cdot \dots \cdot 35$
- iv)  $Y \cdot Y \cdot Y \cdot Y \cdot \dots \cdot Y$  [produto de “n” fatores iguais]
- v)  $1^1 \cdot 2^2 \cdot 3^3 \cdot 4^4 \cdot 5^5 \cdot \dots \cdot n^n$
- vi)  $a^2 \cdot a^3 \cdot a^4 \cdot \dots \cdot a^k$
- vii)  $X_1 \cdot X_2 \cdot \dots \cdot X_7$
- viii)  $a_1 X_1 \cdot a_2 X_2 \cdot a_3 X_3 \cdot \dots \cdot a_n X_n$
- ix)  $X_1^{F_1} \cdot X_2^{F_2} \cdot \dots \cdot X_m^{F_m}$

Mostrar que:

- i)  $\prod_{i=1}^k c x_i^a = c^k \left( \prod_{i=1}^k x_i \right)^a$ ;
- ii)  $\prod_{i=1}^n i = n!$ ;
- iii)  $\log \prod_{i=1}^n x_i = \sum_{i=1}^n \log x_i$ ;
- iv)  $\prod_{i=1}^4 2^i = 1024$ ;

Desenvolver e calcular:

- i)  $\prod_{i=2}^5 (2i - 1)$ ;
- ii)  $\prod_{i=1}^4 3i$ ;
- iii)  $\prod_{j=1}^6 \left( 2 - \frac{1}{j} \right)$ ;
- iv)  $\prod_{i=3}^6 \frac{3i}{i-2}$ ;
- v)  $\prod_{i=1}^3 (3i - 1) i^3$ ;

Verificar que:

$$\text{i) } \prod_{i=3}^7 i = \frac{\prod_{i=1}^7 i}{\prod_{i=1}^2 i}$$

$$\text{ii) } \prod_{i=1}^m i = m!$$

### MÉTODOS UTILIZADOS PARA SE OBTER ESTIMADORES

Quando, em um estudo científico, se faz referência ao processo de estimar um parâmetro, se está buscando um método que permitirá conhecer um valor para a quantidade fixa  $\theta$  que é desconhecida. Para se realizar esse processo são usadas expressões, as quais são funções das observações amostrais denominados de estimadores. A seguir são apresentados alguns métodos para a obtenção de tais expressões. Além disso, procura-se mostrar ao leitor que tais estimadores são frutos de métodos cientificamente desenvolvidos. Muitos são os métodos existentes para obtê-las. Entretanto, são apresentados somente aqueles mais importantes.

### TIPOS DE MÉTODOS DE ESTIMAÇÃO

Dentre os métodos de estimação pontuais mais usuais destacam-se:

#### a) MÉTODO DOS MOMENTOS (MM)

Método mais simples de estimação o qual utiliza a igualdade do parâmetro com o estimador;

#### b) MÉTODO DOS MÍNIMOS QUADRADOS (MMQ)

São usualmente utilizados no âmbito da regressão linear;

#### c) MÉTODO DA MÁXIMA VEROSSIMILHANÇA (MV)

É provavelmente o método mais importante. Geralmente, os estimadores de máxima verossimilhança gozam das propriedades desejáveis num bom estimador: são os mais eficientes e consistentes. Embora, usualmente, não sejam centrados costumam ser assintoticamente não enviesados. Uma forma de obter uma estimativa pontual de um parâmetro da população (por exemplo:  $\mu$ ,  $\sigma$ ,  $\sigma^2$  e  $p$ ) é retirar uma amostra aleatória representativa dessa população e calcular o valor da estatística correspondente (por exemplo:  $\bar{x}$ ,  $s$ ,  $s^2$ ,  $\bar{p}$ ).

### DESCRIÇÃO DOS MÉTODOS

#### a) MÉTODO DOS MOMENTOS (MM)

Este método de estimação é um dos mais simples e mais antigos para obter estimadores de um ou mais parâmetros de uma distribuição. A idéia base é utilizar os momentos da amostra para estimar os correspondentes momentos da população, e, a partir daí, estimar os parâmetros de interesse (MURTEIRA et al., 2001). Seja  $X_1, X_2, \dots, X_n$  uma amostra aleatória de uma dada população com função (densidade) de probabilidade f. d. p.,  $f(x; \theta_1, \theta_2, \dots, \theta_k)$  que depende de  $k$  parâmetros. Admitindo que existam os momentos ordinários,  $\mu_r$  da população  $x$ , estes são funções dos  $k$  parâmetros,

$$\mu_r' = E[X^r] = \begin{cases} \sum_{i=1}^N x^r f(x; \theta_1, \theta_2, \dots, \theta_k) & \text{para distribuições discretas} \\ \int_{-\infty}^{+\infty} x^r f(x; \theta_1, \theta_2, \dots, \theta_k) dx & \text{para distribuições contínuas} \end{cases}$$

Os correspondentes momentos amostrais são dados por  $m_r' = \frac{1}{n} \sum_{i=1}^n x_i^r$

O método dos momentos consiste em considerar que os estimadores dos momentos ordinários são dados pelos momentos ordinários amostrais, ou seja,  $\hat{\mu}_r' = m_r', r = 1, \dots, k$ .

#### b) MÉTODO DOS QUADRADOS MÍNIMOS (MMQ)

Um outro importante método de estimação que aparece na teoria estatística é o dos quadrados mínimos. Grande parte da teoria em que se fundamenta a estatística experimental e baseada nesse método. Para apresentar o conceito e ilustrar o método, considerar o modelo linear a seguir para cada observação da amostra aleatória  $X_1, X_2, \dots, X_n$ . A observação amostral é modelada como resultante da soma de dois componentes básicos: um componente fixo (constante)  $\mu$  e outro de natureza aleatória  $e_i$ .

$X_i = \mu + e_i$ , em geral, o componente  $e_i$  é suposto normal com média 0 e variância constante  $\sigma^2$  para todo valor de  $i$ ,  $i = 1, 2, \dots, n$ .

Pela observação do modelo linear apresentado, verifica-se que  $\mu$  e  $\sigma^2$  são os parâmetros desconhecidos desse modelo. O parâmetro  $\sigma^2$  não é, entretanto, mencionado explicitamente nesse modelo.

O método de estimação de quadrados mínimos baseia-se na minimização da soma de quadrados da variável aleatória denominada erro ou resíduo,  $e_i$ . Assim, isolando-se  $e_i$  no modelo anterior resulta em:

$$e_i = X_i - \mu$$

Tomando-se a soma dos quadrados (SQ) de  $e_i$  para as  $n$  observações amostrais tem-se:

$$SQ = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (X_i - \mu)^2$$

Para se obter o estimador de quadrados mínimos e necessário minimizar o valor dessa soma de quadrados (SQ). Para se obter esse mínimo, é necessário derivar em relação a cada parâmetro, igualar as derivadas a

zero e resolver o sistema de equações formado. Nesse modelo, em que se está exemplificando o método, a soma de quadrados SQ só depende de um parâmetro ( $\mu$ ), então,

$$\frac{dSQ}{d\mu} = -2 \sum_{i=1}^n (X_i - \mu)$$

Igualando a zero a derivada primeira de SQ em relação ao parâmetro  $\mu$  tem-se:

$$-2 \sum_{i=1}^n (X_i - \hat{\mu}) = 0$$

$$-\sum_{i=1}^n X_i + \sum_{i=1}^n \hat{\mu} = 0$$

$$n\hat{\mu} = \sum_{i=1}^n X_i \quad \therefore \hat{\mu} = \frac{\sum_{i=1}^n X_i}{n} = \bar{X}$$

A segunda derivada de SQ em relação a  $\mu$  é:

$$\frac{d^2SQ}{d\mu^2} = 2n > 0$$

Esse resultado indica que  $\mu$  = representa um ponto de mínimo da função SQ. Consequentemente, o estimador de quadrados mínimos da média  $\mu$  = . O estimador de momentos da variância  $\sigma^2$ , não viesado, é:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \hat{X}_1)^2$$

Em que,  $\hat{X}_1 = \hat{\mu} = \bar{X}$  é o preditor de quadrados mínimos de  $X_1$  logo,  $s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$ .

### c) MÉTODO DA MÁXIMA VEROSSIMILHANÇA (MV)

O método da máxima verossimilhança foi introduzido por Ronald Aylmer Fisher em 1922. Sua introdução, em muitos aspectos, determinou o começo da teoria estatística moderna. Para apresentar o seu conceito, considere que  $X_1, X_2, \dots, X_n$  seja uma amostra aleatória de uma população com densidade  $f(x)$ , determinada pelos parâmetros  $\theta_i$ ,  $i = 1, 2, \dots, k$ . Inicialmente, é considerada a situação específica de apenas um parâmetro  $\theta$  ( $k=1$ ), por facilitar o entendimento do conceito a ser apresentado. Para uma amostra particular  $X_1, X_2, \dots, X_n$ , o estimador de máxima verossimilhança ( $\hat{\theta}$ ) do parâmetro  $\theta$  é aquele que maximiza a densidade conjunta de  $X_1, X_2, \dots, X_n$ .

Em razão do fato de os valores amostrais  $X_1, X_2, \dots, X_n$  serem independentes é possível definir a densidade conjunta ou função de verossimilhança ( $L$ ) pelo produto das densidades de cada  $X_i$  ( $i = 1, 2, \dots, n$ ). Assim, a função de verossimilhança,  $L$ , é definida pela seguinte equação.

$$L = f(x_1)f(x_2)f(x_3)\dots f(x_n) = \prod_{i=1}^n f(x_i)$$

O estimador de máxima verossimilhança é aquele que maximiza o valor de  $L$  na equação anterior.

Para obter o estimador de máxima verossimilhança ( $\hat{\theta}$ ), basta tomar a primeira derivada de  $L$  em relação ao parâmetro  $\theta$ , igualar a zero e resolver para  $\theta$ . A solução é o estimador de máxima verossimilhança.

**Nem** sempre uma solução explícita existe e métodos numéricos são utilizados para se obterem as estimativas. Quando se tem mais de um parâmetro, tomam-se as derivadas parciais de  $L$  com respeito a cada um deles. Igualam-se cada derivada a zero e resolve-se o sistema formado obtendo-se os estimadores de máxima verossimilhança dos parâmetros.

Algumas propriedades matemáticas da função  $L$  garantem a possibilidade de usar a função  $S = \ln(L)$  em seu lugar, uma vez que apresentam o máximo para o mesmo valor de  $\theta$ . Isso é feito para tornar mais fácil a obtenção do máximo, uma vez que o produto se transforma em somatório. Essa função é denominada de função suporte.

Na estimação paramétrica, este método só pode ser aplicado se a distribuição da população for conhecida.

A Definição desse método pode ser explicada da seguinte maneira:

Seja  $X_1, X_2, \dots, X_n$ , uma amostra aleatória de uma dada população com função (densidade) de probabilidade, f.d.p.,  $f(x; \theta)$ . Então a f.d.p. conjunta das variáveis que constituem a amostra é dada por:

$$f(x_1, x_2, \dots, x_n; \theta) = f(x_1; \theta)f(x_2; \theta)\dots f(x_n; \theta) \prod_{i=1}^n f(x_i; \theta)$$

Considerando uma amostra em concreto, designa-se por função de verossimilhança a função de  $\theta$  e da amostra tal que:

$$L(x_1, x_2, \dots, x_n; \theta) = L(x; \theta) = \prod_{i=1}^n f(x_i; \theta).$$

O Método da Máxima Verossimilhança pode ainda ser descrito da seguinte forma:

Consiste em encontrar o estimador  $\hat{\theta}$  que maximiza o valor a função de verossimilhança para uma determinada amostra, ou seja, o valor de  $\hat{\theta}$  que torna aquela amostra concreta mais provável, isto é, mais verossímil.

Para se obter o estimador de máxima verossimilhança devem-se fazer as seguintes operações:

- i) Determinar a função de verossimilhança  $L(x; \theta)$ ;
- ii) Se necessário, aplicar a transformação logarítmica à função de verossimilhança  $\ln L(x; \theta)$ . Geralmente, esta transformação torna o problema da maximização mais simples.
- iii) Determinar os pontos onde a 1ª derivada da função ( $L(x; \theta)$  ou  $\ln L(x; \theta)$ ) em relação a  $\theta$  se anula (condição de primeira ordem):

$$\frac{\partial L(x; \theta)}{\partial \theta} = 0 \quad \text{ou} \quad \frac{\partial \ln L(x; \theta)}{\partial \theta} = 0$$

- iv) Verificar se a 2ª derivada da função em relação a  $\theta$  é negativa (condição de segunda ordem):

$$\frac{\partial^2 L(x; \theta)}{\partial \theta^2} < 0 \quad \text{ou} \quad \frac{\partial^2 \ln L(x; \theta)}{\partial \theta^2} < 0$$

O método da máxima verossimilhança pode ser utilizado para estimar mais do que um parâmetro em simultâneo.

**EXEMPLOS**

1. Suponha que uma variável aleatória  $X$  representa o número de avarias ou defeitos de um motor usado em irrigação por aspersão durante um período de tempo de 3 meses e que obedece a uma lei de Poisson de parâmetro  $\lambda$  desconhecido. Para este parâmetro foram sugeridos dois estimadores.

$$\hat{\lambda} = \frac{X_1 + L + X_n}{n} \text{ e } \hat{\lambda} = \frac{X_1 + X_n}{n}$$

- i) Compare-os quanto ao enviesamento.
- ii) Deduza a variância para cada um deles.
- iii) Qual dos dois estimadores é mais eficiente? Justifique a sua escolha.
- iv) Estude os dois estimadores quanto à consistência.

A Resolução se dá da seguinte maneira:

$$i) E[\hat{\lambda}] = E\left[\frac{X_1 + L + X_n}{n}\right] = \frac{1}{n}\{E[X_1] + E[X_2] + L + E[X_n]\} = \frac{1}{n}[\mu + \mu + L + \mu] = \frac{1}{n}n\mu = \mu$$

$$E[\hat{\lambda}] = E\left[\frac{X_1 + X_n}{n}\right] = \frac{1}{2}\{E[X_1] + E[X_2]\} = \frac{1}{2}[\mu + \mu] = \mu$$

Portanto, ambos os estimadores são centrados.

$$ii) Var[\hat{\lambda}] = Var\left[\frac{X_1 + \dots + X_n}{n}\right]_{X_i \text{ independentes}} = \frac{1}{n^2}\{Var[X_1] + Var[X_2] + \dots + Var[X_n]\} = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$$

$$Var[\hat{\lambda}] = Var\left[\frac{X_1 + X_n}{n}\right]_{X_i \text{ independentes}} = \frac{1}{2^2}\{Var[X_1] + Var[X_n]\} = \frac{1}{4}[\sigma^2 + \sigma^2] = \frac{\sigma^2}{2}$$

iii)  $Var[\hat{\lambda}] < Var[\hat{\lambda}]$ , para  $n > 2$ . Portanto  $\hat{\lambda}$  é mais eficiente.

iv)  $\hat{\lambda}$  é um estimador consistente, pois  $E[\hat{\lambda}] = \mu$  e  $Var[\hat{\lambda}] = \frac{\sigma^2}{n} \rightarrow 0$ , quando  $n \rightarrow \infty$ .

$\hat{\lambda}$  não é um estimador consistente, pois  $E[\hat{\lambda}] = \mu$  mas  $Var[\hat{\lambda}] = \frac{\sigma^2}{n}$ , seja qual for o valor de  $n$ .

2. Seja  $(X_1, X_2, \dots, X_n)$  uma amostra aleatória de uma distribuição Normal,  $X \cap N(\mu, \sigma)$ . Estime os parâmetros  $\mu$  e pelo método:

- i) Dos momentos.
- ii) Da Máxima verossimilhança.

A resolução se dá da seguinte maneira:

i) Sabe-se que

$$\mu'_1 = E[X] = \mu$$

$$\mu'_2 = E[X^2] = Var[X] + E^2[X] = \sigma^2 + \mu^2.$$

Para obter os estimadores  $\hat{\mu}$  e  $\hat{\sigma}^2$  pelos métodos dos momentos, é preciso resolver o seguinte sistema de equações:

$$\begin{cases} \mu'_1 = m'_1 = \frac{1}{n} \sum_{i=1}^n X_i \\ \mu'_2 = m'_2 = \frac{1}{n} \sum_{i=1}^n X_i^2 \end{cases}$$

$$\begin{cases} \mu = \frac{1}{n} \sum_{i=1}^n X_i \\ \sigma^2 + \mu^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 \end{cases}$$

$$\begin{cases} \mu = \bar{X} \\ \sigma^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 \end{cases}$$

portanto, os estimadores obtidos foram:  $\begin{cases} \hat{\mu} = \bar{X} \\ \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 \end{cases}$

ii) Função densidade de probabilidade (f.d.p.):

$$f(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \text{ para } \sigma > 0$$

$$\text{Função de verossimilhança: } L(\mu, \sigma^2, x) = \prod_{i=1}^n f(x_i; \mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2} =$$

$$= \frac{1}{\left[\sqrt{2\pi\sigma^2}\right]^n} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i-\mu)^2} = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i-\mu)^2}$$

Logaritmo da Função de Verossimilhança:

$$\ln L(\mu, \sigma^2; x) = -\frac{n}{2}(\ln 2 + \ln \pi + \ln \sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

Condição de 1ª ordem:

$$\begin{cases} \frac{\partial \ln L(\mu, \sigma^2; x)}{\partial \mu} = 0 & \left\{ -\frac{1}{2\sigma^2} \left( -2 \sum_{i=1}^n x_i + 2n\mu \right) = 0 \right. \\ \frac{\partial \ln L(\mu, \sigma^2; x)}{\partial \sigma^2} = 0 & \left\{ -\frac{n}{2} \frac{1}{\sigma^2} + \sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma^4} = 0 \right. \end{cases} \therefore \begin{cases} \sum_{i=1}^n x_i - n\mu = 0 \\ -n\sigma^2 + \sum_{i=1}^n (x_i - \mu)^2 = 0 \end{cases} \therefore$$



$$\begin{cases} \mu = \sum_{i=1}^n \frac{x_i}{n} \\ \sigma^2 = \sum_{i=1}^n \frac{(x_i - \mu)^2}{n} \end{cases} \therefore \begin{cases} \mu = \bar{x} \\ \sigma^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n} \end{cases} \therefore \begin{cases} \mu = \bar{x} \\ \sigma^2 = S^2 \end{cases}$$

Condição de 2ª ordem:

$$\begin{cases} \frac{\partial^2 \ln L(\mu, \sigma^2; x)}{\partial \mu^2} = -\frac{1}{2\sigma^2} 2n < 0 \\ \frac{\partial^2 \ln L(\mu, \sigma^2; x)}{\partial \sigma^4} = \frac{n}{2} \frac{1}{\sigma^4} - \sum_{i=1}^n \frac{(x_i - \mu)^2}{2} \frac{2\sigma^2}{\sigma^8} < 0 \end{cases}$$

Portanto, os estimadores de máxima verossimilhança obtidos foram:

$$\begin{cases} \hat{\mu} = \bar{X} = \sum_{i=1}^n \frac{X_i}{n} \\ \hat{\sigma}^2 = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{n} = \frac{n}{n-1} S^2 \end{cases}$$

3. Considere uma população com distribuição de Bernoulli, com parâmetro p (0 ≤ p ≤ 1).

i) Derive o estimador de máxima verossimilhança para o parâmetro p, sabendo-se que foi obtida uma amostra de dimensão n = 3, cujos valores observados foram (1,1,0).

ii) Esboce o gráfico da função de verossimilhança e interprete-o.

iii) Forneça uma estimativa para p com base no método da máxima verossimilhança.

A resolução é processada da seguinte forma:

i) Função de probabilidade (f.p.): f(x, p) = p<sup>x</sup>(1 - p)<sup>1-x</sup>, para x = 0, 1 e 0 ≤ p ≤ 1.

Função de verossimilhança:

$$L(p; x) = \prod_{i=1}^n f(X_i, p) = \prod_{i=1}^n p^{X_i} (1-p)^{1-X_i} = p^{\sum_{i=1}^n X_i} (1-p)^{n - \sum_{i=1}^n X_i}$$

Logaritmo da Função de Verossimilhança:  $\ln L(p; x) = \sum_{i=1}^n X_i \ln p + \left( n - \sum_{i=1}^n X_i \right) \ln(1-p)$

Condição de 1ª ordem:  $\frac{\partial \ln L(p; X)}{\partial p} = 0 \therefore \sum_{i=1}^n X_i \frac{1}{p} + \left( n - \sum_{i=1}^n X_i \right) \frac{-1}{1-p} = 0 \therefore$

$$\therefore (1-p) \sum_{i=1}^n X_i - p \left( n - \sum_{i=1}^n X_i \right) = 0 \Leftrightarrow \sum_{i=1}^n X_i - pn = 0 \Leftrightarrow p = \sum_{i=1}^n \frac{X_i}{n}$$

Condição de 2ª ordem:

$$\frac{\partial^2 \ln L(p; x)}{\partial p^2} = -\frac{\sum_{i=1}^n X_i}{p^2} + \frac{\left( n - \sum_{i=1}^n X_i \right) (-1)}{(1-p)^2} = -\frac{\sum_{i=1}^n X_i}{p^2} - \frac{\left( n - \sum_{i=1}^n X_i \right)}{(1-p)^2} < 0$$

Pois X<sub>i</sub> ≥ 0, p<sup>2</sup> ≥ 0, n > 0, (1 - p)<sup>2</sup> ≥ 0 e n ≥ ∑<sub>i=1</sub><sup>n</sup> X<sub>i</sub> pois X<sub>i</sub> = 0 ou 1.

