



**APLICAÇÕES ESTATÍSTICAS
NAS ÁREAS DAS CIÊNCIAS BIO-MÉDICAS**

Desenvolvimento estatístico:

Manuel Ayres

Professor Emérito da Universidade Federal do Pará

Manuel Ayres Jr. Ph.D.,

Universidade de Maryland

Desenvolvimento do software

Daniel Lima Ayres

Engenheiro de Computação

Universidade de Illinois

Alex de Assis Santos dos Santos

Tecnólogo em Processamento de Dados

Instituto Bioestatístico de Ciência e Tecnologia

BELÉM – PARÁ – BRASIL

ANO 2007

Manuel Ayres, Manuel Ayres Júnior, Daniel Lima Ayres, Alex de Assis Santos dos Santos e Lucas Lima Ayres dedicam esta versão – 5.0 – do BioEstat a IZA DO AMARAL CORRÊA AYRES, com um abraço carinhoso, repassado com admiração e apreço.

Ao longo de mais de dez (10) anos ininterruptos, você IZA tem nos proporcionado apoio incondicional na elaboração de nosso trabalho científico, participando direta e indiretamente com sugestões, na escolha de textos e imagens que enriquecem nosso livro, e no apoio material e espiritual desmedido para o aperfeiçoamento de uma ferramenta imprescindível para todos aqueles que desenvolvem pesquisa em nosso País. Sua bondade e humanidade nos tem sido preciosos e você tem sido a grande responsável por nossa alegria e energia nesse labor, não havendo qualquer exagero nessa afirmação. A unanimidade dentre aqueles que tem sorte de conhecê-la de perto não deixa margem a equívocos quanto ao estofo de que é feita a sua personalidade, sendo uma criatura singular – difícil de encontrar nesse mundo conturbado em que vivemos. Você é a Mãe admirável, a Mulher incomparável e a amiga mais carinhosa que poderíamos desejar.

Um muito obrigado de todos os seus fiéis admiradores:

Manuel Ayres

Manuel Ayres Júnior

Daniel Lima Ayres

Alex Santos dos Santos

Lucas Lima Ayres

Prefácio da 5ª Edição

O universo encontra-se em constante evolução. A idéia de que o mundo seria imutável, com toda a sua monotonia, foi alterada por outra, de uma dinâmica incessante. Da mesma maneira, a família Ayres e seu fiel companheiro Alex Santos dos Santos também não param. Estamos agora diante da quinta versão de um manual que já provou sua imensa utilidade junto à comunidade científica brasileira. Tendo como origem texto publicado há 20 anos atrás, as diferentes versões do BioEstat estão sempre procurando manter-se alertas às necessidades desta comunidade, em uma área que tradicionalmente é considerada difícil por alunos e profissionais da biologia e biomedicina.

Esta natural (será mesmo natural?) aversão de biólogos e biomédicos à matemática e à estatística está sendo vencida, e boa parte do progresso alcançado vincula-se à revolução causada pela eletrônica. Com a internet houve uma considerável democratização do saber. Informações que antes levavam meses para transitar entre o Primeiro e o Terceiro Mundos atualmente podem ser obtidas de maneira quase instantânea. Da mesma forma, os pacotes estatísticos tornaram imensamente mais fácil a tarefa dos neófitos, como pode ser perfeitamente testemunhado por todos os usuários das versões anteriores do BioEstat.

O que pode ser encontrado nas mesmas, bem como na versão atual? Bem, ali (e aqui) estão matérias presentes também em qualquer tratado de bioestatística, como probabilidades; amostragem; construção de gráficos e tabelas; fórmulas mais empregadas; estimação de parâmetros clássicos; correlação e regressão; e análise de variância. Mas também há tópicos mais especializados e de grande valor, como análise de sobrevivência; “bootstrap”; ou meta-análises diversas.

O que há de novo? O teste de Mantel; acréscimo aos testes de risco relativo; índices de diversidade; e diversos outros aplicativos que complementam de maneira importante aspectos de processos considerados anteriormente.

Qual seria a mensagem final desta breve introdução? A de que devemos agradecer que cérebros como os dos autores deste livro estejam se dedicando à difícil tarefa de colocar à disposição de seus colegas ou de qualquer outro interessado as ferramentas básicas necessárias à utilização de conceitos e interpretação de processos científicos. Em uma época em que lado a lado com os fantásticos progressos da ciência pululam por todo o mundo movimentos anticiência é salutar encontrar obras como esta: uma aragem revigorante que parte do ambiente quase sempre escaldante da nossa Amazônia!

Francisco M. Salzano

Professor Emérito

Universidade Federal do Rio Grande do Sul

Prefácio da 1^a e 2^a Edições

“O alívio da pobreza na Amazônia só pode ser feito se diminuirmos a migração da zona rural para as grandes cidades. Para mantermos a população humana na zona rural, além de darmos infra-estrutura escolar, sanitária e médica, precisamos manejar os recursos naturais de forma sustentável com base científica sólida e assim manter a estrutura dos ecossistemas através da conservação dos processos ecológicos e evolutivos”.

José Márcio Ayres

APRESENTAÇÃO

O pacote estatístico *BioEstat*, aqui apresentado, tem como objetivo propiciar aos acadêmicos de diversas áreas do conhecimento um instrumento de grande praticidade e de fácil manuseio na avaliação de informações originadas através de pesquisa, sendo composto por este manual e de um aplicativo. Este estudo representa na verdade uma evolução bastante significativa dos mesmos autores de *Aplicações Estatísticas em Basic*, publicado em 1987 pela editora McGraw-Hill e, através de observações ao longo dos anos, oriundas do curso de Bioestatística fornecido pelo autor principal na Universidade Federal do Pará.

Este trabalho foi realizado por três gerações dos Ayres: a primeira está representada por Manuel Ayres, professor de Estatística da Universidade Federal do Pará; a segunda, por Manuel Ayres Jr., PhD em engenharia civil pela Universidade de Maryland (EUA); e a terceira, pelo estudante do segundo grau, Daniel Lima Ayres, com 15 anos de idade, que desenvolveu o aplicativo, juntamente com Alex de Assis Santos dos Santos, em Visual Basic, e a diagramação deste manual.

BioEstat é dirigido especialmente aos estudantes de graduação e pós-graduação das áreas médica e biológica que possuam noções básicas de estatística. O pacote é bastante facilitado pelo emprego do aplicativo e deste manual, onde a escolha do teste adequado poderá ser feita de maneira prática de acordo com a natureza dos dados, número de amostras e tipo de experimento. Ainda, para facilitar seu uso, cada teste tem seu respectivo exemplo que auxilia o emprego deste pacote estatístico. Os exemplos aqui

fornecidos podem ter sido originados de situações práticas reais ou puramente hipotéticas.

Embora os testes estatísticos aqui apresentados sejam empregados nas mais diversas áreas da ciência, a escolha foi baseada na demanda levantada ao longo dos últimos anos por estudantes das áreas médicas e de diversos segmentos da área biológica. Poderão ser observadas pequenas divergências de resultados quando comparados a outros pacotes desse gênero. Essas diferenças, porventura detectadas, podem estar relacionadas ao emprego de distintos algoritmos.

Sugestões para melhoria do pacote e possíveis erros encontrados durante o seu uso serão bem recebidos pelos autores. Essas modificações também poderão ser inseridas em futuras versões.

J. Márcio C. Ayres

MPEG/CNPq

Wildlife Conservation Society –
NY

Belém, Pará, Maio de 1998.

In memoriam: o Autor desta Apresentação faleceu em 7 de março de 2003 e, se vive fosse, o segundo parágrafo desta Apresentação teria certamente o seguinte teor:

Este trabalho foi realizado por três gerações dos Ayres: a primeira está representada por Manuel Ayres, ex-professor de Estatística da Universidade Federal do Pará; a segunda, por Manuel Ayres Jr., PhD em engenharia civil pela Universidade de Maryland (EUA); e a terceira, por Daniel Lima Ayres, engenheiro de computação pela Universidade de Illinois (EUA), que desenvolveu o aplicativo, juntamente com Alex de Assis Santos dos Santos, e a diagramação deste manual, e, mais recentemente, Lucas Lima Ayres, estudante de medicina veterinária da Universidade Federal de Minas Gerais, o qual contribuiu no desenvolvimento do Capítulo 18 sobre Estatísticas Circulares.

Prefácio da 3ª Edição

A conservação da biodiversidade é componente essencial para qualquer política de desenvolvimento sustentável. A falta de uma política consistente para este fim poderá levar a contínua eliminação de diferentes formas de vida, perda de recursos genéticos e a modificação de processos ecológicos e evolutivos que são essenciais para a própria sobrevivência da nossa própria espécie. Como demonstrado por inúmeros estudos, a perda da diversidade biológica terá custos enormes para toda a humanidade, tanto se utilizarmos os indicadores econômicos tradicionais, como se considerarmos também os valores sociais, estéticos e espirituais que fazem parte da nossa cultura.

A Conservation International acredita que conservação da biodiversidade deve ser baseada em ciência de boa qualidade. Por isso, estamos estabelecendo, através dos nossos Centros de Conservação da Biodiversidade, parcerias diversas para o desenvolvimento de pesquisas e capacitação em todas as áreas do conhecimento relevantes para a ação de conservação. Independente da área, o conhecimento da estatística é fundamental.

Uma necessidade comum no Brasil, tanto para os cientistas mais novos como para os que já possuem anos de experiência, é a existência de programas de computador que sejam efetivos para as análises estatísticas. É nesse contexto, que temos a honra de apoiar a publicação da terceira edição do livro BIOESTAT, escrito com toda a competência por Manuel Ayres, Manuel Ayres Júnior, Daniel Lima Ayres e Alex Santos dos Santos. Há poucos cientistas no Brasil que não conhecem as versões anteriores e certamente são milhares os estudantes de graduação e pós-graduação no país que iniciaram ou aprofundaram seus estudos de estatística utilizando este livro. O sucesso das duas versões anteriores foi tão grande, que os autores tiveram que trabalhar bastante para atender a uma demanda enorme para a atualização e publicação periódica de um livro tão importante para todos.

Em nome de todos os que se dedicam ao estudo e a conservação da biodiversidade brasileira e mundial, desejamos agradecer o extraordinário esforço dos autores para produzir uma obra tão importante para todos, assim como reconhecer a importância do trabalho realizado por nosso grande amigo José Márcio Ayres, cujo trabalho fenomenal para demonstrar de forma inequívoca a indissociabilidade entre ciência e conservação.

Russel A. Mittermeier

Presidente

Conservation International

Gustavo A. B. da Fonseca

Vice-Presidente Executivo

Conservation International

Prefácio da 4ª Edição

A Estatística tem tido grande importância ao longo de sua história, que começa Antes de Cristo, com os registros egípcios “de presos de guerra”, a partir da organização de bancos de dados que possibilitaram o estabelecimento de uma relação-verdade entre o passado, o presente e o futuro.

O *BioEstat* trilha a égide que fortalece o conhecimento da Estatística como ciência, desvendando modelos matemáticos e procedimentos de análises e, assim, possibilitando maior facilidade na interpretação dos fenômenos que se manifestam espontaneamente ou são induzidos à natureza dos seres vivos, sobretudo pelo próprio Homem.

A evolução deste trabalho didático deve-se, entre outros fatores, às fases evolutivas dos conhecimentos estatísticos. A princípio houve a preocupação de introduzir aqueles aplicativos de uso mais geral, e assim o *BioEstat* 1.0 se apresentou como ferramenta destinada a oferecer aspectos voltados sobretudo para estudantes de graduação e pós-graduação das áreas médicas e biológicas. Na segunda etapa, o *BioEstat* 2.0 não perde seu foco inicial, mas evolui com a inclusão de novos modelos de análises e ultrapassa as fronteiras brasileiras e alcança o público técnico-científico da língua espanhola.

Na fase seguinte, com a adição de novos aplicativos, a sugestão da escolha dos testes estatísticos e o cálculo do tamanho das amostras, o *BioEstat* 3.0 solidifica-se como ferramenta didática quase obrigatória devida, sobretudo, à facilidade de sua aplicação e interpretação dos fenômenos biológicos, sejam de ordem observacional, quer de natureza experimental. Tem assim início, e de forma definitiva, a empatia entre usuários e autores, cada um interpretando as necessidades de desenvolvimento do outro.

Por fim, ocorre a interdependência de diversas áreas do conhecimento das ciências biomédicas com a Biometria – o *BioEstat* 4.0 chega com nova roupagem, com mais 15 aplicativos, sem se despir de sua trajetória histórica, mas atendendo as exigências dos usuários, ávidos por ampliar seus conhecimentos e se coloca, com seus 184 testes, como alternativa didática para outras áreas das ciências, com sugestões dos próprios usuários.

Esta edição do *BioEstat* 4.0 consolida a relação de dependência com o seu público fiel, onde as trocas se igualam à vontade de evoluir, com ganhos de conhecimentos e confiabilidade, diante da necessidade de aprender a ensinar ensinando. Desperta interesse maior e permite a sua inserção, como instrumento de consultas, por estudiosos de outras áreas de conhecimento técnico-científico, que se auxiliam do *BioEstat* nas avaliações de seus estudos experimentais e observacionais.

Do mesmo modo como a Estatística surgiu e ganhou reconhecimento, o projeto *BioEstat* vem evoluindo a cada lançamento. Esta evolução só se tornou possível pelo forte empenho e competência de seus autores, os quais tem se dedicado ao longo de 20 anos no aperfeiçoamento e na produção desta obra.

Ana Rita Pereira Alves

Diretora

AGRADECIMENTOS

Os autores penhoradamente agradecem o desvelo da Professora JUSSARA MORETTO MARTINELLI, da Universidade Federal do Pará, na revisão minuciosa e com primor do manuscrito do *BioEstat* 5.0. O interesse inextinguível da ilustre Professora na execução dessa árdua tarefa, trouxe grande satisfação, incentivo e alegria aos participantes do trabalho ora publicado.

Ao LUCAS LIMA AYRES, aluno de Medicina Veterinária da Universidade Federal de Minas Gerais, outro componente da família Ayres engajado nesta tarefa, agradecemos seu incentivo na introdução dos índices de Shannon-Wiener e de Simpson no *BioEstat* 5.0 e, sobretudo, no desenvolvimento das Estatísticas Circulares, ainda de menor utilização em nosso meio acadêmico.

NOTA: Manuel Ayres, Manuel Ayres Júnior, Daniel Lima Ayres e Alex Assis Santos dos Santos têm a satisfação em lançar a mais nova edição do *BioEstat* (5.0), com 210 aplicativos voltados sobretudo para as áreas de ciências biológicas e médicas, resultantes do aperfeiçoamento das versões anteriores e de sugestões recebidas dos Usuários deste *software*, esclarecendo, todavia, que possíveis incorreções no programa e no texto são de inteira responsabilidade dos seus autores.

Agradecemos, também, a todas as pessoas que colaboraram direta ou indiretamente na construção do *BioEstat*, tornando público o nosso reconhecimento àquelas indicadas em seguimento e, ainda, aos docentes e discentes das Universidades brasileiras que muito contribuíram com suas sugestões no aperfeiçoamento deste trabalho:

Ademar Ayres do Amaral	Jeffrey Shaw
Alberto Cargnelutti Filho	José Márcio Corrêa Ayres *
Alex Fiúza de Melo	José Maria Cardoso da Silva
Altino Pinheiro	José Maria de Azevedo Barbosa *
Ana Rita Pereira Alves	José Maria de Sousa
Anís Rassi Júnior	José Sílvia Govone
Antonio Carlos Simões Pião	Jussara Moretto Martinelli
Arno Holf Hamel	Lucimar Macedo Rodrigues
Bernardo Beiguelman	Lúcio Flávio Pinto
Cecília Banhara Marigo	Luiz Claudio Marigo
Cristovam Diniz	Marcos Porpino de Oliveira
Deborah Magalhães Lima	Maria Aparecida Pion Abuabara
Edson Elias Andrade Barbary	Miguel Petrere Jr.
Elon Lages Lima	Oswaldo Frota-Pessoa
Fábio Pereira Alves	Paulo A. Otto
Fernanda Sagica	Peter Toledo
Francisco Mauro Salzano	Renata Gomes G. Oliveira
Giorgini Augusto Venturieri	Roberto B. Cavalcanti
Gustavo A. B. da Fonseca	Russel A. Mittermeier
Helder Lima de Queiroz	Sérgio Cacela Alves
Horácio Schneider	Sídia M. Callegari Jacques
Ima Vieira	Suely Fraiha
Israel Roisenberg	Victoria J. Isaac
Iza do Amaral Corrêa Ayres	William D. Hamilton *
J. C. Voltolini	

* : *In memorian*

Os autores estendem seus agradecimentos às seguintes Instituições:

Conselho Nacional de Pesquisa - CNPq
Conservation Internacional - CI
Instituto de Desenvolvimento Sustentável Mamirauá
Imprensa Oficial do Estado do Pará - IOEPA
Museu Paraense “Emílio Goeldi” - MPEG
Ministério da Ciência e Tecnologia – MCT
Universidade Federal do Pará - UFPA
Wildlife Conservation Society - WCS

SUMÁRIO

CAPÍTULO 1 Usando o BioEstat	1
1.1 Requerimentos de sistema, 1	1.7 Imprimindo, 6
1.2 Instalando o BioEstat, 1	1.8 Escolhendo o teste, 6
1.3 Expressão de casas decimais, 1	1.9 Executando o teste, 6
1.4 Entrando com os dados, 3	1.10 Configurando o BioEstat, 7
1.5 Salvando os dados, 4	1.10 Obtendo ajuda, 7
1.6 Abrindo arquivos, 5	1.12 Capacidade do geral do sistema, 7
CAPÍTULO 2 Amostragem	9
2.1 Introdução, 9	2.5 Amostra por Conglomerados, 11
2.2 Amostra Aleatória Simples, 10	2.6 Amostragem dos Escores do <i>Grid</i> , 13
2.3 Amostra Sistemática, 10	2.7 Aumento da amostra, 14
2.4 Amostra Estratificada, 11	
CAPÍTULO 3 Análise Multivariada	17
3.1 Introdução, 17	3.6 Distância Euclidiana, 30
3.2 Análise de Conglomerados, 17	3.7 Distância Penrose-Mahalanobis, 31
3.3 Análise Discriminante, 20	3.8 Teste de Bartlett, 33
3.4 Componentes Principais, 23	3.9 Teste de Hotelling, 33
3.5 Correlação Canônica, 25	3.10 Teste de Mantel, 34
CAPÍTULO 4 Análise de Sobrevida	39
4.1 Introdução, 39	4.6 Log-Rank Test, 47
4.2 Atuarial, 40	4.7 Gehan (Wilcoxon Generalizado), 51
4.3 Modelo de Cox-Mantel, 41	4.8 Mantel-Haenszel, 52
4.4 Cox: Modelo de Risco Proporcional, 42	4.9 Pirâmide Populacional, 53
4.5 Kaplan-Meier, 46	

CAPÍTULO 5 Análise de Variância **59**

- | | |
|--|---|
| 5.1 Introdução, 59
5.2 ANOVA: um critério, 59
5.3 ANOVA: dois critérios, 64
5.4 ANOVA: fatorial a x b, 65 | 5.5 ANOVA: fatorial a x b x c, 67
5.6 Teste de Friedman, 68
5.7 Teste de Kruskal-Wallis, 70
5.8 Teste Q de Cochran, 72 |
|--|---|

CAPÍTULO 6 Bootstrap – Tira de Bota **75**

- | | |
|---|--|
| 6.1 Introdução, 75
6.2 Dados Dicotômicos, 75
6.3 Dados Contínuos, 77
6.4 IC da Correlação Linear, 78
6.5 IC da Média, 79 | 6.6 IC da Proporção, 80
6.7 Teste t, 2 amostras independentes, 81
6.8 ANOVA, um critério, 82
6.9 Correlação Linear, 83
*IC: Intervalo de Confiança |
|---|--|

CAPÍTULO 7 Correlação **85**

- | | |
|--|--|
| 7.1 Introdução, 85
7.2 Correlação Linear de Pearson, 85
7.3 Matriz de Correlação, 86
7.4 Correlação Parcial, 88
7.5 Correlação Linear: $\rho \neq 0$, 89 | 7.6 Coeficiente de Contingência C, 90
7.7 Concordância de Kendall, 91
7.8 Correlação de Kendall, 92
7.9 Coeficiente Phi: r_ϕ , 93
7.9 Correlação de Spearman: r_s , 94 |
|--|--|

CAPÍTULO 8 Distribuição de Probabilidades **95**

- | | |
|---|--|
| 8.1 Introdução, 95
8.2 Distribuição Binomial, 95
8.3 Distribuição Hipergeométrica, 96
8.4 Distribuição de Poisson, 97
8.5 Distribuição Exponencial, 98
8.6 Distribuição Normal, 99
8.7 Crivo (<i>Screening Test</i>), 102
8.8 Curva ROC, 103 | 8.9 Ponto de Corte, 105
8.10 Regra de Bayes, 106
8.11 Sensibilidade/Especificidade, 107
8.12 Distribuição F , 107
8.13 Distribuição do Qui-Quadrado, 108
8.14 Distribuição t de Student, 109
8.15 Distribuição Z, 110 |
|---|--|

CAPÍTULO 9 Uma Amostra 111

- | | |
|---|---|
| 9.1 Introdução, 111 | 9.8 Teste G, 119 |
| 9.2 Teste t , dados amostrais, 111 | 9.9 Teste de Kolmogorov-Smirnov, 120 |
| 9.3 Teste t , resumo amostral, 113 | 9.10 Teste de Lilliefors, 121 |
| 9.4 Teste Z, dados amostrais, 114 | 9.11 Teste do Qui-Quadrado, iguais, 121 |
| 9.5 Teste Z, resumo amostral, 115 | 9.12 Teste do Qui-Quadrado, desiguais, 122 |
| 9.6 Teste Binomial, 116 | 9.13 Teste de Poisson, 123 |
| 9.7 Iterações (runs test), 117 | |

CAPÍTULO 10 Duas Amostras Independentes 125

- | | |
|--|---|
| 10.1 Introdução, 125 | 10.10 Teste U , de Mann-Whitney, 135 |
| 10.2 Teste t , dados amostrais, 126 | 10.11 Teste da Mediana, 137 |
| 10.3 Teste t , resumo amostral, 128 | 10.12 Teste do Qui-Quadrado, 138 |
| 10.4 Teste Z, dados amostrais, 129 | 10.13 Odds Ratio (OR), 139 |
| 10.5 Teste Z, resumo amostral, 130 | 10.14 Risco Relativo (RR), 140 |
| 10.6 Teste Binomial, 131 | 10.15 Taxa de Incidência, 143 |
| 10.7 Teste Exato de Fisher, 132 | 10.16 Coeficiente de Variação, 144 |
| 10.8 Teste G, 133 | 10.17 Teste de Poisson, 145 |
| 10.9 Teste de Kolmogorov-Smirnov, 134 | |

CAPÍTULO 11 Duas Amostras Relacionadas 147

- | | |
|--|------------------------------------|
| 11.1 Introdução, 147 | 11.6 Teste de McNemar, 152 |
| 11.2 Teste t , dados amostrais, 147 | 11.7 Teste dos Sinais 154 |
| 11.3 Teste t , resumo amostral, 149 | 11.8 Teste Kappa, 156 |
| 11.4 ANOVA, 150 | 11.9 Teste de Wilcoxon, 157 |
| 11.5 Correlação Intraclassa, 151
(dados contínuos) | |

CAPÍTULO 12 Estatística Descritiva 159

- | | |
|--|--|
| 12.1 Introdução, 159 | 12.3.1 Índice de Shannon-Wiener, 162 |
| 12.2 Variáveis quantitativas, 160 | 12.3.2 Índice de Simpson: k amostras, 163 |
| 12.3 Variáveis qualitativas, 162 | |

CAPÍTULO 13 Estimação de Parâmetros 165

- | | |
|--|--|
| 13.1 Introdução, 165 | 13.9 Fourier (densidade), 173 |
| 13.2 Da Média, 165 | 13.10 Impacto Populacional, PIN-ER-t, 174 |
| 13.3 Da Proporção, 167 | 13.11 Matriz de Crescimento Populacional, Modelo de Leslie, 175 |
| 13.4 Diferença entre Duas Médias, 167 | 13.12 Tamanho (simples, direto), 176 |
| 13.5 Diferença entre Duas Proporções, 169 | 13.13 Tamanho (simples, inverso), 177 |
| 13.6 Densidade e Tamanho, 170 | 13.14 Tamanho (múltipla, fechada), 177 |
| 13.7 Equilíbrio de Hardy-Weinberg, 171 | 13.15 Tamanho (múltipla, aberta), 179 |
| 13.8 Fenótipos e Genótipos, 172 | |

CAPÍTULO 14 Meta-Análise 181

- | | |
|---|---|
| 14.1 Introdução, 181 | 14.7 Qui-Quadrado (independência), 192 |
| 14.2 Teste de DerSimonian-Laird, 182 | 14.8 Regressão para k amostras, 194 |
| 14.3 Dados Contínuos, 185 | 14.9 Risco Relativo (k Tabelas), 195 |
| 14.4 Teste de Mantel-Haenszel, 187 | 14.10 Teste G (aderência), 197 |
| 14.5 P-Valor (<i>pw</i> combinado), 190 | 14.11 Várias Correlações, 198 |
| 14.6 Qui-Quadrado (aderência), 191 | 14.12 Várias Proporções, 199 |

CAPÍTULO 15 Normalidade 203

- | | |
|--|---|
| 15.1 Introdução, 203 | 15.5 Teste de Lilliefors, 205 |
| 15.2 Teste D'Agostino, 203 | 15.6 Teste de Shapiro-Wilk, 206 |
| 15.3 Teste D'Agostino-Pearson, 204 | 15.7 Valores Extremos (desvios), 207 |
| 15.4 Teste de Kolmogorov-Smirnov, 205 | 15.8 Valores Extremos (quartis), 209 |

CAPÍTULO 21 Gráficos e Tabela		259
21.1	Introdução, 259	21.11 Gráfico de Linha, 275
21.2	Gráfico de Área, 260	21.12 Gráfico Máximo e Mínimo, 276
21.3	Box-Plot (Mediana e quartis), 261	21.13 Pirâmide populacional, 277
21.4	Box-Plot (Média e desvios), 262	21.14 Gráfico Polar, 281
21.5	Caule e Folha (<i>Stem and Leaf</i>), 263	21.15 Gráfico de setor, 282
21.6	Gráfico de Colunas, 265	21.16 Manipulação de Gráficos, 283
21.7	Diagrama de Dispersão, 267	21.17 Tabelas de Freqüências, 285
21.8	Diagrama IC (95% e 99%), 268	21.18 Gráficos de Resultados, 287
21.9	Diagrama Pontual (<i>Dot plot</i>), 271	21.19 Configurações Avançadas, 287
21.10	Histograma, 272	
CAPÍTULO 22 Fórmulas Estatísticas		289
22.1	Introdução, 289	22.10 Estatística Descritiva, 307
22.2	Análise Multivariada, 289	22.11 Estimação de Parâmetros, 309
22.3	Análise de Sobrevivência, 292	22.12 Meta-Análise, 315
22.4	Análise de Variância, 294	22.13 Normalidade, 318
22.5	Correlação, 296	22.14 Qui-Quadrado, 319
22.6	Distribuição de Probabilidades, 297	22.15 Regressão, 321
22.7	Uma Amostra, 300	22.16 Tamanho das Amostras, 325
22.8	Duas Amostras Independentes, 301	22.17 Transformação de dados, 327
22.9	Duas Amostras Relacionadas, 305	22.18 Distribuição circular, 327
CAPÍTULO 23 Escolha do Teste Estatístico – Sugestões		331
23.1	Introdução, 331	23.5 Duas Amostras Relacionadas, 335
23.2	Número de Amostras, 332	23.6 k Amostras Independentes, 336
23.3	Uma Amostra, 332	23.7 k Amostras relacionadas, 337
23.4	Duas Amostras Independentes, 334	

Glossário	339
-----------	-----

Referências Bibliográficas	351
----------------------------	-----

Índice	359
--------	-----

CAPÍTULO 1

USANDO O *BIOESTAT*

1.1 – REQUERIMENTOS DE SISTEMA

O *BioEstat* 5.0 requer o sistema operacional Microsoft Windows XP ou posterior e um processador compatível com o padrão Intel Pentium com no mínimo 128MB de memória RAM.

1.2 – INSTALANDO O *BIOESTAT*

- Coloque o CD-ROM do *BioEstat* no *drive* apropriado;
- O programa de instalação deve inicializar após alguns instantes. Caso isso não ocorra, clique no botão Iniciar, em seguida escolha Executar e digite: D:\SETUP.EXE, onde “D” é a letra do *drive* apropriado;
- Após o programa de instalação inicializar, siga as instruções contidas na tela;
- Quando a instalação estiver terminada, você pode rodar o *BioEstat* clicando em Iniciar, Programas, *BioEstat* 5.0 e, por fim, ao lado, em *BioEstat* 5.0.

1.3 – EXPRESSÃO DE CASAS DECIMAIS

No *BioEstat*, para que seja mantida a compatibilidade com outros *softwares* e acompanhando a norma aceita internacionalmente nas calculadoras eletrônicas, todos os valores que possuam casas decimais devem ser grafados usando o ponto decimal. A vírgula, portanto, não é usada no *BioEstat* como indicador da separação entre a parte inteira e a decimal, conforme os exemplos abaixo:

VALOR DESEJADO	GRAFIA CORRETA	GRAFIA INCORRETA
Um e meio	1.5	1,5
Dois inteiros e 75 centésimos	2.75	2,75
Cinco mil e um centésimo	5000.01	5000,01

1.4 – OUTRAS INFORMAÇÕES

Para usar o *BioEstat* é preciso entender alguns termos básicos do programa e do padrão Microsoft Windows XP®:

Menu superior

Figura 1.1

Janela do *BioEstat*.

Barra de ferramentas

Grid geral de entrada de dados

Tela de resultados

Barra de status

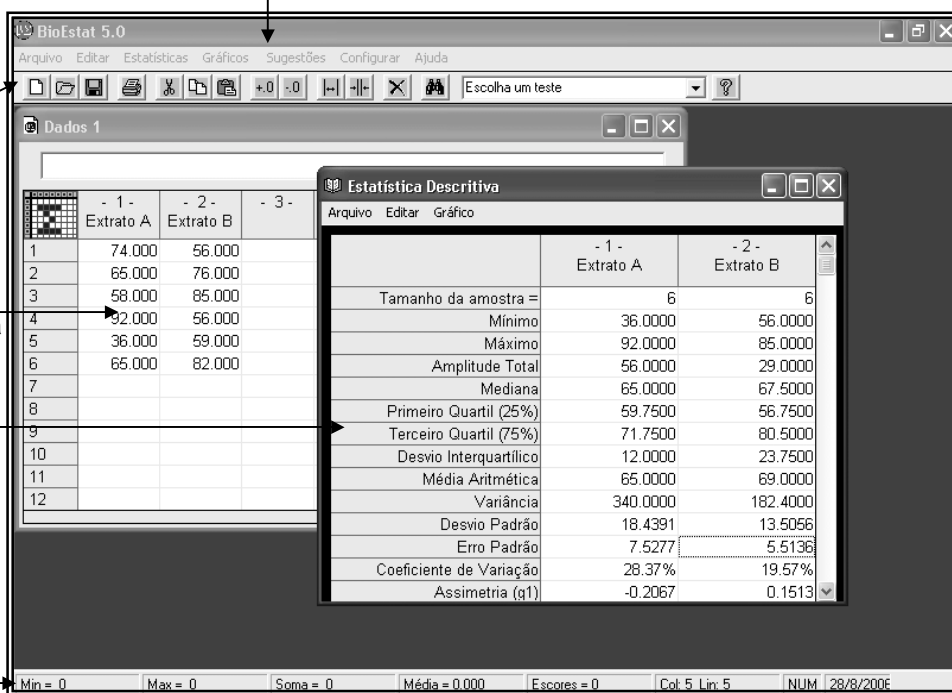
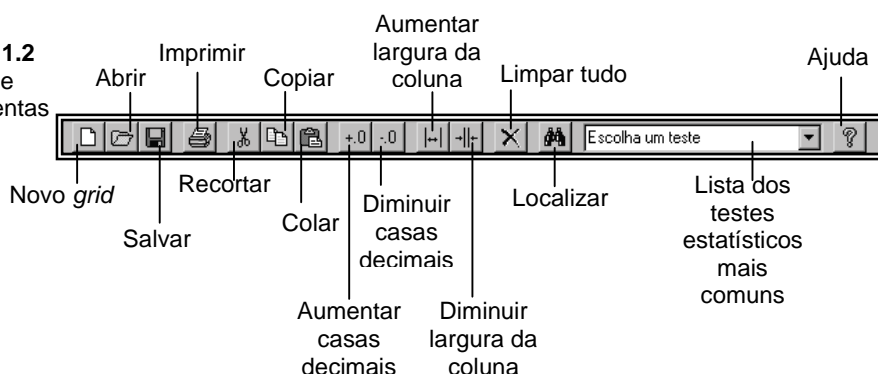


Figura 1.2

Barra de ferramentas



1.5 – ENTRANDO COM OS DADOS

Para entrar com os dados no *grid* geral (planilha), clique com o *mouse* na célula desejada e entre com o valor ou texto, se for o caso, passando para outra célula com as setas direcionais do teclado e inserindo o próximo dado, e assim por diante.

A denominação de uma coluna pode ser efetuada de uma das seguintes maneiras:

- Pressione a tecla F2, preencha o nome desejado na caixa de texto e pressione a tecla *Enter*;
- Clique em *Configurar* no menu superior e, a seguir, na opção *Nome da coluna*, preencha a caixa de texto com o nome desejado e pressione a tecla *Enter*.

Se há interesse em copiar os dados do *grid*, adote as seguintes condutas:

- Clique em uma das células já preenchidas e, com o botão pressionado, arraste o *mouse* sobre os dados que deseje copiar;
- Clique no botão *Copiar* da barra de ferramentas ou pressione *Editar* do menu superior, clicando em *Copiar* ou utilizando as teclas *Ctrl + C*.

Para recortar os dados do *grid* adote os seguintes procedimentos:

- Clique em uma das células já preenchidas e, com o botão pressionado, arraste o *mouse* sobre os dados que deseje recortar;
- Clique no botão *Recortar* da barra de ferramentas ou pressione *Editar* do menu superior, clicando em *Recortar* ou pressione as teclas *Ctrl + X*.

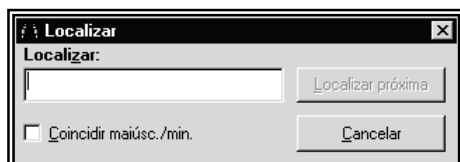
Quando for necessário colar os dados, adote as seguintes condutas:

- Copie ou recorte os dados que deseje colar;
- Clique na célula desejada e pressione o botão *Colar* na barra de ferramentas, ou em *Colar* do menu *Editar* ou, ainda, apertar *Ctrl + V*.

Para localizar um valor ou um texto do *grid* geral, adote os seguintes procedimentos:

- Clique no botão *Localizar* na barra de ferramentas ou na opção *Localizar* no menu *Editar* ou, ainda, pressione as teclas *Ctrl + L*, quando surgirá a seguinte tela:

Figura 1.3
Localizar



- Preencha a caixa de texto com a palavra ou o valor a ser localizado, pressione em *Localizar próxima*;
- Após fechar-se a janela *Localizar*, pode-se repetir a última busca utilizando a tecla F3 ou a opção *Localizar próxima* do menu *Editar*.

Caso queira, pode-se apagar todos os dados presentes no *grid* clicando no botão *Limpar tudo* da barra de ferramentas ou na opção *Limpar tudo* do menu *Editar*.

1.6 – SALVANDO OS DADOS

O arquivamento para posterior recuperação dos dados contidos no *BioEstat* pode ser efetuado no *grid* geral – a maioria – e, em alguns casos, em *grids* específicos.

1.6.1 – DO GRID GERAL (Matriz de entrada de dados)

Quando for necessário salvar os dados, siga as seguintes instruções:

- Clique na opção *Salvar* (tecla *Ctrl + B*) no menu *Arquivo* ou pressione o botão *Salvar* na barra de ferramentas;
- Caso seja a primeira vez que esteja salvando os dados, entre com o nome do arquivo desejado na caixa de texto e, em seguida, pressione o botão *Salvar*.

Caso deseje salvar um arquivo com outro nome, siga o seguinte procedimento:

- Clique na opção *Salvar Como* no menu *Arquivo*;
- Preencha a caixa de texto com o nome do arquivo desejado e, em seguida, pressione o botão *Salvar*.

É possível também exportar os dados para o formato *CSV* (delimitado por vírgula), que pode ser lido por diversos outros aplicativos, tal como o Microsoft Excel. Para exportar, siga as seguintes condutas:

- Clique na opção *Exportar* do menu *Arquivo*;
- Preencha a caixa de texto com o nome do arquivo desejado e, em seguida, pressione o botão *Salvar*.

1.6.2 – DE GRIDS ESPECÍFICOS

Em alguns testes o *BioEstat* apresenta *grids* próprios, a saber:

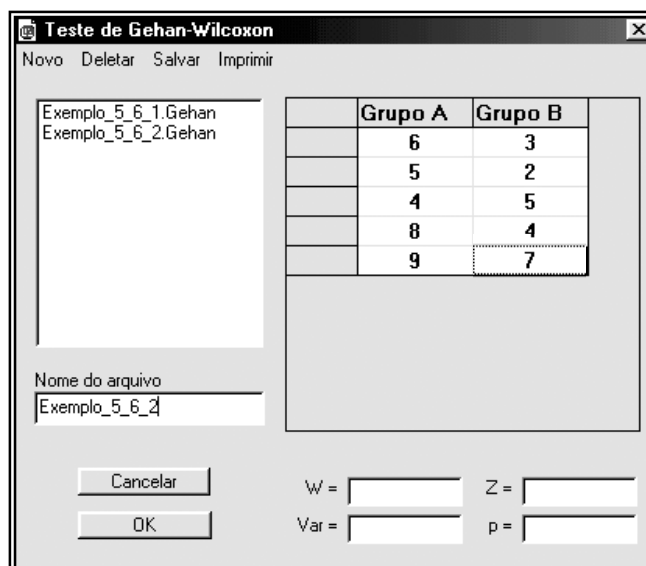
- Teste de Cox-Mantel;
- Teste de Gehan (Wilcoxon generalizado);
- ANOVA Fatorial: $a \times b$;
- ANOVA Fatorial: $a \times b \times c$;
- Correlação: Teste de homogeneidade para vários valores de r ;
- Qui-Quadrado: Homogeneidade (aderência);
- Teste G de Homogeneidade (aderência).

Clique no menu *Estatística* e escolha, conforme o caso, um dos testes acima indicados. Após introduzir os escores nas *caixas de texto*, como especificado na Figura 1.4, digite em *Nome do arquivo* a designação que servirá para identificar seus dados.

Depois desses procedimentos, o nome do arquivo aparecerá na lista dos itens salvos na parte superior esquerda do respectivo *grid* com a denominação do teste em questão. A Figura 1.4 mostra um exemplo do salvamento de um teste de Análise de Sobrevivência – teste de Gehan. Na lista daqueles que foram salvos aparecem dois exemplos: Exemplo_5_6_1.Gehan e Exemplo_5_6_2.Gehan.

Para recuperar os dados, basta clicar no nome do item contido na lista, pressionando o *mouse*, a seguir, em *Executar* para obter os resultados.

Figura 1.4
Salvando os dados de um *grid* específico.



Na parte superior do *grid* consta o seguinte: Novo, Deletar, Salvar e Imprimir, que podem ser utilizados pelo usuário, conforme o caso.

1.7 – ABRINDO ARQUIVOS

Quando for necessário abrir um arquivo, siga as seguintes instruções:

- Clique na opção *Abrir* (tecla *Ctrl + A*) no menu *Arquivo* ou pressione o botão *Abrir* na barra de ferramentas;
- Selecione o arquivo desejado e pressione o botão *Abrir*.

Para abrir um arquivo que foi recentemente utilizado, clique no menu *Arquivo* e selecione aquele desejado na parte inferior do menu.

O *BioEstat* contém os dados dos vários exemplos encontrados neste manual; para acessá-los, siga os seguintes passos:

- Clique na opção *Abrir* (tecla *Ctrl + A*) no menu *Arquivo* ou pressione o botão *Abrir* na barra de ferramentas;
- Dê um duplo clique na pasta *Exemplos*, que se encontra na lista de arquivos;
- Selecione um dos vários exemplos disponíveis e pressione o botão *Abrir*.

O programa também permite importar dados do formato *CSV* (delimitado por vírgula) e, para isso, siga as seguintes condutas:

- Clique na opção *Importar* do menu *Arquivo*;
- Selecione o arquivo desejado e pressione o botão *Abrir*.

1.8 – IMPRIMINDO

Para imprimir dados, siga os seguintes passos:

- Clique na opção *Imprimir* (*Ctrl + P*) do menu *Arquivo* ou no botão *Imprimir* da barra de ferramentas;
- Escolha se deseja imprimir o título ou não, o número de cópias e a orientação da página, sendo que a opção *Automática* utiliza a orientação mais adequada;
- Caso seja necessário, configure a impressão apertando no botão *Configurações* ou através da opção *Configurar impressão* do menu *Arquivo*.
- Quando esses procedimentos estiverem definidos, pressione o botão *Imprimir*.

1.9 – ESCOLHENDO O TESTE

Quando o usuário desejar escolher um teste estatístico que seja adequado para os seus dados, siga, a título de sugestão, os seguintes procedimentos:

- Pressione *Sugestões* no menu superior;
- Clique na célula correspondente ao número de amostras do seu trabalho: uma amostra, duas amostras independentes, duas amostras pareadas, k amostras independentes ou k amostras relacionadas;
- Clique na célula relacionada com o número de variáveis da(s) amostra(s): uma, duas ou k variáveis;
- Clique na célula relativa ao tipo de dados amostrais: categóricos, ordenativos ou numéricos, surgindo, nesse momento, na parte inferior da tela, exemplos desses modelos de dados e, na coluna seguinte, os testes sugeridos pelo programa;
- Ao passar o cursor sobre as sugestões do *BioEstat*, surgirão, na parte inferior da tela, um resumo e a fórmula estatística do teste correspondente. Se julgar adequado, pressione o teste desejado.

A qualquer momento é possível retornar à coluna anterior pressionando o botão correspondente: *Dados*, *Variáveis* ou *Amostras*, conforme o caso ou, ainda, em *Sair* para retornar ao *grid* geral.

1.10 – EXECUTANDO O TESTE

O *BioEstat 5.0* dispõe de 210 métodos estatísticos, entre procedimentos de amostragem, descritivos e inferenciais; para executá-los, adote as seguintes condutas:

- Abra um dos arquivos de Exemplos do *BioEstat* ou digite seus próprios dados no *grid* principal;
- Clique no menu *Estatísticas* e escolha um teste dentre as diversas opções;
- Na janela *Seleção de Amostras*:
 - Selecione as colunas a serem utilizadas no teste, clicando-as. Para selecionar todas as colunas disponíveis para o teste, aperte o botão “>>”;
 - Caso queira remover alguma coluna que havia sido selecionada, dê um duplo clique nela mesma e, para remover todas as colunas selecionadas, pressione o botão “<<”;
- Quando as colunas desejadas estiverem selecionadas, clique no botão *Executar Estatística*.

Para tornar a executar o último teste, clique na primeira opção do menu *Estatísticas* ou pressione as teclas *Ctrl + U*.

Você também pode acessar mais facilmente os testes estatísticos mais comuns utilizando a lista dos mesmos na barra de ferramentas.