Portanto, o estimador de máxima verossimilhança é  $\hat{p} = \sum_{i=1}^n \frac{X_i}{n}$

ii)

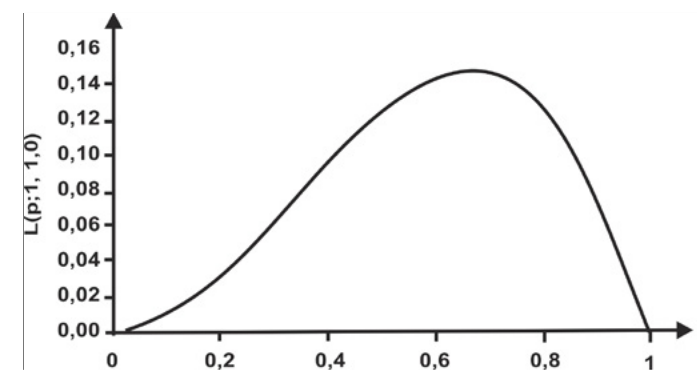


Figura: Gráfico representativo do estimador de máxima verossimilhança da proporção  $\hat{p}$ , com o tamanho da amostra n=3.

iii)  $\hat{p} = \frac{2}{3} = 0,6667$

iv) Obter o estimador de máxima verossimilhança do parâmetro λ da distribuição exponencial, a partir de uma amostra aleatória de tamanho n.

A função de verossimilhança da distribuição exponencial é:

$$L = \prod_{i=1}^n f(X_i) = \prod_{i=1}^n \lambda e^{-\lambda X_i} = \lambda^n e^{-\lambda \sum_{i=1}^n X_i}$$

Tomando-se o logaritmo neperiano de L:

$$\ln L = n \ln(\lambda) - \lambda \sum_{i=1}^n X_i$$

Deriva-se ln L em relação ao parâmetro λ, iguala-se a zero, resolve-se a equação formada e obtém-se o estimador de máxima verossimilhança:

$$\frac{d \ln L}{d \lambda} = \frac{n}{\lambda} - \sum_{i=1}^n X_i$$

$$\frac{n}{\hat{\lambda}} - \sum_{i=1}^n X_i = 0 \Rightarrow \frac{n}{\hat{\lambda}} = \sum_{i=1}^n X_i$$

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^n X_i} = \frac{1}{\bar{X}}$$

Logo,  $\hat{\lambda} = \frac{1}{\bar{X}}$ , é o estimador de máxima verossimilhança do parâmetro  $\lambda$  da distribuição exponencial.

**- PROPRIEDADES DOS ESTIMADORES**

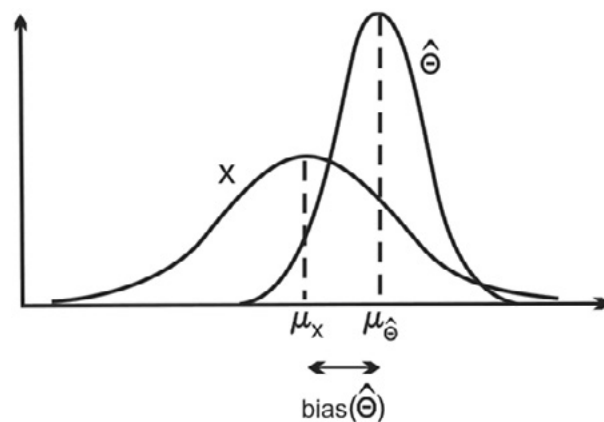
Algumas propriedades dos estimadores são desejáveis no processo de inferência. A escolha de um estimador de um parâmetro  $\theta$  qualquer em detrimento de outro, depende de uma criteriosa avaliação dessas propriedades.

As seguintes propriedades são consideradas desejáveis num bom estimador.

**a) JUSTO, NÃO VIESADO, NÃO VICIADO OU NÃO-TENDENCIOSO**

Um estimador justo é uma estatística amostral cujo valor esperado é igual ao parâmetro que está sendo estimado, ou melhor.

Seja “X” uma variável aleatória (população) que dependa do parâmetro  $\theta$ , e seja  $\hat{\theta}$  um estimador de  $\theta$ . Logo  $\hat{\theta}$  é justo se  $E[\hat{\theta}] = \theta$ , ou seja, um estimador é não viesado se, em repetidas amostragens aleatórias da população, o estimador (Estatística) correspondente da distribuição amostral teórica, for igual ao parâmetro da população  $\theta$ .



**Figura:** Curvas da distribuição amostral do estimador  $\hat{\theta}$  e da população X, mostrando o enviesamento do estimador  $\hat{\theta}$ .

**-Exemplo 1**

A média da Amostra aleatória  $\bar{X}$  é um estimador justo ou não viciado da média da população  $\mu$ , pois  $E[\bar{X}] = \mu$ , ou seja,.

$$E[\bar{X}] = E\left[\frac{\sum_{i=1}^n x_i}{n}\right] = \frac{1}{n} E[X_1 + X_2 + \dots + X_n] = \frac{1}{n} \{E[X_1] + E[X_2] + \dots + E[X_n]\} =$$

$$E[\bar{X}] = \frac{1}{n} \{\mu + \mu + \dots + \mu\} = \frac{1}{n} \cdot n \cdot \mu = \mu, \text{ portanto } E[\bar{x}] = \mu$$

**-Exemplo 2**

A variância  $S^2 = \frac{\sum (x - \bar{x})^2}{n}$ , não é um estimador justo de  $\sigma^2$ , pois  $E(S^2) < \sigma^2$ . Para se tornar justa tem-se que multiplicar  $\frac{\sum (x - \bar{x})^2}{n}$  por  $\frac{n}{n-1}$  onde, a esperança matemática é dada pela seguinte expressão, E

$$\left(\frac{n}{n-1} \cdot S^2\right) = E\left(\frac{\sum (x - \bar{x})^2}{n-1}\right) = \sigma^2, \text{ é uma estimativa justa de } \sigma^2. \text{ Aliás, a variância } \sigma^2 \text{ é por definição } \sigma^2 =$$

$$\frac{\sum (x - \mu)^2}{n} \text{ e como parâmetro } \mu, \text{ é desconhecido, é substituída pela estimativa } \bar{x}, \text{ correspondendo a isso}$$

a perda de um grau de liberdade.

**b) CONSISTENTE OU COERENTE**

Um estimador  $\hat{\theta}$  de um parâmetro  $\theta$  é consistente se a condição abaixo for satisfeita.

$$\lim_{n \rightarrow \infty} \text{Pr ob} \{|\theta - \hat{\theta}| > \varepsilon\} = 0, \text{ para todo } \varepsilon > 0, \text{ ou ainda,}$$

$$\lim_{n \rightarrow \infty} \text{Pr ob} \{|\theta - \hat{\theta}| < \varepsilon\} = 1, \text{ onde } \varepsilon \text{ é um valor positivo } (\varepsilon > 0).$$

Na prática esse critério não é muito simples de aplicarmos, para que possamos verificar a consistência de um estimador  $\hat{\theta}$ . Sendo assim na prática usa-se os seguintes critérios ou regras.

**i) Se  $E(\hat{\theta}) = \theta$ , isto é se  $\hat{\theta}$  é justo ou não viesado.**

**ii) Se  $\lim_{n \rightarrow \infty} \text{Var}(\hat{\theta}) = 0$**

Então  $\hat{\theta}$  é um estimador consistente de  $\theta$ .

Ou ainda um estimador  $\hat{\theta}$  de um parâmetro  $\theta$  de uma população é consistente quando sua distribuição se torna mais concentrada à medida que o tamanho da amostra “n” tende a infinito. Um estimador  $\hat{\theta}$  com consistência simples é aquele que converge em probabilidade para o parâmetro  $\theta$  à medida que n tende a infinito, ou seja, além de não ser viesado, sua variância tende a zero com o aumento de n ( $\lim_{n \rightarrow \infty} \sigma_{\hat{\theta}}^2 = 0$ ).

**- Exemplo**

Para os estimadores de  $\mu$ ,  $M_d$  e  $\bar{X}$ , de uma população normal verificar a consistência simples e fazer um gráfico comparativo de ambas as situações em função de n.

Para a média  $\bar{X}$ , temos que:  $\lim_{n \rightarrow \infty} \sigma_{(\bar{X})}^2 = \lim_{n \rightarrow \infty} \frac{\sigma^2}{n} = 0$

Para a mediana  $M_d$ , temos que:  $\lim_{n \rightarrow \infty} \sigma_{(M_d)}^2 = \lim_{n \rightarrow \infty} \left(\frac{\sigma^2}{n} \cdot \frac{\pi}{2}\right) = 0$

Ambos são estimadores consistentes de  $\mu$ , pois se  $\bar{X}$  tem um erro padrão dado por  $\frac{\sigma}{\sqrt{n}}$ , e quando

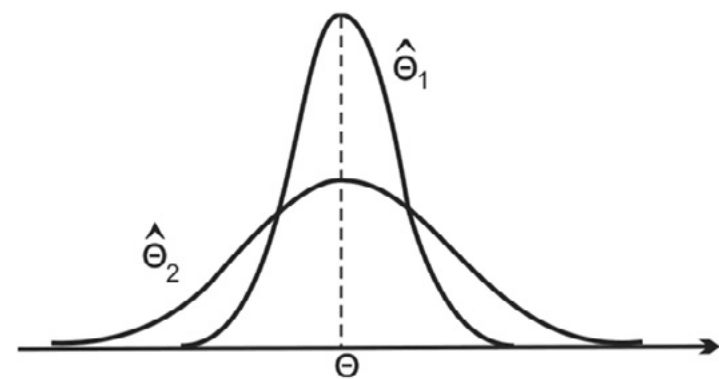
$n \rightarrow \infty$ , então  $\frac{\sigma}{\sqrt{n}} \rightarrow 0$ , e assim a distribuição de  $\bar{X}$  torna-se mais centrada em torno de  $\mu$ .

c) EFICIENTE

Sejam  $\hat{\theta}_1$  e  $\hat{\theta}_2$  dois estimadores centrados de um mesmo parâmetro  $\theta$ ; baseados no mesmo número de observação,  $\hat{\theta}_1$  será mais eficiente que  $\hat{\theta}_2$  quando e somente quando a  $Var[\hat{\theta}_1] < Var[\hat{\theta}_2]$ .

A eficiência relativa do primeiro estimador relativamente ao segundo é dada por:

$$\text{Eficiência Relativa [ER]} = \frac{Var(\hat{\theta}_1)}{Var(\hat{\theta}_2)}$$



**Figura:** Curvas das distribuições amostrais dos estimadores  $\hat{\theta}_1$  e  $\hat{\theta}_2$ , mostrando a eficiência de estimadores não viesados  $[Var(\hat{\theta}_1) < Var(\hat{\theta}_2)]$ .

- Exemplo

Sejam os estimadores  $\bar{X}$  e  $M_d$  (média e mediana) da média de uma população normal ( $\mu$ ) obtidos em uma amostra aleatória de tamanho  $n$ . Verificar qual dos estimadores é o mais eficiente, sabendo que devido a simetria da distribuição normal, temos que:

$$E(M_d) = E(\bar{X}) = \mu$$

Para determinar a eficiência relativa dos estimadores é necessário conhecer suas variâncias. A variância de  $\bar{X}$  é conhecida e dada por  $\sigma^2_{(\bar{x})} = \frac{\sigma^2}{n}$  mas a variância de  $M_d$  não é conhecida de forma exata, tendo

apenas uma boa aproximação. Essas variâncias estão apresentadas a seguir:

$$\sigma^2_{(\bar{x})} = \frac{\sigma^2}{n} \text{ e } \sigma^2_{(Md)} \cong \frac{\sigma^2}{n} \cdot \frac{\pi}{2}$$

A razão entre  $\sigma^2_{(\bar{x})}$  e  $\sigma^2_{(Md)}$  é a eficiência relativa (ER). Assim temos que,

$$ER = \frac{\sigma^2_{(\bar{x})}}{\sigma^2_{(Md)}} = \frac{\frac{\sigma^2}{n}}{\frac{\sigma^2}{n} \cdot \frac{\pi}{2}} = \frac{2}{\pi} \cong 0,64 = 64\%$$

A eficiência relativa de  $M_d$  em relação à média  $\bar{X}$  é de apenas 64%. Isso poderia ser interpretado em termos de tamanho amostral, que seria necessário para se estimar  $\mu$  usando a mediana, em detrimento daquele que seria necessário, usando a média como estimador. Se o tamanho da amostra fosse igual a

n para se estimar  $\mu$  com  $\bar{X}$ , então o tamanho de uma amostra para estimá-la, usando  $M_d$  seria  $1,57n$ , para se alcançar a mesma precisão. Isso significa que uma estimativa da média da população “ $\mu$ ” de 64 observações, utilizando-se a média da Amostra  $\bar{X}$  dá a mesma informação que a de 100 observações usando-se a mediana.

Se, entretanto, a população estivesse “contaminada”, a média ( $\bar{X}$ ) teria uma maior chance de apresentar maiores desvios de  $\mu$ . A mediana  $M_d$ , por ser menos afetada por esses elementos discrepantes da amostra, poderia ser um estimador mais adequado.

d) SUFICIENTE

Um estimador  $\hat{\theta}$  é suficiente (são os mais desejáveis) se para uma dada amostra ele fornece toda a informação possível ou relevante a respeito do parâmetro  $\theta$ , independentemente do auxílio de qualquer outra estimativa.

Um estimador  $\hat{\theta}$  diz-se suficiente para um parâmetro  $\theta$  se a distribuição condicional da amostra ( $X_1, X_2, \dots, X_n$ ) dado o valor observado  $\hat{\theta} = t$ , não depende de  $\theta$ .

Um estimador  $\hat{\theta}$  diz-se suficiente se extrai da amostra toda a informação que esta contém sobre o parâmetro  $\theta$  de tal maneira que, dado o valor observado  $\hat{\theta} = t$ , o conhecimento dos valores observados para os elementos da amostra nada acrescenta sobre  $\theta$ . Os estimadores suficientes gozam da propriedade de retirar da amostra toda a informação relevante sobre o parâmetro (Murteira et al., 2001).

- Exemplo

Um estimador de máxima verossimilhança será um estimador suficiente.

Seja ( $X_1, X_2, \dots, X_n$ ) uma amostra aleatória simples de uma distribuição Normal ou Gaussiana,  $X \cap N(\mu, \sigma)$ . Dessa forma estimar os parâmetros  $\mu$  e  $\sigma$  pelo método da Máxima verossimilhança, está se utilizando estimadores suficientes, senão vejam a justificativa a seguir: A resolução se dá da seguinte maneira: Seja a função densidade de probabilidade (f.d.p.):

$$f(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \text{ para } \sigma > 0$$

$$\text{Função de verossimilhança: } L(\mu, \sigma^2, x) = \prod_{i=1}^n f(x_i; \mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2} =$$

$$= \frac{1}{\left[\sqrt{2\pi\sigma^2}\right]^n} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2} = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2}$$

Logaritmo da Função de Verossimilhança:

$$\ln L(\mu, \sigma^2; x) = -\frac{n}{2} (\ln 2 + \ln \pi + \ln \sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

Condição de 1ª ordem:

$$\begin{cases} \frac{\partial \ln L(\mu, \sigma^2; x)}{\partial \mu} = 0 \\ \frac{\partial \ln L(\mu, \sigma^2; x)}{\partial \sigma^2} = 0 \end{cases} \begin{cases} -\frac{1}{2\sigma^2} \left( -2 \sum_{i=1}^n x_i + 2n\mu \right) = 0 \\ -\frac{n}{2} \frac{1}{\sigma^2} + \sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^4} = 0 \end{cases} \begin{cases} \sum_{i=1}^n x_i - n\mu = 0 \\ -n\sigma^2 + \sum_{i=1}^n (x_i - \mu)^2 = 0 \end{cases} \therefore$$

$$\begin{cases} \mu = \sum_{i=1}^n \frac{x_i}{n} \\ \sigma^2 = \sum_{i=1}^n \frac{(x_i - \mu)^2}{n} \end{cases} \therefore \begin{cases} \mu = \bar{x} \\ \sigma^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n} \end{cases} \therefore \begin{cases} \mu = \bar{x} \\ \sigma^2 = S^2 \end{cases}$$

Condição de 2ª ordem: 
$$\begin{cases} \frac{\partial^2 \ln L(\mu, \sigma^2; x)}{\partial \mu^2} = -\frac{1}{2\sigma^2} 2n < 0 \\ \frac{\partial^2 \ln L(\mu, \sigma^2; x)}{\partial \sigma^4} = \frac{n}{2} \frac{1}{\sigma^4} - \sum_{i=1}^n \frac{(x_i - \mu)^2}{2} \frac{2\sigma^2}{\sigma^8} < 0 \end{cases}$$

Portanto, os estimadores de máxima verossimilhança obtidos foram:

$$\begin{cases} \hat{\mu} = \bar{X} = \sum_{i=1}^n \frac{X_i}{n} \\ \hat{\sigma}^2 = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{n} = \frac{n}{n-1} S^2 \end{cases}$$

**PROPRIEDADES, CARACTERÍSTICAS, TEOREMAS, COROLÁRIOS E OBSERVAÇÕES SOBRE A DISTRIBUIÇÃO TEÓRICA DE PROBABILIDADE NORMAL, NORMAL DOS ERROS, DE GAUSS (GAUSSIANA), SIMÉTRICA, MESOCÚRTICA, EM FORMA DE SINO, EM FORMA DE CHAPÉU DE NAPOLEÃO, EM FORMA CAMPANULAR OU DE CÂMPANULA:**

- i) É uma curva simétrica em relação a  $x = \mu$  ou  $z = 0$ , isto é  $\mathcal{Y}_1 = 0$  (coeficiente de assimetria igual a zero);
- ii) Tem forma de sino, campanular ou de chapéu de Napoleão;
- iii) A curva Normal tem o grau ou coeficiente de curtose (achatamento) do tipo mesocúrtica, sendo o valor do coeficiente de curtose deste modelo igual a 3;
- iv) Tem um máximo no ponto  $x = \mu$  (média), ou seja, a moda  $= \mu$  e o valor da ordenada máxima é  $\frac{1}{\sigma\sqrt{2\pi}}$ ;
- v) As três medidas de posição (média, mediana e moda) se confundem no ponto de máximo da curva, isto é  $\mu = M_d = M_o$  em  $x = \mu$  ou  $z = 0$ ;
- vi) Tem dois pontos de inflexão cujas abscissas valem  $\mu + \sigma$  e  $\mu - \sigma$ , e cujos pontos (X, Y) são os de coordenadas,  $\left[ (\mu - \sigma); \frac{1}{\sigma\sqrt{2\pi}e} \right]$  e  $\left[ (\mu + \sigma); \frac{1}{\sigma\sqrt{2\pi}e} \right]$ , que compreendem 68,27 % do total de dados;

Observação importante: Pontos de inflexão: Os pontos onde a função muda a forma da concavidade chamam-se de inflexão e são caracterizados pelas condições de que as derivadas segunda da função sejam nulas ou iguais a zero e diferente de zero, ou seja,  $y''(x_0) = 0$  e  $y'''(x_0) \neq 0$ . Dependendo do valor da

primeira derivada da função em relação a um ponto  $x_0$  isto é,  $y'(x_0) = 0$ , e dos valores das derivadas primeira, segunda e terceira, têm-se quatro situações diferentes. a e b é quando  $y'(x_0) = 0$ ;  $y''(x_0) = 0$  e  $y'''(x_0) \neq 0$ , o c é quando  $y'(x_0) > 0$ ;  $y''(x_0) = 0$  e  $y'''(x_0) \neq 0$  e finalmente o caso d é quando  $y'(x_0) < 0$ ;  $y''(x_0) = 0$  e  $y'''(x_0) \neq 0$ . No caso do estudo da função da distribuição normal ou Gaussiana, dada por:

$$y(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x^2}$$

Esta função  $y(x)$  é derivável em  $\mathbb{R}$  (conjunto dos reais) e suas derivadas de 1ª e 2ª ordem são dadas por  $y'(x) = \frac{1}{\sqrt{2\pi}} \cdot (-x)e^{-\frac{1}{2}x^2}$  e  $y''(x) = \frac{1}{\sqrt{2\pi}} \cdot (x^2 - 1)e^{-\frac{1}{2}x^2}$ , os pontos de máximo e ou

mínimos dessa função, ou eventualmente de inflexão, se houver, são obtidos fazendo-se  $y'(x_0) = 0$ , isto é,

$$y'(x) = \frac{1}{\sqrt{2\pi}} \cdot (-x)e^{-\frac{1}{2}x^2} = 0, \text{ o que implica } x_0 = 0. \text{ A natureza do ponto } x_0 = 0 \text{ é obtida quando se determina}$$

o sinal de  $y''(x_0)$  isto é,  $y''(0) = -\frac{1}{\sqrt{2\pi}} < 0$ , e, portanto  $x_0 = 0$  é um ponto de máximo relativo de  $y(x)$ ,

cujos valor é dado por  $y(0) = \frac{1}{\sqrt{2\pi}}$ . Note que esta função não apresenta ponto de mínimo relativo e

inflexão do tipo preconizado no caso a e b. Para se determinar os pontos de inflexão sugeridos pelos casos c e d, deve-se impor  $y''(x_0) = 0$ , isto é,

$$y''(x) = \frac{1}{\sqrt{2\pi}} \cdot (x^2 - 1)e^{-\frac{1}{2}x^2} = 0, \text{ o que implica que } x^2 - 1 = 0 \text{ ou } x \neq \pm 1. \text{ A derivada de terceira ordem de}$$

$y(x)$  nesses pontos é diferente de zero, isto é,  $y'''(1) \neq 0$  e  $y'''(-1) \neq 0$ , o que garante que estes pontos são de inflexão. Por outro lado, se  $y'(-1) > 0$  e  $y'(1) < 0$ , no ponto -1, a inflexão é do tipo c e em 1, do tipo d. As funções constante, linear e afim não apresentam pontos de máximo, mínimo ou de inflexão. A função quadrática apresenta ponto de máximo ou mínimo, dependendo da forma de sua concavidade, porém, não de inflexão. As funções cúbicas apresentam, pelo menos, um ponto de inflexão, podendo ou não ter pontos de máximo ou mínimo relativo. As funções exponencial  $a^x$  e  $a^{-x}$  e logarítmica  $\log x$  ou  $\ln x$  não apresentam pontos de máximo, mínimo ou de inflexão. A função logística apresenta somente um ponto de inflexão.

Ou ainda Em cálculo diferencial, um ponto de inflexão ou simplesmente inflexão, é um ponto sobre uma curva na qual a curvatura (a derivada de segunda ordem) troca o sinal. A curva muda de ter curvatura côncava para cima (positiva) para concavidade para baixo (curvatura negativa), ou vice-versa. Pode-se comparar com a condução de um veículo ao longo de uma estrada sinuosa, sendo o ponto de inflexão aquele em que o volante é momentaneamente endireitado quando a curva muda da esquerda para a direita ou vice-versa.

- vii) A área total compreendida pela curva e o eixo dos  $x$  é igual a 1 ou 100%, como qualquer função densidade de probabilidade (f.d.p.) que é um modelo estatístico do tipo probabilístico, estocástico, casual, aleatório, randômico ou não determinístico, utilizado para caracterizar as distribuições teóricas contínuas de densidade de probabilidades;



- viii) A curva é assintótica em relação ao eixo das abscissas, ou seja, o eixo dos x. À medida que cresce ou decresce o valor de x, decresce o valor de y [f(x)] sem chegar a zero. Isto quer dizer que ambos extremos da curva se aproximam do eixo horizontal sem chegar a tocá-lo, ou seja,  $\lim_{x \rightarrow \pm\infty} f(x) = 0$ ;
- ix) A distribuição fica perfeitamente definida se for conhecido a média  $\mu$  e o desvio padrão  $\sigma$ , isto é, somente com estes parâmetros estatísticos pode-se calcular a altura da curva em qualquer ponto do eixo horizontal;
- x) A área sob a curva entre dois pontos é a probabilidade de uma variável normalmente distribuída tomar um valor entre esses pontos. As probabilidades referem-se sempre a intervalos.
- xi) Se  $X_i$  tem distribuição normal ou gaussiana, qualquer múltiplo de  $X_i$  também o terá. Esta demonstração é evidente, lembrando que, quando se multiplica uma variável por uma constante a sua média fica multiplicada pela constante e a sua variância pelo quadrado da constante. Ora, como a distribuição normal pode se apresentar com vários valores combinados diferentemente para a média e variância, um múltiplo de uma variável  $X_i$  com distribuição normal também se distribuirá normalmente;
- xii) Se  $X_i$  tem distribuição normal de média  $\mu$  e variância  $\sigma^2$ , qualquer função linear de  $X_i$  terá distribuição normal, ou se distribuirá normalmente, Seja  $X_i \in N(\mu; \sigma^2)$  e  $g(X_i) = A + B \cdot X_i$ , então  $g(X_i) = N[(A + B\mu); (B^2 \cdot \sigma^2)]$ , veja a demonstração a seguir: Como  $g(X_i) = A + B \cdot X_i$ , aplicando o operador esperança matemática vem:  $E[g(X_i)] = E(A + B \cdot X_i) = E(A) + B \cdot E(X_i) = A + B \cdot \mu$ . A variância de  $g(X_i) = E[g(X_i)^2] - \{E[g(X_i)]\}^2$ , ou

$$\begin{aligned} Var[g(X_i)] &= E[(A + B \cdot X_i)^2] - [A + B \cdot \mu]^2 = E[A^2 + B^2 X_i^2 + 2ABX_i] - [(A^2 + B^2 \mu^2 + 2AB\mu)] = \\ &= E(A^2) + B^2 E(X_i^2) + 2AB \cdot E(X_i) - A^2 - B^2 \mu^2 - 2AB\mu = \\ &= A^2 + B^2 \cdot E(X_i^2) + 2AB\mu - A^2 - B^2 \mu^2 - 2AB\mu, \end{aligned}$$

Que simplificando resulta:

$$Var[g(X_i)] = B^2 [E(X_i^2) - \mu^2] = B^2 \sigma^2$$

$$\text{Logo: } f[g(X_i)] = \frac{1}{B\sigma\sqrt{2\pi}} \cdot e^{-\frac{[g(X_i) - E[g(X_i)]]^2}{2B^2\sigma^2}}$$

$$f[g(X_i)] = \frac{1}{B\sigma\sqrt{2\pi}} \cdot e^{-\frac{[A + B \cdot X_i - A - B\mu]^2}{2B^2\sigma^2}}$$

$$f[g(X_i)] = \frac{1}{B\sigma\sqrt{2\pi}} \cdot e^{-\frac{[B(X_i - B\mu)]^2}{2B^2\sigma^2}} = \frac{1}{B\sigma\sqrt{2\pi}} \cdot e^{-\frac{B^2(X_i - \mu)^2}{2B^2\sigma^2}}$$

$$f[g(X_i)] = \frac{1}{B\sigma\sqrt{2\pi}} \cdot e^{-\frac{(X_i - \mu)^2}{2\sigma^2}} = \frac{1}{B} \left[ \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(X_i - \mu)^2}{2\sigma^2}} \right] \text{ ou } f[g(X_i)] = \frac{1}{B} \cdot f(X_i)$$

Como  $X_i \in N(\mu; \sigma^2)$  em função do item i resulta  $g(X_i) \in N(A + B\mu); (B^2\sigma^2)$ ;

- xiii) Soma de variáveis normais independentes: Se  $X = X_1 + X_2 + X_3 + \dots + X_n$ , uma variável aleatória formada pela soma de n variáveis normais independentes. Então se  $X_j \in N(\mu_j; \sigma_j^2)$ ,  $j = 1, 2, 3, \dots, n$ , e se

$$f(X_j) = \frac{1}{\sigma_j\sqrt{2\pi}} \cdot e^{-\frac{(X_j - \mu_j)^2}{2\sigma_j^2}} \text{ segue-se que } X \in N\left(\sum_{j=1}^n \mu_j; \sum_{j=1}^n \sigma_j^2\right), \text{ isto é, a soma de qualquer número}$$

de variáveis independentes com distribuição normal tem também distribuição normal. O Cálculo da média e da variância de X é mostrado a seguir:

$$E(X) = E(X_1 + X_2 + X_3 + X_4 + \dots + X_n) = E(X_1) + E(X_2) + E(X_3) + E(X_4) + \dots + E(X_n),$$

$$\text{ou } E(X) = \mu_1 + \mu_2 + \mu_3 + \mu_4 + \dots + \mu_n = \sum_{j=1}^n \mu_j \text{ e a variância é dada por,}$$

$$\begin{aligned} Var(X) &= Var(X_1 + X_2 + X_3 + X_4 + \dots + X_n) = \\ &= Var(X_1) + Var(X_2) + Var(X_3) + Var(X_4) + \dots + Var(X_n) + 2 \sum_{i \neq j} Cov(X_i; X_j) \end{aligned}$$

Como as variáveis são independentes, segue-se, por definição, que a covariância é nula ou. Seja,  $Cor(X_i; X_j) = 0$  e logo:  $Var(X) = \sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \sigma_4^2 + \dots + \sigma_n^2 = \sum_{j=1}^n \sigma_j^2$ .

- xiv) Um caso particular mostrado no item iii anterior de grande importância em problemas de inferência refere-se ao caso em que as amostras têm mesma média  $\mu$  ( $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \dots = \mu_n$ ) e mesma variância  $\sigma^2$  constante; ( $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \dots = \sigma_n^2$ ); esta última hipótese é conhecida com o nome de hipótese de homocedasticidade. Dentro dessas duas hipóteses,  $X \in N(n, \mu; n, \sigma^2)$  e é evidente que tem distribuição normal de acordo com o resultado do item i anterior.

- xv) Como qualquer função linear de uma variável com distribuição normal é também normal, conclui-se dos itens ii e iii anteriores que uma função linear  $g(X) = A + A_1 X_1 + A_2 X_2 + A_3 X_3 + A_4 X_4 + \dots + A_n X_n$  de variáveis normais independentes tem também distribuição normal.

- xvi) Se  $X = X_1 + X_2 + X_3 + \dots + X_n$  é formada pela soma de n variáveis normais independentes de mesma média  $\mu$  e mesma variância (constante)  $\sigma^2$ , então a média amostral

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} = \frac{1}{n} \sum_{i=1}^n X_i \text{ também se distribuirá normalmente com média da } \mu \text{ população e variância } \frac{\sigma^2}{n}. \text{ A demonstração da distribuição normal para a média amostral } \bar{X} \text{ é evidente, bastante que}$$

se recorde dos resultados das proposições i e v anteriores. O cálculo da média e da variância de  $\bar{X}$  é dado a seguir:

$$E(\bar{X}) = E\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} [X_1 + X_2 + X_3 + \dots + X_n]$$

$$E(\bar{X}) = \frac{1}{n} [E(X_1) + E(X_2) + E(X_3) + \dots + E(X_n)]$$

$$E(\bar{X}) = \frac{1}{n} \left[ \underbrace{\mu + \mu + \mu + \dots + \mu}_{n \text{ vezes}} \right] \text{ ou } E(\bar{X}) = \frac{1}{n} \cdot n\mu = \mu$$

A variância de  $\bar{X}$  é calculada como:

$$Var(\bar{X}) = E[\bar{X} - \mu]^2 = Var\left[\frac{\sum_{i=1}^n X_i}{n}\right] = \frac{1}{n^2} Var[X_1 + X_2 + X_3 + \dots + X_n]$$

$$Var(\bar{X}) = \frac{1}{n^2} \left[ Var(X_1) + Var(X_2) + Var(X_3) + \dots + Var(X_n) + 2 \sum_{i \neq j} Cov(X_i; Y_j) \right]. \text{ Como pela}$$

hipótese de independência das variáveis aleatórias normais  $Cov(X_i; Y_j) = 0 \quad \forall i \neq j$ , e  $Var(X_i) = \sigma^2 = \text{constante} \quad \forall i$ , resulta:

$$Var(\bar{X}) = \frac{1}{n^2} \left[ \underbrace{\sigma^2 + \sigma^2 + \sigma^2 + \dots + \sigma^2}_{n \text{ vezes}} \right] = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

xvii) Teorema do Limite Central: Quaisquer que sejam as distribuições das variáveis aleatórias independentes, distribuições estas conhecidas ou desconhecidas e sujeitas a certas condições muito gerais, a soma dessas variáveis é assintoticamente normal, isto é, converge para a normalidade.

xviii) Distribuição Teórica de Probabilidade Log Normal: Se  $X_i$  é uma variável aleatória que assume valores apenas positivos e se  $Log X_i = Y_i \cap N(\mu; \sigma^2)$ , então  $X_i = Y_i \cap \log N(\mu; \sigma^2)$ .

Outros autores anunciam os teoremas e corolário como mostrado a seguir:

xix) Teorema da aditividade: distribuição normal.

Se  $X_i, i = 1, 2, \dots, n$ , são variáveis aleatórias independentes e  $X_i \cap N(\mu_i, \sigma_i) \quad \forall_{i \in \{1, 2, \dots, n\}}$ , então

$$\sum_{i=1}^n a_i X_i \cap N\left(\sum_{i=1}^n a_i \mu_i, \sqrt{\sum_{i=1}^n a_i^2 \sigma_i^2}\right)$$

ix.1) Corolário: se  $X_i, i = 1, 2, 3, \dots, n$ , são variáveis aleatórias independentes e possui distribuição normal com média e desvio padrão isto é,  $X_i \cap N(\mu, \sigma)$ , então:

$$\sum_{i=1}^n X_i \cap N(n\mu, \sigma\sqrt{n}), \text{ sendo que } \sum_{i=1}^n X_i = S_n = X_1 + X_2 + X_3 + \dots + X_n$$

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \cap N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Na prática n suficiente grande é quando  $n \geq 30$ .

xx) Uma combinação linear de variáveis aleatórias independentes e com distribuição normal é ainda uma variável aleatória com distribuição normal. Caso particular:

$$X \sim N(\mu, \sigma^2) \Rightarrow Y = a + bX \sim N(a + b\mu, b^2\sigma^2)$$

xxi) Teorema do limite central (TCL) ou Teorema central do limite:

Se  $X_1, X_2, \dots, X_n$ , são variáveis aleatórias independentes e identicamente distribuídas (i.i.d.)

$$E(X_i) = \mu \quad e \quad Var(X_i) = \sigma^2 \quad \forall_{i \in \{1, 2, \dots, n\}} \Rightarrow, \text{ então } \sum_{i=1}^n X_i \cap N(n\mu, n\sigma^2)$$

Na prática n suficiente grande é quando  $n \geq 30$ .

**RELAÇÃO NOMINAL DE 330 PROGRAMAS, PACOTES, APLICATIVOS OU SOFTWARES UTILIZADOS ISOLADOS OU QUANDO REQUEREM OUTROS PACOTES PARA FUNCIONAR, APLICADOS NAS ÁREAS DE ESTATÍSTICA, AUDITORIA, CONTABILOMETRIA, ENGENHARIA, ATUÁRIA, ADMINISTRAÇÃO, ECONOMETRIA, GEOESTATÍSTICA, CÁLCULO E MATEMÁTICA, GENÉTICA, NA CONSTRUÇÃO DE MAPAS E GRÁFICOS, NA EDIÇÃO DE BANCO DE DADOS, DE IMAGENS E DE FOTOS, EM SIMULAÇÃO, E COMO EDITOR DE TEXTOS. VALE RESALTAR QUE ALGUNS DESSES PACOTES POSSUEM VERSÕES PARA ESTUDANTES, DISPONÍVEIS A PREÇOS SUBSTANCIALMENTE REDUZIDOS OUTROS SÃO GRATUÍTOS OU LIVRES, DE CÓDIGO ABERTO E OUTROS SÃO PAGOS OU COMERCIAIS, SENDO QUE MUITOS DELES POSSUEM VERSÕES DEMONSTRATIVAS E VERSÕES PRÓPRIAS PARA ESTUDANTES, PODENDO O USUÁRIO OBTER INFORMAÇÕES OU BAIXAR (FAZER DOWNLOAD) DO PROGRAMA POR MEIO DE QUALQUER SITE DE BUSCA NA INTERNET TAIS COMO O GOOGLE, O YAHOO, DENTRE OUTROS. COMO ESSES PACOTES SÃO ATUALIZADOS REGULARMENTE, ACONSELHA-SE AO USUÁRIO CONSULTAR O VENDEDOR OU AUTORES A RESPEITO DA MAIS RECENTE VERSÃO.**

1. SAS® (VERSÃO 9.2)-SAS INSTITUTE INC.: SAS INSTITUTE INC., P.O. BOX 8000, SAS CIRCLE, CARY, NC 27511-8000. (919) 467-8000, ou SAS Institute Inc., SAS Campus Drive, Cary, NC 27513 – 2414, U. S. A. [ <http://www.sas.com.br>]. SAS INSTITUTE: Sistema SAS em muitas plataformas, JMP, compilador C. Leque variado de material do SAS, vista geral do software, e serviços de suporte incluindo uma base de dados pesquisável por SAS Notes e documentos de suporte técnico;

2. SPSS®-STATISTICAL PACKAGE FOR THE SOCIAL SCIENCES - PACOTE ESTATÍSTICO PARA AS CIÊNCIAS SOCIAIS): SPSS INC., 444 N. MICHIGAN AVE., CHICAGO, IL 60611. (312) 329 3600. (ESTADOS UNIDOS);

3. MINITAB'15- STATISTICAL SOFTWARE: MINITAB, 3081 ENTERPRISE DRIVE, STATE COLLEGE, PA 16801. (814) 238-3280. (ESTADOS UNIDOS) ou Minitab Inc. , 3081 Enterprise Drive, state College, PA 16801 – 3008, U. S. A. [<http://www.minitab.com>];
4. S-PLUS<sup>®</sup> Insightful Corporation, 1700 Westlake Avenue North, Suite 500, Seattle, WA 98109 – 3044, U. S. A. [<http://www.insightful.com/products/splus/splus2000/splusstdintro.html>];
5. STATISTICA-STATSOFT<sup>®</sup>;
6. R-CRAN.R-PROJECT.ORG- THE COMPREHENSIVE R ARCHIVE NETWORK;
7. VISTA;
8. OPENSTAT;
9. WINIDAMS;
10. SSP (SMITH'S STATISTICAL PACKAGE);
11. DATAPLOT;
12. WEB SAT;
13. REGRESS<sup>+</sup>;
14. SCILAB;
15. STATISTICAL SOFTWARE;
16. IRRISTAT;
17. STATLETS;
18. STATGRAPHICS PLUS – STSC INC. , 2115 E. JEFFERSON ST. , ROCKVILLE, MD 20852. (800) 592 0050. SGC–STATISTICAL GRAPHICS CORPORATION - STATGRAPHICS: Pacote da análise estatística para MS Windows;
19. STAT SIMPLE;
20. QUICKSTAT;
21. INSTAT;
22. NCSS-2000;
23. STATISTICAL ANALYSIS SYSTEM AND PASS-2000 POWER AND SAMPLE SIZE PROGRAM;
24. PRISM;
25. SAMPLE - SOFTWARE USADO JUNTO COM O PACOTE MINITAB PARA ANÁLISE DE DADOS OBTIDOS POR MEIO DE AMOSTRAGEM PROBABILÍSTICA, UFMG, BRASIL;
26. GEOEAS;

27. GLIM - GENERALISED LINEAR INTERACTIVE MODELLING. Numerical Algorithms Group, NAG ltd, Wilkinson House, Jordan Hill Road, Oxford, OX2 8DR, United Kingdom [<http://www.nag.co.uk>];
28. SPRING - (SISTEMA DE PROCESSAMENTO DE INFORMAÇÕES GEOREFERENCIADAS), PARA PLATAFORMA WINDOWS E UNIX, UTILIZADO PARA SISTEMA DE INFORMAÇÃO GEOGRÁFICA COM FUNÇÕES DE PROCESSAMENTO DE IMAGENS E ANÁLISE ESPACIAL. INCLUI FUNÇÕES PARA ANALISE DE PADRÕES PONTUAIS E GEOESTATÍSTICA. ELABORADO PELO INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS (INPE), DIVISÃO DE PROCESSAMENTO DE IMAGENS;
29. SPRING WEB;
30. TERRALIB;
31. TERRAVIEW-SOFTWARE DE VISUALIZAÇÃO DE MAPAS, DESENVOLVIDOS NO PROJETO DENOMINADO DE SAUDÁVEL;
32. WINMAT;
33. GRAPHMATICA;
34. WINPLOT;
35. MATHGV;
36. STATIFIT;
37. JMP-JUMP VERSÃO 7;
38. GQMOL-UFV;
39. GBOL-UFV;
40. SYSTAT – SYSTAT INC., 1800 SHERMAN AVE., EVANSTON, IL 60201. (708) 864-5670. (ESTADOS UNIDOS), ou ainda SPSS Science, 233 S. Wacker Drive, 11th Floor, Chicago, IL 60606 – 6307, U. S. A. [<http://www.spssscience.com/SYSTAT/> ou <http://www.systat.com/>];
41. GEOR – BIBLIOTECA DE ANÁLISE GEOESTATÍSTICA, REQUER O PACOTE R DE AUTORIA DO PROFESSOR PAULO JUSTINIANO RIBEIRO JÚNIOR DO DEPARTAMENTO DE ESTATÍSTICA DA UFPR, AND PETER DIGGLE DA LANCASTER UNIVERSITY, INGLATERRA;
42. GEORGLM–EXTENSÃO DO GEOR PARA MODELOS LINEARES GENERALIZADOS ESPACIAIS, DE AUTORIA DE OLÉ CHRISTENSEN DA LANCASTER UNIVERSITY AND PAULO JUSTINIANO RIBEIRO JÚNIOR DO DEPARTAMENTO DE ESTATÍSTICA DA UFPR;
43. PROGRAMA ESTABILIDADE-UFLA, LAVRAS, MG, (PROFESSOR DR. DANIEL FURTADO FERREIRA);
44. PROLIN-UFV;
45. PESTALIS VERSÃO 1.0;