1.11 – CONFIGURANDO O BIOESTAT

O menu *Configurar* apresenta uma série de opções de configuração:

- a) *Aumentar Casas Decimais*: aumenta o número de casas decimais visíveis da coluna atual, podendo também ser acessada através de seu botão na barra de ferramentas;
- b) *Diminuir Casas Decimais*: diminui o número de casas decimais visíveis da coluna atual. Esta opção pode também ser acessada através de seu botão na barra de ferramentas;
- c) *Aumentar Largura*: aumenta a largura da coluna atual. Também pode ser acessada através do botão do mesmo nome na barra de ferramentas;
- d) *Diminuir Largura*: diminui a largura da coluna atual. Ainda pode ser acessada através do botão similar da barra de ferramentas;
- e) *Nome da coluna (F2)*: denomina uma coluna;
- f) *Barra de Ferramentas*: habilita ou desabilita a barra de ferramentas;
- g) *Barra de Status*: esconde ou torna visível a barra de *status*;
- h) *Aumentar o número de colunas*: basta caminhar com o cursor para à direita, utilizando a seta direcional do teclado.

1.12 – OBTENDO AJUDA

Para obter ajuda enquanto estiver no *BioEstat*, aperte o botão *Ajuda* na barra de ferramentas ou clique na opção *Ajuda do BioEstat* (tecla F1) no menu *Ajuda*.

1.13 – CAPACIDADE GERAL DO SISTEMA

A Capacidade de armazenamento e processamento de dados do *BioEstat* é definida de acordo com os aspectos referidos em seguimento.

1.13.1 – TAMANHO MÁXIMO DA MATRIZ DE DADOS

O *BioEstat* suporta, no máximo, 35 mil células distribuídas entre linhas e colunas.

1.13.2 – ANÁLISE DE CONGLOMERADOS

Capacidade máxima de 50 variáveis, com até 500 observações cada.

1.13.3 – TESTE EXATO DE FISHER

A soma de todas as células pode atingir 600 escores.

CAPÍTULO 2

AMOSTRAGEM

2.1 – INTRODUÇÃO

Amostragens ou sondagens são procedimentos de seleção de subconjuntos de populações – *as amostras* –, destinados ao conhecimento de uma ou várias características dos universos investigados. Quando se estuda uma população por sondagens, é necessário selecionar suas amostras representativas, a fim de que os resultados obtidos possam ser inferidos para o conjunto objeto da investigação.

As amostragens dividem-se em dois grupos:

- a) *Amostragens probabilísticas*:
 - i – amostras aleatórias simples;
 - ii – amostras sistemáticas;
 - iii – amostras estratificadas;
 - iv – amostras por conglomerados;
 - v – amostras aleatórias: dados do *grid*;
 - vi – aumento do tamanho da amostra;
 - vii – reamostragem “Bootstrap”.
- b) *Amostras não-probabilísticas*:
 - i – intencional;
 - ii – por conveniência;
 - iii – por cotas.

No *BioEstat* serão abordadas as amostragens probabilísticas, nas quais cada indivíduo tem uma chance conhecida, diferente de zero, de ser incluído na amostra e o método de seleção deve ser compatível com a probabilidade de cada unidade.

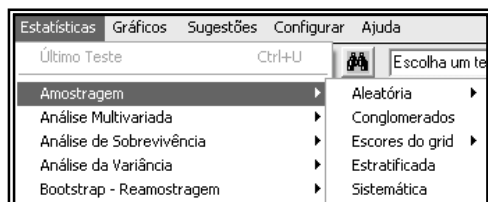
Algumas etapas devem ser observadas na seleção amostral:

- a) Definição da população-alvo;
- b) Alistamento de todos os indivíduos desse universo: 1, 2, 3, ..., n ;
- c) Determinação do tamanho da amostra;
- d) Escolha do modelo de amostragem;
- e) Seleção das unidades da população.

2.1.1 – INSTRUÇÕES

- a) Dirigir o *mouse* para o *menu* superior e clicar em *Estatística*;
- b) Apontar o *mouse* em *Amostragem*, quando surgirá a figura abaixo, comum para todos os modelos de seleção amostral do *BioEstat*:

Figura 2.1
Menu de
Amostragem.



2.2 – AMOSTRAS ALEATÓRIAS SIMPLES

São aquelas cujos indivíduos foram obtidos ao acaso da população, tendo cada amostra a mesma chance de ser escolhida dentre todas aquelas possíveis do mesmo tamanho. Os termos aleatório, randômico, casual e ao acaso são sinônimos. As amostras aleatórias simples podem ser obtidas por sorteio, em tabelas de números aleatórios e por computação.

2.2.1 – EXEMPLO

Para determinar a estatura média de uma população de 350 alunos da segunda série do Ensino Fundamental, tomou-se amostra aleatória de 40 estudantes, precedida da listagem numérica de todos os discentes desse universo.

2.2.2 – INSTRUÇÕES

- Apontar o *mouse* em *Aleatória*, para escolha de um dos modelos: *Com* ou *Sem reposição*.
- Clicar o *mouse* em *Sem reposição*, modelo escolhido neste exemplo; preencher as *caixas de texto* com os tamanhos da população e da respectiva amostra, pressionando, a seguir, o botão *Executar*.

Figura 2.2
Amostragem
Aleatória
(Sem
Reposição).

Valores aleatórios obtidos (sem reposição):									
3	9	23	26	33	35	37	40	77	82
87	118	127	142	144	153	159	167	170	177
189	191	199	205	210	219	228	242	244	245
252	254	263	283	286	312	319	323	334	340

2.3 – AMOSTRAS SISTEMÁTICAS

Nestas amostras o primeiro elemento é escolhido ao acaso e os demais, a partir dessa unidade, são selecionados a *intervalos sistemáticos* (k) previamente definidos até ser atingido o número total de indivíduos. Os intervalos são calculados pela razão:

$$k = \text{tamanho da população } (N) / \text{tamanho da amostra } (n).$$

A primeira unidade amostral pode variar de um (1) até o limite máximo do intervalo sistemático (k).

2.3.1 – EXEMPLO

Do mesmo alunado do exemplo anterior, retirar amostra sistemática de tamanho 50.

2.3.2 – INSTRUÇÕES

- Clicar o *mouse* em *Sistemática*;
- Preencher as *caixas de texto* com os tamanhos da população e da respectiva amostra, pressionando, a seguir, o botão *Executar*.

Figura 2.3
Amostragem
sistemática.

Amostragem: Sistemática

Imprimir

Entrada de dados

Tamanho da população: 350 Tamanho da amostra: 50

Executar Cancelar

Resultados

Intervalo sistemático = 7
Primeiro número sorteado = 4
Amostra sistemática:

4	11	18	25	32	39	46	53	60	67
74	81	88	95	102	109	116	123	130	137
144	151	158	165	172	179	186	193	200	207
214	221	228	235	242	249	256	263	270	277
284	291	298	305	312	319	326	333	340	347

2.4 – AMOSTRAS ESTRATIFICADAS

Trata-se de amostras obtidas de subconjuntos de uma população designados por *estratos* ou *grupos*, sem superposição de unidades entre eles, retirando-se amostra randômica de cada um. Há maior homogeneidade dos elementos dentro de cada grupo e maior diversidade entre os estratos, possibilitando maximizar as informações sobre todo o universo investigado. Deve-se definir cuidadosamente a quantidade de estratos, os limites de cada um e o número de variáveis a investigar.

2.4.1 – EXEMPLO

A população de uma cidade de 600 habitantes foi dividida em dois estratos: *urbano* e *rural*, com a finalidade de se estudar a incidência de malária em cada grupo de seus residentes. O estrato urbano possui 400 habitantes e o rural, 200. O tamanho total da amostra deve compreender 60 pessoas. Deve-se notar que o tamanho da amostra retirada de cada subconjunto é proporcional ao tamanho de cada estrato em relação ao tamanho da população (ver Figura 2.4).

2.4.2 – INSTRUÇÕES

- Clicar o mouse em *Estratificada*;
- Preencher as *caixas de texto* com o número de estratos, o tamanho total da amostra e o tamanho da população em cada estrato, pressionando, a seguir, o botão *Executar*.

2.5 – AMOSTRAS POR CONGLOMERADOS

Caracterizam-se pelo fato de que os elementos simples destas amostras são obtidos de unidades coletivas denominadas conglomerados ou *clusters*. Os conglomerados apresentam acentuada similitude no que diz respeito à variação de seus indivíduos, como são exemplos as colméias, os blocos residenciais, as aldeias de um mesmo grupo indígena, etc. Por esse procedimento é possível a listagem de todos os *clusters*, a escolha randômica de alguns, a relação completa das unidades simples daqueles já sorteados e, finalmente, a obtenção casual dos indivíduos dos conglomerados

selecionados. A retirada da amostra pode ser efetuada em dois ou mais estágios: duplo, triplo, múltiplo, conforme o caso.

Figura 2.4
Amostragem Estratificada.

Amostragem: Estratificada

Imprimir

Entrada de dados

Número de estratos:

Tamanho total da amostra:

População	
Estrato 1	400
Estrato 2	200

Executar

Cancelar

Resultados

Estrato: 1 Tamanho da amostra = 40

8	15	22	44	49	58	66	110	115	129
156	161	168	177	186	191	194	205	206	207
224	225	226	228	231	238	255	265	271	274
275	291	319	321	322	355	367	377	382	397

Estrato: 2 Tamanho da amostra = 20

4	8	12	15	29	37	39	60	66	79
87	93	103	106	108	128	141	145	167	168

2.5.1 – EXEMPLO

Calcular o peso médio de estudantes da quinta série do ensino fundamental das escolas públicas de um município. O total de estabelecimentos de ensino é de 52 escolas (conglomerados do 1º estágio), cada uma com 10 turmas da quinta série (conglomerados do 2º estágio), e cada turma apresentando 40 alunos matriculados (unidades simples, 3º estágio). Foram selecionadas 5, 4 e 8 unidades dos estágios 1º, 2º e 3º, respectivamente, constituindo, ao final, amostra de 160 discentes (5 x 4 x 8).

2.5.2 – INSTRUÇÕES

- Clicar o *mouse* em *Conglomerados*;
- Preencher a *caixa de texto* com o número de estágios (3), com a denominação de cada um, o número de unidades contidas nos respectivos *clusters* e, por fim, a quantidade de unidades determinadas para sorteio em cada conglomerado. Pressionar, em seguida, o botão *Executar*.

Figura 2.5
Amostragem por Conglomerados.
(parte dos resultados)

Amostragem: Conglomerados

Imprimir

Número de estágios:

Estágio	Denominação do estágio	No. de unidades do estágio	No. de unidades para sorteio
1o.	Escolas do 1o grau	52	5
2o.	Turmas da 5a série	10	4
3o.	Alunos por turma	40	8

Executar

Cancelar

Escolas do 1o grau...: 51
Turmas da 5a série...: 8
Alunos por turma (amostra)...: 17 5 35 32 16 26 8 15

Escolas do 1o grau...: 51
Turmas da 5a série...: 10
Alunos por turma (amostra)...: 30 31 27 25 32 14 1 6

2.6 – AMOSTRA ALEATÓRIA DOS ESCORES DO GRID – SEM REPOSIÇÃO

Pode-se retirar amostras aleatórias de escores inseridos no *grid* geral do *BioEstat*.

2.6.1 – EXEMPLO

Estão inseridos no *grid* 140 escores (Figura 2.6), referentes à estatura de estudantes do sexo masculino de idades compreendidas entre 16 e 18 anos, de uma escola de ensino fundamental. Tendo em vista que os dados compreendem todos os alunos matriculados nessa instituição, com as características referidas, esses escores constituem, portanto, uma população, com estatura média igual a 169.45 cm. Retirou-se amostra aleatória desse conjunto para calcular a média amostral da variável *estatura* no sentido de verificar se ela é representativa do universo em questão.

Figura 2.6
Dados
parciais do
Exemplo
2.6.1.

	- 1 -	- 2 -
	Estatura: N	
1	159	
2	159	
3	159	
4	160	
5	160	

2.6.2 – INSTRUÇÕES

- Introduzir os dados no *grid* padrão do *BioEstat*, conforme a Figura 2.6;
- Pressionar o *mouse* em *Aleatória dos escores do grid*;
- Selecionar a coluna, neste exemplo é a de número 1;
- Pressionar em *Executar*;

Informar na *caixa de texto* o tamanho da amostra desejada (10, neste caso) e Pressionar em *OK*, para resultados (Figuras 2.7 e 2.8).

Figura 2.7
Tamanho da
Amostra.

Tamanho da Amostra?

OK

Cancel

10

Figura 2.8
A Coluna
"Amostra"
indica os
escores
selecionados
de maneira
aleatória.

	- 1 -	- 2 -	-
	Estatura: N	Amostra	
1	159	160	
2	159	163	
3	159	168	
4	160	171	
5	160	174	
6	160	178	
7	161	185	
8	161	187	
9	162	188	
10	162	190	
11	162		
12	163		

Efetuada-se a estatística descritiva dos escores amostrais, com o cálculo da média e do erro padrão, e comparando-os com a média paramétrica, obtêm-se:

Dados da amostra: $\bar{X} = 176.4$; $s_{\bar{X}} = 3.4$;

Dado da população: $\mu = 178.5$;

Viés = $(\mu - \bar{X}) = 178.5 - 176.4 = 2.1$

Pode-se concluir que a média amostral ($\bar{X} = 176.4$) é um estimador não enviesado do parâmetro ($\mu = 178.5$), eis que o viés (2.1) é inferior a um erro padrão, representado por $s_{\bar{X}} = 3.4$.

A média amostral é, assim, representativa do universo em questão.

2.7 – AUMENTO DO TAMANHO DA AMOSTRA

É um método computacional de geração de dados a partir de valores observados, com o objetivo de gerar novas e maiores amostras, cujo tamanho será decidido pelo pesquisador. É possível, assim, aumentar o poder do teste, sobretudo quando as amostras iniciais são pequenas por indisponibilidade de indivíduos.

2.7.1 – EXEMPLO

A partir de dados hipotéticos foram obtidas duas amostras independentes de tamanho $n = 4$, cujos escores medidos em escala ordinal expressam resultados de análise histológica, sendo identificados conforme o grau da lesão:

- Escore 1: leve;
- Escore 2: moderada;
- Escore 3: alta.

A figura abaixo contém os escores observados no experimento. Quando o teste de *Mann-Whitney* foi aplicado o *p*-valor não foi significativo ($p = 0.1266$).

Figura 2.9
Dados do
Exemplo
2.7.1

	- 1 - Grupo A	- 2 - Grupo B	- 3 -
1	1.000	2.000	
2	1.000	2.000	
3	2.000	3.000	

2.7.2 – INSTRUÇÕES

- a) Inserir no *grid* os dados obtidos;
- b) Selecionar no menu *Amostragem: Escores do Grid e Aumento da Amostra*;
- c) Na janela *Seleção de Amostras*, escolha apenas uma coluna de cada vez;
- d) Clique em *Executar Estatística*;
- e) Informe o tamanho da amostra a ser gerada. Neste exemplo adotamos 40 como o tamanho da nova amostra;
- f) O *BioEstat 5.0* informará em qual coluna será inserida a reamostragem;
- g) Repita o procedimento para a segunda amostra;
- h) Execute o teste em questão com as novas amostras obtidas.

Figura 2.10
Exemplo
2.7.1,
já com a
reamostragem
(visão parcial).

	- 1 - Grupo A	- 2 - Grupo B	Amostra: Col 1	Amostra: Col 2
1	1.000	2.000	2.000	3.000
2	1.000	2.000	2.000	3.000
3	2.000	3.000	1.000	2.000
4			2.000	2.000
5			2.000	3.000
6			1.000	3.000
7			2.000	2.000
8			1.000	3.000
9			2.000	2.000
10			2.000	2.000

Neste exemplo, com amostras geradas de tamanho = 40, foi aplicado o teste de *Mann-Whitney* o qual mostrou diferença muito significativa, com *p-valor* <0.0001. Este procedimento poderá ser aplicado para *k* amostras, como na Análise de Variância.

Figura 2.11
Resultado do
teste de
Mann-Whitney
para a
reamostragem

Mann-Whitney: amostras ...	
Arquivo Editar Gráfico	
	Resultado
n1 =	40
n2 =	40
R1 =	968.50
R2 =	2271.50
U =	148.50
Z(U) =	6.2691
(p) =	< 0.0001

CAPÍTULO 3

ANÁLISE MULTIVARIADA

3.1 – INTRODUÇÃO

A maioria dos testes estatísticos é concernente ao exame de uma única variável, a qual é considerada importante na descrição ou no estudo analítico dos dados amostrais. Em algumas situações, porém, torna-se importante considerar a análise de diversas variáveis aleatórias de modo simultâneo, para atingir outros aspectos não abrangidos por um exame mais simplificado. Embora se deva a Francis Galton (1886) as primeiras análises multivariadas, somente com o desenvolvimento dos computadores foi possível construir pacotes estatísticos com diferentes modelos de análises multivariadas, possibilitando seu uso em escala muito mais ampla. Do ponto de vista matemático, além dos conhecimentos básicos de álgebra, há necessidade de alguma noção sobre *matrizes algébricas*, para a efetivação da análise multivariada.

A atual versão do *BioEstat* (5.0) dispõe dos seguintes testes:

- Análise de Conglomerados;
- Análise Discriminante;
- Componentes Principais;
- Correlação Canônica;
- Distância Multivariada (Euclidiana);
- Distância Multivariada (Penrose e Mahalanobis);
- Teste de Bartlett (duas amostras);
- Teste de Hotelling (uma amostra);
- Teste de Mantel.

3.1.1 – INSTRUÇÕES

- Introduzir os dados no *grid* padrão ou específico, conforme o teste;
- Dirigir o *mouse* para o *menu* superior e clicar em *Estatística*;
- Apontar o *mouse* em *Análise Multivariada*, conforme demonstra a Figura 3.1:

Figura 3.1
Menu de
análise
multivariada.

Análise Multivariada	▶	Análise de Conglomerados
Análise de Sobrevivência	▶	Análise Discriminante
Análise da Variância	▶	Componentes Principais
Bootstrap - Reamostragem	▶	Correlação Canônica
Correlação	▶	Distância Multivariada (Euclidiana)
Distribuição de Probabilidades	▶	Distância Multivariada (Penrose/Mahalanobis)
Uma Amostra	▶	Teste de Bartlett
Duas Amostras Independentes	▶	Teste de Hotelling
Duas Amostras Relacionadas	▶	Teste de Mantel

3.2 – ANÁLISE DE CONGLOMERADOS

O objetivo desta análise é agrupar os indivíduos em conglomerados ou *clusters*. Os conglomerados são identificados como grupos que compõem a população ou a amostra que está sendo analisada. As unidades de um grupo apresentam muitas

semelhanças entre si e muitas diferenças em relação às unidades dos demais grupos. É, pois, uma verdadeira classificação ou taxonomia, muito usada por biólogos. Dessa forma é um método exploratório, não um teste estatístico, contendo uma série de algoritmos para classificação dos objetos segundo o grau de similaridade, na tentativa de estabelecer uma estrutura, mesmo sendo inexistente. A análise pode ser realizada segundo alguns critérios de procedimentos, os quais são selecionados conforme as características das variáveis a serem analisadas, sendo apresentados a seguir:

Método de Agregação

A medida da distância é a base da análise de conglomerados. Originalmente, cada objeto é único e, portanto, as distâncias entre os objetos são óbvias. Entretanto, à medida que os objetos são agrupados, deve-se decidir qual a forma de medir a distância entre um grupo (*cluster*) e um objeto, ou entre dois grupos. Cada método de agregação fornece um critério diferente para definir, no espaço multidimensional, o ponto dentro do grupo que servirá de referência para calcular as distâncias entre esse grupo e outros objetos ou grupos.

Tipo de Distância

A distância na análise de conglomerados é um parâmetro para medir o grau de similaridade entre dois objetos. Cada variável corresponde a uma dimensão, e cada objeto é representado por um ponto no espaço multidimensional resultante. Quanto menor a distância mais similar serão os objetos e, portanto, aumenta a probabilidade de pertencerem ao mesmo grupo. As fórmulas que caracterizam os tipos de distâncias que podem ser aplicadas à análise de conglomerados podem ser obtidas no Capítulo 22 e podem ser escolhidas durante o procedimento deste aplicativo.

Estandarização das variáveis

Os resultados da análise de conglomerados poderão ter algum viés, se os dados são medidos em escalas diferentes. Dessa forma, por exemplo, se uma das variáveis for a altura do indivíduo medida em metros e a outra for a massa corporal medida em quilograma, aqueles terão um peso menor quando se compara a mesma informação medida em centímetros. A maneira de tornar as variáveis “comparáveis” será realizar a transformação z , em outras palavras, estandardizar todas essas variáveis.

3.2.1 – EXEMPLO

Procurou-se verificar similitudes entre a forma escrita de diversos idiomas. As comparações têm como base as duas primeiras letras da grafia dos números de 1 a 10 em cada idioma. Para cada letra do alfabeto foi atribuído um valor numérico. Para a letra A (primeira letra do alfabeto) foi atribuído o valor 1, para a letra B (segunda letra do alfabeto) foi atribuído 2 e assim sucessivamente até a letra Z que corresponde ao número 26 por ser a vigésima sexta letra (foi incluída a letra *w*).

Observe a Figura 3.2, na primeira linha da primeira coluna está localizado o grupo que representa o idioma Português. Na Figura 3.3, somente as duas primeiras letras de

cada *número* e de cada *idioma* foram substituídas pelos respectivos valores numéricos das letras. Em Português o número **um** recebeu o escore 2113, pois a letra inicial da palavra **um** (u) é a vigésima primeira (21^a) letra do alfabeto e a segunda letra (m) é a décima terceira (13^a), e assim por diante.

Figura 3.2
Números
escritos por
extenso.
Exemplo
3.2.1A.

- 1 - ** Idiomas	- 2 - ** Nº 1	- 3 - ** Nº 2	- 4 - ** Nº 3	- 5 - ** Nº 4	- 6 - ** Nº 5	- 7 - ** Nº 6	- 8 - ** Nº 7	- 9 - ** Nº 8	- 10 - ** Nº 9	- 11 - ** Nº 10
Português	um	dois	tres	quatro	cinco	seis	sete	oito	nove	dez
Inglês	one	two	three	four	five	six	seven	eight	nine	ten
Norueguês	en	to	tre	fire	fem	seks	sju	atte	ni	ti
Dinamarquês	en	to	tre	fire	fem	seks	syv	otte	ni	ti
Holandês	een	twee	drie	vijf	vier	zes	zeven	acht	negen	tien
Alemão	ein	zwei	drei	vier	funf	sechs	sieben	acht	neum	zehn
Francês	un	deux	trois	quatre	cinq	six	sept	huit	neuf	dix
Espanhol	uno	dos	tres	cuatro	cinco	sieis	siete	ocho	nuefe	diez
Húngaro	egy	ketto	harom	negy	ot	hat	het	nyolc	kilenc	tiz

Figura 3.3
Valores que
representam
as duas
primeiras
letras de
cada idioma.
Exemplo
3.2.1B.

- 1 - **	- 2 - - 1 -	- 3 - - 2 -	- 4 - - 3 -	- 5 - - 4 -	- 6 - - 5 -	- 7 - - 6 -	- 8 - - 7 -	- 9 - - 8 -	- 10 - - 9 -	- 11 - - 10 -
Por	2113	415	2018	1721	39	195	195	159	1415	45
Ing	1514	2023	208	615	69	199	195	59	149	205
Nor	514	2015	2018	69	65	195	1910	120	149	209
Din	514	2015	2018	69	65	195	1925	1520	149	209
Hol	55	2023	418	229	229	265	265	13	145	209
Ale	59	2623	418	229	621	195	199	13	145	265
Fra	2114	45	2015	1721	39	199	195	821	145	49
Esp	2114	415	2018	321	39	199	199	153	1421	49
Hun	57	115	81	145	1520	81	85	1425	119	209

Deseja-se efetuar a Análise de Conglomerados para que sejam identificados e agrupados os idiomas conforme as similitudes e as diferenças existentes entre eles.

3.2.2 – INSTRUÇÕES

- Clicar no *menu* Análise Multivariada e na opção Análise de Conglomerados;
- Selecionar as colunas referentes às variáveis do *grid* geral;
- Clicar em *Executar o Teste*, para obter a Figura 3.4:

Figura 3.4
Critérios
usados na
análise.

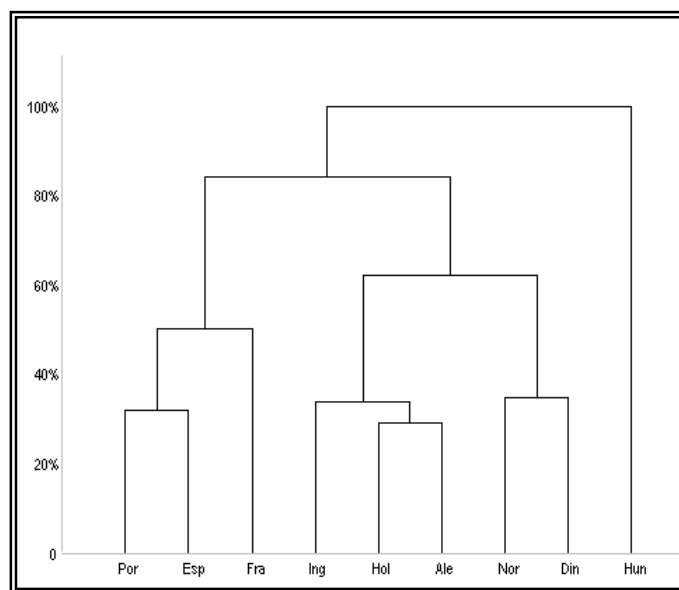
Método de Agregação <input type="radio"/> Ward (Variância Mínima) <input type="radio"/> Ligação Simples <input checked="" type="radio"/> Ligação Completa <input type="radio"/> Média do Grupo <input type="radio"/> McQuitty <input type="radio"/> Gower (Mediana) <input type="radio"/> Centróide	Tipo de Distância <input checked="" type="radio"/> Euclidiana <input type="radio"/> Quadrado da Dist. Euclidiana <input type="radio"/> Absoluta <input type="radio"/> Chebychev <input type="radio"/> Bray-Curtis Estandarização das Variáveis <input checked="" type="radio"/> Com <input type="radio"/> Sem
---	--

- d) Selecione as opções:
- *Ligação Completa*, como Método de Agregação;
 - *Euclidiana*, como Tipo de Distância;
 - *Com Estandarização das Variáveis*;
- e) Clicar em *Executar*, para obter o Dendograma da Figura 3.5.

O dendograma (Figura 3.5) demonstra:

- Similitude entre o idioma Português e o Espanhol e proximidade com o Francês;
- Similitude entre o Holandês e o Alemão e proximidade com o Inglês;
- Similitude entre o Norueguês e o Dinamarquês;
- O idioma Húngaro está separado dos demais.

Figura 3.5
Resultado
do Exemplo
3.2.1.



Nota: no *BioEstat* as distâncias são normalizadas em termos relativos à maior distância calculada.

3.3 – ANÁLISE DISCRIMINANTE

É uma técnica de Análise Multivariada que determina a separação de grupos de indivíduos de acordo com os valores de suas variáveis. Na Análise Discriminante pode-se obter:

- A determinação das funções discriminantes de *Fisher* (Y_1 e Y_2) capazes de separar ou discriminar os grupos de indivíduos. Y_1 e Y_2 são duas variáveis adimensionais, criadas de modo a permitir melhor separação entre os grupos que estão sendo analisados.
- O diagrama que identifica os grupos e permite a visualização das separações e aproximações grupais.
- Através das funções discriminantes de *Fisher*, torna-se possível a classificação de novos indivíduos, em um dos grupos determinados, sendo o cálculo semelhante ao utilizado na Regressão Linear Múltipla.

Destaque-se, por exemplo, que na área médica é comum em psiquiatria a classificação de pacientes dentre categorias grupais, tais como: psicose afetiva, depressão, esquizofrenia, bipolar e outras, com apoio no exame clínico e nos sintomas observados durante o exame inicial de rotina. Dados laboratoriais posteriores, o seguimento dos enfermos (*follow-up*) e novas informações familiares podem resultar na alocação em uma nova categoria ou grupo, diminuindo, tanto quanto possível, o número de classificações equivocadas. A Análise Discriminante constitui um instrumento importante nesse modelo de abordagem.

3.3.1 – EXEMPLO (dados hipotéticos)

Foram coletadas amostras de três espécies de *Iris*, medindo-se o comprimento e a largura das sépalas e respectivas pétalas. Os Grupos e as respectivas Variáveis foram introduzidas no *grid* do *BioEstat* na seguinte ordem:

- Grupo 1: Variáveis da *Iris sp.1*, contém 20 unidades;
- Grupo 2: Variáveis da *Iris sp.2*, contém 20 unidades;
- Grupo 3: Variáveis da *Iris sp.3*, contém 20 unidades.

Cada unidade de cada grupo está representada com as seguintes variáveis:

- X1 = comprimento das sépalas;
- X2 = largura das sépalas;
- X3 = comprimento das pétalas;
- X4 = largura das pétalas.

Figura 3.6
Visão parcial
dos dados
do Exemplo
3.3.1

	- 1 - Grupo	- 2 - X1	- 3 - X2	- 4 - X3	- 5 - X4
1	1.000	6.200	3.200	5.900	2.400
2	1.000	5.700	2.600	5.000	1.800
3	1.000	7.000	2.900	5.800	2.000
4	1.000	6.200	2.800	5.500	1.700

3.3.2 – INSTRUÇÕES

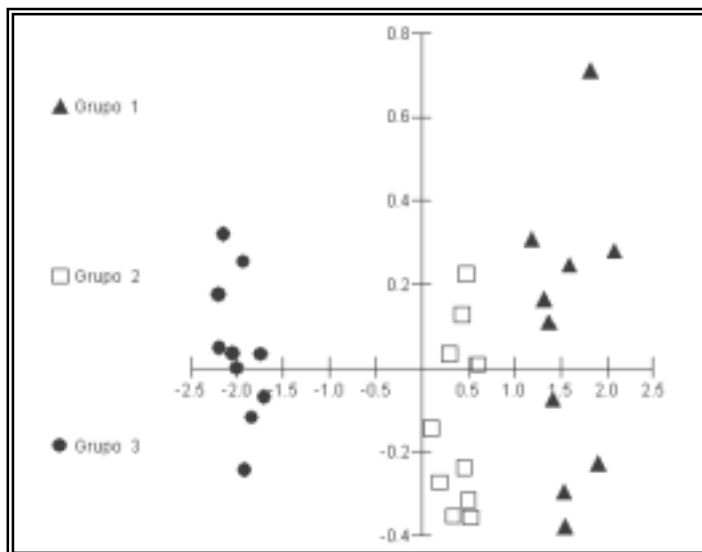
- Clicar no *menu* Análise Multivariada e selecionar em Análise Discriminante;
- Selecionar as colunas referentes às variáveis do *grid* geral;
- Clicar em *Executar o Teste*, para obter a Figura 3.7;
- Pressionar em *OK*, nessa Figura, surgindo as funções discriminantes de *Fisher* (Y1 e Y2), com os respectivos estimadores das variáveis X1, X2, X3 e X4.

Figura 3.7
As funções
discriminante
de *Fisher*.
Exemplo
3.3.1.

Tamanho da Amostra	Quantidade de Grupos	Número de Variáveis
60	3	4
<input type="button" value="Gráfico"/> <input type="button" value="OK"/>		
Funções Discriminantes de Fisher		
Y1 =	-0.3370 X1 -0.4105 X2 +0.7083 X3 +0.4651 X4	
Y2 =	-0.5049 X1 +0.7847 X2 +0.1460 X3 +0.3287 X4	

- e) Pressionar o botão *Gráfico*, na parte superior à esquerda da Figura 3.7, para obter o diagrama representado pela Figura 3.8.

Figura 3.8
Diagrama da função discriminante. Exemplo 3.3.1.



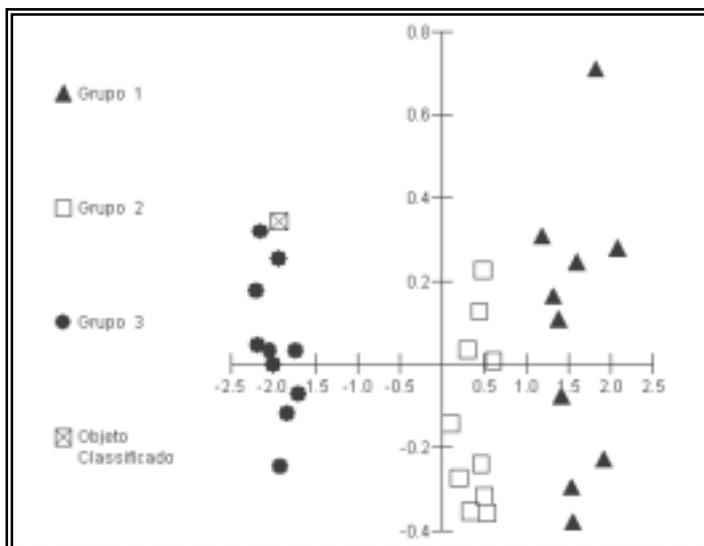
Observe os agrupamentos: o Grupo 2 está ao centro, entre os Grupos 1 e 3. Quando desejamos classificar uma nova unidade, devemos preencher o valor de cada uma de suas variáveis e pressionar em *Classificar*. Por exemplo, foram introduzidos os valores $X_1 = 4.7$, $X_2 = 3.5$, $X_3 = 1.3$ e $X_4 = 0.22$, (Figura 3.9), pressionando-se o botão *Classificar*, resulta em uma unidade do Grupo 3 (*Íris sp.3*).

Figura 3.9
Entrada de dados das variáveis para obter a classificação de um novo indivíduo. Exemplo 3.3.1.

Tamanho da Amostra	Quantidade de Grupos	Número de Variáveis		
60	3	4		
Gráfico		OK		
Funções Discriminantes de Fisher				
Y1 = -0.3370 X1 -0.4105 X2 + 0.7083 X3 + 0.4651 X4				
Y2 = -0.5049 X1 + 0.7847 X2 + 0.1460 X3 + 0.3287 X4				
Entrada de Dados				
X1 4.7	X2 3.5	X3 1.5	X4 0.22	Var 5
Var 6	Var 7	Var 8	Var 9	Var 10
Var 11	Var 12	Var 13	Var 14	Var 15
Classificar		0 indivíduo pertence ao grupo 3		

Para obter um novo diagrama, pressione o botão *Gráfico* (Figura 3.9), sendo gerado novo diagrama com destaque para a nova unidade (Figura 3.10).

Figura 3.9
A classificação
de um novo
indivíduo.
Exemplo
3.3.1.



Percebe-se que a nova unidade (Objeto Classificado) está localizada entre as unidades do Grupo 3, *Iris sp.3*, conforme resultados n Figura 3.9.

3.4 – 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 através de índices designados de Z_p , de tal maneira que a variância $var(Z_1) \geq$ variância $var(Z_2) \geq$ variância $var(Z_3) \geq \dots \geq$ variância $var(Z_p)$. 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.

3.4.1 – EXEMPLO

Foram examinadas 40 andorinhas, mensurando-se cinco (5) variáveis: X_1 , X_2 , X_3 , X_4 e X_5 , representando o comprimento total, comprimento das asas, comprimento do bico, perímetro cefálico e comprimento do esterno, respectivamente. Determinar os componentes principais dentre essas variáveis, cujos valores foram inseridos no *grid* geral como demonstrado a seguir:

Figura 3.11
Dados
parciais do
Exemplo
3.4.1.

	- 1 - X1	- 2 - X2	- 3 - X3	- 4 - X4	- 5 - X5
1	158.000	245.000	31.600	18.500	20.500
2	154.000	240.000	30.400	17.900	19.600
3	153.000	240.000	31.000	18.400	20.600
4	153.000	236.000	30.900	17.700	20.200
5	155.000	243.000	31.500	18.600	20.300
6	162.000	247.000	32.000	19.000	20.900
7	157.000	238.000	31.900	18.400	21.300
8	156.000	239.000	31.800	18.600	20.200
9	164.000	248.000	32.700	19.100	21.100
10	158.000	238.000	31.000	18.800	22.000
11	158.000	240.000	31.300	18.600	22.000
12	160.000	244.000	31.100	18.600	20.500

3.4.2 – INSTRUÇÕES

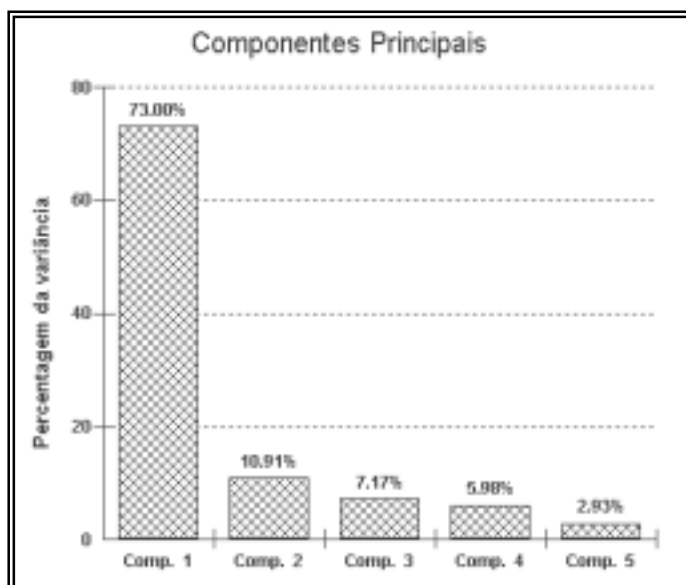
- f) Pressionar o *mouse* em Componente Principal;
- g) Selecionar as colunas referentes às variáveis do *grid* geral;
- h) Clicar em *Executar o Teste*, para obter os resultados (ver Figura 3.12).

Figura 3.12
Resultados do
Exemplo 3.4.1
Os resultados
demonstram
que os índices
1 e 2
representam
84% da
variação total
(73% + 11%)
dos indivíduos
testados,
constituindo
os Componentes
Principais nesse
Exemplo.

Arquivo Editar Gráfico					
Matriz de Correlação	Coluna 1	Coluna 2	Coluna 3	Coluna 4	Coluna 5
Coluna 1	1.0000	---	---	---	---
Coluna 2	0.7448	1.0000	---	---	---
Coluna 3	0.7078	0.6893	1.0000	---	---
Coluna 4	0.6511	0.7685	0.7773	1.0000	---
Coluna 5	0.6164	0.5230	0.5188	0.5984	1.0000
	Autovalores	% total da variância	Autovalor acumulado	% acumulada	---
Componente 1 =	3.6502	73.0032%	3.6502	73.0032%	---
Componente 2 =	0.5453	10.9062%	4.1955	83.9094%	---
Componente 3 =	0.3587	7.1750%	4.5542	91.0843%	---
Componente 4 =	0.2992	5.9842%	4.8534	97.0685%	---
Componente 5 =	0.1466	2.9315%	5.0000	100.0000%	---
	---	---	---	---	---
Coefs. Autovetores	Coef. X1	Coef. X2	Coef. X3	Coef. X4	Coef. X5
Componente 1 =	0.4569	0.4598	0.4555	0.4677	0.3918
Componente 2 =	0.0462	-0.2914	-0.3213	-0.1837	0.8809
Componente 3 =	-0.7025	-0.3029	0.3207	0.5337	0.1649
Componente 4 =	0.3278	-0.6230	0.6446	-0.2941	-0.0495
Componente 5 =	-0.4338	0.4731	0.4134	-0.6134	0.2022

- d) Pressionar em Gráfico, na barra da Figura 3.12, para visualizar a seguinte imagem:

Figura 3.13
Gráfico do
Exemplo
3.4.1.



3.5 – CORRELAÇÃO CANÔNICA

Trata-se de um modelo de Análise Multivariada que permite investigar o relacionamento entre duas séries de variáveis: um grupo de variáveis X e outro de variáveis Y . A associação entre os grupos de variáveis tem como resultante a correlação (R_c) – denominada *Canônica*.

Assim, na área médica, um pesquisador poderá relacionar um grupo de *sintomas* com um conjunto de *fatores de risco* em relação a determinada doença.

OBJETIVOS

- Determinar a magnitude da relação entre dois conjuntos de variáveis;
- Deduzir os pesos de cada grupo de variáveis de forma que a correlação de cada composição linear seja maximizada;
- Avaliar a contribuição relativa de cada variável na correlação canônica, no sentido de explicar a natureza do relacionamento;
- Para propósitos descritivos e inferenciais, devendo, neste caso, as variáveis apresentar distribuição normal.

LIMITAÇÕES

- A Correlação Canônica reflete a variação explicada pela composição linear das variáveis;
- Os valores canônicos derivados constituem amostra-dependente;
- As variáveis canônicas derivadas maximizam a Correlação e não a variância extraída;
- As interpretações estatísticas ainda não estão bem desenvolvidas.

3.5.1 – EXEMPLO 1

Suponha que em um estudo epidemiológico envolvendo enfermeiros que atuam em unidades de saúde de urgência e emergência, tem como objetivo determinar o nível de correspondência entre um conjunto de características das atividades desenvolvidas e a satisfação do profissional de saúde pública. As respostas a um questionário foram interpretadas e os dados introduzidos no *grid* geral conforme a Figura 3.14.

Há três variáveis relacionadas com as *características da atividade*:

- Varição: Nível de variação envolvida nas tarefas, medida em percentual;
- Feedback: É a resposta, em percentual, da satisfação das pessoas usuárias do serviço de saúde onde foi efetuada a pesquisa;
- Autonomia: Percentual de liberdade admitido na realização das tarefas.

As variáveis associadas à *satisfação do profissional* também são três:

- Satisfação: Nesta variável cada funcionário aponta o seu percentual de satisfação com as perspectivas que a carreira lhe oferece;
- Supervisão: Satisfação do empregado com a chefia imediata, o estilo gerencial do supervisor e o padrão de comunicação praticado no ambiente de trabalho, medida em percentual;
- Salário: Nível de satisfação do funcionário com o salário e outros benefícios, usando uma escala ordinal que varia de 1 a 10, sendo 1 = insatisfeito e valor 10 = excelente.

Figura 3.14
Dados do
Exemplo
3.5.1.

	X1 - Variação	X2 - Resposta	X3 - Autonomia	Y1 - Satisfação	Y2 - Supervisão	Y3 - Salário
1	18	18	80	54	40	6
2	77	22	87	63	76	7
3	83	63	73	96	31	9
4	82	75	97	96	98	6
5	28	68	97	84	94	6
6	28	24	68	66	10	5
7	64	32	75	42	37	8
8	19	15	50	45	14	2
9	33	13	70	42	18	6
10	23	14	90	79	74	4
11	37	13	70	39	12	2
12	23	74	53	54	35	3
13	45	58	83	60	75	5
14	22	67	53	63	45	5

3.5.1.1 – INSTRUÇÕES

- Clicar no *menu Análise Multivariada* e na opção *Correlação Canônica*;
- Selecionar todas as colunas referentes às variáveis do *grid* geral;
- Clicar em *Executar o Teste*, para obter o diálogo mostrado na Figura 3.15;
- Indique a quantidade de variáveis do Grupo X_i ;
- Indique a quantidade de variáveis do Grupo Y_i ;
- Pressionar em *Executar*, para obter os resultados.

Figura 3.15
Definição dos grupos de variáveis do Exemplo 3.5.1.

Quantidade de Variáveis X (Grupo do lado esquerdo)	Quantidade de Variáveis Y (Grupo do lado direito)
<input type="text" value="3"/>	<input type="text" value="3"/>
<input type="button" value="Cancelar"/>	<input type="button" value="Executar"/>

A figura acima define que as variáveis serão analisadas em dois grupos, onde o primeiro grupo é formado por três colunas, o segundo grupo de variáveis também é formado por três colunas. Esta especificação obedece a ordem na qual as colunas foram selecionadas na interface “Seleção de Amostras para Correlação Canônica”.