46. ESTATD<sup>+</sup>-PARTE DO LIVRO DE ESTATÍSTICA DA EDITORA ÉRICA-LTDA. INTITULADO ESTATÍSTICA DESCRITIVA DE AUTORIA DE CARLOS TAKEO AKANIME & ROBERTO KATSUHIRO YAMAMOTO, 1998, SÃO PAULO,;
47. STAROFFICE;
48. SURFER;
49. IDRISI;
50. EPI INFO- CRIAÇÃO DE BANCO DE DADOS: CDC-CENTERS FOR DISEASE CONTROL AND PREVENTION, COM DISTRIBUIÇÃO É GRATUITA;
51. MICROSOFT OFFICE EXCEL 2007<sup>M</sup>- ©2007 MICROSOFT CORPORATION;
- 52.AXUM, MATHSOFT;
- 53.FIG. P, BIOSOFT, FERGUSON;
- 54.GRAPHER, GOLDEN SOFTWARE;
- 55.ORIGIN, MICROCAL SOFTWARE INC.,;
- 56.MATLAB – MATH WORKS INC. , 20 N. MAIN ST. , SHERBORN, MA 01770. (617) 653 – 1415. (ESTADOS UNIDOS);
57. MAPLE, <http://www.maplesoft.com/>;
58. MATHEMATICA, <http://www.wolfram.com/>;
59. LAB FIT-UFCG, PB, BRASIL;
60. VTFIT;
61. ECLIMA-UFV;
62. SEAF-UFMG;
63. ALEA-UFMG;
64. CURVE EXPERT: (WINDOWS) É SOFTWARE QUE AJUSTA CURVAS EM CONJUNTO DE PONTOS NO PLANO (POR EXEMPLO, COLETA DE DADOS NUMÉRICOS), VIA MODELOS DE REGRESSÃO-LINEAR E NÃO-LINEAR- E DIFERENTES INTERPOLAÇÕES;
65. SIGMAPLOT, JANDEL SCIENTIFIC;
66. BIOESTAT-FUNDAÇÃO MAMIRAUÁ, PARÁ, BRASIL;
67. SESTATNET;
68. TABLE CURVE 2D- JANDEL SCIENTIFIC;
69. TABLE CURVE 3D- JANDEL SCIENTIFIC;
70. CURVE FITTING;

71. SISVAR-UFLA, MG, BRASIL;
72. SAEG-UFV;
73. SANEST-UFPEL;
74. GENES-UFV;
75. PCTEX:LINGUAGEM LATEX- COMPILADOR PRÓPRIO;
76. CORELDRAW-© 2006- COREL CORPORATION;
77. AUTOCAD 2008-AUTOCAD® 2008;
78. TNT MIPS-TNT MAP AND IMAGE PROCESSING SYSTEM;
79. MICROSIRIS;
80. MACANOVA;
81. LISP-STAT;
82. KYPLOT;
83. INSTAT<sup>+</sup>
84. ADE-4;
85. TANAGRA;
86. DEMETRA;
87. JMULTI;
88. G7;
89. TRAMO-SEATS;
90. MATRIXER;
91. FACTOR;
92. EASYREG INT.;
93. EZANOVA;
94. ARC;
95. ASSISTAT;
96. CMAP;
97. EPI DATA: CRIAÇÃO DE BANCOS DE DADOS UTILIZA OS MESMOS COMANDOS DO EPI INFO, SÓ QUE É CONSTRUÍDO NO AMBIENTE WINDOWS. A INICIATIVA PARA A CONSTRUÇÃO DESSE PROGRAMA FOI DE JENS M.LAURITSEN, SENDO, ENTÃO, LANÇADO PELA ORGANIZAÇÃO (SEM FINS LUCRATIVOS) “THE EPIDATA ASSOCIATION”, DE ODENSE, ALEMANHA. AS FACILIDADES



DE EXPORTAÇÃO (PARA SAS, SPSS, EXCEL E OUTROS) E DE UTILIZAÇÃO DO PROGRAMA SÃO SEMELHANTES AO EPI INFO, EMBORA, ATÉ A VERSÃO 2.1A, NÃO HOUVESSE SIDO IMPLEMENTADAS FERRAMENTAS ESTATÍSTICAS PARA A ANÁLISE DE DADOS;

98. EULER;
99. ESTA+;
100. STATEASY;
101. STATCALC;
102. SCILAB;
103. SALSTAT (BETA);
104. MONDRIAN;
105. STATIST;
106. STATISTICAL LAB;
107. PS;
108. MOREPOWER;
109. EASYSAMPLE;
110. AM;
111. WINSTATS;
112. STATTUCINO;
113. XLSTAT;
114. WEKA;
115. ROSETTA;
116. STATISTICS 101;
117. RUNDOM PRO;
118. BUGS -WINBUGS É UM SOFTWARE DE DOMÍNIO PÚBLICO PARA A ANÁLISE BAYESIANA DE MODELOS COMPLEXOS USANDO MÉTODOS MCMC, DISPONÍVEL NA PÁGINA DO MEDICAL RESEARCH COUNCIL, BIOSTATISTICS UNIT;
119. MACANOVA;
120. STATIBOT;
121. MICROSOFT OFFICE WORD 2007<sup>M</sup>- ©2007 MICROSOFT CORPORATION: FERRAMENTA MICROSOFT EQUATION;

122. MICROSOFT MATHTYPE- ©2007 MICROSOFT CORPORATION;
123. WINSHELL (FREEWARE), UTILIZA COMO COMPILADOR O MIKTEX;
124. WINEDT (SHAREWARE), UTILIZA COMO COMPILADOR O MIKTEX;
125. TEXNICCENTER (FREEWARE) UTILIZA COMO COMPILADOR O MIKTEX;
126. EMACS + AUCTEX—O EMACS É UM EDITOR DE TEXTO BASTANTE FLEXÍVEL (É PROGRAMÁVEL NUMA VARIANTE PRÓPRIA DO LISP) SENDO QUE O AUCTEX É UM MODO (LA) TEX MUITO INTERESSANTE;
127. GNUPLOT—PROGRAMA PARA A CONSTRUÇÃO DE GRÁFICOS DE FUNÇÕES (2D E 3D) COM PRODUÇÃO DE ARQUIVOS DE SAÍDA EM MODO LATEX, ENTRE OUTROS;
128. XFIG, QFIG, TEXCAD—PROGRAMAS DE DESENHO ESQUEMÁTICO COM PRODUÇÃO DE ARQUIVOS DE SAÍDA EM MODO LATEX, ENTRE OUTROS;
129. WINEDT-WWW.WINEDT.COM;
130. GMC 2002: GMC@FORP.USP.BR;
131. PRIMER OF BIOESTATISTICS THE PROGRAM;
132. STAT;
133. EGGRAPHER-EQUATION GRAPHER WITH REGRESSION ANALYZER INSTALLATION;
134. EMACS SPEAKS STATISTICS-EDITOR DE TEXTO:HTTP://ESS.STAT.WISC.EDU/
135. XGOBI: SISTEMA PARA GRÁFICOS:HTTP://WWW.RESEARCH.ATT.COM/AREAS/STAT//XGOBI/;
136. TERRASTAT-APLIATIVO PARA VISUALIZAÇÃO E ANÁLISE ESTATÍSTICA DE DADOS;
137. SKATER (SPATIAL K LUSTER ANALYSIS BY TREE EDGE REMOVAL);
138. PAGER MAKER-ADOBE PAGER MAKER: SOFTWARE DE EDITORAÇÃO ELETRÔNICA;
138. PHOTOSHOP-ADOBE PHOTOSHOP:EDITOR DE IMAGEM
139. PROGRAMA GS+5: GSBETA-PROGRAMA PARA ANÁLISE GEOESTATÍSTICA;
140. BAYES-PARA PLATAFORMA WINDOWS: SOFTWARE PARA INFERÊNCIA BAYESIANA BASEADA EM TÉCNICAS MCMC. AJUSTA MODELOS LINEARES GENERALIZADOS COM COMPONENTES ESPACIAIS E TEMPORAIS. CUJOS AUTORES SÃO: ANDREAS BREZGER, THOMAS KNEIB AND STEFAN LANG (LUDWIG MAXIMILIANS UNIVERSITY MUNICH);
141. STATPRO - PENTON SOFTWARE INC., 420 LEXINGTON AVE., SUITE 2846, NEW YORK, NY 10017. (800) 211-3414. (ESTADOS UNIDOS);
142. STATA - COMPUTING RESOURCE CENTER, 10801 NATIONAL BLVD., 3<sup>RD</sup> FLOOR, LOS ANGELES, CA 90064. (800) STATAPC. (ESTADOS UNIDOS) ou Stata Corporation, 4905 Lakeway Drive, College station, Texas 77845, U. S. A. [<http://www.stata.com>];

143. SORITEC - THE SORITEC GROUP INC., P.O. BOX 2939, 8136 OLD KEENE MILL ROAD, SPRINGFIELD, VA 22152. (703) 569-1400. (ESTADOS UNIDOS);
144. SHAZAM - KENNETH J. WHITE, DEPARTMENT OF ECONOMICS, UNIVERSITY OF BRITISH COLUMBIA, VANCOUVER, BC V6T 1Y2. (604) 228- 5062 (CANADÁ);
145. RATS - VAR ECONOMETRICS, P.O. BOX 1818, EVANSTON, IL 60624-1818. (708) 864-8772. (ESTADOS UNIDOS), RATS: <http://www.estima.com/ratsmain.shtml>;
146. PC-TSP - TSP INTERNATIONAL, P.O. BOX 61015, PALO ALTO, CA 94306. (415) 326-1927. (ESTADOS UNIDOS);
147. PC-GIVE - UNIVERSITY OF OXFORD, INSTITUTE OF ECONOMICS AND STATISTICS, ST. CROSS BUILDING, MANOR RD, OXFORD OX1 3UL. U.K. (REINO UNIDO);
148. MICRO TSP - QUANTITATIVE MICRO SOFTWARE, 4521 CAMPUS DRIVE, SUITE 336, IRVINE, CA 92715. (714) 856-3368. (ESTADOS UNIDOS);
149. BMDP/PC - BMDP STATISTICAL SOFTWARE INC., 1440 SEPULVEDA BLVD., SUITE 316, LOS ANGELES, CA 90025. (213) 479-7799. (ESTADOS UNIDOS);
150. DATA FIT - OXFORD ELETRONIC PUBLISHING COMPANY, OXFORD UNIVERSITY PRESS, WALTON STREET, OXFORD OX2 6DP, U.K. (REINO UNIDO);
151. ESP - ECONOMIC SOFTWARE PACKAGE, 76 BEDFORD ST., SUITE 33, LEXINGTON, MA 02173. (617) 861-8852. (ESTADOS UNIDOS);
152. ET - WILLIAM H. GREENE, STERN GRADUATE SCHOOL OF BUSINESS, NEW YORK UNIVERSITY, 100 TRINITY PLACE, NEW YORK, NY. (212) 285-6164. (ESTADOS UNIDOS);
153. GAUSS - APTECH SYSTEMS INC., 26250 196TH PLACE SE, KENT, WA 98042. (206) 631-6679. (ESTADOS UNIDOS);
154. LIMDEP - WILLIAM H. GREENE, STERN GRADUATE SCHOOL OF BUSINESS, NEW YORK UNIVERSITY, 100 TRINITY PLACE, NEW YORK, NY. (212) 285-6164. (ESTADOS UNIDOS);
155. AGROMET[UNIX / WINDOWS/ C++ CODES] (V-K-C-2D);
156. COSIM [WINDOWS / FORTRAN] (S-2D);
157. EXPLOSTAT [WINDOWS] (V-K-O-C-2D-G);
158. E{Z}-KRIGING [WINDOWS] (V-K-2D);
159. GCOSIM3D [WINDOWS / C CODES] (S-3D);
160. GEO-EAS [WINDOWS] (V-K-2D);
161. GEOPACK [WINDOWS] (V-K-C-2D);
162. GEODA [WINDOWS] (O-2D);
163. GEOSTATISTICAL TOOLBOX [DOS] (V-K-C-3D);

164. GMT [UNIX / C CODES] (O-2D);
165. GRNN [WINDOWS] (O-2D);
166. GSLIB [FORTRAN 77] (V-K-C-3D-S);
167. GSTAT [LINUX, WINDOWS / C CODES/ R] (V-K-C-3D-S);
168. ISIM3D [WINDOWS / C CODES] (S-3D);
169. KRIGING [UNIX / C CODES] (K-2D);
170. SADA [WINDOWS] (V-K-O-3D-G);
171. SAGA GIS [WINDOWS] (V-K-O-2D-G);
172. SGS [LINUX, C CODES] (K-S-2D);
173. S-GEMS [WINDOWS, LINUX, C++ CODES] (V-K-C-3D-S);
174. SPHEREKIT [UNIX, C CODES] (K-O-2D);
175. SURFACE III [MAC] (K-O-2D);
176. SURFIT [WINDOWS, C++ CODES] (O);
177. UNCERT [UNIX / C CODES] (V-K-O-C-2D-G);
178. VARIOWIN [WINDOWS] (V-2D);
179. VESPER [WINDOWS] (V-K-2D);
180. DATAMINE;
181. GEOSTOKOS TOOLKIT & ECOSSE;
182. GEOVARIANCES' ISATIS;
183. GEOSTAT SYSTEMS INT. INC.;
184. GS+ FROM GAMMADESIGN;
185. SURFER FROM GOLDEN SOFTWARE;
186. SAGE2001;
187. LYNX GEOSYSTEMS;
188. GEOSTATSOFFICE;
189. SURGE;
190. CRIMESTAT - SOFTWARE PARA PLATAFORMA WINDOWS, UTILIZADO NA ANÁLISE DE INCIDENTES DE CRIMES REFERENCIADOS ESPACIALMENTE, DE AUTORIA DE NED LEVINE & ASSOCIATES;

191. CLUSTER - PROGRAMA EM DOS, PARA DETECÇÃO DE CLUSTERS ESPACIAIS E ESPAÇO-TEMPORAIS, DE AUTORIA DA AGENCY FOR TOXIC SUBSTANCES AND DISEASE REGISTRY (ATSDR);

192. DISMAPWIN - SOFTWARE PARA PLATAFORMA WINDOWS, EMPREGADO EM ANÁLISE ESTATÍSTICA DE MAPAS EPIDEMIOLÓGICOS, DE AUTORIA DE PETER SCHLATTMANN (FREE UNIVERSITY OF BERLIN);

193. GAM: CLUSTER HUNTING SOFTWARE (GEOGRAPHICAL ANALYSIS MACHINE V 2.0) - SOFTWARE PARA PLATAFORMA WINDWOS, UTILIZADO EM ANÁLISE EXPLORATÓRIA ESPACIAL DE DADOS PONTUAIS OU PEQUENAS ÁREAS, CRIADO PELO CENTRE FOR COMPUTATIONAL GEOGRAPHY, DA UNIVERSITY OF LEEDS;

194. GEM (GEOGRAPHICAL EXPLORATION MACHINE) – SOFTWARE COM VERSÃO ON – LINE PARA DETECÇÃO DE CLUSTERS, DE AUTORIA DO STAN OPENSHAW AND IAN TURTON (UNIVERSITY OF LEEDS);

195. GEODA - SOFTWARE PARA PLATAFORMA WINDOWS, DE ANÁLISE EXPLORATÓRIA DE DADOS ESPACIAIS EM LÁTICES, DE AUTORIA DA SPATIAL ANALYSIS LABORATORY, DA UNIVERSITY OF ILLINOIS;

196. GEOSTAT OFFICE – SOFTWARE PARA PLATAFORMA WINDOWS DE ANÁLISE DE DADOS REFERENCIADOS ESPACIALMENTE EM 2D, DE AUTORIA DO NUCLEAR SAFETY INSTITUTE (IBRAE), RUSSIA;

197. SADIE (SPATIAL ANALYSIS BY DISTANCE INDICES) – SOFTWARE PARA PLATAFORMA WINDOWS, EMPREGADO PARA ANÁLISE ESPACIAL EM ECOLOGIA, ESPECIALMETE PARA DADOS DE CONTAGEM, DE AUTORIA DE JOE PERRY (INSTITUTE OF ARABLE CROP RESEARCH, RHOTAMSTED, INGLATERRA);

198. SAGE (SPATIAL ANALYSIS IN A GIS ENVIRONMENT) – SOFTWARE PARA PLATAFORMA SUN WORKSTATION WITH SOLARIS, EMPREGADO PARA GIS INTEGRADO PARA ANÁLISE DE DADOS DA ÁREA DE SAÚDE. INCLUI FERRAMENTAS DE ESTATÍSTICA ESPACIAL, DE AUTORIA DO CENTRE FOR GEOGRAPHIC INFORMATION AND SPATIAL ANALYSIS, DA UNIVERSITY OF SHEFFIELD;

199. SATSCAN - SOFTWARE PARA PLATAFORMA WINDOWS, APLICADO NA ANÁLISE DE DADOS PONTUAIS ESPACIAIS E ESPAÇO-TEMPORAIS, E DETECÇÃO DE CLUSTERS, DE AUTORIA DA MARTIN KULLDORFF DA HARVARD MEDICAL SCHOLL;

200. SCHOOL COP (SCHOOL CRIME OPERATION PACKAGE) – SOFTWARE PARA PLATAFORMA WINDOWS, UTILIZADOS NA ANÁLISE E MAPEAMENTO DE INCIDENTES QUE ACONTECEM EM E AO REDOR DAS ESCOLAS, DE AUTORIA DO NATIONAL INSTITUTE OF JUSTICE (USA) AND ABT ASSOCIATES INC.;

201. SPPA (SPATIAL POINT PATTERN ANALYSIS) – SOFTWARE PARA PLATAFORMA DOS E WINDOWS, UTILIZADO NA ANÁLISE DE PADRÕES ESPACIAIS PONTUAIS BASEADA NA FUNÇÃO K DE RIPLEY DE AUTORIA DE PETER HAASE;

202. TIMEMAP - SOFTWARE PARA PLATAFORMA WINDOWS, USADO NO MAPEAMENTO E ANÁLISE ESPACIAL, DE AUTORIA DO ARCHAEOLOGICAL COMPUTING LABORATORY, DA UNIVERSITY OF SYDNEY;

203. FIELDS – SOFTWARE QUE FUNCIONA REQUERENDO O PACOTE R PARA AJUSTE DE CURVAS E FUNÇÕES COM ÊNFASE EM DADOS ESPACIALMENTE REFERENCIADOS, DE AUTORIA DE DOUGLAS NYCHKA, DO NATIONAL CENTER FOR ATMOSPHERIC RESEARCH;

204. GEOBUGS 1.2 – SOFTWARE QUE REQUER O SISTEMA WINBUGS 1.4, USADO PARA AJUSTAR MODELOS ESPACIAIS, E FORNECE UMA INTERFACE PARA PRODUZIR MAPAS A PARTIR DA SAÍDA DE MODELOS PARA MAPEAMENTO DE DOENÇAS, E CRIA E MANIPULA MATRIZES DE ADJACÊNCIA, DE AUTORIA DE ANDREW THOMAS, NICKY BEST, RICHARD ARNOLD DO IMPERIAL COLLEGE E DAVID SPIEGELHALTER, DO MRC, BIOSTATISTICS UNIT;

205. ROOKCASE – SOFTWARE QUE REQUER O MS-EXCEL, UTILIZADO NA ANÁLISE DE DADOS LÁTICE, DE AUTORIA DE MICHEAL SAWADA DA UNIVERSITY OF OTTAWA, CANADÁ;

206. SGEOSTAT, SOFTWARE QUE REQUER OS PACOTES R E S-PLUS PARA RODAR, USADO EM MODELAGEM GEOESTATÍSTICA, DE AUTORIA DE ALBRECHT GEBHARDT DA UNIVERSITY OF KLAGENFURT;

207. SPATIAL – SOFTWARE QUE REQUER O SISTEMA SAS PARA RODAR, O QUAL É USADO NA ANÁLISE DE PADRÕES ESPACIAIS PONTUAIS, DE AUTORIA DE BARRY MOSER DA LOUISIANA STATE UNIVERSITY, ESTADOS UNIDOS;

208. SPATSTAT – SOFTWARE UTILIZADO EM CONJUNTO COM O PACOTE “R”, PARA ANÁLISE DE PADRÕES ESPACIAIS PONTUAIS, DE AUTORIA DE ADRIAN BADDELEY AND ROLF TURNER DA UNIVERSITY OF WESTERN AUSTRÁLIA;

209. ECONOMETRICS TOOLBOX – SOFTWARE USADO JUNTO COM O PACOTE MATLAB, PARA ANÁLISE DE MODELOS ECONOMÉTRICOS ESPACIAIS, DE AUTORIA DE JAMES P. LESAGE DA UNIVERSITY OF TOLEDO;

210. SPATIAL STATISTICS TOOLBOX – SOFTWARE UTILIZADO JUNTO COM O PACOTE MATLAB, PARA ANÁLISE DE DADOS LÁTICE PARA CONJUNTOS DE DADOS MUITO GRANDES, DE AUTORIA DE KELLEY PACE DA LOUISIANA STATE UNIVERSITY;

211. SPDEP – SOFTWARE USADO JUNTO COM O PACOTE R O QUAL INCLUI FUNÇÕES PARA TESTES DE AUTOCORRELAÇÃO ESPACIAL. ALÉM DISSO, CRIA MATRIZES DE PESOS ESPACIAIS, E AJUSTA MODELOS SAR, DE AUTORIA DE ROGER BIVAND DA NORWEGIAN SCHOOL OF ECONOMICS AND BUSINESS ADMINISTRATION;

212. SPLANCS - SOFTWARE USADO COM OS PACOTES R E S-PLUS, PARA ANÁLISE DE PADRÕES ESPACIAIS E ESPAÇO TEMPORAIS PONTUAIS, DE AUTORIA DE BARRY ROWLINGSON AND PETER DIGGLE DA LANCASTER UNIVERSITY, INGLATERRA;

213. ADS - SOFTWARE UTILIZADO COM O PACOTE R, PARA ANÁLISE DE PADRÕES ESPACIAIS PONTUAIS, DE AUTORIA DE RAPHAËL PÉLISSIER DO INTITUT DE RECHERCHE POUR LE DÉVELOPPMENT;

214. GEOXP - SOFTWARE USADO JUNTAMENTE COM O PACOTE R PARA ANÁLISE EXPLORATÓRIO INTERATIVO DE DAODS ESPACIAIS, DE AUTORIA DE CHRISTINE THOMAS-AGNAN DA UNIVERSITÉ TOULOUSE 1;

215. GSTAT - SOFTWARE USADO JUNTO COM O PACOTE R PARA MODELAGEM GEOESTATÍSTICA PREDIÇÃO E SIMULAÇÃO, DE AUTORIA DE EDZER J. PEBESMA DA UNIVERSITEIT UTRECHT;

216. MAPTOOLS - SOFTWARE UTILIZDO JUNTO COM O PACOTE R, O QUAL POSSUI FERRAMENTAS PARA LER E MANIPULAR OBJETOS ESPACIAIS, DE AUTORIA DE NICHOLAS J. LEWIN KOH AND ROGER BIVAND DA NORWEGIAN SCHOOL OF ECONOMICS AND BUSINESS ADMINISTRATION;

217. PASTECS - SOFTWARE USADO JUNTO COM O PACOTE R PARA ANÁLISE DE SÉRIES ECOLÓGICAS ESPAÇO-TEMPORAIS, DE AUTORIA DE FREDERIC IBANEZ DO LABORATOIRE D'OCÉANOGRAPHIE DE VILLEFRANCHE-SURMER;

218. SPATCLUS, SOFTWARE USADO JUNTO O PACOTE R PARA DETECÇÃO DE CLUSTERS ESPACIAIS, DE AUTORIA DE CHRISTOPHE DEMATTEI DA UNIVERSITÉ MONTPELLIER I;

219. SPATIAL COVARIANCE - SOFTWARE USADO JUNTO COM O PACOTE R PARA CALCULAR MATRIZES DE COVARIÂNCIA ESPACIAIS PARA DADOS EM UNIDADES RETANGULARES, DE AUTORIA DE DAVID CLIFFORD DA THE UNIVERSITY OF CHICAGO, ESTADOS UNIDOS;

220. SPBAYES - SOFTWARE USADO JUNTO COM O PACOTE R PARA MODELAGEM ESPACIAL UNIVARIADA E MULTIVARIADA USANDO MCMC, DE AUTORIA DE ANDREW FINLEY SUDIPTO BANERJEE AND BRADLEY CARLIN DA UNIVERSITY OF MINNESOTA, ESTADOS UNIDOS;

221. SPGWR - SOFTWARE USADO JUNTO COM O PACOTE R PARA REGRESSÃO GEOGRAFICAMENTE PONDERADA, DE AUTORIA DE ROGER BIVAND DA NORWEGIAN SCHOOL OF ECONOMICS AND BUSINESS ADMINISTRATION;

222. GENELAND - SOFTWARE USADO JUNTO COM O PACOTE R PARA DETECÇÃO DE ESTRUTURA ESPACIAL A PARTIR DE DADOS GENÉTICOS GEO-REFERENCIADOS, DE AUTORIA DE GILLES GUILLOT DO INSTITUT NATIONAL AGRONOMIQUE PARIS-GRIGNON, FRANÇA;

223. TESS - SOFTWARE USADO JUNTO COM O PACOTE R PARA DETECÇÃO BAYESIANA DE CLUSTERS USANDO TESSELAÇÕES E MODELOS MARKOVIANOS EM GENÉTICA POPULACIONAL ESPACIAL, DE AUTORIA DE OLIVIER FRANÇOIS DO INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE;

224. RCITRUS - SOFTWARE USADO JUNTO COM O PACOTE R PARA ANÁLISE ESTATÍSTICA DE DADOS DE INCIDÊNCIA DE DOENÇAS EM PLANTAS, DE AUTORIA DE ELIAS T. KRAINSKI E PAULO JUSTINIANO RIBEIRO JÚNIOR (UFPR) E RENATO B. BASSANEZI (FUNDECITRUS), BRASIL;

225. GEOVISUAL - O SISTEMA GEOVISUAL, NA ATUAL VERSÃO 2.2 DE 2004, É COMPOSTO POR 16 PROGRAMAS COM OS SEGUINTE OBJETIVOS, 1. APOIAR O ENSINO DE GEOESTATÍSTICA APLICADA NO INSTITUTO DE GEOCIÊNCIAS, EM NÍVEL DE GRADUAÇÃO, PÓS-GRADUAÇÃO E EXTENSÃO UNIVERSITÁRIA. A MAIOR PREOCUPAÇÃO FOI O DESENVOLVIMENTO DE UM CONJUNTO DE PROGRAMAS, SOB A PLATAFORMA WINDOWS, QUE FOSSEM AMIGÁVEIS E DE

FÁCIL UTILIZAÇÃO. ALÉM DISSO, HOUVE UMA PREOCUPAÇÃO PEDAGÓGICA NO SENTIDO DE FACILITAR O ENTENDIMENTO DE CERTAS OPERAÇÕES MATEMÁTICAS ROTINEIRAS EM GEOESTATÍSTICA, TAIS COMO A MONTAGEM DO SISTEMA DE EQUAÇÕES LINEARES, OS PONTOS UTILIZADOS, RESULTADOS PARCIAIS (PESOS DA KRIGAGEM ORDINÁRIA), A LOCALIZAÇÃO DOS VIZINHOS PRÓXIMOS DENTRE OUTROS. TRATA-SE, PORTANTO, DE UM SOFTWARE EDUCACIONAL. 2. ATENDER AS NECESSIDADES DE PESQUISA DO AUTOR E DE SEU GRUPO, FOI DESENVOLVIDO PELO GRUPO DE GEOESTATÍSTICA APLICADA DO LABORATÓRIO DE INFORMÁTICA GEOLÓGICA INSTITUTO DE GEOCIÊNCIAS DA UNIVERSIDADE DE SÃO PAULO, BRASIL.

226. TABWIN;

227. INFANVIEW;

228. XNVIEW;

229. GIMP;

230. DIA;

231. GNUMERIC;

232. OPENOFFICE;

233. OPENSTAT;

234. SALSTAT STATISTICS;

235. PAIPIX 7;

236. FITYK;

237. OCTAVE;

238. BLENDER;

239. KILE;

240. LYX;

241. DIA;

242. GRI;

243. LABPLOT;

244. QTIPLLOT;

245. SCIGRAPHICA;

246. WINGEOM(WINDOWS): SOFTWARE QUE PERMITE CONSTRUÇÕES GEOMÉTRICAS BIDIMENSIONAIS E TRIDIMENSIONAIS;



247. GEOPLAN(WINDOWS): SOFTWARE DE CONSTRUÇÃO EM GEOMETRIA QUE TRABALHA OS CONCEITOS ANALÍTICOS DA GEOMETRIA EM UM SISTEMA DE COORDENADAS CARTESIANAS. DESENVOLVIDO PELO CENTRE DE RECHERCHE ET D'EXPÉRIMENTATION POUR L'ENSEIGNEMENT DES MATHÉMATIQUES (CREEM);
248. EUKLID(WINDOWS): SOFTWARE DE CONSTRUÇÕES GEOMÉTRICAS COM RÉGUA E COMPASSO E GEOMETRIA DINÂMICA. SEMELHANTE AO CABRI E AO SKETCHPAD;
249. WINMAT(WINDOWS): SOFTWARE QUE PERMITE QUE SE CONSTRUA MATRIZES E OPERE COM ELAS. CALCULA A INVERSA, TRANSPOSTA, DETERMINANTE E ENCONTRA INCLUSIVE O POLINÔMIO CARACTERÍSTICO DA MATRIZ;
250. MATHGV(WINDOWS):SOFTWARE QUE PERMITE QUE SE CONSTRUA GRÁFICOS A PARTIR DE FUNÇÕES ELEMENTARES. POSSIBILITA QUE SE CONSTRUA GRÁFICOS EM DUAS E TRÊS DIMENSÕES E EM COORDENADAS POLARES;
251. GENSTAT: Numerical Algorithms Group, NAG Ltd., Wilkinson House, Jordan Hill Road, Oxford, OX2 8DR, united Kingdom [<http://www.nag.co.uk>];
252. MLWIN: Multilevel Models project, Institute of Education, 20 Bedford Way, London, WC1H OAL, united Kingdom [<http://multilevel.ioe.ac.uk/index.html>];
253. STATXACT AND LOGXACT: Cytel Software Corporation, 675 Massachusetts Avenue, Cambridge, MA 02139, U. S. A. [<http://www.cvtel.com/index.html>];
254. SPHINX BRASIL<sup>®</sup>: Software para coleta e análise de dados [<http://www.sphinxbrasil.com/po/>];
255. VISTA<sup>™</sup>: The visual statistics System: software para estatística visual [<http://forrest.psych.unc.edu/research/index.html>];
256. SOFTWARE RISK: Simulação:add-on para o excel;
257. SOFTWARE CRYSTAL BALL: Simulação:add-on para o excel;
258. SOFTWARE PERTMASTER: Simulação:add-on para o MS-Project;
259. SOFTWARE EXTEND: Simulação industrial e outras;
260. SOFTWARE MEDMODEL:Simulação em serviço de saúde;
261. SELEGEN: O software Selegen foi desenvolvido para ambiente IBP/PC – DOS. É fundamentado em algoritmos que maximizam a eficiência do processo seletivo, em função de diferentes situações experimentais. Os métodos de seleção que esse software possui recomendáveis para o melhoramento de espécies perenes;
262. PROBABILITAS 1.0: Calculadora que realiza a estatística completa das tabelas 2x2, incluindo razão de verossimilhança, razão de chance, probabilidades pré e pós-teste, prevalência, sensibilidade, especificidade, falso positivo, falso negativo, valor preditivo, teste qui-quadrado, taxa de alarme falso e taxa de garantia falsa;
263. ESBStats: Statistical Analysis Software 2.2.0;

264. CLIMA: Programa Computacional para organização e análise de dados meteorológicos, publicação acompanhada de um CD-Rom para instalação de software. O sistema foi criado pelos pesquisadores Rogério Teixeira de Faria e Paulo Henrique Caramori, das Áreas de Engenharia Agrícola e Agrometeorologia, com desenvolvimento de programação dos bolsistas Eduardo Y. Chibana, Agnaldo Koiti Nakamura, Aline Rodrigues e Luciano Roberto de Souza Brito, do Consórcio Café e Programa de Iniciação Científica do CNPq/IAPAR. O programa possibilita cadastrar estações meteorológicas, importar dados climáticos, fazer a consistência dos dados importados, estimar dados para períodos com erros ou falhas de coleta, calcular evapotranspiração potencial por diferentes métodos (FAO24-Penman, Penman-Monteith, Ritchie-Priestley & Taylor e Thornthwaite), determinar médias, desvio-padrão e probabilidades para períodos decendiais, mensais e anuais, a partir de dados diários, e apresentar os resultados em forma de gráficos e tabelas, que podem ser copiados para uso em outros aplicativos computacionais ou comparados com os resultados de outras estações já cadastradas no sistema.A análise de dados pode ser feita com uma ou mais estações e possibilita a obtenção de gráficos diários decendiais e mensais gerados a partir de dados climáticos num período pré-definido. Quando da análise de uma única estação é possível realizar a impressão e exportação das valores climáticos diários e de suas médias mensais.O sistema é programado em linguagem Delphi 5.0<sup>™</sup>, em ambiente Windows<sup>™</sup> e base de dados Access 97<sup>™</sup>, para uma configuração mínima recomendada de processador Pentium 166, com 32 MB de memória RAM, leitor de CD, monitor de resolução 800 x 600 “high color” 16 bits, espaço em disco de 30 MB e Windows 9x, Me ou NT ;
265. MICROARRAY;
266. E-VIEWS (V.6.0): O E-Views permite aos seus utilizadores todas as ferramentas para análise econométrica de séries temporais num ambiente Windows. Previsão, análise estatística, gráficos, simulações, estimativas e gestão de dados, poderá encontrar tudo nesta poderosa interface. VIEWS: <http://www.eviews.com/eviews4/eviews4/eviews4.html>;
267. OXMETRICS: É o nome de um conjunto de módulos de software. Trata-se de uma solução integrada de análises econométricas de séries temporais, previsão, criação de modelos econométricos na área financeira, análises estatísticas de cross-section e panel data;
268. RATS (V.7.2): Programa para análise de séries temporais e criação de modelos econométricos. Possui gráficos de excelente qualidade;
269. SHAZAM (V.9.0): Software para as áreas de econometria, estatística, psicometria, sociometria e outras áreas que utilizam técnicas estatísticas. O Shazam possui uma linguagem de comando com grande flexibilidade, fornecendo capacidades de programação consideráveis. O Shazam tem uma interface para o package GNUPLOT por forma a elaborar gráficos de grande qualidade;
270. UNISTAT (V.5.5): Sistema integrado de estatística com excelentes gráficos e de utilização bastante acessível. Pode ser utilizado como add-in do Microsoft Excel. UNISTAT: <http://www.unistat.com/>;
271. ABACUS CONCEPTS: Package Estatístico da StatView e softwareSuperANOVA para o Mac;
272. APTECH SYSTEMS: Home Page do GAUSS, um pacote de software matemático e estatístico incluindo uma poderosa língua de programação, para o DOS, o Windows, o OS/2 e plataformas Unix. <http://www.aptech.com/>;

273. BMDP STATISTICAL SOFTWARE: Sistema BMDP, BMDP/DIAMOND, SOLO, SOLO power analysis, StatXact, algum freeware como um calculador de probabilidade SOLO. BMDP Inc. é agora propriedade da SPSS Inc;
274. DATA DESCRIPTION: Data Desk: software visual de análise de dados para o Apple Macintosh e Power-Mac;
275. FORECASTX: ForecastX é um componente ActiveX para a previsão e análise estatística.
276. GRAPHPAD: GraphPad Prism para Windows;
277. IDEA WORKS: Vários sistemas de software, incluindo o Statistical Navigator para DOS/Win;
278. IVATION DATASYSTEMS: Beyond 20/20 software para Windows;
279. JANDEL: Software científico incluindo o SigmaStat (Windows 32bit);
280. MATHSOFT: Produtos de Softwares: MathCAD, Axum, S-Plus;
281. MATHWORKS: Software MATLAB;
282. NAG–THE NUMERICAL ALGORITHMS GROUP: Software numérico e bibliotecas, Fortran 90, GenStat, Glim, MLP, AXIOM, Iris Explorer;
283. NCSS STATISTICAL SOFTWARE: Cópia gratuita de avaliação e calculador de probabilidade da NCSSgratuito;
284. NUMERICAL RECIPES: A arte de computação científica, fonte de códigos em várias linguagens de programação com algoritmos numéricos, com loja on-line, livros on-line, e mais;
285. NWP ASSOCIATES - STATLETS: Java applets para análise estatística e gráficos, gratuito, mas com versões acadêmicas disponíveis limitadas;
286. PROCESS BUILDER - STRATEGY FOR WINDOWS: Software para planejamento de experiências, software para o Windows 32bit (por Dave Doehlert - Experiment Strategies Foundation);
287. PROGAMMA SOFTWARE DISTRIBUTION CATALOG: oferece um catálogo de software por categoria com descrições de SIByl (banco da informação do software), cobre um variado leque de software comercial especializado;
288. WAIS SEARCH SAS 6 USAGE NOTES: Search a versão 6 SAS Notes por palavra-chave ou título;
289. SAS TIPS AND SAMPLES: Por PW Consulting;
290. STAT-EASE: Design-Ease e Design-Expert para Windows, software para o planejamento de experiências;
291. STATISTICAL SOLUTIONS: Distribuidor de software estatístico, incluindo Solas para análise de dados omisso enQuery Advisor para planejamento de experiências (versões de demonstração disponíveis);
292. STAT/TRANSFER: Stat/Transfer, utilitário para transferir dados entre programas diferentes, versão de demo disponível para download;
293. STATWARE: Software Statit em várias plataformas UNIX;
294. VISUAL NUMERICS: Produtos: PV-WAVE, IMSL (bibliotecas numéricas e gráficas de C e Fortran), Stanford Graphics;
295. ARFIT: Coleção gratuita de rotinas Matlab, funcionando em Unix, por de T.Schneider e A.Neumaier;
296. B/D: The Bayes Linear Programming Language–Software Baysiano gratuito, linguagem de programação iterativa, funcionando em Win3.1 e Win95, por David Wooff e Michael Goldstein;
297. DECISION ANALYST – STATS: Software Estatístico gratuito para pesquisa de mercado, funcionando em IBM PC Windows;
298. CLICK & LEARN REGRESSION: Software iterativo para análise de regressão linear, versão de demonstração disponível;
299. EZQUANT: Folha de cálculo para análise estatística, versão apenas para DOS;
300. MD Anderson Cancer Center Biomathematics: Variado software estatístico especializado, freeware disponível, programas DOS ou Mac, incluído software STPLAN;
301. NCSS PROBABILITY CALCULATOR FOR WINDOWS: Freeware por NCSS. Cálculo de probabilidades e quantis de 14 distribuições da Beta à Weibull;
302. PROPHET: (do National Computing Resource for Life Science Research). O PROPHET system é um package de software UNIX. Análise de dados, gráficos, análise estatística, modelação matemática e análise sequencial;
303. PROVALIS RESEARCH: Algum software estatístico como o SimStat (para DOS e Windows), StatItem, Easy Factor Analysis e MVSP;
304. TEXASOFT: É oferecido shareware e a versão completa do KwikStat para DOS, e WinKS, a versão Windows do KwikStat;
305. XTREMES: Software estatístico para análise de dados extremos (versões para Windows), incluído no livro Statistical Analysis of Extreme Data, by;
306. AUTOBOX: <http://www.autobox.com/>;
307. SPHYNX: <http://www.sphynx.com/>;
308. XPLORE: [http://www.xplore-stat.de/index\\_js.html](http://www.xplore-stat.de/index_js.html);
309. C 4.5: Programs for Machine Learning. San Francisco: Morgan Kaufmann, 1993. Autoria de QUINLAN, R;
310. WEKA SOFTWARE.–Weka 3: Data Mining with Open Source Machine Learning Software in Java. <http://www.cs.waikato.ac.nz/~ml/weka/index.html>;
311. DATA MINING: Practical machine learning tools with java implementations. Morgan Kaufmann: San Francisco, 2000. autoria de WITTEN, I. H.; FRANK, E;
312. GEOGEBRA: Software para Matemática Software livre. Autoria de Markus Hohenwarter e Yves Kreis;
313. Z.U.L.: Software para Matemática, Software livre. Autoria de R. Grothmann;

314. GEONEXT: Software para Matemática, Software livre. Autoria da Universität Bayreuth. Mais recentemente um conceito antigo foi adaptado para a Geometria Dinâmica. A modelação 3D em computador foi já amplamente utilizada nas mais diversas áreas. A equipa Cabri desenvolveu o Cabri 3D onde aplica a modelação 3d ao estudo da geometria do espaço;
315. CABRI 3D V2: Software para Matemática, Versões de avaliação. Autoria de Jean-Marie Laborde e Eric Bainville. Existem outros aplicativos com objetivos mais específicos, como estudar secções produzidas por planos em diversos polígonos (s3D SecBuilder), ou analisar poliedros (Poly Pro) ou estudar homotetias no plano (Tess);
316. ALPHAGEN;
317. LATANOVA;
318. ALPHANAL;
319. MSTAT;
320. MAPGEN;
321. ESTAT versão 2.0 - Software desenvolvido pelo Pólo Computacional e Departamento de Exatas da UNESP, Campus de Jaboticabal.
322. ASSISTAT 7.5 BETA. (Campina Grande, Paraíba, Brasil);
323. ACL–Audit Command Language - Software para análise de dados em auditoria;
324. IDEA - Interactive Data Extraction and Analysis - Software para análise de dados em auditoria;
325. EZ-R Stats–Site que disponibiliza diversos softwares para análise de dados gratuitamente;
326. EZ Quant–Software Análise Estatística;
327. ActiveData - Suplemento para Excel destinado à realização de testes de auditoria;
328. TopCAATs - Suplemento para Excel semelhante ao ActiveData;
329. RAT-STATS - Software para seleção de amostras desenvolvido pelo Health & Human Services dos EUA;
330. Picalo - Software Open Source para Análise de Dados baseado na Linguagem de Programação Python.