Figura 3.16
Resultados da correlação canônica. Exemplo 3.5.1.

Arquivo Editar Matriz das Correlações					
R canônico =	0.9352				
Qui-quadrado =	26.7996				
Graus de liberdade =	9				
p-valor =	0.0015				
Grupo de Variáveis X =	3				
Grupo de Variáveis Y =	3				
Autovalores (R2)	0.8747	0.3203	0.0854		
Correlações canônicas	0.9352	0.5660	0.2923		
U1 =	0.0611 X1	0.4994 X2	0.7892 X3		
V1 =	0.2565 Y1	0.7158 Y2	0.2283 Y3		
U2 =	1.1462 X1	-0.3475 X2	-0.4842 X3		
V2 =	-0.0246 Y1	-0.5516 Y2	1.0027 Y3		
U3 =	0.0658 X1	0.8430 X2	-0.6132 X3		
V3 =	1.3368 Y1	-0.9066 Y2	-0.4094 Y3		
Variável Canônica	R canônico	R2 canônico	Qui-quadrado	GL	p-valor
1	0.9352	0.8747	26.7996	9	0.0015
2	0.5660	0.3203	4.9918	4	0.2881
3	0.2923	0.0854	0.9376	1	0.3329
	X1	X2	X3		
Correlação entre U1 e Xi	0.5385	0.5816	0.8573		
	Y1	Y2	Y3		
Correlação entre V1 e Yi	0.8011	0.9390	0.5362		

A Figura 3.16 mostra as correlações canônicas, os autovalores e as probabilidades. A primeira correlação canônica (a correlação entre o primeiro par de variáveis canônicas) é 0.9352. Este valor, também denominado R Canônico, representa a melhor correlação possível entre qualquer combinação linear das variáveis de satisfação do profissional (Grupo Y) com as variáveis da característica da atividade desenvolvida (Grupo X).

A estatística R Canônica expressa a magnitude do relacionamento entre os dois conjuntos de variáveis. Neste exemplo é confirmada a associação entre os conjuntos de variáveis X_i e Y_i pelo Qui-Quadrado (χ^2) igual a 26.7996, com $gl = 9$, e o correspondente p -valor igual a 0.0015, que é muito significativo.

Pode-se, ainda, determinar a correlação entre as variáveis do Grupo X_i e as variáveis do Grupo Y_i , bastando clicar na parte superior da Figura 3.16 (*Matriz das Correlações*), cujos resultados estão dispostos na Figura seguinte (Figura 3.17), onde se constata que há quatro correlações X_i vs. Y_j que são significativas.

Figura 3.17
Correlações
do Exemplo
3.5.1.

	Y1 - Satisfação	Y2 - Supervisão	Y3 - Salário
X1 - Variação	0.437 ns	0.333 ns	0.669 (p < 0.01)
X2 - Resposta	0.574 (p < 0.05)	0.480 ns	0.234 ns
X3 - Autonomia	0.552 (p < 0.05)	0.783 (p < 0.01)	0.435 ns

3.5.2 – EXEMPLO 2

Estrutura populacional dos camarões *Penaeidae* no estuário do rio Caeté, litoral norte do Brasil.

Sendo seis variáveis do grupo X.

- CPUA: índice de captura por área (B/A), sendo B a biomassa em gramas e A corresponde a abertura efetiva da rede embaixo d'água.
- Densidade: medida em indivíduos/m².
- Temperatura.
- Salinidade.
- Oxigênio dissolvido.
- pH.

Variáveis do grupo Y.

- Dimensão 1.
- Dimensão 2.

Obs: As Dimensões representam o escalonamento multidimensional da matriz gerada pelo agrupamento dos dados com a distância.

Os dados estão exibidos parcialmente na Figura 3.18.

Figura 3.18
Visão parcial
dos dados do
Exemplo
3.5.2.

** Tempo	- 1 - Temperatura	- 2 - Salinidade	- 3 - Oxigênio Dissol	- 4 - pH	- 5 - Densidade	- 6 - CPUA	- 7 - Dimensão 1	- 8 - Dimensão 2
Sp1_Jun	28.530	21.320	5.470	7.260	8.670	0.007	0.158	-0.288
Sp1_Jul	29.700	23.800	5.880	7.230	7.880	0.008	-0.088	-0.205
Sp1_Set	27.950	28.730	6.380	7.510	2.860	0.003	-0.504	0.260
Sp2_Out	28.400	30.000	6.490	7.530	10.840	0.009	-0.661	-0.445
Sp2_Nov	35.520	34.000	6.540	7.900	0.760	0.000	-1.097	0.765
Sp2_Jan	26.930	32.250	7.770	8.080	14.340	0.006	-0.902	-0.793
Sp2_Fev	27.900	12.000	6.540	7.130	1.880	0.008	1.085	0.251
Sp2_Abr	27.100	8.000	8.030	7.890	12.640	0.004	1.411	-0.762
Sp2_Mai	27.000	24.000	8.000	8.000	18.280	0.006	-0.169	-1.205
Sp2_Jun	28.730	22.670	5.700	7.090	12.440	0.007	-0.015	-0.639
Sp2_Jul	29.070	26.660	6.250	7.490	13.820	0.007	-0.371	-0.732
Sp2_Set	28.050	29.830	6.440	7.490	13.390	0.007	-0.677	-0.663

3.5.2.1 – INSTRUÇÕES

- Clicar no menu *Análise Multivariada* e na opção *Correlação Canônica*;
- No diálogo *Seleção de Amostras para Correlação Canônica*, selecionar todas as colunas referentes às variáveis do grid geral;
- Clicar em *Executar a Estatística*;
- Indique a quantidade de variáveis dos Grupos X e Y, observe a Figura 3.19.
- Pressionar em *Executar*, para obter os resultados.

Figura 3.19
Definição dos
grupos de
variáveis do
Exemplo
3.5.2.

Quantidade de Variáveis X (Grupo do lado esquerdo)	Quantidade de Variáveis Y (Grupo do lado direito)
<input type="text" value="6"/>	<input type="text" value="2"/>
<input type="button" value="Cancelar"/>	<input type="button" value="Executar"/>

Figura 3.20
Resultados
do Exemplo
3.5.2.

Arquivo Editar Matriz das Correlações						
R canônico =	0.9999					
Qui-quadrado =	302.7856					
Graus de liberdade =	12					
p-valor =	< 0.0001					
Grupo de Variáveis X =	6					
Grupo de Variáveis Y =	2					
Autovalores (R2)	0.9999	0.9988				
Correlações canônicas	0.9999	0.9994				
U1 =	-0.0292 X1	-0.9768 X2	0.0026 X3	-0.0049 X4	-0.1344 X5	0.0010 X6
V1 =	1.0044 Y1	0.0880 Y2				
U2 =	0.1586 X1	0.0692 X2	-0.0089 X3	0.0124 X4	-0.9318 X5	-0.0020 X6
V2 =	0.0065 Y1	1.0006 Y2				
Variável Canônica	R canônico	R2 canônico	Qui-quadrado	GL	p-valor	
1	0.9999	0.9999	302.7856	12	< 0.0001	
2	0.9994	0.9988	130.8639	5	< 0.0001	
	X1	X2	X3	X4	X5	X6
Correlação entre U1 e Xi	-0.3816	-0.9919	-0.1895	-0.2858	-0.1408	0.0875
	Y1	Y2				
Correlação entre V1 e Yi	0.9962	-0.0064				

A correlação foi altamente significativa ($p\text{-valor} < 0.0001$) verificando-se na Figura 3.21 pela matriz de correlação as principais associações lineares entre as variáveis que compõem os dois grupos X e Y.

Figura 3.21
Matriz de
correlação do
Exemplo
3.5.2.

	Dimensão 1	Dimensão 2
Temperatura	-0.421 (p <0.05)	0.467 (p <0.05)
Salinidade	-0.998 (p <0.01)	0.126 ns
Oxigênio Dissol	-0.183 ns	-0.068 ns
pH	-0.299 ns	0.166 ns
Densidade	-0.054 ns	-0.979 (p <0.01)
CPUA	0.150 ns	-0.721 (p <0.01)

3.6 – DISTÂNCIA MULTIVARIADA (Euclidiana)

A *distância Euclidiana* abrange amostra com n indivíduos de espécies ou sub-espécies diferentes, cada uma apresentando p variáveis – X_1, X_2, \dots, X_p – mensuradas a nível intervalar ou de razões, procurando-se determinar a distância, sob o aspecto evolutivo, entre os indivíduos estudados. Os tamanhos das amostras podem ser iguais ou desiguais.

3.6.1 – EXEMPLO (hipotético)

Procurou-se verificar a *distância Euclidiana* de cinco (5) espécies de felinos, medindo-se: X_1 = distância entre o primeiro molar e o último molar; X_2 = comprimento do canino; X_3 = extensão da mandíbula; X_4 = distância entre o primeiro pré-molar e o último molar; X_5 = altura mandibular. Os dados foram inseridos no *grid* geral, de tal modo que as variáveis mensuradas correspondem às colunas X_1, X_2, X_3, X_4 e X_5 , e as espécies de felinos são correlativas às linhas: 1 = espécie A; 2 = espécie B; 3 = espécie C; 4 = espécie D; e 5 = espécie E.

Figura 3.22
Dados do
Exemplo
3.6.1.

	- 1 - Espécies	- 2 - X1	- 3 - X2	- 4 - X3	- 5 - X4	- 6 - X5
1	A	9.600	21.000	19.400	7.700	32.000
2	B	8.100	16.700	18.300	7.000	30.300
3	C	13.500	27.300	26.800	10.500	41.800
4	D	11.500	24.300	24.500	9.300	40.000
5	E	10.700	23.500	21.300	8.500	28.800

3.6.2 – INSTRUÇÕES

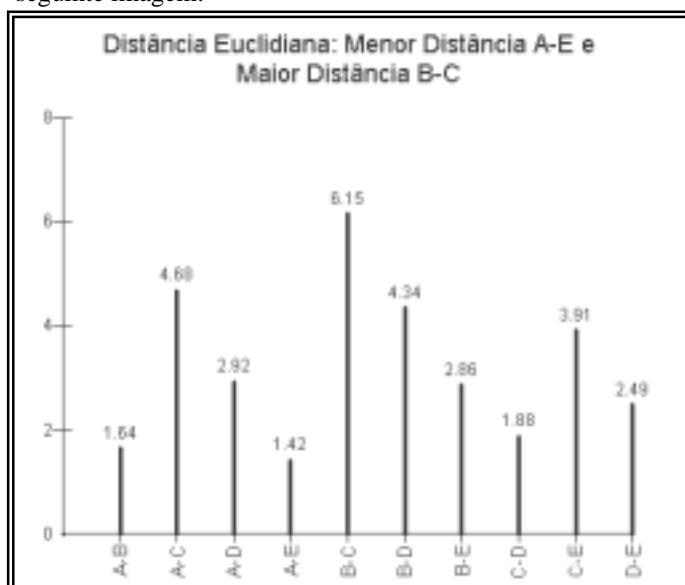
- Pressionar o *mouse* em *Distância Multivariada (Euclidiana)*;
- Selecionar as colunas referentes às variáveis do *grid* geral;
- Clicar em *Executar o Teste* para obter os resultados (ver Figura 3.23).

Figura 3.23
Resultados
do Exemplo
3.6.1.
A distância
observada
entre as
espécies.

	A	B	C	D	E
A	0.0000	---	---	---	---
B	1.6429	0.0000	---	---	---
C	4.6789	6.1539	0.0000	---	---
D	2.9238	4.3375	1.8798	0.0000	---
E	1.4198	2.8630	3.9055	2.4852	0.0000
Menor distância (A e E) =	1.4198	---	---	---	---
Maior distância (B e C) =	6.1539	---	---	---	---

d) Pressionar em Gráfico – barra de ferramentas, Fig. 3.23 – para visualizar a seguinte imagem:

Figura 3.24
Gráfico do
Exemplo
3.6.1.



3.7 – DISTÂNCIA MULTIVARIADA (*Penrose* e *Mahalanobis*)

O método de *Penrose* testa as distâncias multivariadas de duas ou mais populações, levando em consideração, para o cálculo, os dados sobre médias, variâncias e covariâncias. O teste de *Mahalanobis* considera ainda as correlações entre as variáveis, além dos valores numéricos utilizados no procedimento de *Penrose*. Os tamanhos das amostras podem ser iguais ou desiguais.

3.7.1 – EXEMPLO

O exemplo diz respeito a um trabalho (hipotético) realizado por um antropólogo sobre as medidas de quatro (4) variáveis cranianas do homem em cinco diferentes grupos indígenas: Grupo A = variáveis A₁, A₂, A₃, A₄; Grupo B = variáveis B₁, B₂, B₃, B₄; Grupo C = variáveis C₁, C₂, C₃, C₄; Grupo D = variáveis D₁, D₂, D₃, D₄; Grupo E = variáveis E₁, E₂, E₃, E₄. Os dados são introduzidos no *grid* geral abrangendo 20 colunas e 30 linhas, estas representando o número de indivíduos mensurados, conforme demonstrado na Figura 3.25.

Figura 3.25
Dados
parciais do
Exemplo
3.7.1.

	- 1 - A1	- 2 - A2	- 3 - A3	- 4 - A4	- 5 - B1	- 6 - B2	- 7 - B3	- 8 - B4
1	131	138	89	49	124	138	101	48
2	125	131	92	48	133	134	97	48
3	131	132	99	50	138	134	98	45
4	119	132	96	44	148	129	104	51
5	136	143	100	54	126	124	95	45
6	138	137	89	56	135	136	98	52

3.7.2 – INSTRUÇÕES

- Pressionar o mouse em *Distância Multivariada (Penrose/Mahalanobis)*;
- Selecionar as colunas referentes às variáveis do grid geral;
- Clicar em *Executar o Teste*;
- Informar o número de amostras introduzidas. Neste exemplo são cinco (5) amostras;
- Clicar em *OK*, para os resultados.

Figura 3.26

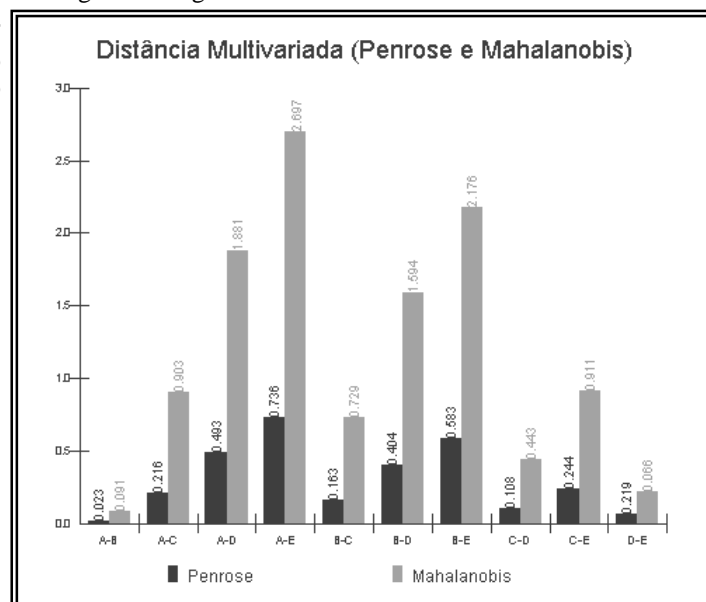
Dados
parciais do
Exemplo
3.7.1.

	Amostra A	Amostra B	Amostra C	Amostra D	Amostra E
Teste de Penrose:	---	---	---	---	---
Amostra A	0.0000	---	---	---	---
Amostra B	0.0228	0.0000	---	---	---
Amostra C	0.2158	0.1630	0.0000	---	---
Amostra D	0.4929	0.4044	0.1081	0.0000	---
Amostra E	0.7356	0.5827	0.2443	0.0663	0.0000
Maior distância (A e E) =	0.7356	---	---	---	---
Menor distância (A e B) =	0.0228	---	---	---	---
Teste de Mahalanobis:	---	---	---	---	---
Amostra A	0.0000	---	---	---	---
Amostra B	0.0910	0.0000	---	---	---
Amostra C	0.9031	0.7289	0.0000	---	---
Amostra D	1.8811	1.5940	0.4431	0.0000	---
Amostra E	2.6968	2.1757	0.9109	0.2193	0.0000
Maior distância (A e E) =	2.6968	---	---	---	---
Menor distância (A e B) =	0.0910	---	---	---	---

- Pressionar em Gráfico – barra de ferramentas da Fig. 3.26 – para visualizar a seguinte imagem:

Figura 3.27

Gráfico do
Exemplo
3.7.1.



3.8 – TESTE DE BARTLETT

O teste de *Bartlett* compara três ou mais amostras – m – multivariadas, obtendo-se um “coeficiente de máxima-verossimilhança” simbolizado por ϕ (phi), cuja significância estatística é calculada pelo Qui-Quadrado com $p(m - 1)$ graus de liberdade. O tamanho das amostras pode ser igual ou desigual.

3.8.1 – EXEMPLO

Os mesmos dados de exemplo da Distância Multivariada (*Penrose e Mahalanobis*) constituem o exemplo para este teste, compreendendo, pois, cinco (5) amostras, cada uma com quatro (4) variáveis (ver Figura 3.25).

3.8.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Teste de Bartlett*;
- Selecionar as colunas referentes às variáveis do *grid* geral;
- Clicar em *Executar o Teste*;
- Informar o número de amostras introduzidas. Neste exemplo são cinco (5) amostras, e clicar em *OK*, para resultados:

Figura 3.28
Resultados do Exemplo 3.8.1. Os resultados apresentam um valor de *Phi* elevado (61.3095), sendo estatisticamente muito significativo, eis que o *p-valor* é menor que 0.0001.

	Var 1	Var 2	Var 3	Var 4
Matriz T				
	3563.8933	-222.8133	-615.1600	426.7333
	-222.8133	3635.1733	1046.2800	346.4667
	-615.1600	1046.2800	4309.2600	-16.4000
	426.7333	346.4667	-16.4000	1533.3333
Matriz W				
	3061.0667	5.3333	11.4667	291.3000
	5.3333	3405.2667	754.0000	412.5333
	11.4667	754.0000	3505.9667	164.3333
	291.3000	412.5333	164.3333	1472.1333
Phi =	61.3095	---	---	---
Graus de liberdade =	16	---	---	---
Número de amostras =	5	---	---	---
Número de variáveis =	4	---	---	---
(p) =	< 0.0001	---	---	---

3.9 – TESTE DE HOTELLING

Destina-se a comparar duas amostras multivariadas, cada uma com o mesmo número de variáveis – duas ou mais –, baseando-se na generalização do *Teste t* de *Student*, mais precisamente no quadrado dessa estatística, sendo representado simbolicamente por T^2 . A probabilidade do teste – *p-valor* – é calculada pela estatística F resultante da transformação de T^2 , conforme demonstrado no capítulo sobre *Fórmulas Estatísticas*. Os dados devem ser mensurados a nível intervalar ou de razões.

3.9.1 – EXEMPLO

Duas amostras foram selecionadas, estudando-se cinco (5) variáveis de cada uma, cujos dados foram introduzidos no *grid* geral, correspondendo às variáveis X_i à

primeira, enquanto as variáveis Y_i representam a segunda, num total de dez (10) colunas (variáveis) e 28 linhas, estas representando o número de indivíduos sorteados para o teste.

Figura 3.29

Visão
parcial dos
dados do
Exemplo
3.9.1.

	- 1 - X1	- 2 - X2	- 3 - X3	- 4 - X4	- 5 - X5	- 6 - Y1	- 7 - Y2	- 8 - Y3
1	156.000	245.000	31.600	18.500	20.500	155.000	240.000	31.400
2	154.000	240.000	30.400	17.900	19.600	156.000	240.000	31.500
3	153.000	240.000	31.000	18.400	20.600	160.000	242.000	32.600
4	153.000	236.000	30.900	17.700	20.200	152.000	232.000	30.300
5	155.000	243.000	31.500	18.600	20.300	160.000	250.000	31.700
6	163.000	247.000	32.000	19.000	20.900	155.000	237.000	31.000
7	157.000	238.000	30.900	18.400	20.200	157.000	245.000	32.200
8	155.000	239.000	32.800	18.600	21.200	165.000	245.000	33.100
9	164.000	248.000	32.700	19.100	21.100	153.000	231.000	30.100
10	158.000	238.000	31.000	18.800	22.000	162.000	239.000	30.300
11	158.000	240.000	31.300	18.600	22.000	162.000	243.000	31.600
12	160.000	244.000	31.100	18.600	20.500	159.000	245.000	31.800

3.9.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Teste de Hotelling*;
- Selecionar as colunas referentes às variáveis do *grid* geral;
- Clicar em *Executar o Teste*, para resultados:

Figura 3.30

Resultados
do Exemplo
3.9.1.
Não há
diferença
estatística-
mente signifi-
cante entre
as duas
amostras,
 $p = 0.5210$.

	Var 1	Var 2	Var 3	Var 4	Var 5
Amostra 1: média =	157.1429	240.9643	31.4571	18.4821	20.8464
variância =	10.2011	20.2579	0.5225	0.2489	0.6367
tamanho =	28	28	28	28	28
Amostra 2: média =	158.4286	241.5714	31.4786	18.4464	20.8393
variância =	15.0688	32.5503	0.7284	0.4344	1.3210
tamanho =	28	28	28	28	28
T2 (Hotelling) =	4.5911	---	---	---	---
F =	0.8502	---	---	---	---
Graus de liberdade (n, d) =	5, 50	---	---	---	---
(p) =	0.5210	---	---	---	---

3.10 – TESTE DE MANTEL

Comparação entre duas matrizes de distâncias, como, por exemplo, da composição genética entre várias populações e respectivos intervalos geográficos ou temporais, obtendo-se valores do Coeficiente de Correlação (r), o qual pode variar de -1 a $+1$, denotando ou não associação matricial.

Há dois procedimentos que podem ser utilizados na entrada de dados para este teste:

- Procedimento A: dispomos das informações para a construção das matrizes;
- Procedimento B: dispomos das matrizes M e E.

3.10.1 – EXEMPLO: PROCEDIMENTO A

MATRIZ GENÉTICA x MATRIZ DE DISTÂNCIA GEOGRÁFICA

Um estudo epidemiológico cujo objetivo era verificar se fatores climáticos estariam associados a variações genéticas dos portadores de anemia de células falciformes. A amostra era composta por 20 indivíduos moradores de cinco (5) cidades, com características climáticas diversificadas. As cidades foram identificadas pelas letras: A, B, C, D e E. Para isso seria necessário comparar as características genéticas e geográficas presentes nos cinco (5) grupos de indivíduos. No primeiro momento determinou-se a frequência de um grupo de quatro (4) genes (Fator Alfa, Fator B, Fator V e IL 10). No segundo momento, obteve-se a distância linear (km) entre as cidades. A Figura 3.31 representa os dados genéticos da amostra, que estão armazenados no arquivo “Ex 03.10.01A – Dados Genéticos”. A Figura 3.32 representa as distâncias geográficas entre as cidades, estando armazenada no arquivo “Ex 03.10.01B – Distâncias geográficas”. Os arquivos estão disponíveis na pasta de Exemplos do Capítulo 3 do *BioEstat*.

$$H_0: r = 0$$

$$H_1: r \neq 0;$$

$$\text{Nível alfa} = 0.05.$$

Figura 3.31
Visão parcial da coleta de dados de características genéticas nas 5 cidades
Ex 3.10.1A.

	- 1 - Fator Alfa (A)	- 2 - Fator B (A)	- 3 - Fator V (A)	- 4 - IL 10 (A)	- 5 - Fator Alfa (B)	- 6 - Fator B (B)	- 7 - Fator V (B)
1	0.000	1.000	1.000	0.000	0.000	0.000	0.000
2	0.000	1.000	0.000	1.000	0.000	1.000	0.000
3	0.000	1.000	0.000	0.000	0.000	1.000	0.000
4	0.000	0.000	1.000	0.000	0.000	0.000	1.000
5	0.000	1.000	0.000	0.000	0.000	0.000	1.000
6	0.000	1.000	1.000	0.000	0.000	1.000	0.000
7	0.000	0.000	1.000	1.000	0.000	1.000	0.000
8	0.000	1.000	1.000	1.000	0.000	1.000	1.000
9	0.000	1.000	1.000	0.000	0.000	0.000	0.000
10	0.000	1.000	0.000	0.000	0.000	1.000	0.000

Figura 3.32
Distâncias geográficas em 5 pontos de coleta em cada município.
Ex 3.10.1B.

	- 1 - Município	- 2 - Distrito 1	- 3 - Distrito 2	- 4 - Distrito 3	- 5 - Distrito 4	- 6 - Distrito 5
A		9.600	21.000	19.400	0.000	32.000
B		8.100	16.700	18.300	7.000	30.300
C		13.500	27.300	26.800	10.500	41.800
D		11.500	24.300	24.500	9.300	40.000
E		10.700	23.500	21.300	8.500	28.800

3.10.1.1 – OBTENÇÃO DAS MATRIZES

- a) Com os dados do “Ex 03.10.1A – Dados Genéticos” efetuar o procedimento de *Penrose–Mahalanobis*, para cinco (5) amostras, conforme descrito no item 3.7.1, para obter as distâncias entre as cinco (5) amostras. A matriz das distâncias de *Penrose* é mostrada no Figura 3.33.

Figura 3.33
Dados da
matriz
Penrose.

	Amostra A	Amostra B	Amostra C	Amostra D	Amostra E
Teste de Penrose:	---	---	---	---	---
Amostra A	0.0000	---	---	---	---
Amostra B	0.0946	0.0000	---	---	---
Amostra C	0.3559	0.2371	0.0000	---	---
Amostra D	0.2114	0.2360	0.1477	0.0000	---
Amostra E	0.0798	0.0273	0.1969	0.2356	0.0000

b) Clicar em *Teste de Mantel* (segmento superior da Figura 3.33), para obter a imagem como na Figura 3.34. Este procedimento insere instantaneamente as Distâncias de *Penrose* (matriz) na entrada de dados do *Teste de Mantel*.

Figura 3.34
Entrada de
dados da
primeira
matriz (dados
genéticos)
para o teste
de *Mantel*.

Teste de Mantel: Matrizes de Distâncias

Configurar Matrizes

Matriz A
Distância Multivariada (Penrose e Mahalanobis)
Ex 03.10.1 - Dados Genéticos

	C1	C2	C3	C4	C5
L1	0.0000	---	---	---	---
L2	0.0897	0.0000	---	---	---
L3	0.4093	0.2464	0.0000	---	---
L4	0.2648	0.2453	0.1477	0.0000	---
L5	0.0749	0.0273	0.2062	0.2448	0.0000

Matriz B

	C1	C2	C3	C4	C5
L1					
L2					
L3					
L4					
L5					

Teste de Mantel Bootstrap ↩

- c) Para fechar as janelas do *Teste de Mantel* e da *Distância Multivariada de Penrose-Mahalanobis*, clicar na seta (parte inferior direita da Figura 3.34), e fechar a janela das Figuras 3.34 e 3.33;
- d) Repetir este procedimento para os dados das Distâncias Geográficas (Figura 3.32), mas utilizando a *Distância Euclidiana*, como se observa na Figura 3.34;

Figura 3.35
Dados da
Matriz
Euclidiana.

	A	B	C	D	E
A	0.0000	---	---	---	---
B	2.4360	0.0000	---	---	---
C	4.9666	5.5292	0.0000	---	---
D	3.6229	3.9576	1.6358	0.0000	---
E	2.6155	2.6184	3.5875	2.4073	0.0000

- e) Repetir os procedimentos do item b), ou seja, clicar em Teste de *Mantel*, na parte superior da Figura 3.33, no sentido de obter a seguinte imagem:

Figura 3.36
As duas matrizes estão prontas para o teste de *Mantel*.

Matriz A							Matriz B						
Distância Multivariada (Penrose e Mahalanobis)							Distância Multivariada (Euclidiana)						
Ex 03.10.1 - Dados Genéticos							Ex 03.10.2 - Distâncias Geográficas						
	C1	C2	C3	C4	C5			C1	C2	C3	C4	C5	
L1	0.0000	—	—	—	—		L1	0.0000	—	—	—	—	
L2	0.0897	0.0000	—	—	—		L2	2.4360	0.0000	—	—	—	
L3	0.4093	0.2464	0.0000	—	—		L3	4.9666	5.5292	0.0000	—	—	
L4	0.2648	0.2453	0.1477	0.0000	—		L4	3.6229	3.9576	1.6358	0.0000	—	
L5	0.0749	0.0273	0.2062	0.2448	0.0000		L5	2.6155	2.6184	3.5875	2.4073	0.0000	

- f) Clicar em Teste de *Mantel* (parte inferior da figura 3.34), para obter os seguintes resultados (Fig. 3.37):

Figura 3.37
Resultado do Teste de *Mantel*

Resultado do Teste de Mantel	
Matrizes	5 x 5
Z: Somatório (Mi x Ei) =	7.3819
Coef. de Correlação (r) =	0.6830
t =	2.6451
Graus de liberdade =	8
p-valor =	0.0294

Os resultados mostram associação entre dados genéticos e distâncias geográficas, com o Coeficiente de Correlação igual a 0.6830 e *p-valor* significativo (0.0294). O valor **Z** representa o somatório do produto das variáveis das matrizes M_i e E_i . Assim temos: $Z = (0.0897 \times 2.4360 + 0.4093 \times 4.9666 + \dots + 0.2448 \times 2.4073) = 7.3819$.

É possível, entretanto, obter outros valores de **Z** permutando-se os valores das matrizes, procedimento conhecido como *bootstrap*. Assim, basta clicar em **Bootstrap**, na parte inferior da Figura 3.34, e preencher as caixas de texto, conforme o desejado, como se depreende da Figura 3.38.

Figura 3.38
Resultados da reamostragem das matrizes.

Bootstrap - p-valor: Correlação Lin...	
Imprimir	
Número de pares	Número de simulações
10	10000
Executar	Cancelar
r (Pearson) prior	r (simulação)
0.6830	0.5484
p-valor (simulação)	
0.0170	

Os resultados são semelhantes aos obtidos diretamente no Teste de *Mantel*, sem a reamostragem, com Coeficiente de Correlação $r = 0.5484$ e *p-valor* = 0.0170. É possível, na reamostragem, a ocorrência de valores positivos ou negativos, mas o *p-valor* terá a significação aproximada do obtido diretamente das matrizes sem *bootstrap*, e que correspondeu a 0.0294 (Figura 3.37).

3.10.2 – EXEMPLO: PROCEDIMENTO B

AS MATRIZES GENÉTICA E GEOGRÁFICA JÁ ESTÃO DISPONÍVEIS

Os dados são semelhantes ao exemplo anterior, isto é, dizem respeito às Distâncias Genéticas e Espaciais, entretanto são preenchidas com outros escores.

$$H_0: r = 0$$

$$H_1: r \neq 0;$$

$$\text{Nível alfa} = 0.05.$$

3.10.2.1 – INSTRUÇÕES

- Introduzir as Matrizes M e E no *grid* geral do *BioEstat*, conforme se verifica na Figura 3.39. É importante destacar que as matrizes devem ter sempre o mesmo tamanho: neste exemplo elas são **5x5**.
- Pressionar o *mouse* em *Análise Multivariada* e, a seguir, em *Teste de Mantel*, seleccionar as colunas do *grid*, clicando, em seguida em *Executar*, para resultados (Figura 3.40).

Figura 3.39

Matrizes M_i e E_i obtidas das Distâncias Penrose e Euclidiana.

- 1 - Matriz M	- 2 - Matriz M	- 3 - Matriz M	- 4 - Matriz M	- 5 - Matriz M	- 6 - Matriz E	- 7 - Matriz E	- 8 - Matriz E	- 9 - Matriz E	- 10 - Matriz E
0.000	---	---	---	---	0.000	---	---	---	---
0.095	0.000	---	---	---	1.643	0.000	---	---	---
0.356	0.237	0.000	---	---	4.679	6.154	0.000	---	---
0.211	0.236	0.148	0.000	---	2.924	4.338	1.880	0.000	---
0.080	0.027	0.197	0.236	0.000	1.420	2.863	3.906	2.485	0.000

Figura 3.40
Resultados do Teste de Mantel.
Exemplo 3.10.2.

Arquivo Editar Bootstrap	
	Resultado
Matrizes	5 x 5
Z: Somatório ($M_i \times E_i$) =	6.7451
Coef. de Correlação (r) =	0.6548
t =	2.4508
Graus de liberdade =	8
p-valor =	0.0398

A correlação obtida, com $r = 0.6548$, é significativa, eis que o *p-valor* é igual a 0.0398, rejeitando-se a hipótese de nulidade. Há, portanto, associação entre as frequências gênicas e as distâncias geográficas. Se houver interesse na reamostragem, pressionar em *Bootstrap*, na parte superior da Figura 3.40, efetuando o mesmo procedimento já referido.

CAPÍTULO 4

ANÁLISE DE SOBREVIVÊNCIA

4.1 – INTRODUÇÃO

A análise de sobrevivência vem sendo enfatizada nos estudos biométricos e destina-se a calcular a probabilidade de ocorrência de eventos ao longo de certo período, o qual é dividido em vários intervalos de tempo. Os eventos podem ser de diversas naturezas, tais como: períodos de remissão de uma doença, tempo de sobrevivência de indivíduos acometidos por neoplasia, estudos epidemiológicos, duração de matrimônios, efeitos de drogas e outros.

Um fato a destacar nas análises de sobrevivência é o de que nem todos os elementos selecionados para o estudo atendem ao seguimento ao longo do período previamente estabelecido, podendo, a qualquer momento, abandoná-lo por motivos diversos. Esses indivíduos recebem a designação geral de *censurados*, os quais, de acordo com o teste escolhido, devem ou não ser registrados em cada tempo t no qual o período total foi dividido.

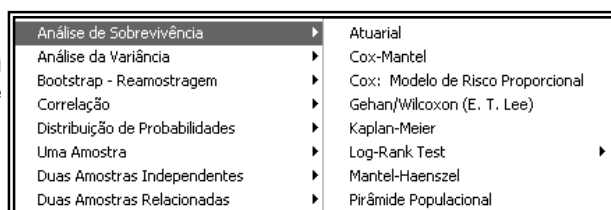
O *BioEstat* oferece os seguintes modelos de Análise de Sobrevivência:

- Atuarial;
- Cox-Mantel;
- Cox: Modelo de Risco Proporcional;
- Kaplan-Meier;
- Log-Rank Test*;
- Gehan (Wilcoxon generalizado);
- Mantel-Haenszel;
- Pirâmide Populacional.

4.1.1 – INSTRUÇÕES

- Introduzir os dados no *grid* padrão do *BioEstat*;
- Dirigir o *mouse* para o *menu* superior e clicar em *Estatística*;
- Apontar o *mouse* em *Análise de Sobrevivência*, conforme demonstra a Figura 4.1

Figura 4.1
Menu
Análise de
Sobrevivência.



4.2 – ATUARIAL

Este procedimento consiste na observação de um grupo de indivíduos em períodos relativamente longos e divididos em vários intervalos constantes – t – de semanas, meses ou anos, registrando-se na 1ª coluna o número de indivíduos vivos, na 2ª, o de ocorrências – remissões, recidivas, mortes, etc. – e na 3ª, o de censurados – 0, 1, 2, ..., n . Este modelo é de largo uso em instituições securitárias.

4.2.1 – EXEMPLO

Um estudo foi efetuado em 87 pessoas operadas de neoplasia do estômago, as quais foram observadas ao longo de cinco anos, com intervalos – t – anuais. Os dados foram introduzidos no *grid* geral, registrando-se nas colunas 1, 2 e 3 os vivos, as *ocorrências* e os *censurados*, respectivamente. As linhas correspondem aos intervalos anuais: 1, 2, 3, 4 e 5 anos.

Figura 4.2
Dados do
Exemplo
4.2.1.

	- 1 - Vivos	- 2 - Ocorrências	- 3 - Censurados
1	87.000	12.000	0.000
2	75.000	9.000	0.000
3	66.000	17.000	1.000
4	48.000	36.000	4.000
5	8.000	6.000	2.000

Medida de tempo

Informe a Unidade de tempo do intervalo constante entre cada amostra
Ex: dia, mês, trimestre, ano, etc...

OK Cancel

ano

4.2.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Atuarial*;
- Selecionar as colunas referentes às variáveis do *grid* geral;
- Clicar em *Executar o Teste*, informando – *caixa de texto* ao lado da Fig. 4.2 – a unidade de tempo representada pelos números 1 a 5 (ano, neste exemplo);
- Pressionar *OK*, para resultados.

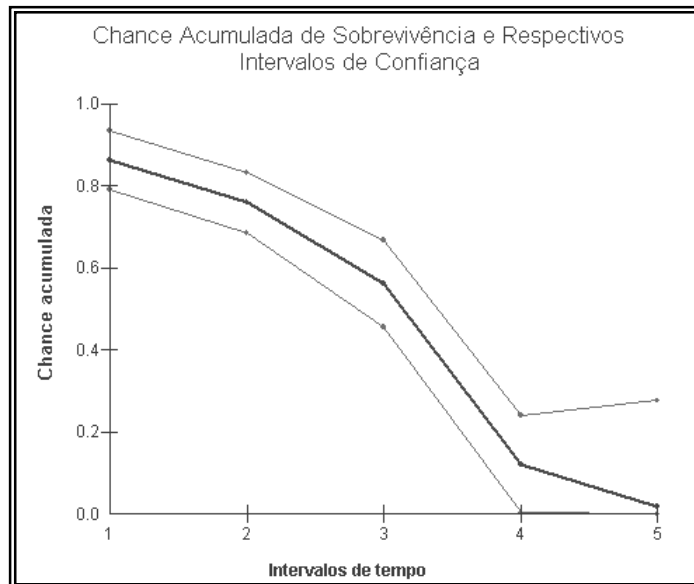
Figura 4.3
Resultados
do Exemplo
4.2.1.

Análise de Sobrevida: Atuarial						
Arquivo Editar Gráfico						
	Vivos	Sob risco	Proporção de risco	Chance de Sobrevida	Chance acm. Sobrevida	IC 95%
Intervalo 1	87	87.00	0.1379	0.8621	0.8621	0.7896 - 0.9345
Intervalo 2	75	75.00	0.1200	0.8800	0.7586	0.6851 - 0.8322
Intervalo 3	66	65.50	0.2595	0.7405	0.5617	0.4556 - 0.6679
Intervalo 4	48	46.00	0.7826	0.2174	0.1221	0.0029 - 0.2413
Intervalo 5	8	7.00	0.8571	0.1429	0.0174	0.0000 - 0.2767
Expectativa de vida	2.82	ano				

Ao final de cinco (50 anos, a chance de sobrevivência é bem exígua (0.0174), aumentando a proporção de risco a cada intervalo t . A expectativa de vida anual é de 2.82.

Pressionando-se em Gráfico, na barra superior desta figura, escolhe-se uma das opções: Frequência, Proporção ou Chance acumulada de sobrevivência com IC (95%). Neste exemplo escolheu-se a terceira opção, surgindo o gráfico da Figura 4.4.

Figura 4.4
Gráfico do
Exemplo
4.2.1.



4.3 – COX-MANTEL

Teste estatístico não-paramétrico para duas amostras independentes com observações censuradas.

4.3.1 – EXEMPLO (hipotético)

Dois grupos de seis (6) mulheres com câncer mamário foram tratados do seguinte modo: o primeiro grupo recebeu quimioterapia e radioterapia, enquanto o outro nada recebeu após mastectomia total. Ao final de três (3) anos, o tempo de remissão foi assinalado em meses. As observações **censuradas** estão assinaladas no *grid* específico com o sinal +.

$$H_0: S_1(t) = S_2(t);$$

$$H_1: S_1(t) > S_2(t);$$

Nível de decisão: alfa = 0.05 (unilateral)

4.3.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Cox-Mantel*;
- Inserir os valores de cada grupo (A e B) no *grid* específico, assinalando com o caractere “+” sucedendo os escores censurados;
- Clicar em *OK* para obter os resultados, conforme disposto na Figura 4.5.

Figura 4.5
Resultados
do Exemplo
4.3.1.

	Grupo A	Grupo B
	21	14
	17+	16
	19+	18
	21+	19
	23+	21
	18	15

Nome do arquivo
Exemplo_4_3_1.CoxMantel

Cancelar Executar

U 3.0621 p (unilateral) 0.0079
Var 1.6119 p (bilateral) 0.0158
Z 2.4118

A diferença entre os tratamentos é muito significativa ($p = 0.0079$), rejeitando-se a hipótese de nulidade, constatando-se que o tempo de sobrevivência foi maior no grupo tratado com quimioterapia e radioterapia.

4.4 – COX: MODELO DE RISCO PROPORCIONAL

É uma técnica de análise estatística que busca explicar o efeito que um grupo de variáveis exerce na sobrevivência do paciente. Permite a análise, além do efeito da droga, de vários outros fatores de risco simultaneamente. As variáveis denominadas *preditoras* (*explanatórias*) podem ser fatores de risco tais como: idade, sexo, peso, glicemia, duração da doença ou variáveis de controle que indicam o tipo de tratamento administrado ao paciente. O modelo de risco proporcional introduzido por Cox é conhecido na literatura como “*Proportional Hazard Regression Analysis*”.

INTERPRETAÇÃO DO COEFICIENTE b

Quando o *Coefficiente de Regressão* (b) de uma das variáveis preditoras é um número positivo ($b > 0$), então entendemos que o risco (*hazard*) para os pacientes é crescente com a grandeza da variável e, por via de consequência, o prognóstico é que a variável em questão diminui o tempo de sobrevivência do paciente. Por outro lado, quando uma especificada variável é caracterizada por um *Coefficiente de Regressão Negativo* ($b < 0$) devemos deduzir que a tal variável melhora o prognóstico de sobrevivência do paciente.

Num exemplo hipotético, o risco estimado com determinada dose de uma droga terapêutica é de $b = -0.41$, a respectiva *Taxa de Risco* é calculada por $\exp(-0.41)$ é igual a 0.66 (66%), ou seja, a droga decresce em 34% o risco de rejeição, óbito, etc.

Em outro exemplo, uma variável preditora, como a idade do paciente, o Coeficiente de Regressão (b), é positivo, igual a 0.70. A *Taxa de Risco* dessa variável é de $\exp(0.70) = 2.01$, isto é, pessoas idosas têm o risco aumentado em duas vezes (2.01).

TAXA DE RISCO (HAZARD RATE)

É definida como a probabilidade, por unidade de tempo, que um paciente tenha sobrevivido ao início do respectivo intervalo e que falhará nesse intervalo. Especificamente é computado como o número de falhas por unidade de tempo no respectivo intervalo, dividido pelo número médio de casos de sobrevivência no ponto médio do intervalo.

4.4.1 – EXEMPLO (dados hipotéticos)

Em um grupo de 48 pacientes portadores de mieloma múltiplo, por randomização, 24 indivíduos receberam tratamento com 75mg de *riluzole* (Droga), e os demais receberam placebo. Ao mesmo tempo, foram registrados, para cada paciente, os valores referentes às seguintes variáveis: Idade, Sexo, Uréia sanguínea, Cálcio sérico, Hemoglobina e Proteína de *Bence-Jones*. A Duração do tratamento (Tempo) e a ocorrência de remissão foram registradas como nas análises de sobrevivência.

O objetivo deste experimento é, além de testar o efeito da droga, verificar a influência dos demais fatores (variáveis preditoras) na remissão da doença.

Figura 4.6
Visão parcial
do Exemplo
4.4.1.