# REFERÊNCIAS

AITKEN, A. C. On least squares and the linear combination of observations. **Proceedings of the Royal Society of Edinburgh**, n. 55, p. 42-48. 1934.

AITKEN, A. C. Studies in practical mathematics: the evaluation of the latent roots and latent vectors of a matrix. **Proceedings of the Royal Society of Edinburgh**, v. 57, p. 269-304. 1937.

ALMEIDA, A. C. **Como são feitas as pesquisas eleitorais e de opinião**. Rio de Janeiro: Editora FGV. 2002. 196 p.

ALMEIDA, V. A. F. **Métodos quantitativos para ciência da computação experimental**. Departamento da Ciência da Computação. Universidade Federal de Minas Gerais, DCC-UFMG. 2005. 34 p. (nota de aula 7).

ALTMAN, D. G. **Practical statistics for medical research**. London: Chapman & Hall. 1991. 624 p.

ALVES, J.F. **Experimentação agrícola**. Fortaleza: Universidade Federal do Ceará-Centro de Ciências Agrárias. Departamento de Fitotecnia, 1986. 69 p.

AMERICAN PSYCHOLOGICAL ASSOCIATION. **Publication manual of the american psychological association**. 2. Ed. Washington, DC: American Psychological Association. 1974.

ANDERSON, D. R.; SWEENEY, D. J.; WILLIAMS, T. A. **Estatística aplicada à administração e economia**. Tradução da 2ª edição norte-americana por Luiz Sérgio de Castro Paiva. São Paulo, SP: Editora pioneira. Thompson Learning, 2002. 646p.

ANDRADE, D. F.; OGLIARI, P. J. **Estatística para as ciências agrárias e biológicas: com noções de experimentação**. 3. ed. Florianópolis: EDUFSC, 2013. 475 p.

ANDRADE, D. F.; OGLIARI, P. J. **Estatística para as ciências agrárias e biológicas: com noções de experimentação**. Florianópolis: Editora da UFSC, 2007. 432 p.

ANDRADE, M.; PINHEIRO, H. P. **Métodos estatísticos aplicados em genética humana**. Campinas: UNICAMP, 2002. 189 p. (Apostila).

ANDRADE, P. C. R.; FERREIRA, D. F. Comparações múltiplas bayesianas em modelos normais homocedásticos e heterocedásticos. **Ciência e Agrotecnologia**, Lavras, v. 34, n. 4, p. 845-852, jul/ago. 2010.

ANJOS, A. **Análise gráfica com uso do r**. Curitiba: Universidade Federal do Paraná-UFPR. 2006. 136 p.

ANSCOMBE, F. J. Graphs in statistical analysis. **The American Statistician**, Georgia, v. 27, n. 1, p. 17-21. 1973.

ANSCOMBE, F. J. The transformation of Poisson, binomial, and negative binomial data. **Biometrika**, **Rothamsted**, Harpenden, Inglaterra, v. 35, n. 3-4, p. 246-254. 1948.



- APPOLINARIO, F. **Dicionário de metodologia científica**: um guia para a produção do conhecimento científico. 3ª Reimpressão. São Paulo: Atlas, 2009. 300 p.
- ASCH, S. E. The doctrine of suggestion, prestige and imitation in social psychology. **Psychological Review**, v. 55, n. 5, p. 250-276.1948.
- ASSIS, F. N.; ARRUDA, H. V.; PEREIRA, A. R. **Aplicações de estatística a climatologia**: teoria e prática. Pelotas: Editora Universitária. UFPel. 1996. 161 p.
- ASSIS, J. P. et al. **Estatística descritiva**. Piracicaba: Fundação de Estudos Agrários Luiz de Queiroz. 2016. 394 p.
- AYRES, M. et al. **BioEstat. versão 5.0**. Belém-PA: Sociedade Civil Mamirauá, MCT-CNPq. 2007.
- AYRES, M. et al. **Bioestat 4.0**: Aplicações estatísticas nas áreas das ciências biológicas e médicas (manual). 4. Ed. Sociedade Civil Mamirauá/MCT/Imprensa Oficial do Estado do Pará. Com CD - Rom. Belém-PA, 2005. 337 p.
- AYRES, M. et al. **Bioestat 5.0**: Aplicações estatísticas nas áreas das ciências biológicas e médicas (manual). 5. Ed. Instituto de Desenvolvimento Sustentável Mamirauá - IDSM/MCT/CNPq. Com CD - Rom. Belém-PA. 2007. 383 p.
- BABBIE, Earl. **Métodos de pesquisas de survey**. Belo Horizonte: Editora da UFMG, 2001. 519 p.
- BAILY, T.C. ; GATRELL, A. C. **Interactive spatial data analysis**. London: Longman, 1996. 413 p.
- BANZATTO, D. A.; KRONKA, S. N. **Experimentação agrícola**. 4. ed. Jaboticabal: FUNEP, 2006. 237p.
- BARBIN, D. **Componentes de variância**: teoria e aplicações. 2. Ed. Piracicaba: FEALQ, 1993. 120 p.
- BARNETT, V. **Elements of sampling theory**. London: The English Universities Press Ltda, 1974. 152 p.
- BARTLETT, N. S. The use transformations. **Biometrics**, v.3, p. 39-52. 1947.
- BASTOS, J.; ROCHA, C. **Análise de sobrevivência**: conceitos básicos. Notas metodológicas: arquivos de medicina. Serviço de Higiene e Epidemiologia da Faculdade de Medicina da Universidade do Porto; Faculdade de Ciências da Universidade de Lisboa. p.185-187. 2006.
- BATISTA, J. L. F. **Levantamentos florestais**. Departamento de ciências florestais. Piracicaba: ESALQ/ USP, 1998. 45 p. (Apostila).
- BATISTA, J. L. F. **Mensuração de árvores**: uma introdução à dendrometria. Departamento de Ciências Florestais, Escola Superior de Agricultura Luiz de Queiroz. Piracicaba: Universidade de São Paulo. 1998. 45 p.
- BATISTA, J. L. F. **METRVM-GLOSSARIVM QVANTITATIVM SILVARVM**. Disponível em: <<http://cmq.esalq.usp.br/wiki/doku.php?id=publico:metrvm:glossarivm:start>>. Acesso em: 25 ag. 2009.
- BATISTA, J. L. F.; COUTO, H.T. Z. **O Estéreo**. METRVM, 2002. 19 p.
- BATSCHULET, E. **Introdução a matemática para biocientistas**. São Paulo: Editora da EDUSP, 1978. 596 p.

- BAYES, T. An essay towards solving a problem in the doctrine of chances. **Philos. Trans. R. Soc.**, London, v. 53, p. 370-418. 1763.
- BEAVER, R.; MENDENHALL, W.; REINMUTH, J. **Study guide for statistics for management and economics**. 2. ed. Massachusetts: Wadsworth, 1974. 441 p.
- BEIGUELMAN, B. **Curso prático de bioestatística**. 5. ed. Ribeirão Preto: FUNPEC, 2002. 274 p.
- BELLO, J. L. de P. **Metodologia científica**: sumário. Rio de Janeiro, 2004. Disponível em: <<http://www.pedagogiaemfoco.pro.br/meto1.htm>>. Acesso em: 16 abr. 2009.
- BERNOULLI, J. **Ars conjectandi**. Basel: Thurnisiorum Fratrum, 1713. 306 p.
- BERQUÓ, E. S. et al. **Bioestatística**. São Paulo: EDUSP, 1980. 325 p.
- BESAG, J. Spatial interaction and the statistical analysis of lattice systems. **Journal of the Royal Statistical Society**. Series B (Methodological), v. 36, n. 2. p. 192-236, 1974.
- BICKEL, P. J., DOBSUM, K. A. **Mathematical statistics**: basic ideas and selected topics. San Francisco: Holden-Day, Inc, 1977. 556 p.
- BITTERLICH, W. **The relascope idea**: relative measurement in forestry. Slough, England: Commonwealth Agricultural Bureaux, 1984. 242 p.
- BLACKWELL, D. Conditional expectation and unbiased sequential estimation. **The Annals of Mathematical Statistics**, v. 18, p. 105-112, 1947.
- BOLFARINE, H.; BUSSAB, W. O. **Elementos de amostragem**. São Paulo: Edgard Blusher, 2005. 269p.
- BOLFARINE, H.; SANDOVAL, M. C. **Introdução à inferência estatística**. Coleção matemática aplicada. Sociedade Brasileira de matemática-SBM. Rio de janeiro, RJ., 2001. 132 p.
- BOX, G. E. P.; COX, D. R. An analysis of transformations. **Journal of the Royal Statistical Society**, London, Série B (Statistical Methodological), v. 26, n. 2, p. 211-252. 1964.
- BRASIL ADMINISTRAÇÃO. Disponível em: <<http://www.bradministracao.blogspot.com.br/2008/09/termos-estatisticos-mais-utilizados.html>>. Acesso em: 04 abr. 2015.
- BRASIL, G. H.; SILVA, A. F. P. Uma introdução à teoria da amostragem com aplicações em pesquisas eleitorais. In: Simpósio Brasileiro de Pesquisa Operacional, 27., 2005. Gramado. **Mini-curso...** Vitória: Universidade Federal do Espírito Santo, 2005. 151 p.
- BRITO, M. **Metodologia da investigação agrária**. Lisboa: Escola Superior Agrária: Instituto Politécnico de Viana do Castelo, 2016. 70 p.
- BROWN, D.; ROTHERY, P. **Models in biology**: mathematics, statistics and computing. Chichester, England: John Wiley & Sons Inc. Ltd., 1994. 708 p.
- BUSSAB, W. O.; MORETTIN, P. A. **Estatística básica**. 4. Ed. São Paulo: Atual Editora, 1987. 332 p.
- BUSSAB, W. O.; MORETTIN, P. A. **Estatística básica**. 5. Ed. São Paulo: Saraiva. 2002. 544 p.

- BUSSAB, W. O.; MORETTIN, P. A. **Estatística básica**. 5. Ed., São Paulo: Saraiva, 2002. 526p.
- CALADO, V.; MONTGOMERY, D. C. **Planejamento de experimentos usando Statistica**. Rio de Janeiro: E-Papers Serviços Editoriais, 2003. 260 p.
- CÂMARA, F. P. **Glossário de análise multivariada**. Rio de Janeiro: Instituto de microbiologia de pesquisa e pós-graduação-Universidade Federal do Rio de Janeiro. IMPPG-UFRJ, 2008. 4 p. Disponível em: <<http://www.microbiologia.ufrj.br/Gloss%20de%20An%20lise%20Multivariada.pdf>> Acesso em: 03 dez. 2008.
- CAMPOS H. **Amostragem I**. Piracicaba: Universidade de São Paulo-Escola Superior de Agricultura Luiz de Queiróz-ESALQ/USP, 1980. 150 p. (Notas de aula)
- CAMPOS, H. **Amostragem II**. Piracicaba: ESALQ/USP, 1981. 156 p. (Notas de aula)
- CAMPOS, H. **Estatística experimental não paramétrica**. 4. ed. Piracicaba: ESALQ/ USP, 1983. 349 p.
- CAMPOS, K. A.; PAIXÃO, C. A.; MORAIS, A. R. Transformação de dados como alternativa a análise de variância univariada. **Sigmae**, Alfenas, v.2, n.3, p. 57-64, 2013.
- CARVALHO, M. S. et al. **Análise de sobrevivência: teoria e aplicações em saúde**. Rio de Janeiro: Editora Fiocruz, 2005. 396 p.
- CASELLA, G.; BERGER, R. L. **Statistical inference**. 2. ed. Duxbury: Thomson Learning. Wadsworth Pub Co, 2002. 660 p.
- CATALUNHA, M. J. et al. Aplicação de cinco funções densidade de probabilidade a séries de precipitação pluvial no estado de Minas Gerais. **Revista Brasileira de Agrometeorologia**, Santa Maria, v. 10, n.1, p.153-162, 2002.
- CAZORLA, I. M.; SILVA, C. B. **Educação estatística-Glossário in: Sócio estatística**. Bahia. Disponível em <<http://www.socio.estatistica.com.br/Edestatistica/glossario.htm>>. Acesso em: 15 abr. 2008.
- CECON, P. R.; BARROS, W. S. **Métodos estatísticos I**. Viçosa: Universidade Federal de Viçosa-Centro de Ciências Exatas e Tecnológicas, Departamento de Informática, 2005. 144 p. (Apostila).
- CENTENO, A. J. **Curso de estatística aplicada à biologia**. 2. ed. Goiânia: Editora da UFG. 1990. 189 p.
- CHARNET, R. et al. **Análise de modelos de regressão linear com aplicações**. Campinas: Editora da Unicamp, 1999. 366 p.
- CHAVES, J. B. P. **Avaliação sensorial de alimentos (métodos de análises)**. Universidade federal de Viçosa, Centro de Ciências Exatas e Tecnológicas, Departamento de Tecnologia de Alimentos. Viçosa: Imprensa Universitária da Universidade federal de Viçosa, 1980. 77 p.
- CHOW, V. T. **Handbook of applied hydrology**. New York: McGraw-Hill, 1964. 1495 p.
- CHOW, V. T.; MAIDMENT, D. R.; MAYS, L. W. **Applied hydrology**. New York: McGraw-Hill, 1988. 572 p.
- CHRISTMANN, R. U. **Estatística aplicada**. São Paulo: Edgard Blücher, 1978. 192 p.

- CLARK, A. B.; DISNEY, R. L. **Probabilidade e processos estocásticos**. Rio de Janeiro: Livros Técnicos e Científicos Editora S/A, 1979. 352 p.
- COCHRAN, W. G. **Sampling techniques**. 3. ed. New York: John Wiley & Sons Inc, 1977. 555 p.
- CONAGIN, A., **Princípios de técnica experimental e análise estatística de experimentos**. Campinas: Instituto agrônomo de Campinas, v. 1 e v. 2, 1961. 236 p.
- COSTA NETO, P. L. O. **Estatística**. 2. ed. São Paulo: Editora Edgard Blucher, 2002. 266 p.
- COSTA NETO, P. L. O. **Estatística**. São Paulo: Editora Edgard Blucher Ltda, 1977. 276 p.
- COX, D. R.; REID, N. Parameter orthogonality and approximate conditional inference (with discussion). **Journal of the Royal Statistical Society, Series B**, 49, 1-39. 1987.
- CRAMER, H. **Mathematical methods of statistics**. New Jersey: Princeton University Press, 1966. 245 p.
- CURI, P. R. **Metodologia e análise da pesquisa em ciências biológicas**. 2. ed. Botucatu: Gráfica e Editora Tipomic, 1998. 276 p.
- DAGNELLI, P. **Estatística: teoria e métodos**. Portugal: Publicações Europa-America, v. 1, 1975. 440 p.
- DAGNELLI, P. **Estatística: teoria e métodos**. Portugal: Publicações Europa-America, v. 2. 1975. 534 p.
- DANTAS, C. A. B. **Probabilidade: um curso introdutório**. 2. ed. São Paulo: EDUSP, 2004. 256 p. Coleção Acadêmica.
- DE FINETTI, B. La prévision: ses lois logiques, ses sources subjectives. **Annales de l'Institut Henri Poincaré**, v. 7, p. 1-68, 1937.
- DEMING, W. E. **Saia da crise: as 14 lições definitivas para controle de qualidade**. Rio de Janeiro: Futura, 2003. 502 p.
- DE MUNTER, P. Sur les transformations des variables aléatoires. **Bulletin Society Belge of Statistics**, n. 97, p. 111, 1958.
- DE VRIES, P. G. **Sampling theory for forest inventory: a teach-yourself course**. Berlin: Springer-Verlag, 1986. 399 p.
- DEVORE, J. L. **Probability and statistics for engineering and the sciences with Infotrac**. Hardcover: Thomson learning, 2003. 764 p.
- DI DIO, R. A. T. **Estatística: instrução programada**. São Paulo: Editora Pedagógica e Universitária Ltda. EPU, 1979. 200 p.
- DIAS, L. A. S.; BARROS, W. S. **Biometria experimental**. Viçosa: Universidade Federal de Viçosa, 2009. 420 p.
- DICIONÁRIO MATEMÁTICO. Disponível em: <<http://www.psiqweb.med.br/site/DefaultLimpo.aspx?area=ES/VerClassificacoes&idZClassificacoes=64>>. Acesso em: 01 jun. 2016.

- DOANE, D. P. Aesthetic frequency classifications. **The American Statistician**, Alexandria, v. 30, n. 4. p. 181-183, nov. 1976.
- DOWNING, D. **Estatística aplicada**. São Paulo: Saraiva, 1999. 474 p.
- DRAPER, N.; SMITH, H. **Applied regression analysis**. 3. ed. New York: John Wiley & Sons, 1998. 736 p.
- EHLERS, R. S. **Análise de séries temporais**. 3. ed. Curitiba: Departamento de Estatística. Universidade Federal do Paraná-UFPR, 2005. 88 p. (Apostila).
- ELLIOTT, J. M. Some methods for the statistical analysis of samples of benthic invertebrates. **Sci. Public. Freshwater Biological Association**, Ambleside, n. 25, 1979. 148 p.
- ELLISON, A. M.; GOTELLI, N. J. **Princípios de estatística em ecologia**. São Paulo: Artmed, 2010. 532 p.
- EKAMBARAM, S. K. **Fundamentos estatísticos da inspeção por amostragem**. São Paulo: Editora Polígono. S/A, 1971. 168 p.
- EVERITT, B. S., **Cambridge dictionary of statistics**. 3. ed. Londres: Editora Cambridge, 2006. 432 p.
- FEDERER, W. T. Augmented (or hoonuiaku) designs. **Hawaiian Planters' Record**, Aiea, v. 55, n. 2, p. 191-208, 1956.
- FELLER, W. **An introduction to probability theory and its application**. 3. ed. New York: John Wiley & Sons, 1967. v. 1. 236 p.
- FELLEGI, I. P. Sampling with varying probabilities without replacement: rotating and non-rotating samples. **Journal of the American Statistical Association**, n. 58, p. 183-201, 1963.
- FERNANDEZ, P. **Introdução à teoria das probabilidades**. Rio de Janeiro: LTC, 1975. Coleção Elementos da Matemática.
- FERREIRA, D. F. **Estatística básica**. Lavras: Editora UFLA, 2005. 664 p.
- FERREIRA, A. B. H. **Dicionário aurélio básico da língua portuguesa**. 2. ed. Rio de Janeiro: Editora Nova Fronteira, 1986. 1832 p.
- FIGUEIREDO, F. et al. **Estatística descritiva e probabilidades-problemas resolvidos e propostos com aplicações em R**. Lisboa: Escolar Editora, 2007. 420 p.
- FIGUEIREDO FILHO, D. B.; SILVA JÚNIOR, J. A. Desvendando os mistérios do coeficiente de correlação de Pearson (r). **Revista Política Hoje**, Recife, v. 18, n. 1, p. 115-146, 2009.
- FISHER, R. A. On an absolute criterion for fitting frequency curves. **Messenger of Mathematics**. v. 41, p. 155-160, 1912.
- FISHER, R. A. On the mathematical foundations of theoretical statistics. **Philosophical Transactions Royal Society**. London. Série A, v. 222, p. 309-368, 1922.
- FISHER, R. A. On the probable error of a coefficient of correlation deduced from a small sample. **Metron**, I, parte 4, p. 1-32. 1921.