- 1 - Droga	- 2 - Idade	- 3 - Sexo	- 4 - Uréia	- 5 - Cálcio sérico	- 6 - Hemoglobina	- 7 - Bence Jones	- 8 - Tempo	- 9 - Remissão
1	63	1	24.0	12.00	14.0	1	15	0
0	76	1	12.0	10.00	13.0	0	50	1
1	54	0	15.0	13.00	12.0	0	8	0
0	65	0	11.0	11.00	9.9	1	40	1
1	70	1	22.0	10.00	14.0	1	11	0
1	60	0	11.0	9.00	11.0	0	6	1
1	53	1	22.0	11.00	13.0	1	63	1
0	61	0	38.0	8.00	12.0	1	12	1
1	68	1	37.0	13.00	7.0	0	11	0
0	72	0	41.0	11.00	10.9	1	14	1
1	63	1	40.0	9.00	11.0	0	15	0
0	54	0	148.0	10.00	10.0	1	5	1
1	61	1	27.0	11.00	7.0	0	58	1

4.4.2 – INSTRUÇÕES

- No menu *Estatísticas* clique em *Análise de Sobrevivência*;
- Selecione a opção *Cox: Modelo de Risco Proporcional*;
- Na janela *Seleção de Amostras*, escolha inicialmente as variáveis: Droga, Idade, Sexo, Uréia, Cálcio sérico, Hemoglobina, Proteína de *Bence-Jones*.
- Ainda na mesma janela selecione as colunas: Tempo e Remissão, as quais devem, obrigatoriamente, ser as duas últimas colunas selecionadas, nessa ordem.
- A janela *Seleção de Amostras* será configurada conforme a visualização a seguir.

Figura 4.7
Seleção de
variáveis
do Exemplo
4.4.1

f) Finalmente, clique em *Executar Estatística* para obter os resultados abaixo, que para fins didáticos foram divididos em duas partes (Figuras 4.8 e 4.9).

Figura 4.8
Parte
superior do
grid de
resultados
do Exemplo
4.4.1.

Arquivo Editar Curva de Sobrevida							
	- 1 -	- 2 -	- 3 -	- 4 -	- 5 -	- 6 -	- 7 -
Tempo de sobrevivência	Tempo						
Variável dependente	Remissão						
	Droga	Idade	Sexo	Uréia sang.	Cálcio sérico	Hemoglobina	Bence Jones
Média aritmética=	0.5417	63.3125	0.5625	31.8542	10.5000	10.3396	0.5208
Desvio padrão =	0.4983	7.0213	0.4961	31.8793	1.6073	2.6371	0.4996
Tamanho da amostra	48						
Null -2 Log Likelihood	189.8304						
Full -2 Log Likelihood	157.2521						
Qui-Quadrado	32.5782	GL = 7	p < 0.0001				
Variáveis preditoras	b	E. Padrão	p-valor	Tx. Risco	IC 95% (Tx.		
Droga	-1.8432	0.4916	0.0002	0.1583	0.0604 - 0.4149		
Idade	-0.0166	0.0304	0.5862	0.9836	0.9267 - 1.0440		
Sexo	-0.1202	0.4423	0.7858	0.8868	0.3727 - 2.1101		
Uréia sang.	0.0054	0.0070	0.4388	1.0054	0.9917 - 1.0193		
Cálcio sérico	-0.4125	0.1454	0.0046	0.6620	0.4978 - 0.8803		
Hemoglobina	-0.2407	0.0796	0.0025	0.7861	0.6725 - 0.9189		
Bence Jones	-0.0173	0.4240	0.9675	0.9829	0.4281 - 2.2564		

A parte superior da tela de resultados da análise de *Cox* apresenta:

- O Qui-Quadrado igual a 32.5782 e o correspondente $p < 0.0001$, o qual indica uma forte evidência do efeito da droga ($p = 0.0002$) e de variáveis preditoras sobre a variável Remissão, destacando-se também os níveis de Cálcio sérico ($p = 0.0046$) e o de Hemoglobina ($p = 0.0025$).
- A droga (*riluzole*) obteve o coeficiente $b = -1.8432$ ($p = 0.0002$), cuja Taxa de Risco correspondente é igual a 0.1583 (15.83%), dando uma proteção de 84.17%.

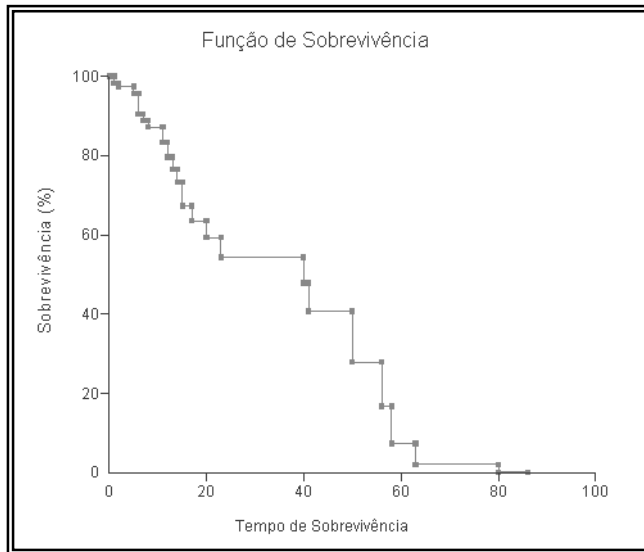
Figura 4.9
Visão parcial
da parte
inferior dos
resultados
do Exemplo
4.4.1.

Tempo	Prop. Sobrev.	Tx. Risco.	Risco Acm.
1	0.983	0.0167	0.0167
2	0.974	0.0093	0.0259
5	0.956	0.0193	0.0452
6	0.904	0.0554	0.1006
7	0.889	0.0170	0.1176
8	0.871	0.0206	0.1382

Na parte superior da Figura 4.8 está localizada a opção que permite visualizar as Curvas de Sobrevida.

O diagrama abaixo representa a *Curva Geral* de sobrevivência.

Figura 4.10
Gráfico
do Exemplo
4.4.1.



Outras curvas podem ser obtidas para ilustrar a influência de cada variável preditora na sobrevivência dos indivíduos que compõem a amostra:

- Droga;
- Idade;
- Sexo;
- Uréia;
- Cálcio;
- Hemoglobina;
- Proteína de *Bence-Jones*.

Figura 4.11
Gráfico
do Exemplo
4.4.1.

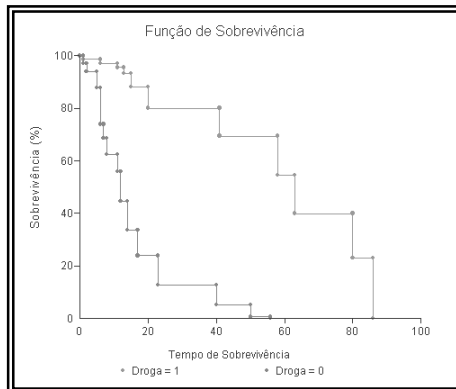


Gráfico da variável Droga. A curva superior corresponde aos pacientes que receberam o tratamento ($p = 0.002$).

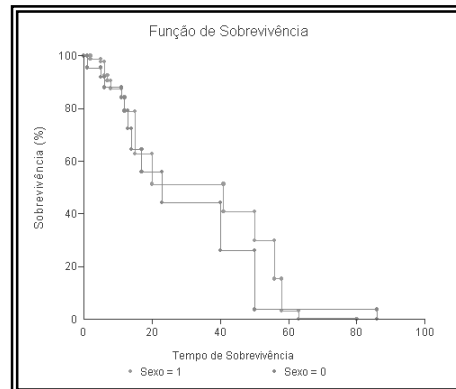


Gráfico da variável Sexo. As curvas se sobrepõem indicando que o sexo não influenciou o tempo de sobrevivência ($p = 0.7858$).

4.5 – KAPLAN-MEIER

Este procedimento é semelhante ao atuarial, porém o tempo aprazado é menor, de dias ou meses, e os intervalos não são, de modo obrigatório, igualmente espaçados, obtendo-se, ainda, em cada tempo t , o intervalo de confiança (IC 95%) respectivo.

4.5.1 – EXEMPLO

180 indivíduos com patologia de alto risco de mortalidade foram submetidos a tratamento com a droga A em altas doses. O período de observação foi de 6 meses, com intervalos mensais de registro das observações. Os dados foram introduzidos no *grid* geral: a Coluna 1 corresponde ao tempo t ; a Coluna 2, aos indivíduos vivos; e a coluna 3, às ocorrências (óbitos). Observa-se que os intervalos mensais não são constantes como ocorre na *Análise Atuarial* (ver Figura 4.12).

Figura 4.12
Dados do
Exemplo
4.5.1.

	- 1 - Tempo t	- 2 - Vivos	- 3 - Ocorrências	-
1	3.000	180.000	4.000	
2	6.000	166.000	7.000	
3	18.000	158.000	16.000	
4	39.000	144.000	22.000	
5	58.000	97.000	16.000	
6	64.000	45.000	6.000	

4.5.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Kaplan-Meier*;
- Selecionar as duas colunas referentes às variáveis do *grid* geral;
- Clicar em *Executar o Teste* para obter os resultados.

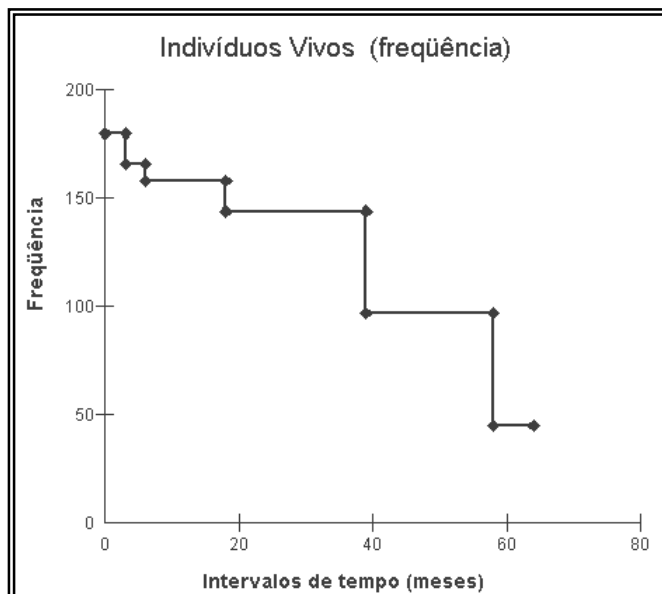
Figura 4.13
Resultados
do Exemplo
4.5.1.

Análise de Sobrevida: Kaplan-Meier							
Arquivo Editar Gráfico							
	Tempo de sobrevivência	Indivíduos vivos	Ocorrências	Sobrevivência no tempo t	Sobrevivência até o tempo t	Estimativa do Risco	IC (95%)
Evento 1	3.00	180	4	0.9778	0.9778	0.0222	0.9562 a 0.9993
Evento 2	6.00	166	7	0.9578	0.9365	0.0422	0.9002 a 0.9729
Evento 3	18.00	158	16	0.8987	0.8417	0.1013	0.7869 a 0.8965
Evento 4	39.00	144	22	0.8472	0.7131	0.1528	0.6453 a 0.7810
Evento 5	58.00	97	16	0.8351	0.5955	0.1649	0.5181 a 0.6728
Evento 6	64.00	45	6	0.8667	0.5161	0.1333	0.4267 a 0.6055
Média	31.3333	131.6667	11.8333				
Desvio padrão	26.3034	51.1338	7.1671				

Probabilidade de sobrevivência em cada período e a chance acumulada até o referido tempo, com IC 95%.

Pressionando-se em *Gráfico*, na barra superior da Figura 4.13, escolhe-se uma das opções: *Frequência*, *Proporção* ou *Chance acumulada de sobrevivência com IC (95%)*. Neste exemplo escolheu-se a primeira opção, surgindo o diagrama da Figura 4.14.

Figura 4.14
Gráfico do
Exemplo
4.5.1.



4.6 – LOG-RANK TEST

Este teste de análise de sobrevivência permite comparar *duas amostras*, cujas observações foram efetuadas em períodos e intervalos idênticos.

4.6.1 – EXEMPLO 1 (sem censurados, método 1)

O tempo de recidiva ao alcoolismo foi observado em dois grupos:

- Grupo A, indivíduos com até 45 anos ($n = 376$);
- Grupo B, com idade superior a 45 anos ($n = 450$).

O objetivo era verificar se a variável *idade* influenciava a ocorrência objeto da investigação. O tempo de observação foi de 365 dias, dividido em intervalos trimestrais, e contadas as ocorrências. Os dados foram introduzidos no *grid* geral.

Figura 4.15
Dados do
Exemplo
4.6.1 (sem
censurados).

- 1 - Tempo (trimestre)	- 2 - ≤ 45 anos	- 3 - > 45 anos
Primeiro	174.000	183.000
Segundo	100.000	120.000
Terceiro	50.000	80.000
Quarto	31.000	37.000
Não recidivaram	21.000	30.000

H_0 : o tempo de recidiva ao alcoolismo é o mesmo nos dois grupos de pessoas: $p_1 = p_2$;

H_1 : o tempo de recidiva ao alcoolismo é dependente do grupo etário dos indivíduos:

$$p_1 \neq p_2;$$

Nível de decisão: $\alpha = 0.05$.

4.6.2 – EXEMPLO 2 (sem censurados, método 2)

Dados hipotéticos foram coletados relativos aos tratamentos A e B. No grupo A foram acompanhados quatorze (14) indivíduos enquanto no grupo B, 35. Os dados estão dispostos na Figura 4.16.

Figura 4.16
Visão parcial
dos dados
do Exemplo
4.6.2.

- 1 - Tempo	- 2 - Óbitos (A)	- 3 - Vivos (A)	- 4 - Óbitos (B)	- 5 - Vivos (B)
5	0	14	2	33
8	0	14	1	31
10	0	14	1	30
13	0	14	1	29
18	0	14	1	28
23	0	14	0	27
24	0	14	1	27
26	0	14	2	26
31	0	14	1	24
35	0	14	1	23
40	0	14	1	22

H_0 : não há diferença no tempo de sobrevivência dos indivíduos submetidos aos tratamentos A e B: $p_1 = p_2$;

H_1 : há diferença no tempo de sobrevivência dos indivíduos submetidos aos tratamentos A e B: $p_1 \neq p_2$;

Nível de decisão: alfa = 0.05.

4.6.3 – EXEMPLO 3 (com censurados)

Dados hipotéticos de duas amostras sobre tempo de sobrevivência de dois conjuntos de indivíduos submetidos a tratamentos diferentes: A, para um grupo e B, para o outro, com seguimento ao longo de seis anos, correspondendo cada valor do *grid* ao respectivo número de indivíduos. Os indivíduos dos Eventos A e B totalizaram 99 e 77, respectivamente.

H_0 : não há diferença no tempo de sobrevivência dos indivíduos submetidos aos tratamentos A e B: $p_1 = p_2$;

H_1 : há diferença no tempo de sobrevivência dos indivíduos submetidos aos tratamentos A e B: $p_1 \neq p_2$;

Nível de decisão: alfa = 0.05.

Figura 4.17
Dados do
Exemplo
4.6.3 (com
censurados).

- 1 - Tempo (dias)	- 2 - Evento A	- 3 - Censur. A	- 4 - Evento B	- 5 - Censur B
40.0	4.0	3.0	7.0	1.0
70.0	12.0	1.0	16.0	3.0
110.0	25.0	0.0	13.0	3.0
125.0	25.0	30.0	20.0	27.0
150.0	21.0	12.0	15.0	20.0
200.0	12.0	5.0	6.0	10.0

4.6.4 – INSTRUÇÕES

- a) Pressionar o *mouse* em *Log-Rank test*, e escolha uma das duas opções: *Sem Censurados* (Figuras 4.15 e 4.16) ou *Com Censurados* (Figura 4.17);

- b) Selecionar, no *grid* geral, as colunas referentes às variáveis conforme o caso;
- c) Clicar em *Executar o Teste*, para obter os resultados.

Figura 4.18
Resultados
do Exemplo
4.6.1.

Arquivo Editar Gráficos	
	Resultados
Observado =	355
Esperado =	340.2306
Variância =	101.3469
Qui-Quadrado =	2.0091
Graus de liberdade =	1
(p) =	0.1564

A diferença no tempo de recidiva ao alcoolismo não foi significativa ($p = 0.1564$).

Figura 4.19
Resultados
do Exemplo
4.6.2.

Arquivo Editar Gráfico	
	Resultados
Observado =	4.0000
Esperado =	10.3443
Variância (VL) =	6.0925
Qui-quadrado (WL) =	6.6064
Graus de liberdade =	1
(p) =	0.0102

A diferença foi significativa ($p = 0.0102$).

Figura 4.20
Resultados
do Exemplo
4.6.3.

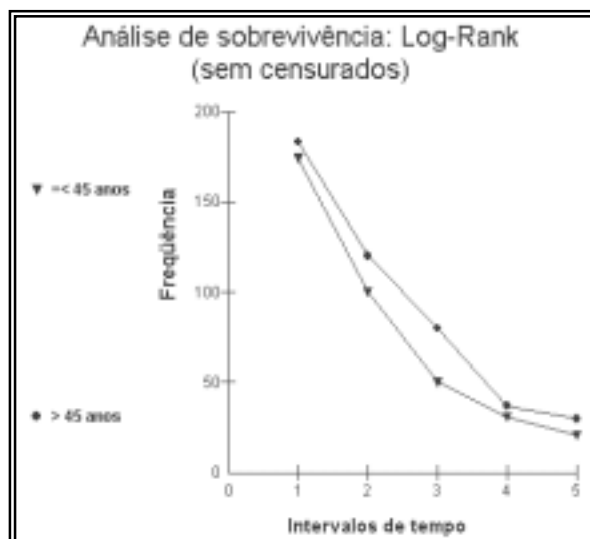
Arquivo Editar Gráficos	
	Resultados
Observado =	99
Esperado =	90.7936
Variância =	33.7218
Qui-Quadrado =	1.7611
Graus de liberdade =	1
(p) =	0.1845

A diferença não foi significativa. ($p = 0.1845$).

4.6.5 – GRÁFICO DO EXEMPLO 4.6.1

Pressionar em Gráfico (Figura 4.18) para visualizar a seguinte imagem:

Figura 4.21
Gráfico do
Exemplo
4.6.1.

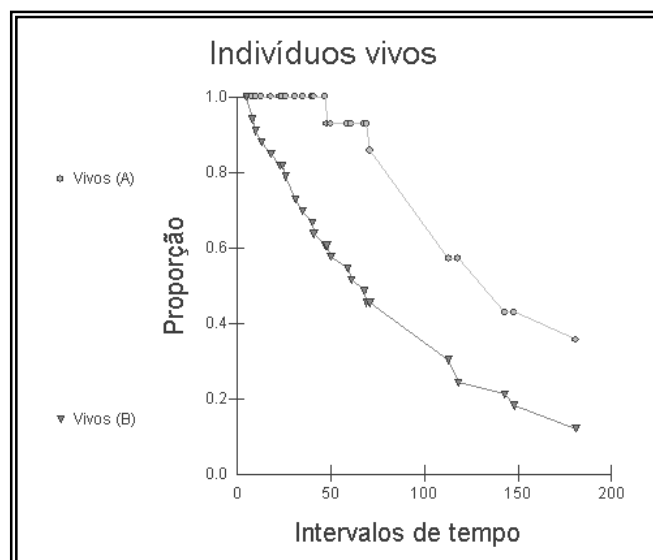


O gráfico desta análise de sobrevivência é compatível com o resultado do teste, o qual não mostrou diferença significativa entre os dois tratamentos. As curvas são muito semelhantes.

4.6.6 – GRÁFICO DO EXEMPLO 4.6.2

Pressionar em Gráfico na barra da Fig. 4.19, para visualizar a seguinte imagem:

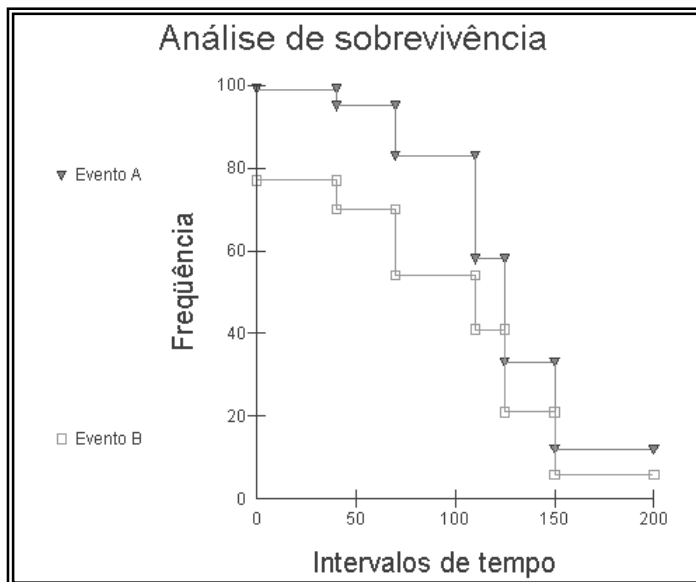
Figura 4.22
Gráfico do
Exemplo
4.6.2.



4.6.7 – GRÁFICO DO EXEMPLO 4.6.3

Pressionar em Gráfico (Figura 4.20) para visualizar a seguinte imagem:

Figura 4.23
Gráfico do
Exemplo
4.6.3.



4.7 – GEHAN (WILCOXON GENERALIZADO)

Teste estatístico semelhante ao anterior para duas amostras independentes com observações censuradas.

4.7.1 – EXEMPLO (hipotético)

Dois grupos de seis (6) mulheres com câncer mamário foram tratados do seguinte modo: o primeiro grupo recebeu quimioterapia e radioterapia, o outro nada recebeu após mastectomia total. Ao final de três (3) anos, o tempo de remissão foi assinalado em meses.

As observações **censuradas** estão assinaladas no *grid* específico com o sinal +.

$$H_0: S_1(t) = S_2(t);$$

$$H_1: S_1(t) > S_2(t);$$

Nível de decisão: alfa = 0.05 (unilateral)

4.7.2 – INSTRUÇÕES

- Pressionar o mouse em *Gehan (Wilcoxon generalizado)*;
- Inserir os valores de cada grupo, assinalando com + os escores censurados;
- Clicar em *OK* para obter os resultados, conforme disposto na Figura 4.24.

Figura 4.24
Resultados
do Exemplo
4.7.1.

Grupo A	Grupo B
22	14
18+	16
20	18
21+	19
24+	21
18	15

Nome do arquivo
Exemplo_4_7_1.Gehan

W = 27.0000 p (unilateral) 0.0085
 Var = 128.1818 p (bilateral) 0.0171
 Z = 2.3848

Cancelar Executar

A diferença entre os tratamentos é muito significativa ($p = 0.0085$), rejeitando-se a hipótese de nulidade, constatando-se que o tempo de sobrevivência foi maior no grupo tratado com quimioterapia e radioterapia.

4.8 – MANTEL-HAENSZEL

O teste de *Mantel-Haenszel* se aplica para duas ou mais amostras cujos dados são dispostos em n tabelas de contingência 2×2 .

4.8.1 – EXEMPLO

Efetuiu-se levantamento de duas investigações (Amostras A e B) em 865 mulheres com idades compreendidas entre 40 e 50 anos, relativo à associação entre neoplasia cervical e positividade para HPV.

Amostra A: 490 mulheres		
	Neoplasia cervical	Sem neoplasia
HPV+	152	120
HPV -	108	110

Amostra B: 375 mulheres		
	Neoplasia cervical	Sem neoplasia
HPV+	143	111
HPV -	45	76

Essas tabelas foram introduzidas nas *caixas de texto* do teste de *Mantel-Haenszel*, como serão vistas a seguir.

H_0 : não há associação entre neoplasia cervical e HPV positivo: $p_0 = p_1$;

H_1 : há associação entre neoplasia cervical e HPV positivo: $p_0 \neq p_1$;

Nível de decisão: alfa = 0.01.

4.8.2 – INSTRUÇÕES

- Pressionar o mouse em *Mantel-Haenszel*;
- Preencher as *caixas de texto* com os dados da Amostra A;
- Clicar em *Nova Tabela* e introduzir os dados da Amostra B;
- Clicar em *Executar*, para obter os resultados.

Figura 4.25
Resultados
do Exemplo
4.8.1.

O teste foi muito significativo ($p = 0.0013$) rejeitando-se a hipótese de nulidade, demonstrando que a neoplasia cervical está associada à presença de HPV. O *Odds Ratio* é calculado conjuntamente com o teste de *Mantel-Haenszel*, concluindo-se que as mulheres com teste positivo para HPV têm mais de uma vez e meia a probabilidade de ter câncer do colo uterino do que aquelas com teste negativo.

4.9 – PIRÂMIDE POPULACIONAL

A maioria dos *gráficos estatísticos* refere-se às amostras. Pode-se, contudo, representar Populações através de *Pirâmides*, constituídas por grupos etários, proporções sexuais, locais de residências, estados civis e outras variáveis.

4.9.1 – EXEMPLO 1

A população do Estado do Pará (IBGE) no ano 2000 foi tabulada em relação aos grupos etários e respectivos sexos.

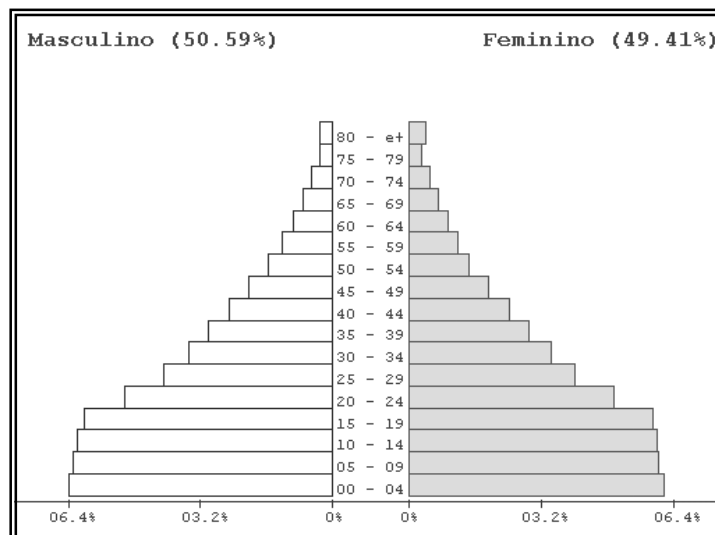
Figura 4.26
Visão parcial
dos dados
do
Exemplo
4.9.1.

	- 1 -	- 2 -	- 3 -	- 4 -
	Limite inferior	Limite superior	Masculino	Feminino
1	0	4	396051	381704
2	5	9	389664	374270
3	10	14	382467	372500
4	15	19	371238	365028
5	20	24	312171	306983
6	25	29	252511	249916
7	30	34	214968	212604
8	35	39	185800	180571
9	40	44	155518	150125
10	45	49	125771	119145
11	50	54	96586	90688
12	55	59	75341	74125

4.9.2 – INSTRUÇÕES

- Inserir no *grid* as informações obtidas do IBGE;
- Pressionar em *Gráficos* e, a seguir, em *Pirâmide Populacional*;
- Clicar em *Executar* para Resultados.

Figura 4.27
Pirâmide
populacional
do
Exemplo
4.9.1.



Há ligeiro predomínio de indivíduos do sexo masculino nos primeiros anos de vida e do sexo feminino após os 80 anos de vida.

4.9.3 – EXEMPLO 2

A população do Estado do Pará (IBGE) no ano 1970 foi tabulada em relação aos grupos etários e locais de residência, urbana e rural.

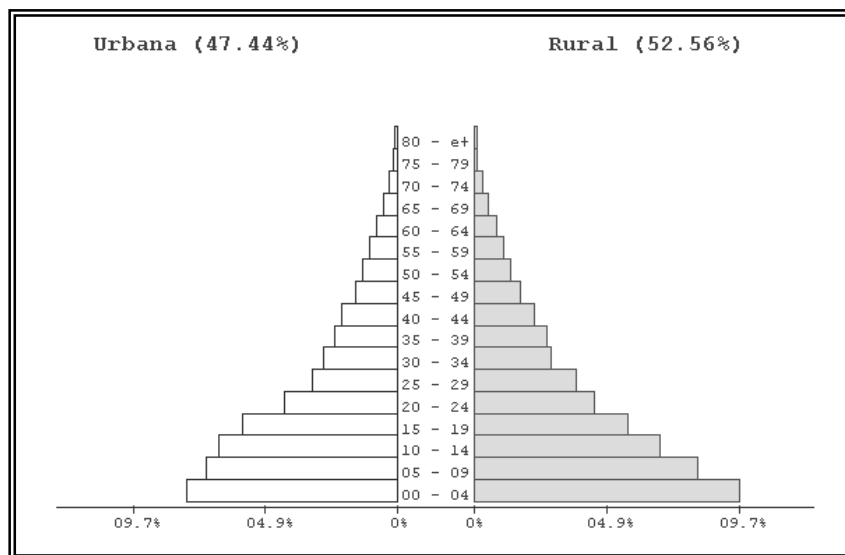
Figura 4.28
Visão parcial
dos dados
do
Exemplo
4.9.3.

- 1 -	- 2 -	- 3 -	- 4 -	
Limite inferior	Limite superior	Urbana	Rural	
0	4	167347	210151	
5	9	151718	177549	
10	14	141683	147341	
15	19	123151	122358	
20	24	90239	95795	
25	29	68277	81272	
30	34	58562	60988	
35	39	50429	57823	
40	44	44196	48076	
45	49	33607	36454	
50	54	27893	29490	
55	59	22580	23304	
60	64	16855	18527	
65	69	11765	11669	
70	74	7274	7064	

4.9.4 – INSTRUÇÕES

- Inserir no *grid* as informações obtidas do IBGE;
- Pressionar em *Gráficos* e, a seguir, em *Pirâmide Populacional*;
- Clicar em *Executar* para Resultados.

Figura 4.29
Pirâmide
populacional
do
Exemplo
4.9.3.



Há ligeira predominância na proporção de residentes na área rural (52.56%) em relação aos da área urbana (47.44%), no ano de 1970.

4.9.5 – EXEMPLO 3

A população do Estado do Pará (IBGE) no ano 2000 foi tabulada em relação aos grupos etários e locais de residência, urbana e rural.

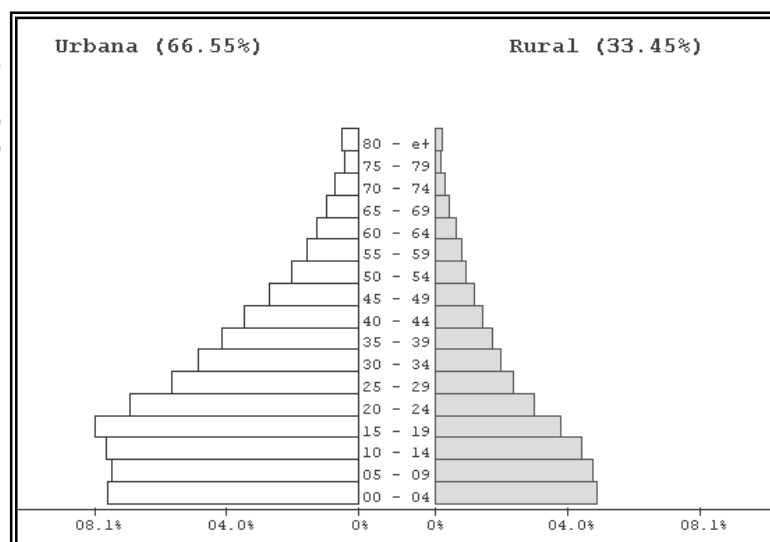
Figura 4.30
Visão parcial
dos dados
do
Exemplo
4.9.5.

	- 1 - Limite inferior	- 2 - Limite superior	- 3 - Urbana	- 4 - Rural
1	0	4	473137	304618
2	5	9	465551	298383
3	10	14	476952	278015
4	15	19	498808	237458
5	20	24	432249	186905
6	25	29	363224	149203
7	30	34	302864	124708
8	35	39	258785	107586
9	40	44	216308	89335
10	45	49	170030	74886
11	50	54	128083	59191
12	55	59	98381	51085
13	60	64	79479	39330
14	65	69	61272	28108
15	70	74	44728	18455

4.9.6 – INSTRUÇÕES

- Inserir no *grid* as informações obtidas do IBGE;
- Pressionar em *Gráficos* e, a seguir, em *Pirâmide Populacional*;
- Clicar em *Executar* para obter os resultados.

Figura 4.31
Pirâmide
populacional
do
Exemplo
4.9.5.



Há predominância de moradores na área urbana (66.55%) quando comparados com os da área rural, denotando uma migração para as cidades em todo o Estado do Pará.

4.9.7 – EXEMPLO 4

A população do Estado do Pará foi tabulada em relação aos grupos etários, sexos e estado civil (solteiros e outros). Os dados são hipotéticos, apenas para demonstrar a possibilidade da introdução de três (3) variáveis.

Figura 4.32
Visão parcial
dos dados
do
Exemplo
4.9.7.

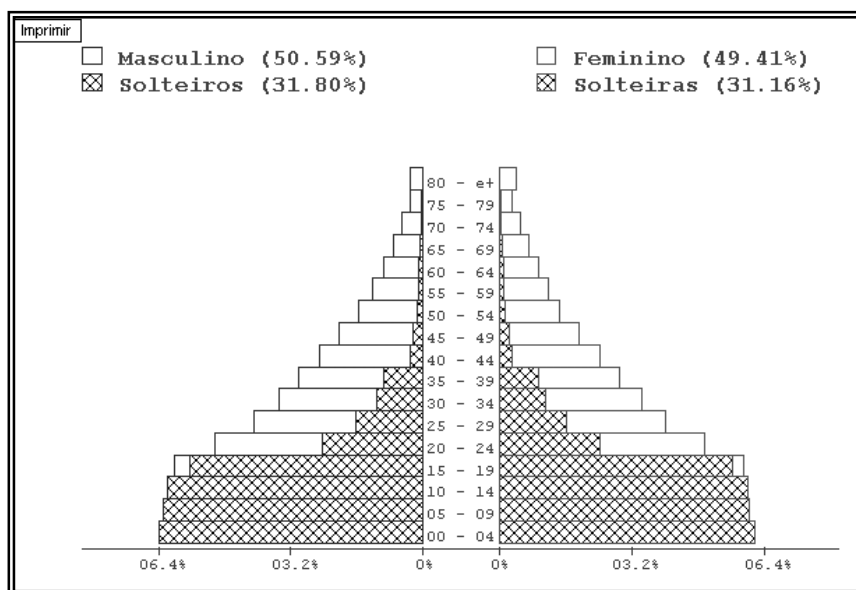
- 1 - Limite inferior	- 2 - Limite superior	- 3 - Masculino	- 4 - Solteiros	- 5 - Feminino	- 6 - Solteiras
0	4	396051	396051	381704	381704
5	9	389664	389664	374270	374270
10	14	382467	382467	372500	372500
15	19	371238	350000	365028	350000
20	24	312171	150000	306983	150000
25	29	252511	100000	249916	100000
30	34	214968	70000	212604	70000
35	39	185800	60000	180571	60000
40	44	155518	20000	150125	20000
45	49	125771	15000	119145	15000
50	54	96586	10000	90688	10000

4.9.8 – INSTRUÇÕES

- Inserir no *grid* os dados hipotéticos;
- Pressionar em *Gráficos* e, a seguir, em *Pirâmide Populacional*;

c) Clicar em *Executar* para Resultados.

Figura 4.33
Pirâmide
populacional
do
Exemplo
4.9.7.



Esta pirâmide representa três variáveis: faixa etária, sexo e estado civil (solteiros e não solteiros).

CAPÍTULO 5

ANÁLISE DE VARIÂNCIA

5.1 – INTRODUÇÃO

A Análise de Variância, introduzida por *R. A. Fisher*, compara a magnitude das variações de mais de duas amostras, decompondo a variância total em duas partes:

- entre as amostras, constituindo o chamado quadrado médio dos tratamentos;
- dentre cada tratamento, compondo o denominado quadrado médio do erro experimental.
- O teste da análise da variância é a *razão* entre esses quadrados médios. Esta análise compreende dois grupos:

a) *Testes paramétricos:*

- ANOVA: um critério;
- ANOVA: dois critérios;
- ANOVA: Fatorial (a x b) com replicação;
- ANOVA: Fatorial (a x b x c) sem replicação.

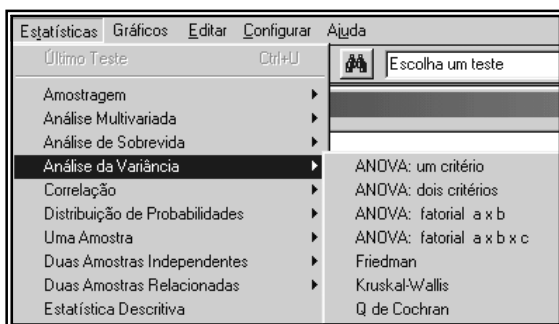
b) *Testes não-paramétricos:*

- Teste de *Friedman*;
- Teste de *Kruskal-Wallis*;
- Teste Q de *Cochran*.

5.1.1 – INSTRUÇÕES

- Introduzir os dados no *grid* geral do *BioEstat*;
- Dirigir o *mouse* para o menu superior e clicar em *Estatística*;
- Apontar o *mouse* em *Análise de Variância*, conforme demonstra a Figura 5.1:

Figura 5.1
Menu de
Análise de
Variância.



5.2 – ANOVA: um critério

A designação ANOVA deriva da expressão inglesa *ANalysis Of VAriance*, chamando-se *F-teste* em homenagem a *Fisher*. Destina-se a comparar mais de duas amostras cujos dados devem ser mensurados em escala intervalar ou de razões. A designação *um critério* é pelo fato de se comparar somente as variações entre os

tratamentos, cujo resultado é traduzido no valor do *F*-teste, complementando-se com o exame, *a priori* (Bonferroni) ou *a posteriori* (Tukey ou teste *t* de Student), das diferenças entre as médias amostrais. As amostras podem ser do mesmo tamanho ou desiguais.

Quando o valor de *F* for significativo e a escolha do teste das diferenças entre as médias for o Bonferroni, deve-se escolher previamente o nível alfa e, neste caso, o aplicativo permite selecionar: 0.1, 0.05, 0.025, 0.01, 0.005, 0.001 ou 0.0005, já indicados como testes bilaterais ($\alpha/2$). Assim, se a escolha for igual a 0.025, significa 2×0.025 .

5.2.1 – EXEMPLO 1 (TUKEY)

Foi efetuada uma investigação em três grupos de estudantes: o grupo A estava constituído por seis (6) alunos não-fumantes (NF); o segundo, por seis (6) discentes que fumavam moderadamente em torno de dez (10) a quinze (15) cigarros por dia (FM); e o terceiro, por cinco (5) estudantes que fumavam mais de 40 cigarros por dia (FI). Mediu-se a função pulmonar através do *fluxo médio expiratório*. Os dados obtidos estão no *grid* geral.

Figura 5.2
Dados do
Exemplo
5.2.1.

	- 1 - NF	- 2 - FM	- 3 - FI
1	3.900	2.870	2.850
2	3.850	2.720	2.690
3	4.000	2.610	2.650
4	3.760	2.650	2.840
5	3.810	2.700	2.750
6	3.770	2.600	

H_0 : o fumo não influencia a função pulmonar medida pelo *fluxo médio expiratório*:

$$\mu_1 = \mu_2 = \mu_3;$$

H_1 : o fumo influencia a função pulmonar medida pelo *fluxo médio expiratório*, havendo diferença, pelo menos, entre duas médias: $\mu_1 \neq \mu_2$ etc.;

Nível de decisão: alfa = 0.01.

5.2.2 – EXEMPLO 2 (DUNNETT)

Três extratos de origem vegetal foram introduzidos em cães por via oral com a finalidade de testar o possível efeito sobre a pressão arterial sistólica desses animais. Os cães foram divididos em grupos de cinco animais, recebendo cada grupo um tipo de extrato – B, C ou D –, além de um grupo controle – A – injetado com placebo. Os dados estão contidos no *grid* geral.

H_0 : a pressão arterial sistólica não é alterada pela introdução oral dos extratos vegetais referidos no experimento: $\mu_1 = \mu_2 = \mu_3 = \mu_4$;

H_1 : a administração desses extratos produz alterações na pressão arterial sistólica dos animais, havendo diferença entre a média do grupo controle com um dos demais grupos submetidos ao experimento: $\mu_1 \neq \mu_i$;

Nível de decisão: alfa = 0.05.

Figura 5.3
Dados do
Exemplo
5.2.2.

	- 1 - Controle A	- 2 - Extrato B	- 3 - Extrato C	- 4 - Extrato D
1	74.000	99.000	100.000	78.000
2	71.000	91.000	95.000	74.000
3	73.000	94.000	97.000	75.000
4	79.000	101.000	99.000	86.000
5	68.000	97.000	98.000	72.000

5.2.3 – INSTRUÇÕES

- Pressionar o mouse em *ANOVA: um critério*;
- Selecionar as colunas referentes aos tratamentos do *grid* geral;
- Clicar em *Executar o Teste*, surgindo a Figura 5.4 naqueles casos em que há diferença estatisticamente significativa do *F* calculado (ver *p-valor*), selecionando-se um dos testes sugeridos para a comparação das médias amostrais, os quais foram o de *Tukey* e o de *Dunnett*. Exemplos 5.2.1 e 5.2.2, respectivamente.

Figura 5.4
Tela de
seleção do
teste para
ANOVA: um
critério,
Exemplo
5.2.1.

ANOVA: um critério

F (tratamentos) = 283.5906
p-valor = 0.0000003

Escolha o teste

☒ Tukey
☐ Teste t (LSD)
☐ Bonferroni
☐ Dunnett

Executar
Cancelar

- Clicar em *Executar* para obter os resultados.

Figura 5.5
Resultados
do Exemplo
5.2.1
(*Tukey*).

Arquivo Editar Gráfico			
FONTES DE VARIAÇÃO	GL	SQ	QM
Tratamentos	2	4.946	2.473
Erro	14	0.122	0.009
F =	283.5906		
(p) =	< 0.0001		
Média (Coluna 1) =	3.8483		
Média (Coluna 2) =	2.6917		
Média (Coluna 3) =	2.7560		
Tukey:	Diferença	Q	(p)
Médias (1 a 2) =	1.1567	30.3407	< 0.01
Médias (1 a 3) =	1.0923	27.3197	< 0.01
Médias (2 a 3) =	0.0643	1.6090	ns

Os valores obtidos rejeitam a hipótese de nulidade, aceitando-se, nesse exemplo, a hipótese alternativa. O valor alfa escolhido *a priori* informa que a probabilidade obtida na comparação entre as médias pelo teste de Tukey deve ser igual ou inferior a 0.01, para que a diferença seja estatisticamente significativa. O *fluxo médio expiratório* é menor entre *fumantes* (FM e FI), quando comparados com os não fumantes (NF), não ocorrendo, contudo, diferença significativa entre os fumantes.

Figura 5.6
Tela de seleção do teste para ANOVA: um critério, Exemplo 5.2.2.

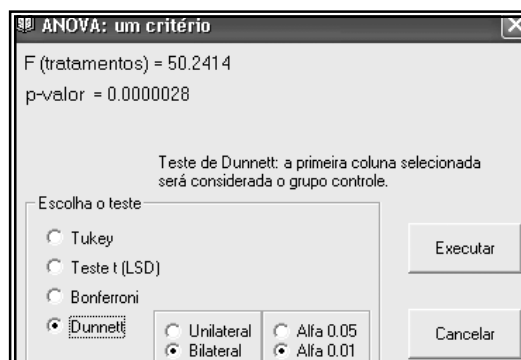


Figura 5.7
Resultados do Exemplo 5.2.2 (Dunnett).

Arquivo Editar Gráfico						
FONTES DE VARIAÇÃO		GL	SQ	QM		
Tratamentos		3	24.9 e+02	828.983		
Erro		16	264.000	16.500		
F =		50.2414				
(p) =		< 0.0001				
Média (Coluna 1) =		73.0000				
Média (Coluna 2) =		96.4000				
Média (Coluna 3) =		97.8000				
Média (Coluna 4) =		77.0000				
Dunnett: Diferença			SE	Q	V Crítico	Conclusão
Colunas 1 e 2		23.4000	2.5690	9.1084	3.2200	p < 0.01
Colunas 1 e 3		24.8000	2.5690	9.6534	3.2200	p < 0.01
Colunas 1 e 4		4.0000	2.5690	1.5570	3.2200	ns

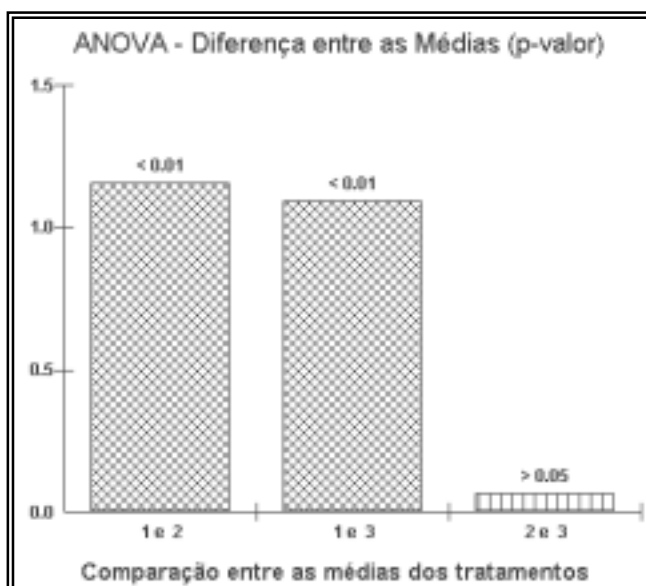
O *F* obtido é altamente significativo ($p < 0.0001$). Pelo teste de *Dunnett* verifica-se que os extratos B (Coluna 2) e C (Coluna 3) quando comparados com o Controle (Coluna 1) aumentaram de maneira significativa a pressão arterial sistólica dos animais submetidos ao experimento ($p < 0.01$), enquanto o extrato D (Coluna 4) não produziu alterações significativas na pressão dos animais investigados (*ns*).

Para se obter os gráficos desses resultados, seguir as instruções abaixo:

5.2.4 – GRÁFICO DO EXEMPLO 5.2.1

- Pressionar, com o *mouse*, em *Gráfico* (barra da Figura 5.5) surgindo a seguinte imagem:

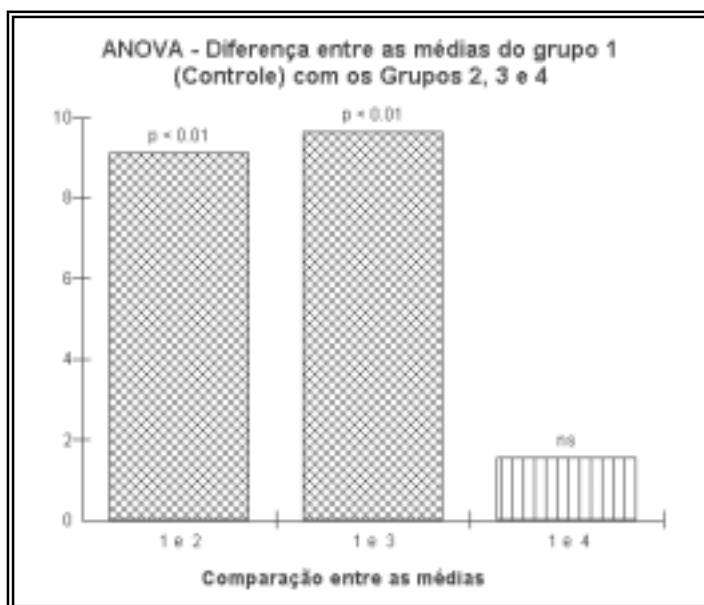
Figura 5.8
Gráfico do
Exemplo
5.2.1.



5.2.5 – GRÁFICO DO EXEMPLO 5.2.2

- b) Pressionar, com o *mouse*, em *Gráfico* (barra da Figura 5.7) surgindo a seguinte imagem:

Figura 5.9
Gráfico do
Exemplo
5.2.2.



5.3 – ANOVA: dois critérios

Este teste compara, além da variação entre os tratamentos, a variabilidade entre os blocos, os quais correspondem às linhas do *grid* geral. As amostras devem ter o mesmo tamanho e os escores mensurados a nível intervalar ou de razões. Há, neste modelo, dois valores de F : dos tratamentos e dos blocos.

5.3.1 – EXEMPLO

Em uma escola do ensino médio efetuou-se levantamento sobre o número de alunos reprovados nas disciplinas Ciências, Matemática e Português (tratamentos), ao mesmo tempo em que se procurou verificar a proporção dos inabilitados nessas matérias nos turnos da Manhã (linha 1), da Tarde (linha 2) e da Noite (linha 3). Em cada turno estavam matriculados 800 discentes, de onde foram retiradas as amostras. Os dados coletados estão inseridos no *grid* geral de entrada de dados.

Figura 5.10
Dados do
Exemplo
5.3.1.

	- 1 - **	- 2 - Ciências	- 3 - Matemática	- 4 - Português	-
1	Manhã	31.000	32.000	50.000	
2	Tarde	84.000	78.000	104.000	
3	Noite	86.000	38.000	112.000	

H_0 : o tipo de disciplina cursada não afeta os índices de reprovação;

H_1 : pelo menos duas médias são diferentes;

H_0 : o turno cursado pelo estudante não afeta os índices de reprovação;

H_1 : pelo menos duas médias são diferentes;

Nível de decisão: alfa = 0.05.

5.3.2 – INSTRUÇÕES

- Pressionar o *mouse* em *ANOVA: dois critérios*;
- Selecionar as colunas referentes aos tratamentos (Ciências, Matemática, Português) do *grid* geral;
- Clicar em *Executar o Teste*, surgindo a Figura 5.11 naqueles casos em que há diferença estatisticamente significativa entre médias dos tratamentos e/ou blocos – p -valor, selecionando-se um dos testes sugeridos para a comparação dos valores médios amostrais, o qual, neste exemplo, foi escolhido o t de *Student* (Blocos).
- Pressionar *Executar*, para resultados.

Figura 5.11
Tela de
seleção do
teste
a posteriori
para ANOVA:
dois critérios.

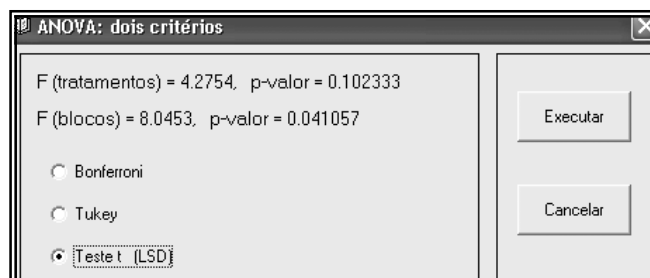


Figura 5.12
Resultados
do Exemplo
5.3.1.

FONTES DE VARIAÇÃO	GL	SQ	QM
Tratamentos	2	2328.667	1164.333
Blocos	2	4382.000	2191.000
Erro	4	1089.333	272.333
F (tratamentos) =	4.2754	---	---
p (tratamentos) =	0.1023	---	---
F (blocos) =	8.0453	---	---
p (blocos) =	0.0411	---	---
Médias (blocos):			
Média (Linha 1) =	37.6667	---	---
Média (Linha 2) =	88.6667	---	---
Média (Linha 3) =	78.6667	---	---
Teste t (LSD)	t	(p)	
Médias (1 e 2) =	3.785	0.0193	---
Médias (1 e 3) =	3.043	0.0383	---
Médias (2 e 3) =	0.742	0.4992	---

O *F*-teste é significativo entre blocos ($F = 8.0453$, $p = 0.0411$), o que não ocorreu com os tratamentos ($F = 4.2754$, $p = 0.1023$). A comparação entre as médias dos blocos exibe diferenças significativas entre os Blocos 1 vs. 2 e 1 vs. 3, o que não foi observado entre os Blocos 2 vs. 3. Pode-se concluir, assim, que as turmas da manhã referentes às três disciplinas apresentaram proporções menores de alunos reprovados, rejeitando-se, conseqüentemente, a hipótese de nulidade entre os blocos (turnos), mas aceitando-se no que diz respeito aos tratamentos (disciplinas).

5.4 – ANOVA: Fatorial a x b (com replicação)

Algumas vezes o estudo procura esclarecer se há interações entre os tratamentos e os blocos, devendo-se, nessa circunstância, efetuar repetições para visualizar esse detalhe da pesquisa. Os resultados desta análise de variância são apresentados sob a forma de três (3) *F*-testes: a) dos tratamentos; b) dos blocos; c) da interação entre tratamentos e blocos, todos com os respectivos valores de p .

5.4.1 – EXEMPLO

Estudo experimental procurou verificar a influência das drogas A, B e C no *peso* de *frangos* e *frangas*, ou seja, procurou-se determinar se, além da *ação ponderal*, haveria influência relacionada com o *sexo* e, ainda, a possibilidade de *interação* entre peso e sexo dos animais determinada pelas drogas, com as seguintes hipóteses:

H_0 : as drogas não afetam o peso médio das aves em crescimento: $\mu_1 = \mu_2 = \mu_3$;

H_1 : pelo menos duas médias são diferentes;

H_0 : a condição sexual não afeta o peso médio das aves em crescimento: $\mu_m = \mu_f$;

H_1 : a condição sexual afeta o peso médio das aves em crescimento: $\mu_m \neq \mu_f$;

H_0 : não há interação significativa entre as variáveis independentes: droga e sexo;

H_1 : há interação significativa entre as variáveis independentes;

Nível de decisão: $\alpha = 0.05$.

5.4.2 – INSTRUÇÕES

- Pressionar o mouse em *ANOVA: Fatorial a x b*;
- Preencher as *caixas de textos* com o número de tratamentos, de blocos e de repetições, e clicar em *Ativar o Grid*.
- Preencher o *grid* específico (Figura 5.13) de acordo com os escores obtidos no experimento, correspondendo cada valor ao peso em gramas de cada animal;
- Clicar em *Executar o teste*, para resultados.

Figura 5.13
Dados do
Exemplo
5.4.1.

Novo Deletar Salvar Imprimir			
Exemplo_5_4_1.Anova2	No. de tratamentos	No. de blocos	No. de repetições
	3	2	4
	Trat.: 1 Bloco.: A	Trat.: 2 Bloco.: A	Trat.: 3 Bloco.: A
	649	587	601
	641	562	572
	632	497	518
	679	498	540
	Trat.: 1 Bloco.: B	Trat.: 2 Bloco.: B	Trat.: 3 Bloco.: B
	681	569	522
	692	560	553
701	521	590	
719	530	580	
Nome do arquivo			
Exemplo_5_4_1.Anova2			
Executar Teste		Cancelar	

Figura 5.14
Resultados
do Exemplo
5.4.1.

FONTES DE VARIAÇÃO	GL	SQ	QM
Tratamentos	2	83780.3333	41890.1667
Blocos	1	2440.1667	2440.1667
Interação	2	2354.3333	1177.1667
Erro	18	16621.0000	923.3889
F (Tratamentos) =	45.3657	---	---
Graus de liberdade =	2, 18	---	---
p (Tratamentos) =	< 0.0001	---	---
F (Blocos) =	2.6426	---	---
Graus de liberdade =	1, 18	---	---
p (Blocos) =	0.1181	---	---
F (Interação) =	1.2748	---	---
Graus de liberdade =	2, 18	---	---
p (Interação) =	0.3035	---	---

O *F-teste* dos tratamentos foi elevado, com *p-valor* < 0.0001, indicando que o efeito das drogas sobre o peso é muito significativo, rejeitando-se sob esse aspecto a hipótese de nulidade e aceitando-se a hipótese alternativa. No que diz respeito à variável *sexo* e à interação entre *drogas* e *sexo* conclui-se, pelos resultados dos respectivos valores de *F*, que as diferenças não são estatisticamente significativas, aceitando-se, assim, a hipótese de nulidade (ver Figura 5.14).

5.5 – ANOVA: Fatorial a x b x c (sem replicação)

Os experimentos fatoriais, como já foi visto no exemplo anterior, incluem a combinação de vários fatores ou conjuntos de tratamentos com as respectivas interações. Neste caso, trata-se dos fatores *a*, *b* e *c*, testados de modo isolado e combinados, como segue:

- i – Fator *a*;
- ii – Fator *b*;
- iii – Fator *c*;
- iv – Fator *a* × Fator *b*;
- v – Fator *a* × Fator *c*;
- vi – Fator *b* × Fator *c*;
- vii – Fator *a* × Fator *b* × Fator *c*.

5.5.1 – EXEMPLO (hipotético)

Procurou-se determinar a acidez de laranjas de 3 variedades, em 3 temperaturas diferentes e em 3 períodos anuais. Os dados estão incluídos no *grid* especial do *BioEstat*, como pode ser visualizado na Figura 5.15, sendo *A* as variedades da espécie, *B* as temperaturas e *C* os períodos anuais.

- H_0 : o teor de acidez é o mesmo nas 3 variedades estudadas;
 - H_1 : o teor de acidez é diferente nas variedades investigadas;
 - H_0 : a temperatura não altera o teor de acidez das variedades estudadas;
 - H_1 : a temperatura altera o teor de acidez das variedades estudadas;
 - H_0 : o período anual não altera o teor de acidez das variedades investigadas;
 - H_1 : o período anual altera o teor de acidez das variedades estudadas;
 - H_0 : não há interação significativa entre as variáveis independentes;
 - H_1 : há interação significativa entre as variáveis independentes.
- Nível de decisão: $\alpha = 0.05$.

Figura 5.15
Dados do
Exemplo
5.5.1.

Novo Deletar Salvar Imprimir

Exemplo_5_5_1.Anova3

Fator A: 3 Fator B: 3 Fator C: 3

	Fator A: 1 Fator B: 1	Fator A: 2 Fator B: 1	Fator A: 3 Fator B: 1
Fator C: 1	30.3	26.5	15.8
Fator C: 2	31.5	27.9	14.3
Fator C: 3	35.6	23.1	12.7
	Fator A: 1 Fator B: 2	Fator A: 2 Fator B: 2	Fator A: 3 Fator B: 2
Fator C: 1	33.4	23.1	13.5
Fator C: 2	34.7	20.6	12.7
Fator C: 3	32.5	22.2	11.1
	Fator A: 1 Fator B: 3	Fator A: 2 Fator B: 3	Fator A: 3 Fator B: 3
Fator C: 1	36.1	18.9	16.2
Fator C: 2	37.8	19.6	15.7
Fator C: 3	35.2	19.4	14.5

Nome do arquivo: Exemplo_5_5_1.Anova3

Executar Teste Cancelar

5.5.2 – INSTRUÇÕES

- Pressionar o *mouse* em ANOVA: Fatorial $a \times b \times c$;
- Preencher as *caixas de textos* com o número de cada fator: variedades, meses do ano e número de anos, e clicar em *Ativar o Grid*.
- Preencher o *grid* específico (Figura 5.15) de acordo com os escores obtidos no experimento;
- Clicar em *Executar teste*, para resultados.

Figura 5.16
Resultados
do Exemplo
5.5.1.

FONTE DE VARIAÇÃO	GL	SQ	QM	F	(p)
Fator A	2	1829.8163	914.9081	---	---
Fator B	2	11.2541	5.6270	---	---
Fator C	2	4.7963	2.3981	---	---
Interação A x B	4	91.8704	22.9676	6.2092	0.0146
Interação A x C	4	10.2681	2.5670	0.6940	0.6183
Interação B x C	4	1.6970	0.4243	0.1147	0.9711
Erro (A x B x C)	8	29.5919	3.6990	---	---

Observa-se pelos resultados contidos na Figura 5.16 que a interação dos Fatores *a* e *b*, variedades vs. temperaturas, respectivamente, é significativa, com *p*-valor da ordem de 0.0146.

5.6 – TESTE DE FRIEDMAN

A *Análise de Variância de Friedman* é um teste não-paramétrico para dados mensurados a nível *ordinal*, abrangendo três ou mais amostras e equivalendo à ANOVA com dois critérios. Este procedimento recebe, também, a designação de *Dupla Análise de Variância por Postos*. Os dados devem ter dupla disposição: as *linhas* representam os indivíduos e as *colunas*, as condições experimentais ou tratamentos. As amostras (tratamentos) devem ter o mesmo tamanho, a fim de permitir a comparação entre todos os indivíduos submetidos ao teste.

5.6.1 – EXEMPLO

Em um experimento, cinco coelhos foram submetidos a quatro estímulos de intensidade crescente: $A < B < C < D$. As respostas de cada animal variaram em uma escala ordinal de 0 a 100 e os escores foram introduzidos no *grid* geral do programa.

Figura 5.17
Dados do
Exemplo
5.6.1.

Estímulo A	Estímulo B	Estímulo C	Estímulo D
20.000	29.000	64.000	98.000
20.000	32.000	57.000	96.000
32.000	30.000	70.000	81.000
33.000	25.000	77.000	77.000
25.000	28.000	68.000	80.000

H_0 : a intensidade dos estímulos não influencia a magnitude de resposta dos animais submetidos ao experimento: $\theta_1 = \theta_2 = \theta_3$;

H_1 : a intensidade dos estímulos influencia a magnitude de resposta dos animais submetidos ao experimento: $\theta_1 \neq \theta_2 \neq \theta_3$;

Nível de decisão: $\alpha = 0.05$.

5.6.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Friedman*;
- Selecionar as colunas referentes aos tratamentos do *grid* geral;
- Clicar em *Executar o Teste*, para resultados:

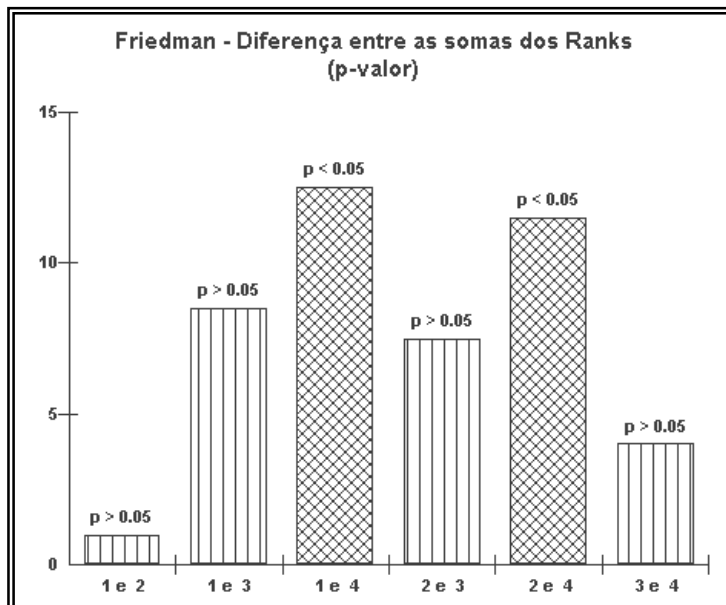
Figura 5.18
Resultados
do Exemplo
5.6.1.

	- 1 -	- 2 -	- 3 -	- 4 -
Soma dos Ranks =	7.0000	8.0000	15.5000	19.5000
Mediana =	25.0000	29.0000	68.0000	81.0000
Média dos Ranks =	1.4000	1.6000	3.1000	3.9000
Média dos valores =	26.0000	28.8000	67.2000	86.4000
Desvio padrão =	6.2849	2.5884	7.3959	9.8133
Friedman (Fr) =	13.0200			
Graus de liberdade =	3			
(p) =	0.0046			
Comparações:	Diferença	(p)		
Ranks 1 e 2 =	1	ns		
Ranks 1 e 3 =	8.5	ns		
Ranks 1 e 4 =	12.5	< 0.05		
Ranks 2 e 3 =	7.5	ns		
Ranks 2 e 4 =	11.5	< 0.05		
Ranks 3 e 4 =	4	ns		

5.6.3 –GRÁFICO DO EXEMPLO 5.6.1

- Pressionar, com o *mouse*, em *Gráfico* (barra da Figura 5.18):

Figura 5.19
Gráfico do
Exemplo
5.6.1.



O teste de Friedman ($Fr^2 = 13.0200$) é muito significativo ($p = 0.0046$), devendo-se, portanto, rejeitar a hipótese de nulidade e aceitar a hipótese alternativa. Com base nesse resultado e na análise dos dados, pode-se concluir que a intensidade da resposta dos animais é diretamente proporcional à intensidade do estímulo provocado. A comparação entre os tratamentos exibe diferença significativa entre os Estímulos A vs. D e B vs. D.

5.7 – TESTE DE KRUSKAL-WALLIS

Teste não-paramétrico, conhecido como *Teste H*, destina-se a *comparar três ou mais amostras independentes do mesmo tamanho ou desiguais*, cujos escores devem ser mensurados, pelo menos, a nível ordinal. O teste de *Kruskal-Wallis* é comparável à ANOVA (um critério).

5.7.1 – EXEMPLO

Consciência Ecológica entre universitários escolhidos ao acaso e pertencentes aos cursos de Biologia, Enfermagem, Odontologia e Administração. A apuração de questionário distribuído entre os estudantes indicaria que, quanto maior o escore, maior o grau de *consciência ecológica* revelado, cujos valores variavam numa escala de 0 a 100 pontos.

H_0 : os escores obtidos dos estudantes independem dos cursos nos quais estão matriculados;

H_1 : a magnitude da *consciência ecológica* é dependente do curso no qual o aluno está matriculado;

Nível de decisão: $\alpha = 0.05$.

Figura 5.20
Dados do
Exemplo
5.7.1.

	- 1 - Biologia	- 2 - Enfermagem	- 3 - Odontologia	- 4 - Administracao
1	80.000	52.000	32.000	70.000
2	76.000	56.000	24.000	52.000
3	82.000	57.000	30.000	84.000
4	75.000	49.000	38.000	63.000
5	81.000	41.000	25.000	

5.7.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Kruskal-Wallis*;
- Selecionar as colunas referentes aos tratamentos do *grid* geral;
- Se a estatística **H** obtiver *p-valor* significativo, o usuário poderá optar entre dois métodos para comparação da média dos postos: *Dunn* ou *Student-Newman-Keuls*. Neste exemplo foi escolhido o método de *Dunn*, o qual é mais conservador, pois retifica o *Nível Alfa* conforme o número de amostras:
 $Nível\ Alfa/(k(k-1))$, sendo k o número de amostras;
- Clicar em *Executar*, para obter resultados;

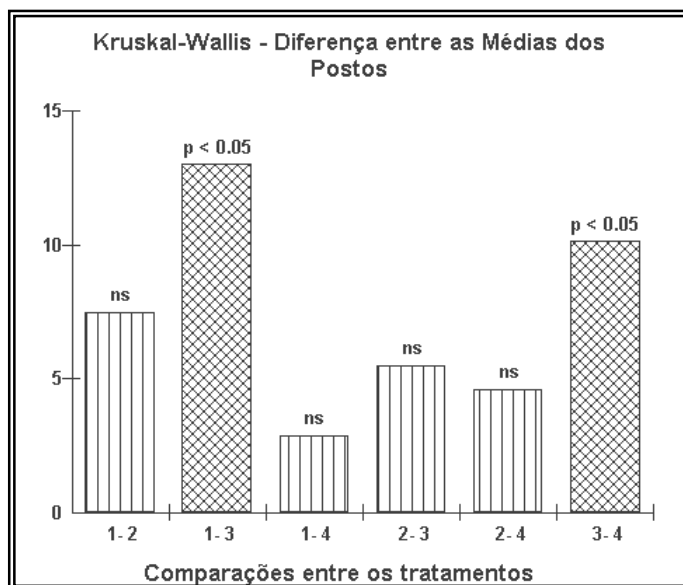
Figura 5.21
Resultados
do Exemplo
5.7.1.

Arquivo Editar Gráfico				
	Resultados			
H =	15.0230	---	---	
Graus de liberdade =	3	---	---	
(p) Kruskal-Wallis =	0.0018	---	---	
R 1 =	80.0000	---	---	
R 2 =	42.5000	---	---	
R 3 =	15.0000	---	---	
R 4 =	52.5000	---	---	
R 1 (posto médio) =	16.0000	---	---	
R 2 (posto médio) =	8.5000	---	---	
R 3 (posto médio) =	3.0000	---	---	
R 4 (posto médio) =	13.1250	---	---	
Comparações (método de Dunn)	Dif. Postos	z calculado	z crítico	p
Postos médios 1 e 2	7.5000	2.1073	2.638	ns
Postos médios 1 e 3	13.0000	3.6527	2.638	< 0.05
Postos médios 1 e 4	2.8750	0.7616	2.638	ns
Postos médios 2 e 3	5.5000	1.5454	2.638	ns
Postos médios 2 e 4	4.6250	1.2252	2.638	ns
Postos médios 3 e 4	10.1250	2.6822	2.638	< 0.05

5.7.3 –GRÁFICO DO EXEMPLO 5.7.1

e) Pressionar, com o mouse, em Gráfico (barra da Figura 5.21):

Figura 5.22
Gráfico do
Exemplo
5.7.1.



O valor de $H = 15.0230$ é altamente significativo ($p = 0.0018$). Rejeita-se a hipótese de nulidade e aceita-se a hipótese alternativa. Verifica-se, de outro lado, que as

diferenças entre os escores médios dos alunos de Biologia vs. Odontologia e Odontologia vs. Administração apresentam valores de p inferiores ao nível alfa de 0.05, sendo considerados significativos ou muito significativos, concluindo-se que os estudantes de Biologia e de Administração apresentaram maior *consciência ecológica* do que os discentes dos demais cursos investigados.

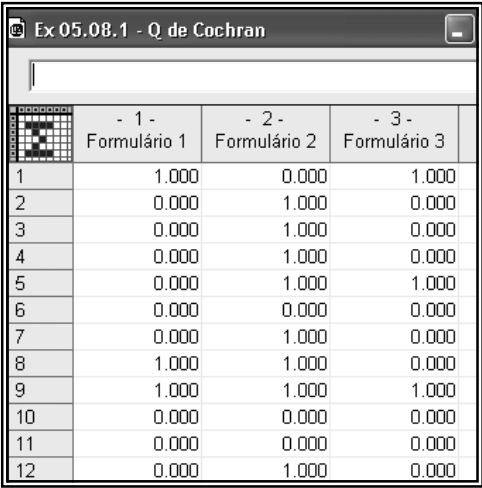
5.8 – TESTE Q DE COCHRAN

Destinado a testar várias amostras relacionadas, nas quais os mesmos indivíduos são observados em três ou mais etapas. Os escores são mensurados a nível nominal ou ordinal, e os resultados apresentam-se dicotomizados: **1** (sucesso) ou **0** (insucesso), por exemplo. Este teste não-paramétrico é considerado uma extensão do teste de *McNemar* referente a duas amostras (ver Capítulo 11), enquanto o de *Cochran* se aplica para três ou mais conjuntos do mesmo tamanho.

5.8.1 – EXEMPLO

Um entrevistador enviou formulário para doze (12) indivíduos sobre a preferência de cada um referente a um produto comercial recentemente lançado à venda, repetindo o procedimento em três oportunidades, com intervalo de dois (2) meses entre cada questionamento. As respostas foram consideradas satisfatórias (**1**) ou não (**0**), de acordo com a avaliação do pesquisador.

Figura 5.23
Dados do
Exemplo
5.8.1.



	- 1 - Formulário 1	- 2 - Formulário 2	- 3 - Formulário 3
1	1.000	0.000	1.000
2	0.000	1.000	0.000
3	0.000	1.000	0.000
4	0.000	1.000	0.000
5	0.000	1.000	1.000
6	0.000	0.000	0.000
7	0.000	1.000	0.000
8	1.000	1.000	0.000
9	1.000	1.000	1.000
10	0.000	0.000	0.000
11	0.000	0.000	0.000
12	0.000	1.000	0.000

H_0 : a probabilidade da resposta favorável (**1**) é a mesma em todas as três etapas da investigação;

H_1 : a probabilidade da resposta favorável (**1**) difere em cada etapa do questionamento;

Nível de decisão: alfa = 0.05.

5.8.2 – INSTRUÇÕES

- Pressionar o *mouse* em Q de *Cochran*;
- Selecionar as colunas referentes aos tratamentos do *grid* geral;
- Clicar em *Executar o Teste*, para resultados.

Figura 5.24
Resultados
do Exemplo
5.8.1.

	Resultados
Q =	6.2500
Graus de Liberdade =	2
(p) =	0.0439

O teste Q de *Cochran* exibe um valor estatisticamente significativo ($p = 0.0439$), devendo-se rejeitar a hipótese de nulidade e aceitar a hipótese alternativa. As respostas favoráveis diferem em cada etapa de resposta ao questionário, não havendo, portanto, uniformidade sobre o assunto objeto da pesquisa.

CAPÍTULO 6

BOOTSTRAP

6.1 – INTRODUÇÃO

A técnica empregada para este procedimento consiste em retirar de uma pequena amostra numerosas outras com reposição, por exemplo 500, 1000, 10000 ou mais. Cada uma delas tem a probabilidade de $1/n$ de ser obtida. Chama-se, também, de técnica de simulação ou de reamostragem, cuja denominação na língua inglesa é *bootstrap*, e a tradução tem sido *Tira de Bota*. Esta designação foi atribuída a *Bradley Efron*, como também a afirmação de que a distribuição relativa de repetidas amostras estatísticas é uma estimação da distribuição de amostragem, conforme publicação de seu artigo no *The Annals of Statistics*, em 1979.

Para alguns autores, é requerido, para este procedimento, que as reamostragens nunca sejam inferiores a 500, constituindo o *bootstrap* um método computacional intensivo, de análise estatística por simulação para estimativas de intervalos de confiança, erro padrão e testes de significância, aplicáveis a modelos paramétricos e não paramétricos.

Os aplicativos tipo *Bootstrap – Tira de Bota* - apresentados no *BioEstat* são:

- Dados Qualitativos: Dicotômicos;
- Dados Quantitativos: Contínuos;
- Intervalo de Confiança da Correlação Linear;
- Intervalo de Confiança da Média;
- Intervalo de Confiança da Proporção;
- Teste de hipótese: Teste *t* – Duas Amostras Independentes;
- Teste de hipótese: ANOVA (um critério);
- Teste de hipótese: Correlação Linear.

6.1.1 – INSTRUÇÕES

- Dirigir o *mouse* para o menu superior e clicar em *Estatística*;
- Apontar o *mouse* em *Bootstrap*, quando surgirá a figura abaixo:

Figura 6.1
Menu de
Bootstrap.

Bootstrap - Reamostragem	
Correlação	Dados Qualitativos: Dicotômicos
Distribuição de Probabilidades	Dados Quantitativos: Contínuos
Uma Amostra	Intervalo de Confiança: Correlação Linear
Duas Amostras Independentes	Intervalo de Confiança: Média
Duas Amostras Relacionadas	Intervalo de Confiança: Proporção
Estatística Descritiva	Teste de hipótese: <i>t</i> - duas amostras independentes
Estimação de Parâmetros	Teste de hipótese: ANOVA (um critério)
	Teste de hipótese: Correlação Linear

6.2 – DADOS QUALITATIVOS: DICOTÔMICOS

Os dados dicotômicos ou binários são representados pelos dígitos Um (1) e Zero (0), correspondentes a SIM e NÃO, respectivamente.

Exemplos: casados (1) e não casados (0), homens (1) e mulheres (0), cara (1) e coroa (0), tratados (1) e não tratados (0), tabagistas (1) e não tabagistas (0), etc. No caso de dados usados em certos jogos, constituídos de seis faces, representadas pelos algarismos 1, 2, 3, 4, 5 e 6, pode-se também usar o sistema dicotômico, escolhendo-se qualquer um dos valores, 5 por exemplo, que corresponderá a SIM e os demais (1, 2, 3, 4 e 6), a NÃO.

6.2.1 – EXEMPLO

O lançamento de moeda onze (11) vezes resultou em seis (6) caras, representadas por um (1), e cinco (5) coroas, equivalentes a 0 (zero). Os dados estão contidos no *grid* geral (Figura 6.2).

Figura 6.2
Dados do
Exemplo
6.2.1.

	- 1 - Moeda	- 2 - ** Face
1	1	Cara
2	0	Coroa
3	1	Cara
4	1	Cara
5	0	Coroa
6	1	Cara
7	0	Coroa
8	0	Coroa
9	1	Cara
10	1	Cara
11	0	Coroa

Neste exemplo, a simulação (reamostragem) será aplicada para estimar a proporção de *coroas* na população.

6.2.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Dados Qualitativos: Dicotômicos*;
- Selecionar a coluna *Moeda*, do Exemplo 6.2.1;
- Clicar em *Executar o Teste*, preencher as *caixas de texto*, com os seguintes dados:
 - tamanho da amostra inicial;
 - tamanho da reamostra;
 - valor do evento de interesse, **0** (coroa) no presente caso;
 - número de simulações, 10000 no presente exemplo;
- Pressionar o *mouse* em *Executar*, para resultados.

Figura 6.3
Resultados
do Exemplo
6.2.1.

Reamostragem - Dados dicotômicos

Imprimir

Tamanho da amostra: 11 Valor do evento: 0

Tamanho da reamostra: 30 Número de simulações: 10000

[Executar] [Cancelar]

Ocorrências na amostra: 5 ou 45.45%

Ocorrências na simulação: 13.65 Percentual na simulação: 45.51%

Os dados da reamostragem estão bem próximos do valor da proporção inicial da amostra de coroas da amostra (valor **0** do evento), 45.45% e 45.51%, respectivamente.

6.3 – DADOS QUANTITATIVOS: CONTÍNUOS

Trata-se de um procedimento de *reamostragem* a partir de uma pequena amostra aleatória, de onde se retira um grande número de amostras do mesmo tamanho da amostra inicial, sempre com reposição, tendo por objetivo a estimação da média e da variância da amostra original e das reamostragens. Os resultados alcançados das medidas de tendência central e de variação aproximam-se dos valores paramétricos.

6.3.1 – EXEMPLO

De uma população de cinco (5) mil pacientes, com faixa etária entre 20 e 40 anos, retirou-se amostra de nove (9) indivíduos. Determinou-se a média e o desvio padrão e em seguida, foram efetuadas as reamostragens cujos resultados estão contidos na Figura 6.5.

Figura 6.4
Dados do
Exemplo
6.3.1.

	- 1 -	- 2 -
	Idade (N)	
1	30	
2	26	
3	18	
4	64	
5	45	
6	12	
7	21	
8	25	
9	50	
10	55	

6.3.2 – INSTRUÇÕES

- Introduzir os dados no *grid geral* (Figura 6.4);
- Pressionar o *mouse* em *Dados Quantitativos: Contínuos*;
- Selecionar a amostra (Ex 06.03.1);
- Clicar no botão *Executar Estatística*;
- Informar o *Número de Reamostragens*, neste exemplo 5000 (Figura 6.5);
- Clicar no botão *Executar*.

Figura 6.5
Exemplo
6.3.1.

Dados da População (N = n)	
Média aritmética	34.6000
Desvio padrão	16.6865
Número de reamostragens	5000
Cancelar	Executar
Média aritmética (reamostragem)	32.0124
Desvio padrão (reamostragem)	17.6226

A reamostragem apresentou Média Aritmética = 32.0124 e Desvio Padrão = 17.6226, sendo estes valores considerados estimações dos valores paramétricos.

6.4 – INTERVALO DE CONFIANÇA DA CORRELAÇÃO LINEAR

É um dos testes estatísticos de largo uso, cujos dados amostrais das variáveis X e Y podem ser submetidos à reamostragem, para determinação do intervalo de confiança do coeficiente r (Pearson), como é demonstrado em seguida.

6.4.1 – EXEMPLO

Efetuuou-se levantamento de 30 recém-nascidos de ambos os sexos, no sentido de verificar a associação entre peso (X) e comprimento (Y) dessas crianças. A amostra foi randômica, cujos dados estão contidos no *grid* geral (Figura 6.6).

Figura 6.6
Dados
parciais do
Exemplo
6.4.1.

	- 1 - Peso X	- 2 - Comprim. Y	- 3 -
6	2920	49	
7	3300	48	
8	3350	50	
9	3100	50	
10	2980	48	
11	2920	47	

6.4.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Intervalo de Confiança da Correlação Linear*;
- Selecionar as colunas Peso (X) e Comprimento (Y);
- Clicar em *Executar o Teste*. Serão exibidas *caixas de texto*, as quais devem ser preenchidas com o número de simulações - 12000, no caso - e o Intervalo de Confiança desejado (95%, neste exemplo). Pressionar, a seguir, em *Executar*, para resultados.

Figura 6.7
Resultados
do Exemplo
6.4.1.

Correlação Linear

Número de pares da amostra	Número de simulações	Intervalo de Confiança
30	12000	95.0 %

Colunas selecionadas	r (Pearson) prior
1 e 2	0.4542
r (simulação)	IC r (simulação)
0.4545	0.12 a 0.73

O IC (95%), após 12000 simulações, está compreendido entre 0.12 a 0.73, também mostrado graficamente na figura inferior em linha de cor cinza, assinalando-se, ainda, o traço vertical escuro correspondente ao valor de r da simulação, igual a 0.4545, incluído, como seria de esperar, no respectivo intervalo, próximo do valor de r da amostra inicialmente coletada (0.4542).

6.5 – INTERVALO DE CONFIANÇA DA MÉDIA

De dados amostrais randômicos podem-se obter a média das reamostragens e seu respectivo intervalo de confiança baseado em *percentis*.

6.5.1 – EXEMPLO

De amostra formada por dez (10) insetos, mediu-se o peso de cada indivíduo em gramas. Os resultados foram estes:

Figura 6.8
Dados do
Exemplo
6.5.1.

	- 1 -	- 2 -
	Peso (g)	
1	3.000	
2	4.000	
3	2.000	
4	1.500	
5	2.300	
6	5.100	
7	4.700	
8	3.300	
9	2.900	
10	1.700	

6.5.2 – INSTRUÇÕES

- Pressionar o mouse em *Intervalo de Confiança da Média*;
- Selecionar a coluna do *grid* geral correspondente aos valores amostrais e clicar em *Executar o Teste*;
- Preencher as *caixas de texto* com as informações sobre o número de simulações e o Intervalo de Confiança desejado;
- Pressionar em *Executar*, para resultados.

Figura 6.9
Resultados
do Exemplo
6.5.1.

A janela 'Técnica de Reamostragem (Bootstrap)' contém as seguintes informações:

Entrada de Dados:

Tamanho da amostra	Número de Simulações	Intervalo de Confiança (%)
30	10000	95

Botões: Executar, Cancelar

Resultado Bootstrap:

Intervalo de Confiança		
Limite inferior	Média	Limite superior
2.640	3.050	3.407
2.50 %		97.50 %

Foram efetuadas 10000 simulações, ou seja, a retirada de dez mil amostras da amostra inicial, com reposição, obtendo-se os seguintes valores da Média e do Intervalo de Confiança de 95%:

Média = 3.050 e IC (95%): 2.640 a 3.407 (em termos de *percentis*).

Para efeito de comparação, quando se efetua a estimação de parâmetros somente com base na média e no desvio padrão da amostra original, obtém-se:

Média = 3.05

IC (95%): 2.622 a 3.478 (em termos de desvio padrão)

Obs.: pode-se repetir n vezes o mesmo procedimento, com ou sem alterações do número de Simulações e do Intervalo de Confiança, clicando-se em *Executar*, para resultados.

6.6 – INTERVALO DE CONFIANÇA DA PROPORÇÃO

O procedimento é semelhante ao utilizado para média, como será visto em seguida, cujos Intervalos de Confiança estão baseados nos *percentis*.

6.6.1 – EXEMPLO

Em sondagem eleitoral, efetuada com dois meses de antecedência do pleito, obteve-se amostra de tamanho 200 e a seguinte proporção para o candidato AZ: 0.36 (36%), correspondendo a 72 eventos favoráveis em 200 eleitores. Efetuar 15000 simulações, para calcular a proporção e o Intervalo de Confiança de 99%.

6.6.2 – INSTRUÇÕES

- Pressionar o mouse em *Intervalo de Confiança da Proporção*;
- Preencher as *caixas de texto* com as informações sobre o número de simulações, o Intervalo de Confiança desejado e os valores originais do tamanho da amostra e da respectiva proporção;
- Pressionar em *Executar*, para resultados.

Figura 6.10
Resultados
do Exemplo
6.6.1.

Intervalo de Confiança		
Limite inferior	Proporção	Limite superior
0.275	0.363	0.445
0.50 %		99.50 %

Foram efetuadas 15000 simulações, ou seja, a retirada de quinze mil amostras com reposição da amostra original, obtendo-se os seguintes valores:

Proporção = 0.363;

IC (99%): 0.275 a 0.445 (em termos de *percentis*).

Para efeito de comparação, quando se efetua a estimação de parâmetros somente com base na amostra inicial e no número de eventos favoráveis, obtém-se:

Proporção amostral = 0.36;

IC (99%): 0.272 a 0.448 (em termos de desvio padrão);

Obs.: pode-se repetir n vezes o mesmo procedimento, com ou sem alterações do *Número de Simulações* e do *Intervalo de Confiança*, clicando-se em *Executar*, para resultados.

6.7 –TESTE DE HIPÓTESE: Teste t – Duas Amostras Independentes

Realiza-se a reamostragem de duas amostras independentes – teste t –, as quais podem ser do mesmo tamanho ou desiguais.

6.7.1 – EXEMPLO

Um investigador admite que a estatura dos homens de duas populações de países da Europa Ocidental são diferentes. Foram mensurados 36 indivíduos do país A e 33 do B. Os dados estão inseridos no *grid* geral (Figura 6.11).

Figura 6.11
Dados
parciais do
Exemplo
6.7.1

	- 1 - País A	- 2 - País B
1	1.760	1.720
2	1.740	1.690
3	1.770	1.650
4	1.790	1.770
5	1.760	1.790
6	1.750	1.760
7	1.730	1.750
8	1.800	1.730
9	1.770	1.800
10	1.780	1.770

H_0 : não há diferença entre a estatura média dos homens nos países A e B: $\mu_1 = \mu_2$;

H_1 : há diferença entre a estatura média dos homens nos países A e B: $\mu_1 \neq \mu_2$;

Nível de decisão: alfa = 0.05.

6.7.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Teste de hipótese: teste t - Duas Amostras Independentes*;
- Selecionar as colunas referentes aos países A e B, do Exemplo 6.7.1;
- Informar, em seguida, o número de simulações desejadas, 1000 neste caso;
- Clicar em *Executar*, para resultados.

Figura 6.12
Resultado
do Exemplo
6.7.1.

Tamanho da amostra A	Tamanho da amostra B
36	33
Número de simulações	
1000	
Executar	Cancelar
Média geral prior	Desvio prior
1.7345	0.0342
Média geral simulação	Desvio simulação
1.7310	0.0176
p-valor de T (simulação)	
0.0010	

A *Média geral prior* e o *Desvio prior* são estatísticas obtidas das amostras, sem simulação. O *p-valor* da simulação mostrou resultado altamente significativo (0.0010), rejeitando-se, portanto, a hipótese de nulidade, podendo-se admitir que a estatura média dos homens desses países é diferente.