- FISHER, R. A. The correlation between relatives on the supposition of mendelian inheritance. **Trans. R. Soc. Edinb.**, v. 52, p. 399-433, 1918.
- FISHER, R. A. **The design of experiments**. Edimburgo: Oliver and Boyd, 1935. 252 p.
- FISHER, R. A. **The genetical theory of natural selection**. Oxford: Oxford University Press, 1930. 272 p.
- FISHER, R. A. **Statistical methods for research workers**. 12. ed. New York: Hafner, 1954. 239 p.
- FONSECA, J. S.; MARTINS, G. A.; TOLEDO, G. L. **Estatística aplicada**. 2. ed. São Paulo: Atlas, 1982. 272 p.
- FISHER, R. A.; TIPPETT, L. H. C. Limiting forms of the frequency distribution of the largest or smallest member of a sample. **Proceedings of the Cambridge Philosophical Society**, v. 24, p.180-190, 1928.
- FRANCISCO, W. **Estatística básica: síntese da teoria, exercícios propostos e resolvidos**. 2. ed. Piracicaba: Editora Unimep, 1995. 220 p.
- FRIEDMAN, M. **Essays in positive economics**. Chicago: University of Chicago Press, 328 p. 1953.
- FRISCH, R. **Statistical confluence analysis by means of complete regression systems**. Noruega: Instituto de Economia, Universidade de Oslo, n. 5, 1934. 192 p.
- FURNIVAL, G. M. **An index for comparing equations used in constructing volume tables**. **Forest Science, Washington**, p. 337-341, n. 4, v. 7, 1961.
- GALLIANO, A. G. **O método científico: teoria e prática**. São Paulo: Harbra, 1986. 200 p.
- GALTON, F. **Hereditary genius**. London: Macmillan and Co, 1869. 423 p.
- GALTON, F. Regression towards mediocrity in hereditary stature. **Journal of the Anthropological Institute**, n. 15, p. 246-263. 1886.
- GALTON, F. Statistics by intercomparison, with remarks on the law of frequency of error. **Philosophy Magazine**. v. 49, p. 33-46, 1875.
- GAMA, S.M. A.; PEDROSA, A. C. **Introdução computacional a probabilidade e estatística**. Porto: Coleção Universitária. Porto Editora, 2004. 590 p.
- GEMAN, S.; GEMAN, D. Stochastic relaxation, Gibbs distribution and the bayesian restoration of imagens. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 6, p. 721-741. 1984.
- GEOESTATÍSTICA ON LINE. **Glossário de termos geoestatísticos**. São Paulo: Universidade de São Paulo. Disponível em < [http://www.igc.usp.br/pesquisa/geoestatistica/glossario/glossario\\_frame.html](http://www.igc.usp.br/pesquisa/geoestatistica/glossario/glossario_frame.html) > Acesso em: 03 nov. 2009.
- GIACOMELLI, M. A. **Amostragem I**. Porto Alegre: Universidade Federal do Rio Grande do Sul, 2007. 60 p. (Apostila)
- GIANOLA, D.; FERNANDO, R. L. Bayesian methods in animal breeding theory. **Journal of Animal Science**, v. 63, p. 217-244, 1986.
- GÍL, A. C. **Técnicas de pesquisa em economia**. São Paulo: Editora Atlas S/A, 1988. 180 p. GLOSSÁRIOS. Disponível em: < <http://www.pucrs.br/famat/statweb/> >. Acesso em: 16 abr. 2009.



- GODAMBE, V. P. Conditional likelihood and unconditional optimum estimating equations. **Biometrika**, v. 63, p. 277-284, 1976.
- GODAMBE, V. P. **Estimating functions**. New York: Oxford Science Pub, 1991. 360 p.
- GOLDBERGER, A. S. **A course in econometrics**. Massachusetts: Harvard University Press, Cambridge, 1991. 249 p.
- GOLDENBERG, M. **A arte de pesquisar: como fazer pesquisa qualitativa em ciências sociais**. 2. ed. Rio de Janeiro: Record, 1998. 107 p.
- GOMES, A. M. A. **Medição de árvoredos**. Lisboa: Livraria Sá da Costa, 1957. 413p.
- GOMES, F. P. **Estatística experimental**. 15. ed. Piracicaba: ESALQ, 2009. 451 p.
- GOMEZ, K. A. **Techniques for field experiments with rice**. Los Baños: Internacional Rice Research Institute, 1972. 46 p.
- GOMEZ, K. A., GOMEZ, A. A. **Statistical procedures for agricultural research**. 2. ed. New York: John Wiley & Sons Inc, 1984. 680 p.
- GOTELLI, N. J.; ELLISON, A. M. **Princípios de estatística em ecologia**. São Paulo: Artmed, 2010. 532 p.
- GUEDES, M. L. S.; GUEDES, J. S. **Bioestatística para profissionais de saúde**. Rio de Janeiro: Editora Ao Livro Técnico S/A, 1988. 204 p.
- GUJARATI, D. **Econometria básica**. São Paulo: Editora Makron Books Ltda, 1999. 819 p.
- GUJARATI, D.N. **Econometria básica**. 3. ed. São Paulo: Editora Makron Books Ltda, 2000. 884 p.
- HAIR, JR. J. F. et al. **Análise multivariada de dados**. 5. ed. Porto Alegre: Bookman, 2005. 593 p.
- HANN, C. T. **Statistical methods in hydrology**. Ames: Iowa State University Press, 1977. 378 p.
- HARRIS, D. C. **Análise química quantitativa**. 8. ed. Rio de Janeiro: LTC, 2012. 886 p.
- HASTINGS, N. A. J.; PEACOCK, J.B. **Statistical distributions: a handbook for students and practitioners**. England: Longon Butterworths, 1975. 129p.
- HAYTER, A. J. **Probability and statistics for engineers and scientists**. 2. Ed. Georgia: Duxbury P, 2002. 916 p.
- HARTLEY, H. O.; RAO, J. N. K. Maximum likelihood estimation for the mixed analysis of variance model. **Biometrika**, v. 54, p. 93-108, 1967.
- HAZAN, S. **Fundamentos da matemática elementar: combinatória e probabilidade**. 6. ed. São Paulo: Atual Editora, 1993. 174 p.
- HEATH, O. V. S. **A estatística na pesquisa científica**. São Paulo: EPU/EDUSP, 1981. 95 p.
- HOEL, P. G. **Estatística elementar**. São Paulo: Editora Atlas, 1981. 430 p.
- HOEL, P. G. **Introduction to mathematical statistics**. New York: John Wiley & Sons, 1965. 427 p.

- HOEL, P. G. ; PORT, S. C. ; STONE, C. J. **Introdução a teoria da probabilidade**. Rio de Janeiro: Interciência, 1978. 269 p.
- HOEL, P. G. **Estatística matemática**. 4. ed. Rio de Janeiro: Editora Guanabara, 1980. 373 p.
- HOERL, A. E.; KENNARD, R. W. Ridge ion: Applications to nonorthogonal problems. **Technometrics**, Alexandria, v.12, n.1, p. 69-82, 1970a.
- HOERL, A. E.; KENNARD, R. W. Ridge regression: Biased estimation for nonorthogonal problems. **Technometrics**, Alexandria, v.12, n.1, p. 55-67, 1970b.
- HOFFMAN, R. **Estatística para economistas**. São Paulo: Livraria Pioneira Editora, 1980. 392 p. HOAGLIN, D.C.; MOSTELLER, F.; TUKEY, J.W. **Análise exploratória de dados. técnicas robustas: um guia**. Lisboa: Edições Salamandra, 1992. 446 p.
- IBGE. **Anuário Estatístico do Brasil**. Rio de Janeiro, RJ. 1953.
- IBGE. **Anuário Estatístico do Brasil**. Rio de Janeiro, RJ. 1955.
- IBGE. **Anuário Estatístico do Brasil**. Rio de Janeiro, RJ. 1956.
- IBGE. **Anuário Estatístico do Brasil**. Rio de Janeiro, RJ. 1964.
- IBGE. **Anuário Estatístico do Brasil**. Rio de Janeiro, RJ. 1965.
- IBGE. **Anuário Estatístico do Brasil**. Rio de Janeiro, RJ. 1969.
- IBGE. **Anuário Estatístico do Brasil**. Rio de Janeiro, RJ. 1970.
- IBGE. **Anuário Estatístico do Brasil**. Rio de Janeiro, RJ. 1975.
- IBGE. **Anuário Estatístico do Brasil**. Rio de Janeiro, RJ. 2000.
- IBGE. **Anuário Estatístico do Brasil**. Rio de Janeiro, RJ. 2001.
- IBGE. **Anuário Estatístico do Brasil**. Rio de Janeiro, RJ. 2002.
- IBGE. **Anuário Estatístico do Brasil**. Rio de Janeiro, RJ. 2003.
- IBGE. **Anuário Estatístico do Brasil**. Rio de Janeiro, RJ. 2004.
- IBGE. **Anuário Estatístico do Brasil**. Rio de Janeiro, RJ. 2005.
- IBGE. **Anuário Estatístico do Brasil**. Rio de Janeiro, RJ. 2006.
- IBGE. **Anuário Estatístico do Brasil**. Rio de Janeiro, RJ. 2007.
- IGUE, T.; LASCA, D. H. C. **Experimentação de campo**. Campinas: Coordenadoria de Assistência Técnica Integral, 1986. 32 p. ( Boletim Técnico, 191).
- INMETRO-Instituto Nacional de Metrologia, Normalização, Qualidade Industrial. **SI-Sistema internacional de unidades**. Rio de Janeiro: INMETRO, 2003. 112 p.
- INMETRO-Instituto Nacional de Metrologia, Normalização, Qualidade Industrial. **Vocabulário internacional de termos fundamentais e gerais de metrologia**. Brasília: INMETRO, 2000. 37 p.



- JACCARD, P. Étude comparative de La distribution florale dans une portion des Alpes et du jura. **Bulletin de La Société Vaudoise des Sciences naturelles**, n. 12, 1901.
- JAMES, B. R. **Probabilidade**: um curso em nível intermediário. 2. ed. Rio de Janeiro: IMPA, Projeto Euclides, 1996. 299 p.
- JECKEL, J. F. **Epidemiologia, bioestatística e medicina preventiva**. 2. ed. Porto Alegre: Artmed, 2005. 432 p.
- JEFFREYS, H. **Theory of probability**. Oxford: Clarendon Press, 1961. 459 p.
- JENKINSON, A. F. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. **Quarterly Journal of the Royal Meteorological Society**, v. 81, p.158-171, Apr. 1955.
- JOHNSON, N. L.; KOTZ, S. **Continous univariate distributions**. Boston: Houghton Millin, v. 2, 1970. 112 p.
- JOHNSON, N. L.; KOTZ, S. **Distribution in statistics, continue univariate distribution**. New York: Houghton Mifflin, v. 2, 1970. 328 p.
- KERLINGER, F. N. **Metodologia da pesquisa em ciências sociais**. São Paulo: Editora Pedagógica e Universitária Ltda, 1979. 396 p.
- KMENTA, J. **Elementos de econometria**: teoria econométrica básica. 2. ed. São Paulo: Editora Atlas, 1988. 215 p. v. 2.
- KOLMOGOROV, A. **Grundbegriffeder wahrscheinlichkeitsrechnun**. Berlin: Springer, 1933. 62 p.
- KOLMOGOROV, A. N. Sulla determinazione empirica di una legge di duistributione. **Giornale dell' Istituto Ialiano delgli Attuar**, v. 4, p. 1-11, 1933.
- KREBS, C. J. **Ecological methodology**. Menlo Park: Benjamin Cummings, 1998. 620p.
- KRUSKAL, J. B. Analysis of factorial experiments by estimating monotone transformations of the data. **Journal of the Royal Statistical Society**, London, Série B. n. 27, p. 251-263, 1965.
- KUHN, T. S. **The structure of scientific revolutions**. São Paulo: Perspectiva, 1969. 262 p.
- KVANLI, A. H. **Statistics**: a computer integrated approach publisher. Saint Paul: West Publishing Company, 1988. 935 p.
- LANDIM, P. M. B. **Introdução a geoestatística**. Rio Claro: UNESP, 1989. 69 p.
- LAKATOS, E. M.; MARCONI, M. A. **Metodologia científica**. 2. ed. São Paulo: Editora Atlas, 1991. 231 p.
- LAPLACE, P. **Théorie analytique des probabilités**. Paris: Courcier, 1812. 464 p.
- LARSON, H. J. **Introducion to probability**: theory and statistical inference. 2. ed. New York: Willy International, 1974. 430 p.
- Le CLERG, E. L.; LEONARD, W. H.; CLARK, A. G. **Field Plot technique**. 2. ed. Minneapolis: Burgess Publishing Company, 1966. 600 p.

- LIPSCHUTZ, S. **Probabilidade**. 3. ed. São Paulo: Editora McGraw-Hill do Brasil, 1972. 238 p.
- LIPSCHUTZ, S. **Probabilidade**. São Paulo: Mc Graw-Hill, 1974. 228 p.
- LIPSCHUTZ, S. **Teoria e problemas de probabilidade**. 2. ed. São Paulo: Mc Graw-Hill, 1974. 228 p.
- LOÉVE, M. **Probability theory**. 3. ed. New York: D. Van Nostrand, 1963. 681 p.
- LOI, I. **A mulher**. São Paulo: Jabuti, 1988. 53 p.
- LOPES, L. F. D. Glossário in: **Estatística e qualidade & produtividade**. Santa Maria, RS. Disponível em: <<http://www.felipelopes.com/glossario/glossario.asp>>. Acesso em: 29 mai. 2014.
- MACHADO, S. do A.; FIGUEIREDO FILHO, A. **Dendrometria**. Curitiba: UFPR, 2003. 309 p.
- MADOW, W. G. **Teoria dos levantamentos por amostragem**. Rio de Janeiro: IBGE-Conselho Nacional de Estatística, 1951. 273 p.
- MAGALHÃES, M. N.; LIMA, C. P. **Noções de probabilidade e estatística**. 6. ed. São Paulo: EDUSP, 2005. 416 p.
- MAGURRAN, A. E. **Measuring biological diversity**. Malden: Blackwell, 2004. 256 p.
- MAHALANOBIS, P. C. On the generalized distance in statistics. In: **Proceedings of the National Institute of Sciences of India**. n. 2, v.1, p. 49-55, 1936.
- MALHOTRA, N. K. **Pesquisa de marketing**: uma orientação aplicada. 3. ed. Porto Alegre: Bookman, 2001. 722 p.
- MARCONI, M. A.; LAKATOS, E. M. **Técnicas de pesquisa**. 2. ed. São Paulo: Editora Atlas S/A, 1990. 232 p.
- MARTINS, G.A.; DONAIRE, D. **Princípios de estatística**. 4. ed. São Paulo: Editora Atlas S/A, 1991. 256 p.
- MARTINS, M. E. G. Glossário in: **ALEA-Ação Local Estatística Aplicada**. Lisboa, Portugal. Disponível em: <<http://www.alea.pt/Html/glossar/html/glossar.html>>. Acesso em: 21 abr. 2015.
- MARTINS, M. J. **Simulação** (1ª Parte). Lisboa: Instituto Superior de Agronomia, 2003. 60 p. (notas de Aula).
- MARTINS, J.; BICUDO, M. A. V. **A pesquisa qualitativa em psicologia**: fundamentos e recursos básicos. São Paulo: Educ/Moraes, 1989. 110 p.
- MATHERON, G. Principles of geostatistics. **Economic Geology**, v. 58, n. 1, p. 1246-1266. 1963.
- MATHERON, G. **The theory of regionalized variables**. Paris: École Nationale Supérieure des Mines de Paris, 1971. 211 p. (Les Cahiers du Centre de Morphologie Mathématique de Fontainebleu, 5).
- MATOS, M. A. **Manual operacional para a regressão linear**. Porto: Faculdade de Engenharia da Universidade do Porto, 1995. 26 p. (Apostila).
- McCULLAGH, P.; NELDER, J. A. **Generalized linear models**. 2. ed. London: Chapman & Hall, 1991. 512 p.
- McCULLAGH, P.; NELDER, J. A. **Generalized linear models**. 2. ed. New York: Chapman & Hall, 1989. 511 p. (Monographs on Statistics and Applied Probability, 37).

- McCULLAGH, P.; NELDER, J. A. **Generalized linear models**. 2. ed. London: Chapman and Hall/CRC Press, 1989. 532 p.
- McDANIAL, C.; GATES, R. **Pesquisa de marketing**. São Paulo: Pioneira Thomson Learning, 2003. 562 p.
- MEDRONHO, R. A. et al. **Epidemiologia**. São Paulo: Editora Atheneu, 2006. 790 p.
- MELLO, F. G. **Introdução aos métodos estatísticos**. Lisboa: Cadernos do Instituto de Orientação Profissional 1, 1971. 136 p. v. 1
- MELLO, F. G. **Introdução aos métodos estatísticos**. Lisboa: Cadernos do Instituto de Orientação Profissional 4, 1973. 192 p. v. 2
- MELLO, F. G. **Probabilidades e estatística**. Lisboa: Escolar Editora, 1997. 686 p. v. 2
- MELLO, F. G. **Probabilidades e estatística**. 2. ed. Lisboa: Escolar Editora, 2000. 458 p. v. 1
- MEMÓRIA, J. M. P. **Curso de estatística aplicada à pesquisa científica**. Viçosa: Universidade Federal de Viçosa, 1973. 314 p. (Apostila).
- MEMÓRIA, J. M. P. **Curso de estatística aplicada à pesquisa científica**. Fortaleza: Universidade federal do Ceará. 1960. 246 p. (Mimeografado).
- MENDES, B. V. M. **Introdução à análise de eventos extremos**. Rio de Janeiro: E-papers Serviços Editoriais Ltda, 2004. 232 p.
- MEYER, P. L. **Probabilidade**: aplicações à estatística. Rio de Janeiro: Livro Técnico, 1969. 391 p.
- MEYER, P. L. **Probabilidade**: aplicações à estatística. 1. Ed., 2ª Reimpressão. Rio de Janeiro: Livros Técnicos Científicos Editora S/A, 1978. 405 p.
- MIGON, H. S.; GARMERMAN, D. **Statistical inference**: an integrated approach. London: Arnold, 1999. 261 p.
- MISCHAN, M. M.; PINHO, S. Z. **Experimentação agrônoma**: dados não-balanceados. Botucatu: Fundação do Instituto de Biociências-FUNDIBIO. Universidade Estadual Paulista-UNESP. 1996. 472 p.
- MONTGOMERY, D. C. **Estatística aplicada e probabilidade para engenheiros**. 4. ed. Rio de Janeiro: LTC, 2009. 514 p.
- MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. **Introduction to the theory of statistics**. New York: McGraw-Hill, 1974. 536 p.
- MOORE, D. S.; McCABE, G. P. **Introdução à prática da estatística**. 3. ed. Rio de Janeiro: LTC, 2002. 482 p.
- MORABITO, R. **Modelos probabilísticos aplicados à engenharia de produção**. São Carlos: Universidade federal de São Carlos, 2002. 121 p.
- MORETTIN, L. G. **Estatística básica**: probabilidade. São Paulo: Mc Graw-Hill, 2003. 198 p.
- MORETTIN, P. A. **Introdução à estatística para ciências exatas**. São Paulo: Atual Editora, 1981. 224 p.

- MORETTIN, P. A.; TOLOI, C. M. **Análise de séries temporais**. 2. ed. São Paulo: Editora Edgard Blucher, 2006. 564 p.
- MORETTIN, P. A.; BUSSAB, W. O. **Estatística básica**. 5. ed. São Paulo: Saraiva, 2003. 526 p.
- MUNDSTOCK, E. C. **Amostragem II**. Porto Alegre: Universidade Federal do Rio Grande do Sul. 2006. Série B, nº 53, 57 p. (Trabalho de apoio didático).
- MUNIZ, J. A.; ABREU, A. R. **Técnicas de amostragem**. Lavras: Universidade Federal de Lavras, 1999. 102 p.
- MURTEIRA, B. J. F. **Probabilidades e estatística**. 2. ed. Lisboa: Editora McGraw-Hill, 1990. 413 p. v. 1.
- MURTEIRA, B. J. F. **Análise exploratória de dados**: estatística descritiva. Lisboa: Editora McGraw-Hill de Portugal Ltda, 1993. 341p.
- MURTEIRA, B. et al. **Introdução à estatística**. Portugal: Editora McGraw-Hill, 2001. 688 p.
- NAGHETTINI, M.; PINTO, E. J. A. **Hidrologia estatística**. Belo Horizonte: CPRM- Serviço geológico do Brasil, 2007. 561 p.
- NASCIMENTO, W. A. **Amostragem de conglomerados**. Rio de Janeiro: Escola Nacional de Ciências Estatísticas-IBGE, 1981. 156 p.
- NEDER, H. D. **Amostragem em pesquisas sócio-econômicos**. Uberlândia: Universidade Federal de Uberlândia, 2008. 103 p. (Apostila).
- NELDER, J. A.; WEDDERBURN, W. M. Generalized linear models. **Journal of the Royal Statistical Society**, Series A (General), v. 135, n. 3, p. 370-384, 1972.
- NETER, J.; WASSERMAN, W.; KUTNER, M. H. **Applied linear statistical models**. Homewood: Richard D. Irwin. 1990. 1181p.
- NEUMANN, J. V.; MORGENSTERN, O. **Theory of games and economic behavior**. Princeton: Princeton University Press, 1944. 776 p.
- NUNES, R. P. **Métodos para a pesquisa agrônoma**. Fortaleza: Universidade Federal do Ceará, 1998. 564 p.
- OGLIARI, P. J. **Estatística para as ciências agrárias e biológicas**. Florianópolis: Editora da UFSC, 2006. 650 p.
- O'HAGAN, A. **Bayesian inference**. London: v. 2B of Kendall's Advanced Theory of Statistics. Edward Arnold, 1994. 496 p.
- OLIVEIRA, L. M. **Distribuições estatísticas**. Viçosa: Imprensa Universitária. Universidade Federal de Viçosa-UFV, 1970. 184 p. (Apostila).
- OLIVEIRA, F. E. M. **Estatística e probabilidade**: exercícios resolvidos e propostos. São Paulo: Editora Atlas, 1995. 128 p.
- OLIVEIRA, M. S.; BEARZOTI, E. VILAS BOAS, F. L.; NOGUEIRA, D. A. LUCI; NICOLAU, L. A. **Introdução à estatística**. 2. ed. Lavras: Editora da UFLA, 2014. 461 p.