6.8 – TESTE DE HIPÓTESE: ANOVA (um critério)

Para obter a reamostragem de dados para fins de análise de variância, realiza-se o seguinte procedimento: retirar três ou mais amostras, dos mesmos tamanhos ou desiguais, cujos critérios de comparação são os tratamentos. Procede-se, então, a simulação conforme o exemplo indicado a seguir.

6.8.1 – EXEMPLO

Foram coletadas sementes das espécies A, B, C e D, cujo peso medido em gramas de cada unidade e as respectivas espécies são mostrados na tabela seguinte:

Figura 6.13
Dados do
Exemplo
6.8.1.

	- 1 - Espécie A	- 2 - Espécie B	- 3 - Espécie C	- 4 - Espécie D
1	62.00	63.00	67.00	56.00
2	60.00	67.00	69.00	62.00
3	62.00	71.00	71.00	60.00
4	59.00	64.00	67.00	61.00
5	61.00	65.00	68.00	63.00
6		66.00	68.00	64.00
7			70.00	63.00
8				59.00

H_0 : os pesos das sementes obtidas das espécies em questão não apresentam diferenças significativas: $\mu_1 = \mu_2 = \mu_3 = \mu_4$;

H_1 : os pesos das sementes pelo menos de duas das espécies obtidas apresentam diferenças significativas: $\mu_i \neq \mu_j$;

Nível de decisão: alfa = 0.01.

6.8.2 – INSTRUÇÕES

- Pressionar o mouse em *Teste de hipótese: ANOVA (um critério)*;
- Selecionar as colunas referentes aos tratamentos (Espécies A, B, C e D) do Exemplo 6.8.1;
- Clicar em *Executar o Teste* e informar o *Número de Tratamentos* incluídos no teste (quatro tratamentos). Preencher, em seguida, o *Número de Simulações* desejadas, 10000 neste exemplo, e pressionar em *Executar*, para resultados.

Figura 6.14
Resultados
do Exemplo
6.8.1.

O *p-valor* da simulação mostrou resultado altamente significativo (0.0001), rejeitando-se, portanto, a hipótese de nulidade, podendo-se admitir que, pelo menos, o peso médio das sementes de duas espécies das amostras coletadas são diferentes.

6.9 – TESTE DE HIPÓTESE: CORRELAÇÃO LINEAR

Pelo procedimento de reamostragem, podemos obter os valores de *r* (Pearson) e o respectivo *p-valor* da Correlação Linear.

6.9.1 – EXEMPLO

Os dados contidos na Figura 6.15 foram submetidos ao teste de reamostragem para obtenção de *r* simulado e do respectivo *p-valor*.

H_0 : não há correlação entre as variáveis *X* e *Y*: $r = 0$;

H_1 : há correlação entre as variáveis *X* e *Y*: $r \neq 0$;

Nível de decisão: alfa = 0.01.

Figura 6.15
Dados do
Exemplo
6.9.1.

	- 1 - X	- 2 - Y	- 3 -
1	2.870	5.230	
2	2.160	4.000	
3	3.140	5.150	
4	2.510	4.130	
5	1.800	3.100	
6	2.000	4.500	
7	2.300	4.200	
8	3.200	5.200	
9			

6.9.2 – INSTRUÇÕES

- Pressionar o mouse em *Teste de hipótese: Correlação Linear*;
- Selecionar as colunas do *grid* geral correspondentes aos valores amostrais X e Y (Exemplo 6.9.1) e clicar em *Executar o Teste*;
- Preencher as *caixas de texto* com o número de simulações desejado, 1000 no presente exemplo;
- Clicar em *Executar*, para resultados.

Figura 6.16
Resultados
do Exemplo
6.9.1.

Bootstrap - p-valor: Correlação Linear

Número de pares	Número de simulações
8	1000

Executar Cancelar

r (Pearson) prior	r (simulação)
0.8728	0.5036

p-valor (simulação)

0.0040

O *p-valor* obtido é igual a 0.0040, sendo estatisticamente muito significativo. Os valores de *r*, *prior* e simulado, foram 0.8728 e 0.5036, respectivamente.

CAPÍTULO 7

CORRELAÇÃO

7.1 – INTRODUÇÃO

A Análise de Correlação proporciona um meio de se verificar o grau de associação entre duas ou mais variáveis. Os testes mais utilizados para essa análise são:

a) *Testes paramétricos:*

- i – Correlação linear de *Pearson*;
- ii – Matriz de Correlação;
- iii – Correlação parcial;
- iv – Correlação linear quando o parâmetro é diferente de zero ($r < 0$ ou $r > 0$);

b) *Testes não-paramétricos:*

- i – Contingência C;
- ii – Concordância de *Kendall*;
- iii – Correlação de *Kendall*;
- iv – Coeficiente Phi;
- v – Correlação de *Spearman*.

7.1.1 – INSTRUÇÕES

- a) Introduzir os dados no *grid* geral do *BioEstat*;
- b) Dirigir o *mouse* para o *menu* superior e clicar em *Estatística*;
- c) Apontar o *mouse* em *Correlação*, conforme demonstra a Figura 7.1:

Figura 7.1
Menu de
Correlação.

Correlação	▶	Coef. de Contingência C
Distribuição de Probabilidades	▶	Coef. de Concordância de Kendall
Uma Amostra	▶	Coef. de Correlação de Kendall
Duas Amostras Independentes	▶	Coef. de Correlação Parcial
Duas Amostras Relacionadas	▶	Coef. de Correlação de Pearson
Estatística Descritiva	▶	Coef. de Correlação: $\mu < 0$ ou $\mu > 0$
Estimação de Parâmetros	▶	Coef. Phi
Meta-Análise	▶	Coef. de Correlação de Spearman
Normalidade	▶	Matriz de Correlação

7.2 – CORRELAÇÃO LINEAR DE PEARSON

É um teste de largo emprego em bioestatística, onde os valores das variáveis X e Y são mensurados a nível intervalar ou de razões. O coeficiente de *Pearson* – r – pode variar de -1 a $+1$, e quanto mais próximos desses valores, mais forte a associação das variáveis em exame. O escore zero desse coeficiente indica ausência de correlação.

7.2.1 – EXEMPLO

Efetuuou-se levantamento em adolescentes no sentido de verificar a associação de peso e estatura desses indivíduos. A amostra foi randômica e os dados estão no *grid* geral (ver Figura 7.2).

H_0 : não há associação entre peso e estatura de adolescentes: $r = 0$;

H_1 : há correlação entre as duas variáveis: $r \neq 0$;

Nível de decisão: alfa = 0.05.

Figura 7.2
Dados do
Exemplo
7.2.1.

	- 1 - Peso (X)	- 2 - Estatura (Y)
1	45.000	1.660
2	53.000	1.730
3	41.000	1.580
4	39.000	1.590
5	44.000	1.630
6	47.000	1.690

7.2.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Coef. de Correlação de Pearson*;
- Selecionar as colunas referentes às variáveis do *grid* geral;
- Clicar em *Executar o Teste* para obter resultados.

Figura 7.3
Resultados
do Exemplo
7.2.1.

	Colunas 1 e 2
n (pares) =	6
r (Pearson) =	0.9624
IC 95% =	0.69 a 1.00
IC 99% =	0.45 a 1.00
R2 =	0.9263
t =	7.0878
GL =	4
(p) =	0.0021
Poder 0.05 =	0.9625
Poder 0.01 =	0.8642

Os resultados mostram forte correlação positiva entre Peso e Estatura de adolescentes ($r = 0.9624$ e $p = 0.0021$), rejeitando-se, assim, a hipótese de nulidade e aceitando-se a alternativa. Em outras palavras, à medida que aumenta uma das variáveis, a outra cresce de valor, não havendo, entretanto, relação de dependência de X em relação a Y ou de Y sobre X.

7.3 – MATRIZ DE CORRELAÇÃO

Pode-se calcular diversos valores de r com as respectivas probabilidades e, ao mesmo tempo, obter a *matriz* dos valores da correlação.

7.3.1 – EXEMPLO

Os dados contidos no *grid* geral, apresentados na Figura 7.4, correspondem a valores de quatro (4) amostras, as quais devem apresentar o mesmo tamanho. A Figura 7.5 indica os resultados obtidos e a matriz de correlação.

7.3.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Matriz de Correlação*;
- Selecionar as colunas referentes às variáveis do *grid* geral;

c) Clicar em *Executar o Teste* para obter resultados

Figura 7.4
Dados do
Exemplo
7.3.1.

	- 1 -	- 2 -	- 3 -	- 4 -	- 5
1	2.870	3.230	2.250	5.230	
2	2.160	3.450	3.130	4.000	
3	3.140	2.780	2.440	5.150	
4	2.510	3.770	3.270	4.130	
5	1.800	2.970	2.810	3.100	

Figura 7.5
Resultados
do Exemplo
7.3.1.
Na parte
inferior está
representa-
da a
matriz de
correlação.

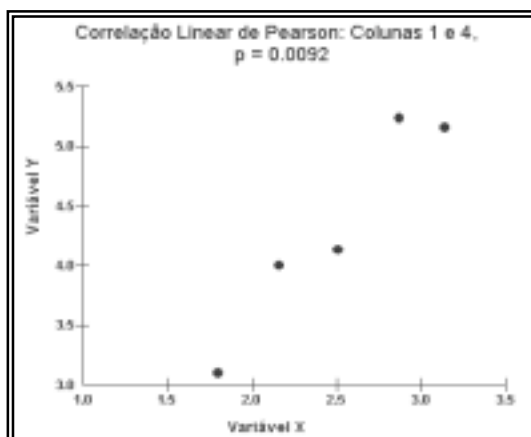
	Colunas 1 e 2	Colunas 1 e 3	Colunas 1 e 4	Colunas 2 e 3	Colunas 2 e 4	Colunas 3 e 4
n (pares) =	5	5	5	5	5	5
r (Pearson) =	-0.2091	-0.5865	0.9609	0.7152	-0.1657	-0.6509
IC 95% =	-0.92 a 0.83	-0.97 a 0.61	0.52 a 1.00	-0.45 a 0.98	-0.91 a 0.84	-0.97 a 0.54
IC 99% =	-0.97 a 0.92	-0.99 a 0.82	0.13 a 1.00	-0.73 a 0.99	-0.96 a 0.93	-0.99 a 0.78
R2 =	0.0437	0.3440	0.9234	0.5115	0.0275	0.4237
t =	-0.3703	-1.2542	6.0135	1.7723	-0.2910	-1.4852
GL =	3	3	3	3	3	3
(p) =	0.7358	0.2986	0.0092	0.1744	0.7900	0.2341
Matriz de Correlação						
	Coluna 1	Coluna 2	Coluna 3	Coluna 4		
Coluna 1	1.0000	---	---	---		
Coluna 2	-0.2091	1.0000	---	---		
Coluna 3	-0.5865	0.7152	1.0000	---		
Coluna 4	0.9609	-0.1657	-0.6509	1.0000		

As variáveis das colunas um (1) e quatro (4) apresentam *Coeficiente de Pearson* igual a 0.9609, com $p = 0.0092$. As demais correlações não são estatisticamente significativas. A matriz está representada na parte inferior da Figura 7.5.

7.3.3 – GRÁFICO DO EXEMPLO 7.3.1

Pressionar em Gráfico na linha superior da Figura 7.5:

Figura 7.6
Gráfico do
Exemplo
7.3.1,
Colunas
1 e 4.



7.4 – CORRELAÇÃO PARCIAL

No estudo da correlação linear é possível estender o conceito medindo-se a associação entre duas variáveis e, ao mesmo tempo, controlando o efeito de uma terceira, a variável Z. Neste caso, chama-se o procedimento de *Correlação Parcial*.

7.4.1 – EXEMPLO

O exemplo que vai ser mostrado é o da influência da renda familiar – variável Z – na correlação entre tempo de estudo dos pais – variável X – e tempo de estudo dos filhos – variável Y. O programa permite que se escolha a variável Z, a qual pode ser qualquer uma das três colunas preenchidas com os escores. No exemplo, a variável a ser fixada está na coluna três (3) do *grid*. Além disso, os resultados da Correlação Linear de *Pearson* das variáveis X e Y são apresentados a título comparativo.

H_0 : a renda familiar (variável Z) não influencia a correlação tempo de estudo dos pais (variável X) e o tempo de escolaridade dos filhos (variável Y);

H_1 : a renda familiar (variável Z) influencia a correlação tempo de estudo dos pais (variável X) e o tempo de escolaridade dos filhos (variável Y);

Nível de decisão: $\alpha = 0.05$.

Figura 7.7
Dados do
Exemplo
7.4.1.

	- 1 - X	- 2 - Y	- 3 - Z	- 4 -
1	12	12	4000	
2	10	8	3000	
3	6	6	2000	
4	16	12	5000	
5	8	10	3500	
6	9	8	3800	
7	12	11	4700	

7.4.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Coef. de Correlação Parcial*;
- Selecionar as colunas referentes às variáveis do *grid* geral;
- Clicar em *Executar o Teste*, surgindo uma tela para a escolha da variável Z;
- Pressionar *Executar*, para resultados.

Figura 7.8
Resultados
do Exemplo
7.4.1.

	Resultados
n (trincas) =	7
Graus de liberdade =	4
r XY.Z =	0.2719
Teste t =	0.5650
(p) =	0.6022
Variável fixa =	Coluna 3
n (pares) =	7
Graus de liberdade =	5
r XY =	0.8295
Teste t =	3.3205
(p) =	0.0210

Há uma associação entre a escolaridade dos pais e dos filhos ($r_{XY} = 0.8295$, $p = 0.021$), a qual deixa de ser significativa ao se introduzir a variável Z ($r_{XY.Z} = 0.2719$, $p = 0.6022$)

7.5 – CORRELAÇÃO LINEAR: $\rho < 0$ ou $\rho > 0$

Algumas vezes há interesse do pesquisador no estudo da associação entre duas variáveis, mas o parâmetro de comparação passa a ser diferente de zero, como se exemplifica a seguir.

7.5.1 – EXEMPLO

Em estudo genético (hipotético) verificou-se que o parâmetro (ρ) de associação entre as variáveis X e Y é igual a **0.50**. Tomou-se amostra aleatória de seis (6) indivíduos, cujas variáveis em questão foram mensuradas, a fim de verificar se estariam de acordo com a distribuição teórica formulada.

$H_0: \rho = 0.50$;

$H_1: \rho \neq 0.50$;

Nível de decisão: $\alpha = 0.05$.

Figura 7.9
Dados do
Exemplo
7.5.1.

	- 1 - X	- 2 - Y	- 3 -
1	5.000	45.000	
2	7.000	8.000	
3	3.000	25.000	
4	9.000	4.000	
5	6.000	7.000	
6	5.000	6.000	

7.5.2 – INSTRUÇÕES

- Pressionar o mouse em *Coef. de Correlação*: $\rho < 0$ ou $\rho > 0$;
- Selecionar as colunas referentes às variáveis do *grid* geral;
- Clicar em *Executar o Teste*, surgindo a Figura 7.10, a qual deve ser preenchida com o valor de ρ assumido da população, **0.5** no caso;

Figura 7.10
Coef. de
correlação da
população (ρ).
(Exemplo
7.5.1).

- Clicar em *OK*, para resultados.

Figura 7.11
Resultados
do Exemplo
7.5.1.

	Resultados
Número de pares x, y =	6
Coef. de correlação na população =	0.5000
Coef. de correlação amostral =	-0.5456
Graus de liberdade =	4
z =	2.0117
(p) =	0.0443

Há diferença estatisticamente significativa entre o coeficiente observado (- 0.5456) e o parâmetro (0.50), sendo o p -valor é igual a 0.044. Rejeita-se, portanto, a hipótese de nulidade.

7.6 – COEFICIENTE DE CONTINGÊNCIA C

Este Coeficiente é indicado para determinar a magnitude de associações de variáveis mensuradas a nível ordinal, dispostas em tabelas de contingência $n \times n$ ($n \geq 2$).

7.6.1 – EXEMPLO

Um pesquisador deseja verificar se há associação entre o grau de escolaridade (X) e o nível salarial (Y). Os dados coletados de cada amostra correspondem ao número de indivíduos e foram inseridos no *grid* geral da seguinte forma:

Figura 7.12
Dados do
Exemplo
7.6.1.

	- 1 - ** Fx salarial	- 2 - Fundamental	- 3 - Médio	- 4 - Superior	- 5 -
1	1 a 2 SM	280	280	280	
2	3 a 5 SM	49	176	275	
3	6 ou mais	3	50	270	
4					

- Colunas 1, 2 e 3 referentes aos níveis de escolaridade: fundamental, médio e superior, respectivamente;
- Linhas 1, 2 e 3: relativas aos níveis salariais: um a dois, três a cinco e mais de cinco salários mínimos, respectivamente;
- Cada casela do *grid* registra o número de indivíduos relativos à escolaridade e ao respectivo nível salarial.

H_0 : não há associação entre níveis de escolaridade e salarial: *Coeficiente C* = 0;

H_1 : há correlação entre as duas variáveis, quanto maior a escolaridade dos indivíduos, maior o salário recebido: *Coeficiente C* \neq 0;

Nível de decisão: alfa = 0.05.

7.6.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Coef. de Contingência C*;
- Selecionar as colunas referentes aos tratamentos do *grid* geral;
- Clicar em *Executar o Teste*, para resultados (Figura 7.13).

Figura 7.13
Resultados
do Exemplo
7.6.1.

	Resultados
Tabela de Contingência =	3 x 3
Qui-quadrado =	312.4659
Coef. de Contingência C =	0.3977
Graus de liberdade =	4
(p) =	< 0.0001

Há forte associação entre o grau de escolaridade e o nível salarial: *Coef. C* = 0.3977 e $p < 0001$, rejeitando-se a hipótese de nulidade e aceitando-se a alternativa.

7.7 – COEFICIENTE DE CONCORDÂNCIA DE KENDALL: *W*

Este coeficiente – *W* – testa a associação entre vários (*k*) conjuntos de postos dos mesmos indivíduos (*n*), cujas variáveis são mensuradas a nível ordinal.

7.7.1 – EXEMPLO

Testar o grau de correlação entre o *nível de desempenho* de três (3) médicos residentes e as *respectivas avaliações* efetuadas por três (3) supervisores.

Figura 7.14
Dados do
Exemplo
7.7.1.

Ex 07.07.1 - Concord de Kendall					
	- 1 -	- 2 -	- 3 -	- 4 -	- 5 -
	** Médico	Supervisor X	Supervisor Y	Supervisor Z	** Total
1	A	9.0	6.0	3.0	18.0
2	B	7.0	5.0	6.0	18.0
3	C	3.0	4.0	11.0	18.0

H_0 : não há concordância entre a avaliação dos supervisores em relação ao desempenho dos médicos residentes: *Coef. de Kendall W* = 0;

H_1 : há concordância entre a avaliação dos supervisores em relação ao desempenho dos médicos residentes: *Coef. de Kendall W* \neq 0;

Nível de decisão: alfa = 0.05.

7.7.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Coef. de Concordância de Kendall*;
- Selecionar as colunas (postos) referentes às variáveis do *grid* geral;
- Clicar em *Executar o Teste*, para resultados.

Figura 7.15
Resultados do
Exemplo
7.7.1.

	Resultados
Número de postos =	3
Número de indivíduos por posto	3
S =	14.0000
W (Coeficiente de concordância) =	0.7778
Graus de liberdade =	2
Qui-Quadrado =	4.6667
(p) =	0.0970

O resultado não é estatisticamente significativo ($W = 0.7778$ e $p = 0.097$), não tendo havido, portanto, concordância (associação) entre os supervisores. Aceita-se, assim, a hipótese de nulidade.

7.8 – COEFICIENTE DE CORRELAÇÃO DE KENDALL: Γ

Mede a associação entre duas variáveis mensuradas a nível ordinal, ou seja, *por postos*.

7.8.1 – EXEMPLO


Dois médicos neonatologistas efetuaram a avaliação de seis recém-nascidos pelos escores de *Apgar*. Os valores foram inseridos no *grid* geral.

H_0 : não há associação entre a avaliação dos neonatologistas em relação aos escores de *Apgar* dos recém-nascidos: *Coef. de Correlação de Kendall* $\Gamma = 0$;

H_1 : há associação entre a avaliação dos neonatologistas em relação aos escores de *Apgar* dos recém-nascidos: *Coef. de Correlação de Kendall* $\Gamma \neq 0$;

Nível de decisão: alfa = 0.05.

Figura 7.16
Dados do
Exemplo
7.8.1.

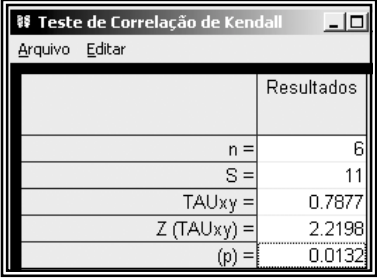


	- 1 - Médico A	- 2 - Médico B	- 3 -
1	9.000	10.000	
2	4.000	4.000	
3	1.000	1.000	
4	10.000	9.000	
5	8.000	9.000	
6	2.000	1.000	

7.8.2 – INSTRUÇÕES

- Pressionar o mouse em *Coeficiente de Correlação de Kendall*;
- Selecionar as colunas referentes às variáveis do *grid* geral;
- Clicar em *Executar o Teste*, para resultados.

Figura 7.17
Resultados do
Exemplo
7.8.1.



Teste de Correlação de Kendall	
Arquivo Editar	
	Resultados
n =	6
S =	11
TAUxy =	0.7877
Z (TAUxy) =	2.2198
(p) =	0.0132

Os resultados revelam um Coeficiente de Correlação de *Kendall* (0.7877) significativo ($p = 0.0132$) tendo havido associação (concordância) dos médicos especialistas em relação aos escores de *Apgar* dos recém-nascidos. Rejeita-se a hipótese de nulidade, aceitando-se a alternativa.

7.9 – COEFICIENTE PHI: r_ϕ

Este teste é aplicado para verificar a associação de variáveis mensuradas a nível nominal e que se apresentam de modo dicotômico em tabela de contingência 2 x 2.

7.9.1 – EXEMPLO

Foi efetuado levantamento em uma universidade para verificar se há associação entre os sexos dos estudantes e os cursos de medicina e biologia. Nesse exemplo, ambas as variáveis apresentam-se dicotômicas e são mensuradas a nível nominal: sexo masculino e feminino de um lado e, de outro, os cursos de medicina e biologia, com os seguintes escores:

- estudantes do sexo masculino cursando biologia: 40;
- estudantes do sexo masculino cursando medicina: 48;
- estudantes do sexo feminino cursando biologia: 55;
- estudantes do sexo feminino cursando medicina: 47;

H_0 : não existe associação entre os sexos dos estudantes e os cursos de medicina e de biologia: $r_\phi = 0$;

H_1 : existe associação entre os sexos dos estudantes e os cursos de medicina e de biologia: $r_\phi \neq 0$;

Nível de decisão: alfa = 0.05.

7.9.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Coeficiente Phi*;
- Preencher o *grid* conforme exemplo abaixo;
- Clicar em *Executar*, para resultados.

Figura 7.18
Coeficiente
Phi.
(Exemplo
7.9.1).

Entrada de Dados	
A	B
40	48
C	D
55	47

Executar

Cancelar

Resultados

$r_\phi = 0.0844$

$\chi^2_{ajustado} = 1.0372$

Graus de liberdade = 1

(p) = 0.3085

Entrada de Dados do *Coeficiente Phi*.

A = estudantes masculinos que cursam biologia;
B = estudantes masculinos que cursam medicina;
C = estudantes femininos que cursam biologia;
D = estudantes femininos que cursam medicina.

Pelos resultados verifica-se que não há diferença estatisticamente significativa entre sexos e escolha dos cursos em questão ($p = 0.3085$). Aceita-se, portanto, a hipótese de nulidade.

7.10 – COEFICIENTE DE CORRELAÇÃO DE SPEARMAN: r_s

É uma prova não-paramétrica com a finalidade de determinar o grau de associação entre duas variáveis mensuradas, pelo menos, a nível ordinal e dispostas em postos ordenados em duas séries: X e Y .

7.10.1 – EXEMPLO

Os mesmos dados relativos à Correlação de *Kendall* sobre os escores de *Apgar* (ver Figura 7.16).

H_0 : não há correlação entre a avaliação dos neonatologistas em relação aos escores de Apgar dos recém-nascidos: $r_s = 0$;

H_1 : há correlação entre a avaliação dos neonatologistas em relação aos escores de Apgar dos recém-nascidos: $r_s \neq 0$;

Nível de decisão: $\alpha = 0.05$.

7.10.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Coef. de Correlação de Spearman*;
- Selecionar as colunas referentes às variáveis do *grid* geral;
- Clicar em *Executar o Teste*, para resultados.

Figura 7.19
Resultados do
Exemplo
7.10.1.

Teste de Correlação de Spearman	
Arquivo Editar	
	Resultados
Coeficiente de Spearman (rs)=	0.8827
t =	3.7573
(p)=	0.0198
Número de pares =	6

Os resultados mostram correlação entre as avaliações efetuadas pelos dois médicos em 6 recém-nascidos, sendo o teste estatisticamente significativo ($p = 0.0198$), confirmando aqueles obtidos pelo teste de Correlação de *Kendall*.

CAPÍTULO 8

DISTRIBUIÇÃO DE PROBABILIDADES

8.1 – INTRODUÇÃO

Quando o procedimento para a coleta de dados, conseqüentemente das variáveis investigadas, é randômico, assumindo cada variável um certo valor e uma determinada probabilidade, diz-se, então, que são *variáveis aleatórias*. O programa apresenta as seguintes distribuições:

- a) *Distribuições discretas:*
 - i – Binomial;
 - ii – Hipergeométrica;
 - iii – Poisson.
- b) *Distribuições contínuas:*
 - i – Exponencial;
 - ii – Normal.
- c) *Probabilidade condicional:*
 - i – Crivo (*screening test*);
 - ii – Curva ROC;
 - iii – Ponto de Corte;
 - iv – Regra de Bayes;
 - v – Sensibilidade/Especificidade.
- d) *Outras distribuições:*
 - i – Distribuição *F*;
 - ii – Distribuição do Qui-quadrado;
 - iii – Distribuição *t* de Student;
 - iv – Distribuição *Z*.

8.1.1 – INSTRUÇÕES

- a) Dirigir o mouse para o menu superior e clicar em *Estatística*;
- b) Apontar o mouse em *Distribuição de Probabilidades*.

Figura 8.1
Menu de
Distribuição
de Probabili-
dades.

Distribuição de Probabilidades	▶	Binomial
Uma Amostra	▶	Exponencial
Duas Amostras Independentes	▶	Hipergeométrica
Duas Amostras Relacionadas	▶	Normal
Estatística Descritiva		Poisson
Estimação de Parâmetros	▶	Probabilidade Condicional ▶
Normalidade	▶	Outras Distribuições ▶

8.2 – DISTRIBUIÇÃO BINOMIAL

Esta distribuição caracteriza-se quando:

- a) o experimento consiste de *n* provas idênticas, cada uma apresentando apenas dois resultados: *sucesso* e *insucesso*;
- b) a probabilidade do *sucesso* é igual a *p*, e a do *insucesso*, a *q*, sendo $q = 1 - p$;

- c) as provas são independentes, procurando-se determinar a probabilidade do número de sucessos – k – em n provas;
- d) o quociente n/N deve ser inferior a 0.05.

8.2.1 – EXEMPLO

No universo de leucócitos da circulação periférica do homem a proporção de linfócitos é igual a 0.36. Contando-se 20 leucócitos, qual é a probabilidade de se obter cinco (5) linfócitos? Considerando-se que a população (N) de leucócitos atinge a casa dos milhões, então o quociente n/N é inferior a 0.05, indicando-se, nesse caso, a aplicação da distribuição binomial para a variável questionada.

8.2.2 – INSTRUÇÕES

- a) Clicar o mouse em *Binomial*;
- b) Preencher as *caixas de texto* com os dados;
- c) Clicar em *Executar*, para obter os resultados.

Figura 8.2
Distribuição
Binomial.
(Exemplo
8.2.1)

Observa-se, além da média (μ), do desvio padrão (σ) e dos valores de p e q , a probabilidade de se obter exatamente cinco (5) linfócitos em 20 células leucocitárias (0.1161), a da obtenção de cinco (5) ou menos linfócitos ($k \leq 5 = 0.2171$) e, ainda, a chance de mais de cinco (5) células linfocitárias (0.7829). A soma desses valores é igual à unidade ($0.2171 + 0.7829 = 1$). A fração 0.1161 está incluída em 0.2171 (igual ou menor que cinco (5) linfócitos).

8.3 – DISTRIBUIÇÃO HIPERGEOMÉTRICA

Em determinadas ocasiões as distribuições de probabilidades de variáveis discretas têm as mesmas características das distribuições binomiais, mas o tamanho da amostra é relativamente grande em relação ao tamanho do universo. Desse modo, a retirada de uma unidade sem reposição altera acentuadamente a probabilidade dos demais constituintes da população e, nesses casos, a distribuição passa a ser *hipergeométrica*. O critério adotado para considerar esse modelo é o valor do quociente n/N , o qual deve ser ≥ 0.05 .

8.3.1 – EXEMPLO

Uma caixa contém quinze (15) vidros (população) de comprimidos de vitamina B₁ e em cada caixa observou-se que três (3) vidros apresentavam dosagem da vitamina

inferior à indicada pelo fabricante, compreendendo uma proporção de $3/15 = 0.2$. Em amostra randômica de cinco (5) unidades, calcular a probabilidade de dois (2) vidros ou menos apresentarem dosagem da vitamina abaixo da indicada pelo produtor. Como o quociente $n/N = 5/15 = 0.333$ é superior a 0.05, indica-se a distribuição hipergeométrica para esse cálculo.

8.3.2 – INSTRUÇÕES

- Clicar o mouse em *Hipergeométrica*;
- Preencher as *caixas de texto* com os dados;
- Clicar em *Executar*, para obter os resultados.

Figura 8.3
Distribuição
Hipergeométrica.
(Exemplo
8.3.1).

Distribuição Hipergeométrica

Imprimir

Dados

N (tamanho populacional) 15 Número de sucessos na população 3

n (tamanho amostral) 5 Número de sucessos na amostra 2

Executar Sair

Resultados

Média = 1.0000 Variância = 0.5714

$p(x) = 2$
0.2198

$p(x) \leq 2$ 0.9780 $p(x) > 2$ 0.0220

A probabilidade de se encontrar na amostra dois (2) vidros com subdosagem de vitamina B₁ é 0.2198, a de 2 vidros ou menos 0.9780 e a de mais de dois vidros 0.0220.

A soma dessas probabilidades é igual à unidade: $0.9780 + 0.0220 = 1$ (o valor 0.2198 está incluído em 0.9780).

8.4 – DISTRIBUIÇÃO DE POISSON

Trata-se de um modelo de distribuição de probabilidade discreta semelhante à binomial, com as seguintes diferenças:

- A probabilidade do evento de interesse – p – se apresenta como *ocorrência rara* e relacionada no *tempo* e no *espaço*.
- O valor de q ($1 - p$) não é considerado nos cálculos.

8.4.1 – EXEMPLO

A hipersensibilidade à penicilina é de 0.0015 em populações humanas. Se em determinada localidade e num período de um ano fossem testadas 4000 pessoas residentes, qual seria a probabilidade de três (3) indivíduos apresentarem a mencionada reação ao antibiótico? Tratando-se de evento raro (0.0015) e relacionado no tempo e no espaço, indica-se a distribuição de *Poisson* para o cálculo da probabilidade dessa distribuição.

8.4.2 – INSTRUÇÕES

- Clicar o mouse em *Poisson*;
- Preencher as *caixas de texto* com os dados;
- Clicar em *Executar*, para obter os resultados.

Figura 8.4
Distribuição
de Poisson.
(Exemplo
8.4.1).

Distribuição de Poisson

Imprimir

Dados

N (tamanho amostral ou populacional): 4000

Número de eventos esperados: 3

Probabilidade do evento: 0.0015

Executar

Sair

Resultados

$P =$ 0.0015

Média = 6.0000

$p(x) =$ 0.0892

$p(x) \leq 3 =$ 0.1512

$p(x) > 3 =$ 0.8488

A probabilidade de ocorrência de três (3) indivíduos com hipersensibilidade à penicilina é 0.0892; de três (3) ou menos, 0.1512, e para mais de 3 pessoas, 0.8488. A soma dessas probabilidades é igual à unidade: $0.1512 + 0.8488 = 1$. No escore 0.1512 está incluído o *p*-valor de (*x*), igual a 0.0892.

8.5 – DISTRIBUIÇÃO EXPONENCIAL

A distribuição exponencial é um modelo para variáveis contínuas referentes a *intervalos de tempo* transcorrido entre eventos raros e discretos, como ocorre na distribuição de *Poisson*. O interesse, no entanto, é o *intervalo temporal*, o qual apresenta distribuição *contínua* e *exponencial*.

8.5.1 – EXEMPLO

Em uma instituição de abrigo aos idosos, a admissão em 12 meses é de 25 pessoas por 1000 habitantes, com valor *lambda* (p) = $25/1000 = 0.025$. Calcular a probabilidade do recebimento de 25 idosos em oito (8) meses.

8.5.2 – INSTRUÇÕES

- Clicar o mouse em *Exponencial*;
- Preencher as *caixas de texto* com os dados;
- Clicar em *Executar*, para obter os resultados.

Figura 8.5
Distribuição
Exponencial.
(Exemplo
8.5.1)

Entrada de dados	
Tamanho da população ou da amostra	1000
Número de eventos observados	25
Unidade de medida temporal dos eventos observados (horas, dias, meses, anos, etc.)	meses
Número de meses nos (nas) quais os eventos foram observados	12
Estimar a probabilidade de ocorrência dos eventos para o seguinte período (em meses)	<input checked="" type="checkbox"/> 8

Resultados		
Média original do evento =	Média do período questionado =	(p) =
0.0250	0.0080	0.1813

Executar Cancelar

A probabilidade do evento ocorrer (25 admissões) em oito (8) meses é inferior a 20% (0.1813).

8.6 – DISTRIBUIÇÃO NORMAL

É um dos mais importantes modelos de distribuição de probabilidades em estatística, sobretudo nas áreas das ciências biológicas e da saúde. A curva desta distribuição apresenta as seguintes características:

- simétrica;
- em forma de sino;
- assintótica: as extremidades se aproximam mas não tocam a linha das abscissas;
- dois pontos de inflexão situados a igual distância da média, um de cada lado desse parâmetro: $\mu \pm 1\sigma$.

Os escores brutos são transformados em escores padronizados (Z), e nessa padronização a média é igual a zero (0) e o desvio padrão igual a um (1). Após o cálculo do valor de Z, o programa determina as probabilidades da distribuição.

A área sob a curva normal apresenta estas proporções, por exemplo:

- $\mu \pm 1.645\sigma = 90\%$ da área, sendo 45% de cada lado da curva;
- $\mu \pm 1.96\sigma = 95\%$ da área, sendo 47.5% de cada lado da curva;
- $\mu \pm 2.33\sigma = 98\%$ da área, sendo 49% de cada lado da curva;
- $\mu \pm 2.58\sigma = 99\%$ da área, sendo 49.5% de cada lado da curva.

8.6.1 – EXEMPLO

Em uma amostra aleatória, o peso de recém-nascidos tem distribuição normal, com média de 3100g e desvio padrão igual a 225g. Considerando-se amostra aleatória de 300 crianças, calcular:

- a) a probabilidade e a respectiva proporção de recém-nascidos pesando entre 2750g e 3200g;
- b) idem, entre 3150g e 3300g;
- c) idem, entre 2659g e 3541g.

8.6.2 – INSTRUÇÕES

- a) Clicar o *mouse* em *Normal*;
- b) Preencher as *caixas de texto*: Tamanho, Média, Desvio padrão e os escores X_1 e X_2 , os quais compreendem o intervalo de valores cuja probabilidade se pretende calcular;
- c) Clicar em *Executar*, para obter os resultados.

Figura 8.6
Distribuição
Normal.
(Exemplo
8.6.1 - a).

Distribuição Normal

Imprimir

Entrada de dados

Tamanho (N ou n)	Média	Desvio padrão
300	3100	225

Área e número de indivíduos entre os escores:

X1	X2
2750	3200

Executar Cancelar

Resultados

$Z_1 = -1.5556$	$Z_2 = 0.4444$
$p(X_1) = 0.0599$	$p(X_2) = 0.3284$
$1 - (p(X_1) + p(X_2)) = 0.6117$	

	$p(X)$	Escore
$p(Z < X_1) =$	0.0599	17.97
$p(X_1 \leq Z \leq X_2) =$	0.6117	183.51
$p(Z > X_2) =$	0.3284	98.52

A Figura acima expressa os valores de Z , as respectivas probabilidades (p) e os escores inferiores a X_1 (17.97), entre X_1 e X_2 (183.51) e superiores a X_2 (98.52). A questão argüida, portanto, corresponde a 184 crianças.

As casas decimais são importantes porque a curva normal representa dados contínuos.

A probabilidade e a respectiva proporção de recém-nascidos pesando entre 3150g e 3300g, conforme argüido no item b, está representada na próxima figura.

Figura 8.7
Distribuição
Normal.
(Exemplo
8.6.1 - b).

Entrada de dados		
Tamanho (N ou n)	Média	Desvio padrão
300	3100	225
Área e número de indivíduos entre os escores:		Executar
X1	X2	Cancelar
3150	3300	
Resultados		
$Z_1 = 0.2222$	$Z_2 = 0.8889$	
$p(X_1) = 0.4121$	$p(X_2) = 0.1870$	
$1 - (p(X_1) + p(X_2)) = 0.4009$		
	$p(X)$	Escores
$p(Z < X_1) =$	0.5879	176.37
$p(X_1 \leq Z \leq X_2) =$	0.2251	67.53
$p(Z > X_2) =$	0.1870	56.10

A Figura 8.7 apresenta os valores de Z , as respectivas probabilidades (p) e os escores inferiores a X_1 (176.37), entre X_1 e X_2 (67.53) e superiores a X_2 (56.10). A resposta à pergunta formulada, assim, corresponde a 68 crianças.

Figura 8.8
Distribuição
Normal.
(Exemplo
8.6.1 - c)

Entrada de dados		
Tamanho (N ou n)	Média	Desvio padrão
300	3100	225
Área e número de indivíduos entre os escores:		Executar
X1	X2	Cancelar
2659	3541	
Resultados		
$Z_1 = -1.9600$	$Z_2 = 1.9600$	
$p(X_1) = 0.0250$	$p(X_2) = 0.0250$	
$1 - (p(X_1) + p(X_2)) = 0.9500$		
	$p(X)$	Escores
$p(Z < X_1) =$	0.0250	7.50
$p(X_1 \leq Z \leq X_2) =$	0.9500	285.00
$p(Z > X_2) =$	0.0250	7.50

A Figura 8.8 exibe os valores de Z , as respectivas probabilidades (p) e os escores inferiores a X_1 (7.50), entre X_1 e X_2 (285) e superiores a X_2 (7.50). A questão formulada (item c), dessa maneira, compreende 285 crianças. Registre-se que os escores escolhidos – 2659 e 3541 – correspondem a valores de Z de -1.96 e $+1.96$, cujas probabilidades de distribuição situam-se nas faixas de 0.025 para cada lado da curva, aquém e além dos escores X_1 e X_2 , respectivamente, totalizando $0.025 + 0.025$

= 0.05 (5%). A área compreendida entre esses valores é igual a $1 - 0.05 = 0.95$ (95%), confirmando, portanto, a informação antes mencionada, ou seja: $\mu \pm 1.96\sigma = 95\%$ da área, sendo 47.5% de cada lado da curva.

8.7 – CRIVO (SCREENING TEST)

Emprega-se em estudos epidemiológicos e em clínica médica como estimadores de determinadas doenças. O mais simples é aquele em que a pessoa testada pode ser classificada como *doença provável* (teste positivo) ou *doença improvável* (teste negativo). A classificação final de cada indivíduo poderá depender, entretanto, do seguimento dos casos, de exames complementares, de procedimentos cirúrgicos e de resultados de necropsia. O aperfeiçoamento cada vez maior dos testes laboratoriais tem aumentado o grau de certeza nos resultados dos *screening tests*.

Há vários estimadores de desempenho de um teste:

- Sensibilidade*: probabilidade de, se o teste for positivo, a pessoa ter a doença;
- Especificidade*: probabilidade de, se o teste for negativo, a pessoa esteja sadia;
- Valor preditivo de um teste positivo*: probabilidade do indivíduo com teste positivo ter a doença;
- Valor preditivo de um teste negativo*: probabilidade do indivíduo com teste negativo estar sadio.

8.7.1 – EXEMPLO

Efetuuou-se um *screening test* para diabetes em 630 indivíduos de ambos os sexos, com idade compreendida entre 18 e 25 anos, considerando-se o teste como positivo se a glicemia fosse igual ou superior a 155 mg/dL, determinada uma hora após refeição padrão. Os resultados foram os seguintes:

Teste positivo em pessoas diabéticas = 61;

Teste positivo em pessoas não diabéticas = 35;

Teste negativo em pessoas diabéticas = 12;

Teste negativo em pessoas não diabéticas = 522.

Figura 8.9
Screening
Test,
(Exemplo
8.7.1).

Entrada de dados		
	Doentes (+)	Não-doentes (-)
Teste (+)	61	35
Teste (-)	12	522

Resultados	
Sensibilidade	= 83.56%
Especificidade	= 93.72%
Falso-positivos (Erro Tipo I)	= 6.28%
Falso-negativos (Erro Tipo II)	= 16.44%
Prevalência	= 0.116 ou 11.59%
Valor preditivo do teste positivo	= 63.54%
Valor preditivo do teste negativo	= 97.75%
Acurácia	= 0.93 ou 92.54%
+LR (Likelihood Ratio) Positiva	= 13.30
-LR (Likelihood Ratio) Negativa	= 0.18

Houve mudança na prevalência ?
☐ SIM ☐ NÃO

Definições

Posicione o mouse sobre os valores para obter as definições.

8.7.2 – INSTRUÇÕES

- Selecionar Probabilidade Condicional e clicar em *Crivo (Screening Test)*;
- Preencher as *caixas de texto* com os dados;
- Clicar em *Executar*, para obter os resultados.

A especificidade do teste é maior que a sensibilidade, ou seja, o valor preditivo do teste negativo é de 97.75%. Portanto, se em uma pessoa o teste for negativo, há quase 100% de chance de não ser diabética. De outro lado, o valor preditivo de um teste positivo é de apenas 63.54%.

Pode haver, contudo, mudança ulterior na prevalência da doença. Neste caso, mantidos os valores da sensibilidade e da especificidade, o valor preditivo será alterado de modo diretamente proporcional à alteração da prevalência. Se esta aumentar, o valor preditivo do teste positivo crescerá, e vice-versa. Suponhamos que a prevalência posterior subiu para 0.25: pressiona-se o *mouse* no botão *SIM*, introduz-se na *caixa de texto* o atual valor desse índice, clicando outra vez em *Executar*.

Figura 8.10
(Exemplo
8.7.1 com
nova
prevalência).

Prevalência = 0.250 ou 25.00%
Valor preditivo do teste positivo = 81.59%
Valor preditivo do teste negativo = 94.48%
Acurácia = 0.91 ou 91.18%

8.8 – CURVA ROC

A Curva ROC (*Receiver Operating Characteristic Curve*) é um método estatístico e gráfico para determinar o melhor ponto de corte (*cutoff point*) de um teste diagnóstico. Representa-se no eixo das ordenadas (Y) a sensibilidade e no eixo das abscissas a proporção de falsos positivos, ou seja, 1-especificidade. O ponto mais elevado da curva, correspondente ao ângulo superior esquerdo do gráfico, representa 100% de sensibilidade e 0% de falsos positivos ($d = 0$), sendo, nesse caso, o valor ideal de uma prova diagnóstica, chamado *padrão ouro*. Podemos testar até várias curvas por este método, para avaliar o ponto de corte de cada uma, medindo-se, para isto, as respectivas áreas e os valores de d .

8.8.1 – EXEMPLO

Uma investigação efetuada para determinar qual é o melhor teste, A ou B, para enfermidade neoplásica, mostrou os resultados abaixo, em exames efetuados em 45 pessoas com a enfermidade e em 45 indivíduos considerados normais:

Figura 8.11
Dados do
Exemplo
8.8.1.

	- 1 - Sens. (A)	- 2 - Espec. (A)	- 3 - Sens. (B)	- 4 - Espec. (B)
1	0.800	0.600	0.900	0.700
2				
3				
4				

8.8.2 – INSTRUÇÕES

- Introduzir os dados no grid geral do *BioEstat*, conforme mostrado na Figura 8.11;
- Selecionar com o *mouse* em *Probabilidade Condicional* e pressionar em Curva ROC;
- Selecionar as colunas do *grid* geral e, em seguida, pressionar *Executar*;
- Informar o número de indivíduos doentes e o número de indivíduos sadios de cada amostra (Figura 8.12) e pressionar o botão *Executar*.

Figura 8.12
Número de
doentes e de
sadios em
cada teste,
Exemplo
8.8.1.

ROC (Receiver Operating Characteristic Curve)
Determina o melhor ponto de corte de um teste diagnóstico.

Informe os quantitativos totais da amostra

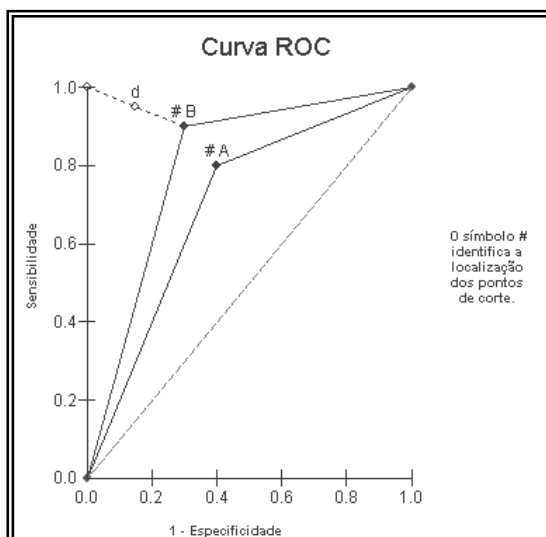
	Doentes	Sadios
Curva A	45	45
Curva B	45	45

Executar Cancelar

Figura 8.13
Resultados
do Exemplo
8.8.1.

Resultados	Curva A	Curva B
Total de 'Doentes' =	45.0	45.0
Total de 'Sadios' =	45.0	45.0
Ponto de corte =	(0.40, 0.80)	(0.30, 0.90)
Distância (d) =	0.45	0.32
Área =	0.7	0.8
Erro padrão =	0.0551	0.0469
IC 95% (Área) =	0.592 a 0.808	0.708 a 0.892

Figura 8.14
A linha
diagonal
representa a
metade da
área. Quanto
mais
próximas
estiverem as
curvas dessa
linha, menor o
desempenho
do teste.



A linha 'd' indica o teste de melhor desempenho, mais próximo do *padrão ouro*, ou seja:

Sensibilidade = 1 e
Especificidade = 0.

O melhor resultado foi obtido com o teste B, com a distância $d = 0.32$.

8.9 – PONTO DE CORTE

Determinar o melhor Ponto de Corte (*Cutoff point*) para um teste diagnóstico.

8.9.1 – EXEMPLO

A Enzima *Alfa-GST* é usada como marcador de lesão hepática causada pela Hepatite C. Foram testados 72 indivíduos, sendo 40 doentes e 32 sadios.

8.9.2 – INSTRUÇÕES

- Introduzir os dados no *grid* (ver Figura 8.15), em duas colunas as quais representam, neste caso, os grupos *Doentes* e *Sadios*, podendo as amostras ter tamanhos diferentes;
- Selecionar no menu principal do *BioEstat*: *Distribuição de Probabilidades*, *Probabilidade Condicional* e clicar em *Ponto de Corte*, nessa ordem;
- O *BioEstat* solicitará que o usuário responda à seguinte pergunta: “O grupo de DOENTES apresenta os valores mais elevados?”. Neste caso a resposta é afirmativa, portanto clicar no botão *Sim* para visualizar a Figura 8.16 que contém a Curva ROC e as especificações: **Ponto de Corte (4.000)** e sua respectiva Distância (d), Sensibilidade, Especificidade e Prevalência.

Figura 8.15
Visão parcial
dos dados
do Exemplo
8.9.1.

- 1 - Doentes	- 2 - Sadios
5.400	3.300
3.700	3.100
6.800	3.200
9.200	3.300

Ponto de Corte



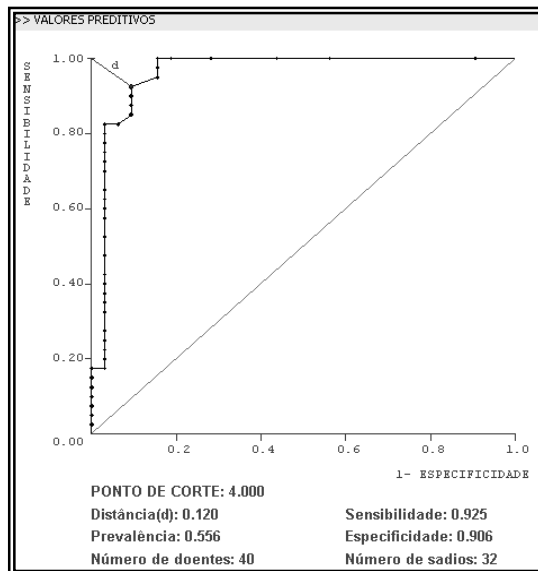
Responda:
O grupo de DOENTES apresenta valores mais elevados

Sim

Não

Será instantaneamente exibido o diagrama abaixo:

Figura 8.16
Curva ROC,
Ponto de
Corte,
Sensibilidade,
Especificidade
e a menor
Distância (d) ao
Padrão Ouro,
Exemplo
8.9.1.



Para obter os Valores Preditivos pressione na parte superior da Figura 8.16.

Figura 8.17
Valores
Preditivos,
Exemplo
8.9.1.

Entrada de Dados (em decimais)		Resultado	
Sensibilidade	0.9250	+LR (Likelihood Ratio) Positiva =	9.84
Especificidade	0.9060	Probabilidade Posterior (+)	92.49 %
Probabilidade prior (Prevalência)	0.5560	Valor Preditivo Positivo (+)	92.49 %
		-LR (Likelihood Ratio) Negativa =	0.08
		Probabilidade Posterior (-)	9.39 %
		Valor Preditivo Negativo (-)	90.61 %
		Acurácia do teste =	91.66 %

Executar Cancelar

8.10 – REGRA DE BAYES

A probabilidade condicional para dois ou mais eventos pode ser obtida pela generalização da regra de Bayes. O programa admite até dez eventos.

8.10.1 – EXEMPLO

Tendo em vista os dados populacionais e as informações clínicas, obtiveram-se os seguintes escores relativos ao sintoma dispnéia (A):

- | | |
|--|---------------------|
| a) Dados da população (prevalência): | b) Dados clínicos: |
| B_1 = Asma brônquica: $\Pr(B_1) = 0.15$ | $\Pr(A B_1) = 0.26$ |
| B_2 = Enfisema pulmonar: $\Pr(B_2) = 0.01$ | $\Pr(A B_2) = 0.75$ |
| B_3 = Câncer do pulmão: $\Pr(B_3) = 0.009$ | $\Pr(A B_3) = 0.43$ |

Dada a ocorrência de dispnéia, calcular as respectivas probabilidades dessas afecções.

8.10.2 – INSTRUÇÕES

- Apontar o mouse em *Probabilidade Condicional* e clicar em *Regra de Bayes*;
- Preencher as caixas de texto da Figura 8.18;

Figura 8.18
Regra de
Bayes
(Exemplo
8.10.1).

Sintoma, sinal ou teste:	Ativar entrada de dados
A =	Dispnéia
Número de eventos:	3
Cancelar	

- Clicar em *Executar*, preenchendo em seguida o *grid* específico com os dados;
- Dirigir o mouse para o menu superior e clicar em *Executar*, para resultados.

Figura 8.19
Exemplo
8.10.1.

Regra de Bayes (A = Dispnéia)				
Novo Imprimir Executar				
Pr	Nome do evento	Pr(Bi) 'a Priori' (em decimais)	Pr(A Bi) 'Condicional' (em decimais)	Resultado: Pr(Bi A) 'a Posteriori'
B1	Asma brônquica	0.15	0.26	0.7743
B2	Enfisema pulmonar	0.01	0.75	0.1489
B3	Câncer do pulmão	0.009	0.43	0.0768

8.11 – SENSIBILIDADE/ESPECIFICIDADE

A probabilidade *posterior* pode ser calculada conhecendo-se a probabilidade *prior* e os novos valores da sensibilidade e da especificidade.

8.11.1 – EXEMPLO

Um determinado evento tem uma probabilidade *prior* 0.45, calculada com informações sobre a sensibilidade e a especificidade de 0.55 e 0.45, respectivamente, obtidas em um *screening test*. Com o aperfeiçoamento dos exames de laboratório, os novos valores da sensibilidade e da especificidade foram, respectivamente, iguais a 0.85 e 0.78. Calcular a probabilidade *posterior* com essas novas informações.

8.11.2 – INSTRUÇÕES

- Apontar o *mouse* em Probabilidade Condicional e clicar em *Sensibilidade/Especificidade* e preencher as *caixas de texto*;
- Clicar em *Executar*, para obter resultado.

Figura 8.20
Sensibilidade/
Especificidade.
(Exemplo
8.11.1).

Entrada de Dados (em decimais)	Resultado
Sensibilidade <input type="text" value="0.8500"/>	+LR (Likelihood Ratio) Positiva = <input type="text" value="3.86"/>
Especificidade <input type="text" value="0.7800"/>	Probabilidade Posterior (+) <input type="text" value="75.97 %"/>
Probabilidade prior (Prevalência) <input type="text" value="0.4500"/>	Valor Preditivo Positivo (+) <input type="text" value="75.97 %"/>
<input type="button" value="Executar"/> <input type="button" value="Cancelar"/>	-LR (Likelihood Ratio) Negativa = <input type="text" value="0.19"/>
	Probabilidade Posterior (-) <input type="text" value="13.60 %"/>
	Valor Preditivo Negativo (-) <input type="text" value="86.40 %"/>
	Acurácia do teste = <input type="text" value="81.15 %"/>

Com os valores da sensibilidade, da especificidade e da prevalência (*prior*), obtêm-se todos os indicadores quantificados na área de resultados da Figura 8.20.

8.12 – DISTRIBUIÇÃO F

A distribuição *F*, como as demais que se seguem, facilita o cálculo do *p-valor* a partir do valor de *F* obtido e de acordo com os graus de liberdade do numerador e do denominador. Trata-se, assim, de um procedimento didático, de vez que o gráfico aproximado da curva respectiva da distribuição é mostrado com destaque para a área de rejeição, o que torna mais fácil a compreensão de um teste *F* pelo usuário.

8.12.1 – EXEMPLO

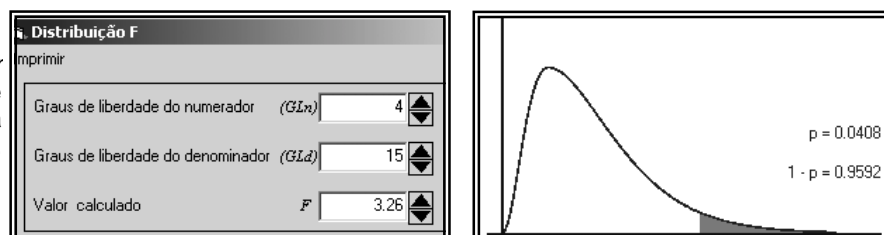
Considerando o valor de $F = 3.26$, obtido em um teste de análise de variância baseada em um critério, com quatro (4) e quinze (15) graus de liberdade do numerador e do denominador, respectivamente, determinar a curva e *p-valor* para esses dados.

8.12.2 – INSTRUÇÕES

- Apontar o *mouse* em *Outras distribuições*;
- Clicar o *mouse* em *Distribuição F*;
- Preencher as *caixas de texto* com os dados do exemplo e clicar em *Executar*.

Figura 8.21

O *p*-valor
obtido é
igual a
0.0408.



8.13 – DISTRIBUIÇÃO DO QUI-QUADRADO

A distribuição do *Qui-Quadrado*, com base no valor de χ^2 obtido e os graus de liberdade, calcula o *p*-valor e mostra, aproximadamente, a respectiva curva.

8.13.1 – EXEMPLO

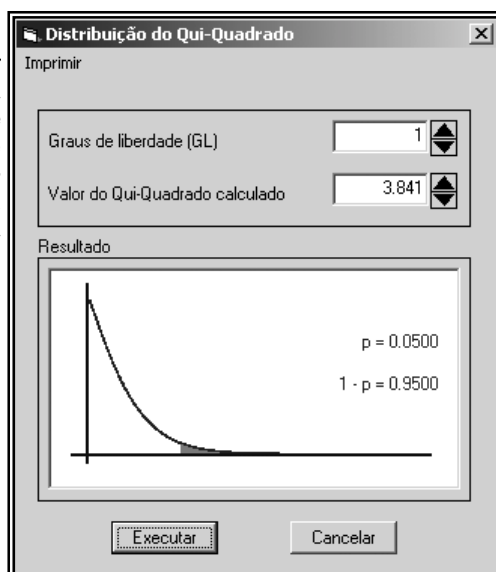
Considerando o valor de $\chi^2 = 3.841$, obtido em uma tabela de contingência 2 x 2, portanto, com um (1) grau de liberdade, calcular o respectivo *p*-valor.

8.13.2 – INSTRUÇÕES

- Apontar o *mouse* em e *Outras distribuições*;
- Clicar o *mouse* em *Distribuição do Qui-Quadrado*;
- Preencher as *caixas de texto* com os dados do exemplo e clicar em *Executar*.

Figura 8.22

O *p*-valor
é igual a
0.05 e
corresponde
ao segmen-
to escuro à
direita da
curva.



8.14 – DISTRIBUIÇÃO t DE STUDENT

A distribuição t de *Student* varia de acordo com o valor de t obtido e os respectivos graus de liberdade, aproximando-se da curva normal à medida que aumentam os graus de liberdade. Os valores alfa unilateral e bilateral são mostrados juntamente com a respectiva curva, esta de modo aproximado.

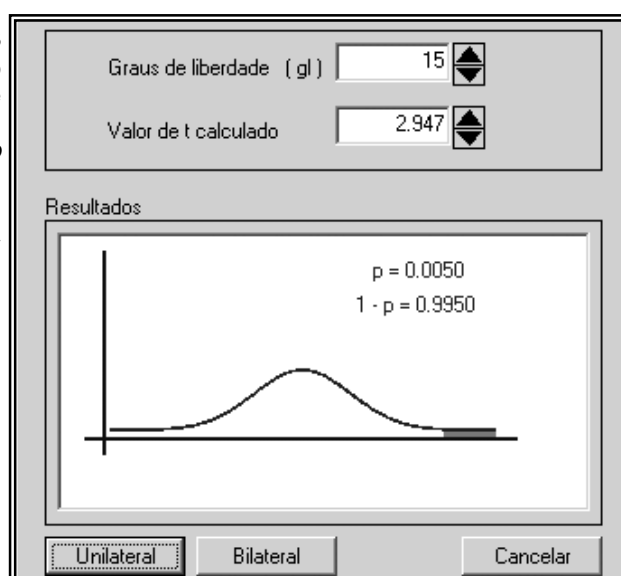
8.14.1 – EXEMPLO

Tendo em vista o valor obtido de $t = 2.947$ e considerando quinze (15) graus de liberdade, calcular o p -valor unilateral.

8.14.2 – INSTRUÇÕES

- Apontar o mouse em *Outras distribuições*;
- Clicar o mouse em *Distribuição t de Student*;
- Preencher as caixas de texto com os dados do exemplo e clicar em *Unilateral*.

Figura 8.23
Distribuição
 t de
Student,
o valor de p
unilateral
obtido é
igual a
0.005.



O resultado unilateral, como era esperado, é igual a 0.005. Se fosse desejado o valor *bilateral*, bastaria pressionar o botão “*Bilateral*”, o qual, com esses dados, o p -valor corresponderia a 0.005 de cada lado da curva, totalizando 0.01 (0.005 + 0.005).

8.15 – DISTRIBUIÇÃO Z

A distribuição Z, como se sabe, corresponde à distribuição normal $N(0,1)$ cujos parâmetros média = 0 e variância = 1. A curva e o *p-valor* são mostrados conforme se verifica na Figura 8.24.

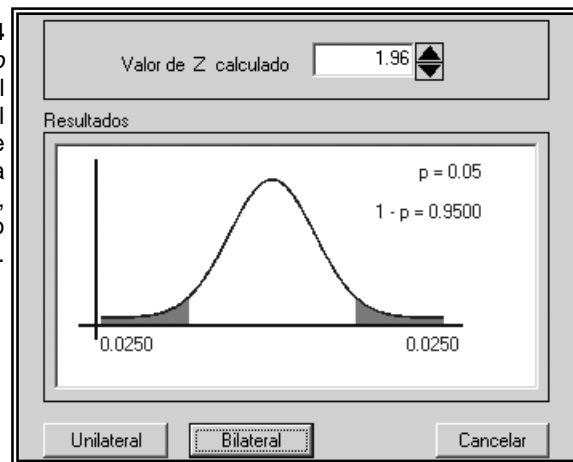
8.15.1 – EXEMPLO

Considerando o valor obtido de $Z = 1.96$, calcular o *p-valor* bilateral.

8.15.2 – INSTRUÇÕES

- Apontar o mouse em *Outras distribuições*;
- Clicar o mouse em *Distribuição Z*;
- Preencher a caixa de texto com o valor de Z e clicar em *Bilateral*.

Figura 8.24
O valor de *p* bilateral obtido é igual a 0.025 de cada lado da curva, totalizando 0.05.



CAPÍTULO 9

UMA AMOSTRA

9.1 – INTRODUÇÃO

Os testes de inferência estatística para uma só amostra envolvem os dados amostrais e os parâmetros da população investigada. Algumas vezes a comparação dos valores numéricos da amostra é efetuada com distribuições teóricas esperadas, recebendo a denominação, nesses casos, de *testes de aderência*.

Estes testes classificam-se em:

- a) *Testes paramétricos*:
 - i – Teste *t*: dados amostrais;
 - ii – Teste *t*: resumo amostral;
 - iii – Teste *Z*: dados amostrais;
 - iv – Teste *Z*: resumo amostral.
- b) *Testes não-paramétricos*:
 - i – Teste Binomial;
 - ii – Teste de Iterações (*Runs test*);
 - iii – Teste *G*;
 - iv – Teste de *Kolmogorov-Smirnov*;
 - v – Teste de *Lilliefors*;
 - vi – Teste do Qui-Quadrado: proporções esperadas iguais;
 - vii – Teste do Qui-Quadrado: proporções esperadas desiguais;
 - viii – Teste de *Poisson*.

9.1.1 – INSTRUÇÕES

- a) Introduzir os dados no *grid* geral do *BioEstat*;
- b) Dirigir o *mouse* para o *menu* superior e clicar em *Estatística*;
- c) Apontar o *mouse* em *Uma Amostra*, conforme demonstra a Figura 9.1:

Figura 9.1
Menu dos aplicativos para uma amostra.

Uma Amostra	Kolmogorov-Smirnov
Duas Amostras Independentes	Lilliefors (k amostras)
Duas Amostras Relacionadas	Qui-Quadrado: Proporções Esperadas Iguais
Estatística Descritiva	Qui-Quadrado: Proporções Esperadas Desiguais
Estimação de Parâmetros	Teste Binomial: Uma Proporção
Meta-Análise	Teste de Poisson
Normalidade	Teste-G (aderência)
Qui-Quadrado	Teste t: Dados Amostrais
Regressão	Teste t: Resumo Amostral
Tamanho das Amostras	Teste Z: Dados Amostrais
Transformação e Ordenação	Teste Z: Resumo Amostral

9.2 – TESTE *t* DE STUDENT: DADOS AMOSTRAIS

Prova paramétrica cuja finalidade é a de comparar a média amostral com a média da população, sendo necessário assumir a média paramétrica do universo em questão, ou

seja, o valor de μ . O teste é efetuado estimando-se a variância da população pela variância amostral (s^2). Esse procedimento é escolhido sobretudo quando o tamanho da amostra é considerado pequeno ($n \leq 30$), porém não há impedimentos para valores amostrais superiores a 30. Deve-se lembrar, todavia, que à medida que aumenta o tamanho amostral, os valores da distribuição t se aproximam dos valores da distribuição Z da curva normal.

9.2.1 – EXEMPLO

Determinaram-se os níveis de fosfatase sérica de 24 crianças com infecção malárica, comparando-se os escores obtidos com o dado paramétrico da população considerado normal para a infância: $\mu = 2.45$ mg/dL.

H_0 : A fosfatase sérica não é alterada durante a infecção malárica: $\mu_1 = \mu_0$;

H_1 : A fosfatase sérica é alterada durante a infecção malárica: $\mu_1 \neq \mu_0$.

Nível de decisão: alfa = 0.01, bilateral.

Figura 9.2
Visão parcial
dos dados do
Exemplo
9.2.1.

	- 1 -	- 2 -
	Fosfatase	
1	3.700	
2	2.800	
3	2.900	
4	2.300	
5	2.400	
6	3.000	
7	2.200	
8	3.400	

9.2.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Teste: Dados amostrais*;
- Selecionar a coluna referente à amostra do *grid* geral;
- Clicar em *Executar o Teste* – surgindo a Figura 9.3 – e preencher então a *caixa de texto* com a média da população.

Figura 9.3
Média da
população.

- Clicar em *Executar* para obter os resultados.

Pelo exame dos resultados (Figura 9.4), conclui-se que a diferença é altamente significativa, eis que o *p-valor* é inferior ao nível alfa previamente estabelecido. Rejeita-se a hipótese de nulidade e aceita-se a alternativa, ou seja, o nível de fosfatase sérica em crianças é alterado - aumentado - pela infecção malárica.

Figura 9.4
Resultados
do Exemplo
9.2.1.

- 1 -	
Tamanho da Amostra	24
Média da População	2.4500
Média Amostral	2.8167
Erro Padrão	0.1063
(t)=	3.4505
Graus de liberdade	23
(p) unilateral =	0.0011
(p) bilateral =	0.0022
Poder (0.05)	0.9645
Poder (0.01)	0.8696
IC 95% (média amostral)=	2.5968 a 3.0365
IC 99% (média amostral)=	2.5184 a 3.1150

9.3 – TESTE *t* DE STUDENT: RESUMO AMOSTRAL

É uma versão do *Teste t de Student* quando dispomos do tamanho, variância e média amostrais e, ainda, o valor da média paramétrica.

9.3.1 – EXEMPLO

Dados coletados da temperatura (graus Centígrados) de 27 caranguejos:

- Tamanho da amostra = 27;
- Média da temperatura dos caranguejos = 24.9°C;
- Variância da amostra = 0.38;
- Média da temperatura do Meio Ambiente (parâmetro) = 25.16°C.

$H_0: \mu = \mu_0$;

$H_1: \mu \neq \mu_0$;

Nível de decisão: $\alpha = 0.05$.

9.3.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Teste t: Resumo Amostral*;
- Introduzir os dados no *grid* Resumo Amostral (Figura 9.5);
- Clicar em *Executar*, para obter os resultados.

Figura 9.5
Entrada de
dados do
Exemplo
9.3.1.

Figura 9.6
Resultados
do Exemplo
9.3.1.

	Resultados
Tamanho da Amostra	27
Média da População	25.1600
Média Amostral	24.9000
Erro Padrão	0.1186
(t) =	-2.1916
Graus de liberdade	26
(p) unilateral =	0.0187
(p) bilateral =	0.0375
Poder (0.05) =	0.7077
Poder (0.01) =	0.4465
IC 95%(média amostral) =	24.6561 a 25.1439
IC 99%(média amostral) =	24.5703 a 25.2297

A diferença é estatisticamente significativa ($p < 0.05$). A temperatura dos animais foi inferior à observada no meio ambiente.

9.4 – TESTE Z: DADOS AMOSTRAIS

Prova paramétrica semelhante à anterior, devendo o procedimento ser escolhido quando a média e a variância da população são conhecidas.

9.4.1 – EXEMPLO

Em um lago criatório foram pesados 50 peixes de determinada espécie, cujos escores estão inseridos no *grid* geral. Em estudos efetuados no hábitat natural, o peso médio dos animais dessa espécie é de 21600g e o desvio padrão igual a 7338g.

Figura 9.7
Visão parcial
dos dados
do Exemplo
9.4.1.

	- 1 -	- 2 -
	Peixes (g)	
1	28000	
2	23200	
3	20800	
4	13300	
5	32600	
6	24600	
7	18700	

H_0 : os animais do lago criatório têm peso igual aos do hábitat natural: $\mu_1 = \mu_0$;

H_1 : os animais do lago criatório têm peso superior aos do hábitat natural: $\mu_1 > \mu_0$;

Nível de decisão: alfa = 0.05, unilateral.

9.4.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Teste Z: Dados amostrais*;
- Selecionar a coluna referente à amostra do *grid* geral;
- Clicar em *Executar o Teste* – surgindo a Figura 9.8 – e preencher então as *caixas de texto* com a média da população e o desvio padrão paramétrico.

Figura 9.8
Média e
desvio padrão
da população
Exemplo
9.4.1.

d) Pressionar *Executar*, para resultados.

Figura 9.9
Resultados
do Exemplo
9.4.1.

Tamanho da amostra	50
Variância	53846244.0000
Desvio Padrão da população	7338.0000
Erro padrão da população	1037.7499
Média amostral	23708.0000
Média da população	21600.0000
(Z)=	2.03
(p) unilateral	0.0212
(p) bilateral	0.0424
Poder (0.05)	0.6504
Poder (0.01)	0.3841
IC 95%(média amostral)=	21674.0102 a 25741.9898
IC 99%(média amostral)=	21030.6052 a 26385.3948

Os resultados revelam que a diferença é significativa ($p_{unilateral} = 0.0212$), inferior, portanto, ao nível de significância previamente estabelecido. Rejeita-se a hipótese de nulidade e aceita-se a alternativa, ou seja, os peixes do lago criatório têm média de peso superior àqueles existentes no habitat natural.

9.5 – TESTE Z: RESUMO AMOSTRAL

É uma versão do *Teste Z* quando dispomos do tamanho e média amostrais e, ainda, os valores da média e do desvio padrão paramétricos.

9.5.1 – EXEMPLO

Dados coletados da altura de uma espécie de planta submetida a determinado fertilizante, comparados com a média da população obtida de plantas da mesma espécie e coletadas no meio ambiente.

- Tamanho da amostra = 53;
- Média da altura das plantas com fertilizante = 54.4 cm;
- Desvio Padrão da população = 8.2 cm;
- Média da População das plantas coletadas no meio ambiente = 51.9 cm.

Hipóteses:

$$H_0: \mu = \mu_0;$$

$$H_1: \mu \neq \mu_0;$$

Nível de decisão: $\alpha = 0.05$.

9.5.2 – INSTRUÇÕES

- Pressionar o *mouse* em *TesteZ: Resumo amostral*;
- Introduzir os dados no *grid* de Resumo amostral (Figura 9.10);
- Clicar em *Executar*, para resultados.

Figura 9.10
Entrada de
dados do
Exemplo
9.5.1.

Média da População	Tamanho da Amostra
51.9	53
Desvio Padrão da População	Média da Amostra
8.2	54.4
<input type="button" value="Cancelar"/> <input type="button" value="Executar"/>	

Figura 9.11
Resultados
do Exemplo
9.5.1

	Resultados
Tamanho da amostra	53
Variância	67.2400
Desvio Padrão da população	8.2000
Erro padrão da população	1.1264
Média amostral	54.4000
Média da população	51.9000
(Z)=	2.22
(p) unilateral	0.0132
(p) bilateral	0.0264
Poder (0.05)=	0.7172
Poder (0.01)=	0.4576
IC (95%) =	52.1923 a 56.6077
IC (99%) =	51.4940 a 57.3060

A diferença é significativa ($p < 0.05$), tendo o fertilizante aumentado o tamanho das plantas submetidas ao tratamento.

9.6 – TESTE BINOMIAL (para uma proporção)

Trata-se de teste não-paramétrico utilizado quando os eventos na população apresentam-se de forma dicotômica ou binária – *sucesso* e *insucesso*. Os valores amostrais são comparados com os dados da população.

9.6.1 – EXEMPLO

Supondo-se que, de 10000 indivíduos com idade entre 50 e 60 anos e com história familiar de câncer gástrico, 350 apresentaram esse tipo de carcinoma. Os estudos demonstraram que a prevalência dessa doença na população é de 0.03. As hipóteses formuladas são estas:

H_0 : a prevalência de câncer gástrico em indivíduos com história familiar dessa neoplasia é igual a da observada na população: $p_1 = p_0$;

H_1 : a prevalência de câncer gástrico em indivíduos com história familiar dessa neoplasia é diferente da observada na população: $p_1 \neq p_0$;

Nível de decisão: alfa = 0.05, bilateral.

9.6.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Teste binomial*;
- Preencher as *caixas de texto* com o tamanho da amostra, o número de *sucessos* e a proporção do evento na população (em decimais).
- Clicar em *Executar*, para resultados.

Figura 9.12
Teste Binomial
(Exemplo 9.6.1).

Entrada de dados		Resultados	
Tamanho da amostra	10000	Z = 2.9311	
No. de eventos na amostra	350	po (popul.)=0.0300	
Proporção do evento na população	0.03	p1 (amostra)=0.0350	
		Unilateral po < p1 ou po > p1	Bilateral po ≠ p1
		p-valor	0.0017
		Poder (0.05)	0.8836

A diferença entre as proporções da amostra e a da população é altamente significativa ($p = 0.0034$) no teste bilateral. Considerando-se que a proporção amostral é superior a da população, rejeita-se a hipótese de nulidade e aceita-se a alternativa, no sentido de que os indivíduos com história familiar de neoplasia gástrica têm maior prevalência dessa enfermidade.

9.7 – ITERAÇÕES (RUNS TEST)

Destina-se a verificar se a amostra em questão foi obtida de maneira aleatória. A técnica deste teste é baseada na *ordem* ou *seqüência* com que os escores originais foram coletados. Os dados podem ser quantitativos ou qualitativos. No caso de binários, como sucesso ou insucesso, sim ou não, masculino ou feminino, sinais + e -, os dados devem ser introduzidos no *grid* com os valores 1 e 0, respectivamente. Este teste está relacionado com a *Mediana*, estabelecendo valores abaixo e acima dessa medida de tendência central. Conhecido também como *Runs test of randomness*.

9.7.1 – EXEMPLO 1

Dados coletados na jogada de uma moeda dezoito (18) vezes: a variável *Cara* foi assinalada no *grid* com o valor um (1), e a variável *Coroa*, com o valor zero (0). Os dados, parciais, podem ser visualizados na Figura 9.12.

H_0 : a seqüência de *Caras* e *Coroas* é aleatória;

H_1 : a seqüência obtida no lançamento da moeda não é aleatória;

Nível de decisão: $\alpha = 0.05$.

Figura 9.12
Visão parcial dos dados do Exemplo 9.7.1.

	- 1 - ** Moeda	- 2 - Dados
1	Coroa	0
2	Cara	1
3	Coroa	0
4	Cara	1
5	Coroa	0
6	Coroa	0
7	Cara	1
8	Cara	1

9.7.2 – INSTRUÇÕES

- Introduzir os dados no *grid* na ordem obtida no experimento (Fig. 9.12);
- Pressionar o mouse em *Iterações (Runs test)*;
- Clicar em *Executar*, para resultados.

Figura 9.13
Resultados
do Exemplo
9.7.1.

Escores	18
n1	10
n2	8
Iterações (r)	13
Valor crítico inferior	5
Valor crítico superior	15
p-valor	ns

Os resultados obtidos revelam que ocorreram dez (10) *Coroas* (n_1) e oito (8) *Caras* (n_2) e treze (13) *Iterações*, sendo o p -valor não significativo (*ns*). A amostra, assim é considerada aleatória. Neste exemplo o tamanho das amostras é menor que 20, note que $n_1=10$ e $n_2 = 8$, neste caso os valores críticos são obtidos a partir da tabela de Frida S. e Eisenhart, C. (1943).

9.7.2 – EXEMPLO 2

Em uma escola observaram-se, uma vez por dia, a discussão entre duas crianças de cada vez, no que diz respeito a determinado conceito social. O grau da intensidade da discussão foi assinalado de 24 pares, totalizando 48 estudantes. O objetivo era determinar se a amostra obtida era ou não aleatória. Os dados coletados poderiam variar numa escala de 0 a 100, conforme a intensidade da discussão de cada grupo de participantes, cujos escores podem ser visualizados na Figura 9.14.

H_0 : a sequência obtida é aleatória;

H_1 : a sequência obtida não é aleatória;

Nível de decisão: $\alpha = 0.05$.

Figura 9.14
Visão parcial
dos dados do
Exemplo
9.7.2.

	- 1 - Intolerância
1	1
2	24
3	37
4	43
5	2
6	42
7	11
8	26
9	40

9.7.2 – INSTRUÇÕES

- Introduzir os dados no *grid* na ordem obtida no experimento (Fig. 9.14);
- Pressionar o mouse em *Iterações (Runs test)*;
- Clicar em *Executar*, para resultados.

Figura 9.15
Resultados
do Exemplo
9.7.2.

Escores	48
n1	23
n2	23
Iterações (r)	33
Valor de Z	2.6839
p-valor	0.0073

Os resultados obtidos revelam que o número de *Iterações* ($r = 33$) foi muito significativo ($p\text{-valor} = 0.0073$), com base na distribuição Z, o que ocorrerá quando um ou os dois escores $n1$ e $n2$ forem superiores a 20 unidades. Conclusão: rejeita-se a hipótese de nulidade, ou seja, a amostra não foi obtida de forma aleatória.

9.8 – TESTE G

Prova não-paramétrica de aderência, semelhante ao teste do Qui-Quadrado, aplicada em amostra com dados mensurados na escala nominal e dispostos em duas ou mais categorias mutuamente exclusivas.

9.8.1 – EXEMPLO

Efetuaram-se 300 lançamentos de um dado cujos escores estão inscritos no *grid* geral. Considerando-se que o dado apresenta seis (6) categorias (ás, duque, terno, quadra, quina e sena) e todas são equiprováveis, então, do ponto de vista teórico deve-se esperar que, em trezentos lançamentos, cada modalidade deve apresentar 50 sucessos. Testar:

H_0 : os valores observados estão de acordo com os teoricamente esperados;

H_1 : os valores esperados diferem da proporção teórica esperada em cada categoria;

Nível de decisão: $\alpha = 0.01$.

Figura 9.16
Dados do
Exemplo
9.8.1.

	- 1 -	- 2 -	- 3 -
	Observado	Esperado	
1	48.000	50.000	
2	57.000	50.000	
3	54.000	50.000	
4	44.000	50.000	
5	45.000	50.000	
6	52.000	50.000	

9.8.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Teste G*;
- Selecionar as colunas do *grid* geral referentes aos dados observados e àqueles teoricamente esperados; e clicar em *Executar*, para resultados:

Figura 9.17
Resultados
do Exemplo
9.8.1.

Número de categorias =	6
Soma das categorias =	300.00
Graus de liberdade =	5
Teste-G =	2.6773
(p) =	0.7496
Teste-G (Williams) =	2.6669
(p) =	0.7512

As proporções observadas concordam com as esperadas, eis que o teste *G ajustado* (Williams) = 2.6669 não é significativo ($p = 0.7512$). O dado pode, assim, ser considerado não-viciado, sem viés.

9.9 – TESTE DE KOLMOGOROV-SMIRNOV

Prova não-paramétrica de aderência destinada a comparar o grau de concordância entre a distribuição acumulada de um conjunto de valores de uma amostra com a distribuição teórica acumulada esperada.

9.9.1 – EXEMPLO

Em uma fazenda registrou-se o grau médio do teor alcóolico de cada safra anual dos vinhos produzidos no período de 1980/89, obtendo-se os seguintes valores: Os dados obtidos em todas as áreas vinícolas da região indicam que o grau médio alcóolico dos vinhos é igual a nove (9), com desvio padrão igual a dois (2).

Figura 9.18
Dados do
Exemplo
9.9.1.

	- 1 - Ano	- 2 - Grau alcóolico	- 3 -
1	1980	12.6	
2	1981	15.4	
3	1982	11.7	
4	1983	8.3	
5	1984	10.2	
6	1985	13.5	
7	1986	14.0	
8	1987	7.9	
9	1988	10.1	
10	1989	12.0	

H_0 : as frequências acumuladas observadas não diferem da distribuição esperada, com base nos dados da população: $F_o(X) = S_n(X)$;

H_1 : as frequências acumuladas observadas diferem da distribuição esperada, com base nos dados da população: $F_o(X) \neq S_n(X)$;

Nível de decisão: alfa = 0.05.

9.9.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Kolmogorov-Smirnov*;
- Selecionar apenas uma coluna do *grid* geral referente aos dados observados, os quais, neste caso, correspondem à coluna 2;
- Clicar em *Executar* e preencher as *caixas de texto* com as informações referentes aos dados paramétricos – média e desvio padrão –, clicando na tecla *OK*, para resultados.

Figura 9.19
Resultados
do Exemplo
9.9.1.

Tamanho da amostra =	10
Desvio máximo =	0.5115
Valor crítico unilateral (0.05) =	0.3690
Valor crítico unilateral (0.01) =	0.4570
p(valor) unilateral	< 0.01
Valor crítico bilateral (0.05) =	0.4090
Valor crítico bilateral (0.01) =	0.4890
p(valor) bilateral	< 0.01

Pelo resultado do teste, rejeita-se a hipótese de nulidade e se aceita a alternativa. O grau alcoólico obtido da referida adega difere do parâmetro da população vinícola da área investigada.

9.10 – TESTE DE LILLIEFORS

Prova não-paramétrica de aderência destinada a comparar o grau de concordância entre a distribuição acumulada de um conjunto de valores de uma amostra com a distribuição teórica acumulada esperada.

9.10.1 – EXEMPLO

O mesmo exemplo anterior – Exemplo 9.9.1 – sobre o grau médio do teor alcoólico de cada safra anual dos vinhos produzidos no período de 1980/89, conforme relação constante na Figura 9.18, desconhecendo-se os parâmetros da população (média e desvio padrão). Este teste pode ser empregado para k amostras de modo simultâneo.

H_0 : as freqüências acumuladas observadas não diferem da distribuição teórica esperada: $F_0(X) = Sn(X)$;

H_1 : as freqüências acumuladas observadas diferem da distribuição teórica esperada: $F_0(X) \neq Sn(X)$;

Nível de decisão: $\alpha = 0.05$.

9.10.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Lilliefors*;
- Selecionar apenas uma coluna do *grid* geral referente aos dados observados, os quais, neste caso, correspondem à coluna 2;
- Clicar em *Executar* para obter resultados.

Figura 9.20
Resultados
do Exemplo
9.10.1

Tamanho da amostra =	10
Desvio máximo =	0.1212
Valor crítico (0.05) =	0.2580
Valor crítico (0.01) =	0.2940
p(valor)	ns

Pelo resultado do teste ($p > 0.05$), aceita-se a hipótese de nulidade e consideram-se os dados amostrais com aderência à normalidade. Pelo modelo de *Lilliefors* o p -valor, dos dados do Exemplo 9.9.1, difere do modelo *Kolmogorov-Smirnov*.

9.11 – TESTE DO QUI-QUADRADO: PROPORÇÕES ESPERADAS IGUAIS

Constitui a prova não-paramétrica de mais largo uso nas áreas de ciências biológicas e médicas. O teste para uma amostra é um procedimento semelhante ao teste G, ou seja, de aderência. A comparação é efetuada entre os escores observados e os esperados, estes iguais, como no exemplo dado a seguir, não sendo necessário introduzi-los no *grid*.

9.11.1 – EXEMPLO

Em 48 filhos de casais constituídos por um cônjuge afetado por anomalia dominante autossômica (Aa) e o outro normal (aa), verificou-se que dezoito (18) descendentes mostravam a anomalia do genitor afetado e, conseqüentemente, 30 não a apresentavam. Pela teoria genética, as proporções esperadas são de 1:1. Os dados devem ser inseridos em uma só coluna do *grid* geral, eis que quando as proporções esperadas são iguais. Testar:

H_0 : as proporções fenotípicas observadas concordam com as esperadas pela Teoria Genética (1:1): $p_1 = p_2$;

H_1 : as proporções fenotípicas observadas não estão de acordo com as esperadas pela Teoria Genética (1:1): $p_1 \neq p_2$;

Nível de decisão: alfa = 0.05.

Figura 9.21

Dados do Exemplo 9.11.1.

	- 1 -	- 2 -
Observado		
1	30.000	
2	18.000	
3		

9.11.2 – INSTRUÇÕES

- Apontar o mouse em *Qui-Quadrado* para *Proporções esperadas iguais*, conforme o Exemplos 9.11.1;
- Selecionar a coluna do *grid* geral e clicar em *Executar o Teste*, para resultados.

Figura 9.22

Resultados do Exemplo 9.11.1.

Soma das Categorias	48.000
Qui-Quadrado	3.00
Graus de Liberdade	1
(p)=	0.0833
Correção de Yates	2.521
(p)=	0.1124

O *Qui-Quadrado* corrigido (Yates) não é significativo ($p = 0.1124$), evidenciando que os valores observados concordam com os esperados pela teoria genética. A diferença existente é, portanto, variação amostral.

9.12 – TESTE DO QUI-QUADRADO: PROPORÇÕES ESPERADAS DESIGUAIS

O teste para uma amostra é um procedimento semelhante ao teste G, ou seja, um teste não-paramétrico de aderência. A comparação é efetuada entre os escores observados e os esperados, estes desiguais, calculados do ponto de vista matemático ou de acordo com alguma teoria.

9.12.1 – EXEMPLO

Em casais com grupos sanguíneos AB x AB, testou-se o fenótipo de 80 descendentes no que diz respeito a esse sistema, observando-se que dezoito (18) pertenciam ao grupo A, 36 ao AB e 26 ao B. Pela teoria genética as proporções esperadas seriam: 1:2:1 (25% : 50% : 25%), dos grupos A, AB e B, respectivamente. Os dados devem ser inseridos em duas colunas do *grid* geral: na primeira coluna os valores observados e na segunda, os esperados, visto que os valores esperados são desiguais.

Figura 9.23
Dados do
Exemplo
9.12.1.

	- 1 -	- 2 -	-
Observado	18.000	20.000	
Esperado	36.000	40.000	
	26.000	20.000	

H_0 : as proporções fenotípicas observadas concordam com as esperadas pela teoria genética (1:2:1): $p_1 = 1/4$; $p_2 = 1/2$; $p_3 = 1/4$;

H_1 : as proporções fenotípicas observadas não estão de acordo com as esperadas pela teoria genética (1:2:1): $p_1 \neq 1/4$; $p_2 \neq 1/2$; $p_3 \neq 1/4$;

Nível de decisão: alfa = 0.05.

9.12.2 – INSTRUÇÕES

- Apontar o mouse em *Qui-Quadrado*, para *Proporções esperadas desiguais*, conforme o Exemplo 9.12.1;
- Selecionar as colunas do grid geral e clicar em *Executar o Teste*, para resultados.