- PACE, L.; SALVAN, A. **Principles of statistical inference from a neo-fisherian perspective**. Singapore: Science & Applied Probability, Uto-print, 1997. 535 p. v. 4.
- PADOVANI, C. R. P.; ARAGON, F. F. Programa computacional para método de discriminante de Fisher. **Energia na Agricultura**, Botucatu, v. 20, n. 1, p. 1-10. 2005.
- PARETO, V. **Cours d'economie politique**. Paris: Rouger, éditeur, 1897. 426 p. v. 2.
- PATTERSON, H. D.; THOMPSON, R. Recovery of inter-block information when block sizes are unequal. **Biometrika**, London, v. 58, p. 545-554, 1971.
- PEARSON, K. Skew variation, a rejoinder. **Biometrika**, n. 4, p. 169-212, 1905.
- PEDROSA, A. C. ; GAMA, S. M. A. **Instrução computacional à probabilidade e estatística**. Porto: Porto Editora Ltda, 2004. 607 p.
- PÉLLICO-NETTO, S.; BRENA, D.A. **Inventário florestal**. Curitiba: Editorado pelos autores, 1997. 316p.
- PEREIRA, A. C. **MEDSTATWEB**. Um curso de estatística médica na web. Serviço de Bioestatística Médica. Faculdade de Medicina da Universidade do Porto. Disponível em: <<http://stat2.med.up.pt/cursop/glossario/index.html>>. Acesso em: 16 jun. 2014.
- PEREIRA, M. G. **Epidemiologia: teoria e prática**. Rio de Janeiro: Editora Guanabara Koogan S/A, 1995. 516 p.
- PEREIRA, W.; KIRSTEN, J.T.; ALVES, W. **Estatística para as ciências sociais: teoria e aplicações**. São Paulo: Editora Saraiva, 1980. 378 p.
- PETRIE, A.; WATSON, P. **Estatística em ciência animal e veterinária**. 2. ed. São Paulo: Editora Roca, 2009. 248 p.
- PETERNELLI, L. A.; MELLO, M. P. **Conhecendo o R uma visão estatística**. Viçosa: Editora da UFV, 2007. 118 p.
- PHILIP, M. S. **Measuring trees and forests**. 2. ed. Wallingford: CAB International, 1994. 310 p.
- PIMENTEL GOMES, F. **Iniciação à estatística**. 5. ed. São Paulo: Livraria Nobel S/A, 1976. 236 p.
- PIMENTEL GOMES, F. **A estatística moderna na pesquisa agropecuária**. Piracicaba: Associação Brasileira para pesquisa da Potassa e do Fosfato. Potafos, 1984. 160 p.
- PIMENTEL GOMES, F. **Curso de estatística experimental**. 15. ed. Piracicaba: FEALQ, 2009. 451 p.
- RAMALHO, M. A. P.; FERREIRA, D. F.; OLIVEIRA, A. C. **Experimentação em genética e melhoramento de plantas**. Lavras: Editora da UFLA, 2000. 326 p.
- RAO, C.R. On a class of arrangements. **Proc. Edinburgh Math. Soc.**, v. 8, p. 119-125, 1949.
- RAVAGNANI, F. R.; CATELAN, L. **Glossário de estatística**. São Paulo: NETRA, 2002. 161p. Disponível em: <[http://www.unibero.edu.br/glossarios\\_def.asp](http://www.unibero.edu.br/glossarios_def.asp)>. Acesso em: 26 maio 2014.
- REA., L. M.; PARKER, R. A. **Metodologia de pesquisa: do planejamento à execução**. 3. ed. São Paulo: Editora Pioneira, 2002. 282 p.

- REIS, M. M. **Conceitos elementares de estatística**. Florianópolis. 2006. 8 p. Disponível em: <<http://www.inf.ufsc.br/~marcelo/intro.html>>. Acesso em: 17 abr. 2015.
- RESENDE, M. D. V. de. **Genética biométrica e estatística no melhoramento de plantas perenes**. Brasília: Embrapa Informação Tecnológica, 2002. 975 p.
- RESENDE, M. D. V. Matemática e estatística na análise de experimentos e no melhoramento genético. Colombo: Editora da Embrapa, 2007. 562 p.
- RICHARDSON, R. J. et al. **Pesquisa social: métodos e técnicas**. 3. ed. São Paulo: Editora Atlas, 2009. 334 p.
- ROBBINS, H. The empirical Bayes approach to statistics. In: **Proc. of the thirdd berkeley symp. on math statistics and probability**. v. I, University of California Press, Berkeley, p. 157-164. 1956.
- RODRIGUES, M. S. **Dicionário brasileiro de estatística**. 2. ed. Rio de Janeiro: Fundação IBGE, 1970. 358 p.
- RODRIGUES, P. C. **Teoria da amostragem**. Lisboa: Universidade Nova de Lisboa, 2006. 54 p. (Notas de Estatística Matemática).
- ROSÁRIO, M. F. 120 anos do nascimento do cientista R. A. Fisher (1890-2010). **Rev. Bras. Biom.**, São Paulo, v. 27, n. 4, p. 659-672, 2009.
- ROUQUAYROL, M. Z.; ALMEIDA FILHO, N. **Epidemiologia & saúde**. 5. ed., Rio de Janeiro: Editora Médica e Científica-MEDSI, 1999. 600 p.
- ROY, S. N. p-statistics and some generalizations in analysis of variance appropriate to multivariate problems. **Sankhyā**, 4, p. 381-396, 1939.
- RUIZ, J. A. **Metodologia científica: guia para eficiência nos estudos**. 6. ed. São Paulo: Editora Atlas, 2006. 184 p.
- RÜMELIN, G. Zur theori der statistic. **ZGS**, n. 19, p. 653-696, 1863.
- SAEG. **Sistema para análises estatísticas e genéticas**. Versão 9.1. Viçosa: Fundação Arthur Bernardes. Universidade Federal de Viçosa, 2007. Disponível em: <<http://www.ufv.br/saeg>> Acesso em: 31 mar. 2015.
- SALVATORE, D. **Estatística e econometria**. São Paulo: Editora McGraw-Hill do Brasil, 1982. 273 p.
- SANTO, A. E. **Delineamentos de metodologia científica**. São Paulo: Editora Loyola, 1992. 171 p.
- SATTERTHWAITE, F. E.; An. approximate distribution of estimates of variance components. **Biometrics Bulletin**, Washington, n. 2, p. 110-114, 1946.
- SCOTT, D. W. On optimal and data-based histograms. **Biometrika**, London, v. 66. n. 3, p. 605-610, 1979.
- SELLTIZ, C.; WRIGHTSMAN, L. S.; COOK, S. W. **Métodos de pesquisa nas relações sociais: Medidas na pesquisa social**. 2. ed. São Paulo: Editora Pedagógica e Universitária, 1987. 152 p. v. 2.
- SELLTIZ, C.; WRIGHTSMAN, L. S.; COOK, S. W. **Métodos de pesquisa nas relações sociais: delineamentos de pesquisa**. 2. ed. São Paulo: Editora Pedagógica e Universitária, 1987. 136 p. v. 1.
- SELLTIZ, C.; WRIGHTSMAN, L. S.; COOK, S. W. **Métodos de pesquisa nas relações sociais: análise de resultados**. 2. ed. São Paulo: Editora pedagógica e universitária, 1987. 88 p. v. 3.



- SEVERINI, T. A. **Likelihood methods in statistics**. Oxford: Oxford Statistical Science Series 22. Pub., Chapter 3, Likelihood, p. 73-104, 2000. 392 p.
- SHEWHART, W. **Economic control of quality of manufactured product**. New York: D. Van Nostrand Company, 1931. 501 p.
- SHIVER, B. D.; BORDERS, B. E. **Sampling techniques for forest resource inventory**. New York: John Wiley & Sons, 1996. 356 p.
- SHULL, G. H. Duplicate genes for capsule form in bursa bursa-pestoris. **Z. I. A.** v. 12, p. 97- 149, 1914.
- SIEGEL, S. **Estatística não-paramétrica para ciências do comportamento**. São Paulo: McGraw-Hill do Brasil, 1981. 288p.
- SIEGEL, S.; CASTELLAN JR, N. J. **Estatística não-paramétrica para ciências do comportamento**. 2. ed. Porto Alegre: Artmed, 2006. 448p.
- SILVA, I. P.; SILVA, J. A. A. **Métodos estatísticos aplicados à pesquisa científica: uma abordagem para profissionais da pesquisa agropecuária**. Recife: Universidade Federal Rural de Pernambuco, 1999. 305 p.
- SILVA, N. N. **Amostragem probabilística: um curso introdutório**. São Paulo: Editora da Universidade de São Paulo, 1998. 128 p.
- SILVA, P. C. R. Introdução à estatística econômica in: **Aula net: curso de estatística on line**. Vitória, ES. Disponível em: <<http://www.ai.com.br/pessoal/indices/2a0.htm>>. Acesso em: 17 abr. 2015.
- SILVA, P. S. L.; SOUSA, R. P. **Técnicas experimentais para o milho**. Mossoró: Fundação Guimarães Duque, DCLV, 1991. 192 p. (Coleção Mossoroense, série C).
- SINHA, B. K. **Teoria unificada de amostragem: uma introdução**. Salvador: Universidade Federal da Bahia, 1979. 136 p.
- SKULL, G. H. Duplicate genes for capsule form in Bursa bursa-pestoris. **Z. I. A.** v. 12, p. 97-149, 1914.
- SMIRNOV, N. On the estimation of the discrepancy between empirical curves of distribution for two independent samples. **Bulletin Mathematique de l'Universite de Moscou**. Moscow, v. 2, n. 2, p. 3-14, 1939a.
- SMIRNOV, N. Sur les ecart de la courbe de distribution empirique. **Recueil Mathematique (Matematiceskii Sbornik)**, v. 6, p. 3-26, 1939b.
- SNEDECOR, G. W. **Calculation and interpretation of analysis of variance and covariance**. Iowa: Ames Collegiate Press, 1934. 96 p.
- SOARES, J. F.; SIQUEIRA, A. L. **Introdução à estatística médica**. 2. ed. Belo Horizonte: COOPMED- Editora Médica, 2002. 300 p.
- SOKAL, R. R.; ROHLF, F. J. **Biometry: the principles of statistics in biological research**. 3. ed. New York: Freeman & Company, 1995. 887 p.
- SOONG, T. T. **Modelos probabilísticos em engenharia e ciências**. Rio de Janeiro: LTC, 1986. 238 p.
- SPIEGEL, M. R. **Estatística**. 2. ed. São Paulo: McGraw-Hill do Brasil, 1985. 463 p. (Coleção Shaum).

- SPIEGEL, M. R. **Probabilidade e estatística**. São Paulo: McGraw-Hill do Brasil, 1978. 528 p.
- SPIEGEL, M. R.; SCHILLER, J.; ALU SRUNIVASAN, E. R. **Probabilidade e estatística**. 2. ed. Porto Alegre: Bookman, 2004. 398p.
- SRIVASTAVA, S. N. Three-dimensional refinement of the structure of anthrone. **Acta Cryst**, n. 17, p. 851-856, 1964.
- STEEL, R. G. D.; TORRIE, J. H. **Introduction to statistics**. New York: McGraw-Hill Book Company, 1976. 387p.
- STEEL, R. G. D.; TORRIE, J. H. **Principles and procedures of statistics**. New York: Mc Graw-Hill Book Company, 1960. 481 p.
- STEEL, R. G. D.; TORRIE, J. H. **Principles and procedures of statistics**. 2. ed. New York: McGrawHill Book Company. Inc. N. Y, 1981. 633 p.
- STEVENSON, W. J. **Estatística aplicada à administração**. 1. Ed. São Paulo: Editora HARBRA. 498 p. 2001.
- STORCK, L. et al. **Experimentação vegetal**. 2. ed. Santa Maria: UFSM, 2006. 198 p.
- TAVARES, M. **Estatística aplicada à administração**. Brasília: Universidade aberta do Brasil, 2007. 142 p.
- TAYLOR, L. R. Aggregation, variance and the mean. **Nature**, Londres, n. 189, p. 732-735, 1961.
- THOMPSON, R. The estimation of heritability with unbalanced data. **Biometrics**, v. 33, p. 485-504, 1977.
- THOMPSON, S. K. **Sampling**. New York: Editora John Wiley & Sons Inc, 2002. 400 p.
- TOFFOLI, S. F. L.; SODRÉ, U. **Ensino fundamental: mini dicionário de matemática elementar**. Londrina, 2006. 19 p. Disponível em: <<http://pessoal.sercomtel.com.br/matematica/fundam/diciomat/Diciomat.htm>>. Acesso em: 23 maio de 2016.
- TOLEDO, G. L.; OVALLE, I. I. **Estatística básica**. 2 ed. São Paulo: Atlas, 1987. 460p.
- THOMPSON, S. K.; SEBER, G. A. F. **Adaptive sampling**. New York: John Wiley. 1996. 265p.
- TRIOLA, M. F. **Introdução à estatística**. 7. ed. Rio de Janeiro: LTC, 1999. 434 p.
- TUKEY, J. Bias and confidence in not quite large samples. **Ann. Math. Statist.**, n. 29, 614 p. 1958.
- UFRJ. **Análise discriminante**. Rio de Janeiro, 2008. 33 p. Disponível em: <<http://www.im.ufrj.br/flavia/mad484/discrim.ppt>>. Acesso em: 13 abr. 2015.
- UNESCO. **Análise discriminante (DISCRAN)**. Nova York. 2016. 8 p. Disponível em: <<http://www.unesco.org/webworld/portal/idams/html/portuguese/P1discra.htm>>. Acesso em: 23 maio 2016.
- VELARDE, L. G. C., MIGON, H. S.; PEREIRA B. B. Space-time modeling of non-negative variables with mass point at zero. **Technical Report # 141: DME-UFRJ**, Rio de Janeiro, RJ. 2001. 20 p.
- VENABLES, W.N.; RIPLEY, B.D. **Modern applied statistics with S-PLUS**. 2. ed. New York: Springer, 1997. 462 p.



- VIALI, L. Glossário. in: **Statweb**. Porto Alegre. 2016. 25 p. Disponível em: <<http://www.pucrs.br/uni/poa/famat/statweb/>>. Acesso em: 23 maio 2016.
- VIEIRA, S. **Bioestatística: tópicos avançados**. Rio de Janeiro: Editora Campus Ltda. 2003. 224 p.
- VIEIRA, S.; HOFFMANN, R. **Estatística experimental**. São Paulo: Atlas, 1989. 179 p.
- Von BORTKIEWICZ, L. **Das gesetz der kleinen zahlen**. Leipzig: B. G. Teubner. Germany, 1898. 136 p.
- Von NEUMANN, J.; MORGENSTERN, O. **Theory of games and economic behavior**. Princeton, N. J.: Princeton University Press, 1944. 776 p.
- WALKER, G.T. Correlation in seasonal variation of weather. III. On the criterion for the reality of relationships or periodicities. **Mem. Indian Meteorol.**, India., v. 21, n. 9. p. 13-15, 1923.
- WATERS, A. C. Note upon estimate of population. In: **Seventieth annual report of the registrar general**. London, 1907. 705 p.
- WAUGH, A. E. **Elements of statistical method**. 3. ed., New York: McGraw-Hill, 1952. 531 p.
- WEDDERBURN, R. W. M. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. **Biometrika**, v. 61, n. 3, p. 439-447, 1974.
- WEISS, L.; WOLFOWITZ, J. Generalized maximum likelihood estimators. **Theory Probab. Appl.**, n. 11, p. 58-81, 1966.
- WIKIPÉDIA. Disponível em: < <http://www.wikipedia.org> > Acesso em: 27 maio 2016.
- WITTSTEIN, T. **Mathematische Statistik**. Hannover: Hanhsche Hofbuchlandlung, 1867. 55 p.
- WISHART, J. The generalized product-moment distribution in samples from a normal population. **Biometrika**, v. 20 A, p. 32-52, 1928.
- WONNACOTT, R. J.; WONNACOTT, T. H. **Fundamentos de estatística**. Rio de Janeiro: L.T.C, 1985. 369 p.
- WONNACOTT, T. H.; WONNACOTT, R. J. **Estatística aplicada à economia e a administração**. Rio de Janeiro: LTC, 1981. 710 p.
- WOOLHOUSE, W. S. B. Explanation of a new method of adjusting mortality tables. **J. I. A.**, n. 15, p. 389, 1870.
- YEO, I. K.; JOHNSON, R. A. A new family of power transformation to improve normality or symmetry. **Biometrika**, v. 87, n. 4, p. 954-959, 2000.
- YEVJEVICH, V. M. Section 8-II statistical and probability analysis of hydrological data. Part II Regression and correlation analysis. In: CHOW, V. T. **Handbook of applied hydrology**. Tucson: McGraw-Hill Company, 1964. 1495 p.
- YOU DEN, W. J. Experimental designs to increase accuracy of greenhouse studies. **Contr. Boyce Thompson Inst.**, v. 11, p. 219-228, 1940.

- YULE, G. U.; KENDALL, M. G. **An introduction to the theory of statistics**. London: Charles Griffin & Co, 1940. 701 p.
- YULE, G. U.; KENDALL, M. G. **An introduction to the theory of statistics**. 14. ed. London: Charles Griffin & Company Limited, 1958. 681 p.
- ZIMMERMANN, F. J. **Estatística aplicada à pesquisa agrícola**. Santo Antônio de Goiás: Empresa Brasileira de Pesquisa Agropecuária-EMBRAPA. Centro Nacional de Pesquisa de Arroz e Feijão-CNPAP, 2004. 402 p.

Editora Universitária da UFERSA (EdUFERSA)  
Av. Francisco Mota, 572  
Compl.: Centro de Convivência  
Costa e Silva - Mossoró/RN - CEP: 59.625-900  
(84) 3317-8267  
<http://edufersa.ufersa.edu.br>  
[edufersa@ufersa.edu.br](mailto:edufersa@ufersa.edu.br)

Formato: PDF  
Números de páginas: 901

A realização desta obra deve-se a necessidade de preencher uma lacuna até hoje sem iniciativas concretas para por no mercado editorial um livro dicionário impresso em português do Brasil, para servir de guia de consultas a todos aqueles que utilizam a estatística como ferramenta de trabalho nas diversas áreas da pesquisa científica, no comércio, no mercado financeiro, na indústria, nas universidades e nas escolas de uma maneira geral, pois em algum momento de suas análises e interpretações necessitamos tirar dúvidas sobre termos, verbetes, símbolos, nomenclaturas estatística, dentre outros.

Neste glossário descreve-se o maior número possível de verbetes e termos mais comuns usados na estatística que achamos razoável para uma obra dessa natureza. O leitor encontrará palavras iniciando desde a vogal A, como por exemplo, a palavra amostra, até a consoante Z como a palavra zona de aceitação, muitas vezes repetidas com outra descrição, mas com o mesmo significado. Os verbetes foram inseridos usando como critério a letra inicial do termo descrito, e não a ordem alfabética, isto porque muitos termos são fórmulas, equações, dentre outros, o que torna sem sentido este critério de inserção ou descrição dos termos do glossário, ou seja, as palavras foram alojadas em capítulos definidos e iniciados pelas letras A, B, C, D, E, ... , dentre outras.

Encontra-se também neste livro um Apêndice com descrições de símbolos, de siglas e de abreviações, simbologias usadas nos testes de hipóteses, glossário matemático, alfabeto grego, bem como, uma relação nominal de 330 programas, pacotes, aplicativos ou softwares utilizados isolados ou quando requerem outros pacotes para funcionar, aplicados nas áreas de estatística, auditoria, contabilometria, engenharia, atuária, administração, econometria, geoestatística, cálculo e matemática, genética, na construção de mapas e gráficos, na edição de banco de dados, de imagens e de fotos, em simulação e como editor de textos.