Figura 9.24
Resultados
do Exemplo
9.12.1.

	Resultados
Soma das Categorias	80.000
Qui-Quadrado	2.40
Graus de Liberdade	2
(p)=	0.3012

O Qui-Quadrado calculado não é significativo ($p = 0.3012$), evidenciando que os valores observados concordam com os esperados pela teoria genética. Aceitar a hipótese de nulidade, eis que a diferença existente é, portanto, variação amostral.

9.13 – TESTE DE POISSON

Este teste deve ser empregado quando os dados são binários, como no teste binomial, e se o evento em questão é considerado raro e ocorrido em determinado espaço de tempo.

9.13.1 – EXEMPLO (hipotético)

Uma investigação foi efetuada sobre a ocorrência de linfomas em dez mil indivíduos que trabalhavam em uma indústria de calçados e que foram seguidos durante um (1) ano. O número registrado de casos foi igual a seis (6), enquanto a taxa dessa neoplasia na população, por 10000 habitantes, no mesmo período, obtida dos órgãos de saúde, foi igual a 4.7. Há diferença significativa entre esses dados?

H_0 : $\mu = \mu_0$;

H_1 : $\mu \neq \mu_0$;

Nível de decisão: alfa = 0.05.

9.13.2 – INSTRUÇÕES

- Pressionar o mouse em *Teste de Poisson*;

b) Preencher a *caixa de texto* e pressionar *Executar*, para resultados.

Figura 9.25
Resultados
do Exemplo
9.13.1.

Teste de Poisson

Imprimir

Taxa na população: 4.7 Taxa na amostra: 6

p-valor = 0.6631

Executar Cancelar

Obs: O fator de multiplicação dos Coeficientes, para obter as taxas, deve ser o mesmo.

As proporções observadas concordam com as esperadas, eis que o teste de *Poisson* não é significativo ($p = 0.6631$). A diferença detectada não é importante do ponto de vista epidemiológico, sendo considerada de natureza amostral.

CAPÍTULO 10

DUAS AMOSTRAS INDEPENDENTES

10.1 – INTRODUÇÃO

Um dos testes mais freqüentes em bioestatística consiste na avaliação da *diferença* entre duas *amostras independentes*, que devem representar as respectivas populações, em que os dados de uma não estão relacionados com os escores da outra.

- a) *Testes paramétricos*:
 - i – Teste *t*: dados amostrais;
 - ii – Teste *t*: resumo amostral;
 - iii – Teste *Z*: dados amostrais;
 - iv – Teste *Z*: resumo amostral.
- b) *Testes não-paramétricos*:
 - i – Binomial para Duas Proporções;
 - ii – Exato de *Fisher*;
 - iii – Teste *G*;
 - iv – *Kolmogorov-Smirnov*;
 - v – *Mann-Whitney*;
 - vi – Mediana;
 - vii – Qui-Quadrado;
 - viii – *Odds Ratio*;
 - viii – Risco Relativo;
 - ix – Taxa de Incidência.
 - x – Coeficiente de Variação;
 - xii – *Poisson*;

10.1.1 – INSTRUÇÕES

- a) Introduzir os dados no *grid* geral do *BioEstat* ou preencher *caixas de texto*;
- b) Dirigir o *mouse* para o *menu* superior e clicar em *Estatística*;
- c) Apontar o *mouse* em *Duas Amostras Independentes*.

Figura 10.1
Menu de
Duas
Amostras
Independentes.

Duas Amostras Independentes	▶	Coeficiente de Variação
Duas Amostras Relacionadas	▶	Exato de Fisher
Estatística Descritiva		Kolmogorov-Smirnov (2 x n)
Estimação de Parâmetros	▶	Mann-Whitney (Wilcoxon rank-sum test)
Meta-Análise	▶	Mediana
Normalidade	▶	Odds Ratio
Qui-Quadrado	▶	Qui-Quadrado (Tabelas de Contingência): L x C
Regressão	▶	Risco Relativo
Tamanho das Amostras	▶	Taxa de Incidência
Transformação e Ordenação	▶	Teste Binomial: Duas Proporções
		Teste de Poisson
		Teste G (Tabelas de Contingência): L x C
		Teste t: Dados Amostrais
		Teste t: Resumo Amostral
		Teste Z: Dados Amostrais
		Teste Z: Resumo Amostral

10.2 – TESTE t DE STUDENT: DADOS AMOSTRAIS

Teste paramétrico de largo uso, sobretudo quando o tamanho das amostras – n_1 e n_2 – é igual ou inferior a 30 unidades e as variâncias paramétricas são desconhecidas, cuja distribuição das diferenças entre as médias ($\mu_1 - \mu_2$) foi descrita por W. Gosset (*Student*). Os dados devem ser mensurados a nível intervalar ou de razões e as amostras podem ser de igual tamanho ou desiguais. As seguintes pressuposições devem ser observadas neste teste:

- amostras randômicas de cada população investigada;
- as variâncias devem ser homogêneas;
- as variáveis das populações de onde as amostras foram selecionadas devem apresentar distribuição aproximadamente normal.

A questão das variâncias amostrais é importante no teste t de *Student*. Como as amostras podem ser de tamanho desiguais, é conveniente ponderar as duas variâncias – s_1^2 e s_2^2 – mesmo que a pressuposição tenha sido feita de que elas não são heterogêneas. Essa ponderação conduz a determinado valor conhecido como *pooled variance*.

Outras vezes uma das variâncias é mais de quatro (4) vezes superior à outra ou, então, n_1 e n_2 são muito diferentes. Nesses casos, as variâncias são consideradas desiguais desde que o teste F , realizado também pelo *BioEstat*, seja significativo ($p \leq 0.05$). O valor de t calculado para as duas médias amostrais – \bar{X}_1 e \bar{X}_2 – levará em consideração esses aspectos.

10.2.1 – EXEMPLO (para dados amostrais)

Um pesquisador admite que a estatura dos homens pertencentes ao grupo indígena A é diferente dos indivíduos do sexo masculino concernentes ao grupo indígena B. Foram mensuradas doze (12) pessoas do primeiro agrupamento e onze (11) do segundo. Testar:

Figura 10.2
Dados do
Exemplo
10.2.1.

	- 1 - Estatura A	- 2 - Estatura B
1	1.760	1.720
2	1.740	1.690
3	1.770	1.650
4	1.790	1.660
5	1.760	1.670
6	1.750	1.680
7	1.730	1.710
8	1.800	1.720
9	1.770	1.640
10	1.780	1.670
11	1.690	1.660
12	1.670	

H_0 : não há diferença entre a estatura média dos indivíduos dos grupos A e B: $\mu_1 = \mu_2$;

H_1 : há diferença entre a estatura média dos indivíduos dos grupos A e B: $\mu_1 \neq \mu_2$;

Nível de decisão: alfa = 0.05.

10.2.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Teste t: dados amostrais*;
- Selecionar as colunas referentes às amostras do *grid* geral;
- Clicar em *Executar o Teste* para obter resultados.

Figura 10.3
Resultados
do Exemplo
10.2.1.

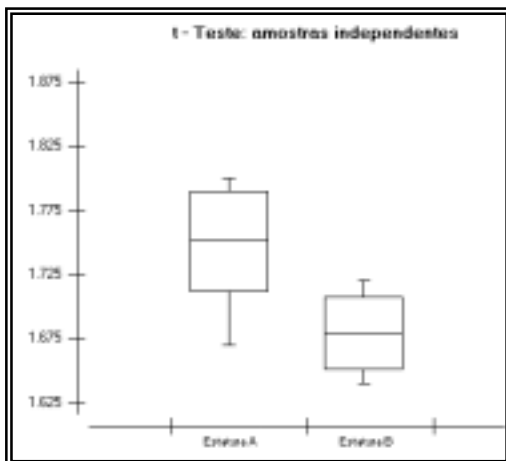
t - Teste: Amostras Independentes			
Arquivo Editar Gráfico			
	- 1 - Estatura A	- 2 - Estatura B	
Tamanho =	12	11	
Média =	1.7508	1.6791	
Variância =	0.0015	0.0008	
	Homocedasticidade		---
Variância =	0.0012		---
t =	5.0647		---
Graus de liberdade =	21		---
p (unilateral) =	< 0.0001		---
p (bilateral) =	< 0.0001		---
Poder (0.05)	0.9997		---
Poder (0.01)	0.9972		---
Diferença entre as médias =	0.0717		---
IC 95% (Dif. entre médias) =	0.0423 a 0.1012		---
IC 99% (Dif. entre médias) =	0.0316 a 0.1118		---

O valor de t (5.0647) é altamente significativo, com p -valor menor que 0.0001 (bilateral). Rejeita-se, assim, a hipótese de nulidade e aceita-se a alternativa: há diferença na estatura dos homens das populações investigadas, sendo em média mais altos aqueles pertencentes ao grupamento indígena A.

10.2.3 – GRÁFICO DO EXEMPLO 10.2.1

Pressionar em *Gráfico* na barra superior da Fig. 10.3.

Figura 10.4
O gráfico é
sugestivo da
diferença
entre as
médias
amostrais
do Exemplo
10.2.1.



10.3 – TESTE t – DUAS AMOSTRAS INDEPENDENTES: RESUMO AMOSTRAL

Trata-se do *Teste t de Student*, quando se dispõe apenas das médias, variâncias e tamanhos amostrais.

10.3.1 – EXEMPLO

Duas amostras de plantas para verificação do crescimento, submetidas a diferentes fertilizantes:

Amostra A: fertilizante XX;

Amostra B: fertilizante YY.

Dados amostrais: contidos no *grid* do Resumo Amostral (Figura 10.5).

$H_0: \mu_1 = \mu_0$;

$H_1: \mu_1 \neq \mu_0$;

Nível de decisão: $\alpha = 0.05$.

10.3.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Teste t: Resumo Amostral*;
- Introduzir os dados nas *caixas de texto* de *Resumo Amostral* (Figura 10.5);
- Clicar em *Executar* para resultados.

Figura 10.5
Entrada de
dados do
Exemplo
10.3.1.

Tamanho da amostra 1:	Tamanho da amostra 2:
11	9
Média 1:	Média 2:
51.4	56.7
Variância 1:	Variância 2:
19.92	9.89
Executar	Cancelar

Figura 10.6
Resultados
do Exemplo
10.3.1.

	Amostra 1	Amostra 2
Tamanho =	11	9
Média =	51.4000	56.7000
Variância =	19.9200	9.8900
F(10, 8) =	2.0142	---
p =	0.3328	---
t =	-2.9988	---
Graus de liberdade =	18	---
p (unilateral) =	0.0038	---
p (bilateral) =	0.0077	---
Poder (0.05) =	0.8631	---
Poder (0.01) =	0.6824	---
Diferença entre as médias =	-5.3000	---
IC 95% (Dif. entre médias) =	-9.0133 a - 1.5867	---
IC 99% (Dif. entre médias) =	-10.3866 a - 0.2134	---

A diferença é muito significativa ($p < 0.01$). O fertilizante YY foi superior ao XX em termos de crescimento das plantas submetidas aos tratamentos.

10.4 – TESTE Z: DADOS AMOSTRAIS

Apresenta características semelhantes ao *Teste t*, mas é indicado para duas amostras independentes quando as variâncias paramétricas são conhecidas e as variáveis nas populações apresentam distribuições normais. Esta última pressuposição pode não ser conhecida, mas o tamanho de cada amostra, nesse caso, deve ser superior a 30 unidades.

10.4.1 – EXEMPLO

Dois produtos comerciais – D1 e D2 –, recomendados para alívio de dores musculares, foram testados em duas amostras independentes de pacientes da mesma idade, do sexo feminino e em condições clínicas semelhantes. Cada grupo estava constituído de 32 indivíduos e o tempo decorrido entre a ingestão das drogas e o alívio do sintoma foi contado em minutos. As variâncias paramétricas dos produtos D1 e D2 são, respectivamente, 1.68 e 1.62. Testar:

H_0 : não há diferença entre as drogas no tempo de alívio da dor muscular: $\mu_1 = \mu_2$;

H_1 : a droga A alivia a dor muscular em um tempo inferior ao da droga B: $\mu_1 < \mu_2$;

Nível de decisão: alfa unilateral = 0.01.

Figura 10.7
Dados
parciais do
Exemplo
10.4.1.

	- 1 - Droga D1	- 2 - Droga D2
1	3.000	7.000
2	4.000	6.000
3	5.000	8.000
4	4.000	7.000
5	6.000	5.000
6	7.000	6.000
7	6.000	5.000
8	5.000	7.000
9	4.000	6.000
10	6.000	8.000
11	5.000	7.000
12	7.000	9.000
13	3.000	7.000
14	8.000	6.000
15	6.000	5.000
16	5.000	4.000
17	4.000	5.000

10.3.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Teste Z*;
- Selecionar as colunas referentes às amostras do *grid* geral e clicar em *Executar o Teste*;
- Preencher as *caixas de texto* (Figura 10.8), com as informações sobre as variâncias paramétricas e pressionar em *Continuar*, para obter resultados (Figura 10.9).

Figura 10.8
Como informar
as variâncias
paramétricas.
Exemplo
10.4.1.

Teste Z - Informe as Variâncias Paramétricas

Variância da Primeira População	Variância da Segunda População
1.68	1.62
Continuar	

Figura 10.9
Resultados
obtidos.
Exemplo
10.4.1.

	- 1 -	- 2 -
Tamanho	32	32
Média amostral	5.0625	6.0000
Variância paramétrica	1.6800	1.6200
Desvio Padrão	1.2961	1.2728
(Z)=	2.9194	---
(p)unilateral =	0.0018	---
(p)bilateral =	0.0035	---
Poder (0.05)	0.8987	---
Poder (0.01)	0.7222	---
Diferença entre as médias =	-0.9375	---
IC 95% (Dif. entre Médias) =	-0.957878	-0.917122
IC 99% (Dif. entre Médias) =	-0.964594	-0.910406

O valor de Z (2.9194) calculado é muito significativo no teste unilateral ($p = 0.0018$), rejeitando-se a hipótese de nulidade e aceitando-se a alternativa. A droga D1 alivia a dor muscular em um período de tempo inferior (5.06 minutos) ao da droga D2. Se desejar o gráfico desse Exemplo, pressionar em *Gráfico* na barra superior da Figura 10.9.

10.5 – TESTE Z – DUAS AMOSTRAS INDEPENDENTES: RESUMO AMOSTRAL

Trata-se do *Teste Z*, quando se dispõe apenas das médias e tamanhos amostrais, e ainda, das variâncias paramétricas.

10.5.1 – EXEMPLO

Duas drogas: A (primeira amostra) e B (segunda amostra) foram testadas em dois grupos de pacientes para alívio de dores nas panturrilhas depois de caminhada de 5000 metros. O tempo decorrido entre a ingestão da droga e o desaparecimento das dores foi registrado em minutos. Os dados amostrais e as respectivas variâncias das populações estão registradas no *grid Resumo Amostral*.

$H_0: \mu_1 = \mu_0$;

$H_1: \mu_1 \neq \mu_0$;

Nível de decisão: $\alpha = 0.05$

10.5.2 – INSTRUÇÕES

- Pressionar o mouse em *Teste z: resumo amostral*;
- Introduzir os dados no *grid* de *Resumo Amostral* (Figura 10.10);
- Clicar em *Executar* para resultados (Figura 10.11).

Figura 10.10
Entrada de
dados do
Exemplo
10.5.1.

Parâmetros	
Variância da Primeira População	Variância da Segunda População
<input type="text" value="5.71"/>	<input type="text" value="6.4"/>
Resumo Amostral	
Tamanho da Primeira Amostra	Tamanho da Segunda Amostra
<input type="text" value="35"/>	<input type="text" value="40"/>
Média da Primeira Amostra	Média da Segunda Amostra
<input type="text" value="5.12"/>	<input type="text" value="6.4"/>
<input type="button" value="Cancelar"/>	<input type="button" value="Executar"/>

Figura 10.11
Resultados
do Exemplo
10.5.1.

	Resultados	
	35	40
Tamanho	35	40
Média amostral	5.1200	6.4000
Variância paramétrica	5.7100	6.4000
Desvio Padrão	2.3896	2.5298
(Z)=	2.2517	---
(p)unilateral =	0.0122	---
(p)bilateral =	0.0243	---
Poder (0.05)	0.6131	---
Poder (0.01)	0.3697	---
Diferença entre as médias =	-1.2800	---
IC 95% (Dif. entre Médias) =	-1.310736	-1.249264
IC 99% (Dif. entre Médias) =	-1.320854	-1.239146

O valor de Z é significativo ($p < 0.05$). O tempo decorrido entre a ingestão das drogas e o alívio do sintoma foi menor com o emprego do medicamento A ($p < 0.05$).

10.6 – TESTE BINOMIAL

Destina-se a testar a diferença entre duas proporções amostrais independentes, procurando-se determinar se a diferença ($p_1 - p_2$) é de tal grandeza que permita rejeitar a hipótese de nulidade. O tamanho de cada amostra deve ser suficiente para a devida aproximação da distribuição binomial à curva normal, ou seja, quando atendidas duas condições: $n_1 \bar{p}_1 \bar{q}_1 \geq 5$ e $n_2 \bar{p}_2 \bar{q}_2 \geq 5$.

10.6.1 – EXEMPLO

Uma pesquisa foi conduzida para estudar o efeito da aspirina no enfarte do miocárdio. Um grupo de 10000 pessoas do sexo masculino, de 40 anos de idade, cada uma recebeu 100 mg da droga por dia e foram observados durante 10 anos. Ocorreram manifestações de enfarte coronariano em quatorze (14). Outro grupo de 9500

indivíduos do mesmo sexo e com a mesma idade foram seguidos por igual período, mas não receberam aspirina, tendo ocorrido 30 casos de enfarte do miocárdio. Testar:

H_0 : a aspirina não influencia a ocorrência da doença coronária: $p_1 = p_2$;

H_1 : a aspirina diminui a incidência da doença coronária: $p_1 < p_2$;

Nível de decisão: alfa unilateral = 0.01.

Os valores $n_1 p_1 q_1 = 10000 \times 0.0014 \times 0.9986 = \mathbf{13.98}$ e $n_2 p_2 q_2 = 9500 \times 0.0032 \times 0.9968 = \mathbf{30.30}$, atendendo, assim, as condições exigidas para este teste.

10.6.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Teste Binomial: duas proporções*;
- Preencher as *caixas de texto* com os dados;
- Clicar em *Executar*, para resultados.

Figura 10.12
Teste
Binomial
(Exemplo
10.6.1).

Entrada de dados		
	Amostra 1	Amostra 2
Tamanho da amostra	10000	9500
No. de sucessos	14	30

Executar Cancelar

Resultados		
Z = -2.5859 p1 = 0.0014 p2 = 0.0032	Unilateral p1 < p2 ou p1 > p2	Bilateral p1 ≠ p2
p-valor	0.0049	0.0097
Poder (0.05)	0.8222	0.7292

A diferença observada é muito significativa ($p_{unilateral} = 0.0049$), rejeitando-se a hipótese de nulidade e aceitando-se a alternativa. O uso prolongado da aspirina diminui a incidência da doença coronária no homem.

10.7 – TESTE EXATO DE FISHER

Prova não-paramétrica com a finalidade de verificar se duas amostras independentes provieram da mesma população. As amostras devem ser classificadas sob dois critérios ou categorias, dispostas em tabelas de contingências 2 x 2, cujos escores amostrais podem ser pequenos, com valor zero (0) inclusive. A probabilidade exata (p -valor) é apresentada sob as formas unilateral e bilateral. A vantagem do *BioEstat* 5.0 é que a soma de todas as caselas pode atingir 600 escores.

10.7.1 – EXEMPLO

Vinte (20) cães, divididos aleatoriamente em dois grupos de dez (10) animais, foram submetidos à gastrectomia experimental, tendo sido efetuada a sutura do coto duodenal em um plano, no primeiro grupo, e em dois, no segundo. O experimento destinou-se a constatar a ocorrência de aderência do epíplo ao duodeno, verificado no pós-operatório imediato:

- a) *primeiro grupo*:
 - i – aderência presente: 9 animais;
 - ii – aderência ausente: 1 animal.
- b) *segundo grupo*:
 - i – aderência presente: 2 animais;
 - ii – aderência ausente: 8 animais.

H_0 : a presença de aderência do epíplo no coto duodenal independe do número de planos de sutura: $p_1 = p_2$;

H_1 : a presença de aderência do epíplo no coto duodenal depende do número de planos de sutura: $p_1 \neq p_2$;

Nível de decisão: alfa bilateral = 0.01.

10.7.2 – INSTRUÇÕES

- a) Pressionar o *mouse* em *Teste Exato de Fisher*;
- b) Preencher as *caixas de texto* com os dados;
- c) Clicar em *Executar*, para resultados.

Figura 10.13
Resultados
do Teste
Exato de
Fisher
(Exemplo
10.7.1).

	1	2	Total / Linha
1	9	1	10
2	2	8	10
Total / Coluna	11	9	20

Executar Sair

Resultados

Unilateral	Bilateral
[p]= 0.0027	[p]= 0.0055

O *p-valor bilateral* é igual a 0.0055, rejeitando-se a hipótese de nulidade e aceitando-se a alternativa. Os animais submetidos ao experimento, cuja sutura foi efetuada em dois planos, apresentaram menor número de aderências ao coto duodenal.

10.8 – TESTE G

Teste não-paramétrico para duas amostras independentes, semelhante em todos os seus aspectos ao do Qui-Quadrado, para dados categóricos. Os escores devem ser mensurados a nível nominal ou ordinal, e as amostras podem apresentar duas ou mais categorias dispostas em tabelas de contingência $l \times c$. Os graus de liberdade são assim calculados: $(l - 1) \times (c - 1)$.

10.8.1 – EXEMPLO

Relação entre sistema secretor ABH em indivíduos normais e com filariose. Foram submetidos ao teste 573 filarióticos, dos quais 472 eram secretores e 101 não-secretores; e 558 controles, sendo 490 secretores e 68 não-secretores.

Figura 10.14
Dados do
Exemplo
10.8.1.

	- 1 -	- 2 -	- 3 -
** Condição	Secretor	Não-Secretor	
1	Filarióticos	472.000	101.00
2	Controles	490.000	68.00

H_0 : a ocorrência de filariose independe do grupo do sistema secretor ABH ao qual pertence o paciente;

H_1 : a ocorrência de filariose está associada ao grupo do sistema secretor ABH ao qual pertence o paciente;

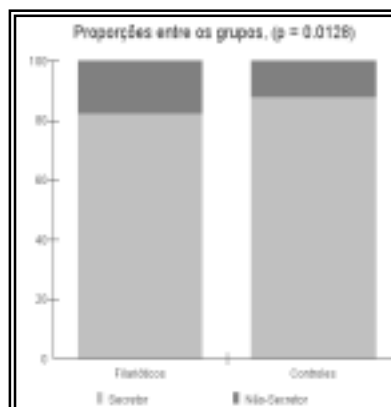
Nível de decisão: alfa = 0.05.

10.8.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Teste G*;
- Selecionar as colunas referentes às amostras do *grid* geral;
- Clicar em *Executar o Teste*, para resultados
- Clicar em *Gráfico* na barra da Figura 10.15.

Figura 10.15
Resultados
do Exemplo
10.8.1 e
respectivo
Gráfico.

Teste-G: Independência	
Arquivo	Editar Gráficos
	Resultados
Tabela de contingência =	2 x 2
Soma das categorias =	1131
Graus de liberdade =	1
Teste-G =	6.6232
(p) =	0.0101
Teste-G (Williams) =	6.6032
(p) =	0.0102
Teste-G (Yates) =	6.1974
(p) =	0.0128



Os resultados mostram associação estatística significativa entre sistema Secretor ABH e filariose, rejeitando-se a hipótese de nulidade e aceitando-se a alternativa.

10.9 – TESTE DE KOLMOGOROV-SMIRNOV

Prova não-paramétrica para duas amostras independentes e várias modalidades, com dados mensurados a níveis nominais, ordinais ou intervalares e retiradas da mesma população ou com a mesma distribuição cumulativa.

10.9.1 – EXEMPLO

Tempo de adaptabilidade à escola de crianças pertencentes às classes Maternal I e Jardim I, com idades médias de três (3) (76 alunos) e quatro (4) anos (63 escolares), respectivamente. A avaliação foi efetuada a nível ordinal, por postos, correspondendo cada linha do *grid* geral a um posto em termos temporais diretamente proporcionais:

Figura 10.16
Dados do
Exemplo
10.9.1.

	- 1 - ** Tempo	- 2 - Maternal	- 3 - Jardim 1	-
1	1 dia	3.000	19.000	
2	3 dias	20.000	20.000	
3	1 semana	12.000	10.000	
4	2 semanas	25.000	12.000	
5	3 semanas	10.000	2.000	
6	4 semanas	6.000	0.000	

H_0 : o período de adaptação à escola independe da idade da criança;

H_1 : o tempo de adaptação à escola depende da idade da criança;

Nível de decisão: $\alpha = 0.05$.

10.9.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Kolmogorov-Smirnov*;
- Selecionar as colunas 2 e 3 do *grid* geral;
- Informar se os dados estão ou não agrupados em classe. Neste exemplo os dados estão agrupados, devendo-se clicar em “*Sim*”;
- Clicar em *Executar o Teste*, para resultados.

Figura 10.17
Resultados
do Exemplo
10.9.1.

Tamanho da primeira amostra	76.000
Tamanho da segunda amostra	63.000
Desvio máximo (bilateral) =	0.3173
Valor crítico (.05) =	0.2317
Valor crítico (.01) =	0.2777
p (bilateral) =	< 0.01
Qui-Quadrado (unilateral) =	13.8678
Graus de liberdade =	2
p (unilateral) =	0.0010

O teste revela que a diferença entre as duas amostras é muito significativa, rejeitando-se a hipótese de nulidade e aceitando-se a alternativa. É mais fácil a adaptabilidade das crianças ao Jardim I que ao Maternal, indicando que a idade influencia o comportamento desses escolares.

10.10 – TESTE DE MANN-WHITNEY

O teste de *Mann-Whitney*, conhecido também como *Teste U* de *Wilcoxon-Mann-Whitney* ou ainda *Wilcoxon rank-sum test*, é uma prova não-paramétrica destinada a comparar duas amostras independentes do mesmo tamanho ou desiguais, cujos escores tenham sido mensurados pelo menos a nível ordinal.

10.10.1 – EXEMPLO

Níveis de fosfatase sangüínea em doze (12) pacientes acometidos de malária por *P. vivax* e em dez (10) parasitados pelo *Plasmodium falciparum*.

H_0 : os níveis de fosfatase sangüínea independem da espécie de *Plasmodium* que parasita o homem;

H_1 : os níveis de fosfatase sangüínea estão relacionados com a espécie de *Plasmodium* encontrada nos doentes de malária;

Nível de decisão: $\alpha = 0.05$.

Figura 10.18
Dados do
Exemplo
10.10.1.

	- 1 - Fosf. (P. vivax)	- 2 - Fosf. (P. falcip)
1	3.70	3.60
2	2.80	2.90
3	2.90	3.00
4	2.30	2.40
5	2.40	2.00
6	3.00	2.50
7	2.20	2.10
8	3.40	2.90
9	2.80	2.70
10	3.20	3.10
11	1.90	
12	3.20	

10.10.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Mann-Whitney*;
- Selecionar as colunas referentes às amostras do *grid* geral;
- Clicar em *Executar o Teste*, para resultados.

Figura 10.19
Resultados
do Exemplo
10.10.1.

Resultado	Amostra 1	Amostra 2
Tamanho da amostra	12	10
Soma dos Postos (R_i)	145.0	108.0
Mediana =	2.85	2.80
U =	53.00	
Z(U) =	0.4616	
p-valor (unilateral) =	0.3222	
p-valor (bilateral) =	0.6444	

O teste U não é estatisticamente significativo ($p = 0.6444$), aceitando-se, portanto, a hipótese de nulidade. Não há diferença nos níveis sangüíneos de fosfatase em pacientes acometidos de malária pelo *P. vivax* ou pelo *P. falciparum*.

10.11 – TESTE DA MEDIANA

O teste da Mediana é um procedimento destinado a comprovar se duas amostras independentes, do mesmo tamanho ou desiguais, provieram da mesma população com o mesmo valor médio, mais precisamente o valor da mediana. Os escores de cada amostra devem ser mensurados pelo menos a nível ordinal.

10.11.1 – EXEMPLO

Após alta hospitalar, doze (12) indivíduos de uma casa de saúde (A) e dez (10) pessoas de outra entidade nosocomial (B) foram avaliados quanto aos cuidados de enfermagem dispensados a cada um nas respectivas instituições. O escore de auto-avaliação de cada paciente, numa escala ordinal de 0 a 100, está contido no *grid* geral do *BioEstat*.

H_0 : o atendimento de enfermagem não difere nos hospitais de onde os indivíduos foram selecionados;

H_1 : o atendimento de enfermagem difere nos hospitais de onde provieram as pessoas selecionadas;

Nível de decisão: $\alpha = 0.05$.

Figura 10.20
Dados do
Exemplo
10.11.1.

	- 1 - Hospital A	- 2 - Hospital B
1	80.0	63.0
2	94.0	57.0
3	92.0	71.0
4	70.0	82.0
5	78.0	70.0
6	83.0	61.0
7	90.0	85.0
8	88.0	66.0

10.11.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Mediana*;
- Selecionar as colunas referentes às amostras do *grid* geral;
- Clicar em *Executar o Teste*, para resultados.

Figura 10.21
Resultados
do Exemplo
10.11.1.

	Amostra 1	Amostra 2
Tamanho =	12	10
Mediana (das amostras) =	81.00	---
Valores > Mediana:	9	2
Valores < Mediana:	3	8
Qui-Quadrado =	4.5833	---
Graus de liberdade =	1	---
(p) =	0.0323	---

O teste da Mediana foi significativo ($p = 0.0323$), rejeitando-se a hipótese de nulidade e aceitando-se a alternativa. Os cuidados de enfermagem do Hospital A foram superiores àqueles oferecidos pelo Hospital B.

10.12 – TESTE DO QUI-QUADRADO

É um dos testes não-paramétricos mais conhecidos e de larga aplicação nas ciências biomédicas, destinando-se a comprovar se duas amostras independentes provieram da mesma população. Os escores devem ser mensurados a nível nominal ou ordinal, e as amostras podem apresentar duas ou mais categorias dispostas em tabelas de contingência $l \times c$.

O Qui-Quadrado sofre a correção de continuidade – *correção de Yates* – quando as amostras apresentam apenas duas categorias (tabelas 2×2), devendo-se escolher outro teste quando o valor de n (soma de todas as modalidades) é inferior a 20 ou se $20 < n < 40$ e qualquer frequência esperada for menor que cinco (5). De acordo com Cochran, se n for igual ou maior que 40, pode-se tolerar valores esperados tão pequenos quanto a unidade (1).

10.12.1 – EXEMPLO

Em rebanho bubalino, parte dos animais recebeu cuidados veterinários, o que não aconteceu com a outra parte. Decorrido um ano, o número de animais vivos e mortos entre os referidos grupos foi o seguinte:

Figura 10.22
Dados do
Exemplo
10.12.1.

	- 1 -	- 2 -	- 3 -
** Condição	C/Cuids. veter.	S/Cuids. veter.	
1	Vivos	88.000	152.000
2	Mortos	12.000	48.000

H_0 : a proporção dos animais vivos e mortos independe dos cuidados veterinários recebidos;

H_1 : a proporção dos animais vivos e mortos depende dos cuidados veterinários recebidos;

Nível de decisão: $\alpha = 0.05$.

10.12.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Qui-Quadrado*;
- Selecionar as colunas referentes às amostras do *grid* geral;
- Clicar em *Executar o Teste*, para resultados;

Figura 10.23
Resultados
do Exemplo
10.12.1.

	Resultados
Tabela de Contingência =	2 x 2
Qui-Quadrado =	6.0
Graus de liberdade =	1
(p) =	0.0143
Correção de Yates =	5.273
(p) =	0.0217

O Qui-Quadrado obtido e devidamente corrigido (*Yates*) é significativo, rejeitando-se a hipótese de nulidade e aceitando-se a alternativa. A frequência de animais vivos ou mortos é dependente dos cuidados veterinários recebidos.

10.13 – ODDS RATIO (OR)

O *Odds Ratio* (OR) é um teste para proporções dispostas em tabela de contingência 2 x 2, ou seja, para duas amostras dicotomizadas, mensuradas a nível nominal, o qual calcula a vantagem (ou desvantagem) de um dos eventos – *sucesso* – em relação ao outro. O programa apresenta também a probabilidade da ocorrência e o intervalo de confiança de 95%. Quando o índice é igual à unidade não há vantagem para quaisquer dos eventos. Indica-se o OR nos estudos de casos nos quais o número de controles é previamente fixado pelo pesquisador, não sendo aconselhável a determinação da incidência do evento.

10.13.1 – EXEMPLO

Efetuuou-se levantamento em um grupo de pessoas do sexo masculino, cujas idades variavam entre 50 e 60 anos, no sentido de verificar a associação de enfisema pulmonar e o hábito do fumo. De 400 indivíduos investigados, durante dez (10) anos, 216 eram não-fumantes (controles), obtendo-se, ao final, os seguintes escores:

Amostra 1 (fumantes): Com enfisema = 120; Sem enfisema = 64;

Amostra 2 (não-fumantes): Com enfisema = 66; Sem enfisema = 150.

Tendo em vista que se trata de um estudo de casos cujos controles foram definidos pelo investigador, pode-se calcular o OR.

10.13.2 – INSTRUÇÕES

- Clicar o mouse em *Odds Ratio*;
- Preencher as *caixas de texto* com os dados;
- Clicar em *Executar*, para obter os resultados.

Figura 10.24
Teste de
Odds Ratio
(Exemplo
10.13.1).

Teste de "Odds Ratio"

Imprimir

Dados

	Sucessos	Insucessos	
Amostra 1	120	64	p1 = 0.6522
Amostra 2 (controle)	66	150	p2 = 0.3056

Executar Cancelar

Resultados

"Odds Ratio" = 4.2614 (p) = < 0.0001

IC 95%
2.8017 ≤ μ ≤ 6.4814

ARR
34.66%

NNH - Número Necessário para causar 1 evento desfavorável no tempo t
3

O Aumento Relativo do Risco é resumido pela sigla ARR e obtido por $(p_1 - p_2)$, todas as vezes que o OR for superior a um (1).

A probabilidade de enfisema pulmonar entre os fumantes é superior a cerca de quatro (4) vezes a dos não-fumantes, com *p-valor* < 0.0001. Quando o *Odds Ratio* calculado atinge valores maiores que um (1), é exibido o Número Necessário para causar um evento desfavorável (NNH). Neste exemplo o NNH = 3. Significa que seriam necessários três indivíduos expostos ao fator de risco para, num período de dez (10) anos, surgir um novo caso da doença (enfisema).

Se os dados fossem introduzidos com inversão da posição das amostras, ter-se-ia:

Figura 10.25
Teste de
Odds Ratio
(Exemplo
10.13.1 com
posição
invertida das
amostras).

Teste de "Odds Ratio"

Imprimir

Dados

	Sucessos	Insucessos	
Amostra 1	66	150	p1 = 0.3056
Amostra 2 (controle)	120	64	p2 = 0.6522

Executar Cancelar

Resultados

"Odds Ratio" = 0.2347 (p) = < 0.0001

IC 95%
0.1543 ≤ μ ≤ 0.3569

ARR
34.66%

NNT - Número Necessário para Tratar e prevenir 1 evento desfavorável no tempo t
3

A leitura passaria a ser diferente: O *Odds Ratio* obtido (0.2347) representa que a chance de enfisema pulmonar ocorrer em não-fumantes é cerca de quatro (4) vezes menor que a dos fumantes.

Quando o *Odds Ratio* obtido é menor que um (1) o *BioEstat* então indica o NNT, ou seja, o Número Necessário para Tratar. Este exemplo, onde o NNT = 3, o resultado deve ser entendido que são necessários 3 indivíduos livres da exposição ao fumo para evitar a ocorrência de um (1) caso de enfisema pulmonar.

10.14 – RISCO RELATIVO (RR)

O *RR* é um teste para duas amostras dicotomizadas e dispostas em tabelas de contingência 2 x 2, no qual se calcula o quociente das proporções p_1 e p_2 . É bastante utilizado para comparar a incidência de eventos em pessoas expostas a determinado fator em relação aos de indivíduos não expostos.

O programa determina também o *p-valor* do *RR* e o intervalo de confiança de 95%. Se o risco relativo se aproxima da unidade, os eventos devem ser considerados independentes, não havendo, por conseguinte, vantagem ou desvantagem entre eles. De outro lado, quanto mais afastado de 1 (um), maior a relação de dependência entre os eventos investigados. Deve ser calculado em ensaios randomizados e estudos de coorte, nos quais um dos grupos de pacientes é submetido a determinado tratamento e o outro sem essa interferência, e que são seguidos durante determinado período de tempo (*follow-up*), a fim de que sejam observadas as ocorrências. O *BioEstat* calcula ainda:

- p_1 : proporção de indivíduos tratados;
- p_2 : proporção de indivíduos controles;
- RR*: **R**isco **R**elativo = p_1/p_2 ;
- RRR*: **R**edução **R**elativa do **R**isco = $1-RR$;
- ARR*: **A**umento **R**elativo do **R**isco: $RR - 1$;
- RAR*: **R**edução **A**bsoluta do **R**isco: $p_2 - p_1$;
- AAR*: **A**umento **A**bsoluto do **R**isco: $p_1 - p_2$;

- h) **NNT**: Número Necessário para Tratar e prevenir 1 (um) evento desfavorável num determinado período de tempo (t) = $1/RAR$ ou $100/(RAR \times 100)$;
- i) **NNH**: Número Necessário para causar 1 (um) evento desfavorável (**Harm**) num determinado período de tempo (t) = $1/AAR$ ou $100/(AAR \times 100)$.

10.14.1 – EXEMPLO (hipotético)

Em um estudo prospectivo realizado numa coorte de dez (10) mil pessoas do sexo masculino, com idade entre 40 e 50 anos, dois grupos foram formados conforme o uso de uma droga do grupo das *Estatinas*: os componentes do grupo A (6500) receberam 10mg diárias do medicamento. Os outros indivíduos (grupo controle, 3500) receberam uma dose diária de placebo. A pesquisa foi efetuada no modelo duplo-cego. Ao final de três (3) anos os resultados sobre a ocorrência de manifestações coronarianas nos dois grupos foi a seguinte:

- a) No grupo A (tratado): a manifestação coronariana ocorreu em 325 indivíduos dentre um total de 6500;
- b) No grupo B (controle): a manifestação coronariana ocorreu em 525 indivíduos dentre um total de 3500.

10.14.2 – INSTRUÇÕES

- a) Clicar o *mouse* em *Risco Relativo*;
- b) Preencher as *caixas de texto* com os dados;
- c) Clicar em *Executar*, para obter os resultados.

Figura 10.26
Risco
Relativo
(Exemplo
10.14.1).

Entrada de dados	Eventos	Tamanho da Amostra	Proporções na Amostra
GRUPO A	325	6500	p1 0.0500
GRUPO B	525	3500	p2 0.1500

Executar Cancelar

Resultados

Risco Relativo: 0.33 (p) = < 0.0001

IC 95%: 0.29 ≤ μ ≤ 0.38

Redução Relativa do Risco: 67.00%

Redução Absoluta do Risco: 10.00%

NNT - Número Necessário para Tratar e prevenir 1 evento desfavorável no tempo t

10

O *Risco Relativo (RR)* de 0.33, muito significativo, determina uma Redução Relativa do Risco (*RRR*) de 67% e uma Redução Absoluta do Risco (*RAR*) de 10%. Desse modo, o Número de pacientes Necessários para Tratar (*NNT*) com *Estatinas* e obter um (1) evento favorável, ou seja, evitar que ocorra um (1) caso de manifestação coronariana, é de dez (10) indivíduos ($NNT = 1/0.1$ ou $100/10$). Em outras palavras, de cada dez (10) indivíduos tratados com *Estatinas* durante três (3) anos, evita-se a manifestação coronariana em 1 (um) deles.

10.14.3 – EXEMPLO (hipotético)

Em um serviço de anestesia foi realizado um levantamento em 95 profissionais observando-se a presença ou ausência do marcador de hepatite B, obtendo-se os seguintes resultados, cujos indivíduos foram divididos conforme o tempo de serviço:

- Grupo com 16 ou mais anos de serviço: 10 apresentaram resultado positivo num total de 20 pessoas;
- Grupo com até 15 anos de serviço: 15 apresentaram resultado positivo entre 75 profissionais.

Deseja-se saber se há associação entre o tempo de serviço em ambiente de Serviços de Anestesia e a presença da hepatite B.

10.14.4 – INSTRUÇÕES

- Clicar o *mouse* em *Risco Relativo*;
- Preencher as *caixas de texto* com os dados;
- Clicar em *Executar*, para obter os resultados.

Figura 10.27
Risco
Relativo
(Exemplo
10.14.3).

Entrada de dados	Eventos	Tamanho da Amostra	Proporções na Amostra
GRUPO A	10	20	p1 0.5000
GRUPO B	15	75	p2 0.2000

Executar Cancelar

Resultados

Risco Relativo: 2.50 (p) = 0.0077

IC 95%
1.33 ≤ μ ≤ 4.69

Aumento Relativo do Risco: 150.00% Aumento Absoluto do Risco: 30.00%

NNH - Número Necessário para causar 1 evento desfavorável no tempo t: 4

O *Risco Relativo (RR)* de 2.5, muito significativo, determina um Aumento Relativo do Risco (*ARR*) de 150% e um Aumento Absoluto do Risco (*AAR*) de 30%. O Número

de profissionais do Serviço de Anestesia Necessários para causar 1 (um) evento desfavorável (*Harm*) é de quatro (4) indivíduos ($NNH = 4$). Em outras palavras, de cada quatro (4) profissionais do Serviço de Anestesia, com mais de dezesseis (16) anos de serviço, pode ocorrer 1 (um) caso de hepatite B.

10.15 – TAXA DE INCIDÊNCIA

Este teste destina-se a verificar se as taxas de incidência (TI) de dois grupos – expostos e não expostos a determinado fator de risco – diferem ou não de modo significativo.

10.15.1 – EXEMPLO

Um levantamento foi efetuado em uma coorte de indivíduos expostos e não expostos à tuberculose no ambiente familiar. Ao final de um ano, obtiveram-se os seguintes dados:

	Número de casos	Número de pessoas no período
Expostos	40	2137
Não Expostos	30	3485

Há diferença nas taxas de incidência desses grupos?

$H_0: TI_1 = TI_2$;

$H_1: TI_1 \neq TI_2$;

Nível de decisão: $\alpha = 0.01$.

10.15.2 – INSTRUÇÕES

- Clicar em *Taxa de Incidência*;
- Preencher as *caixas de texto* com os dados;
- Clicar em *Executar*, para obter os resultados.

Figura 10.28
Resultados
do Exemplo
10.15.1.

	a1	habs / tempo	Total (Linha)
Amostra 1	40	2137	2177
Amostra 2	30	3485	3515
Total (Coluna)	70	5622	5692

Resultados

Valor esperado a1	Valor esperado a2
26.6080	43.3920
Variancia	p-valor
16.4939	0.0015

A diferença é altamente significativa ($p = 0.0015$), havendo maior incidência entre o grupo exposto ao contágio domiciliar com o bacilo responsável pela tuberculose. Os

valores esperados seriam: 27.0 no primeiro grupo e 43.0 no segundo, diferindo de modo acentuado dos valores observados, 40 e 30, respectivamente.

10.16 – COEFICIENTE DE VARIAÇÃO

Este teste permite comparar dois Coeficientes de Variação (CV) de dados mensurados de variáveis diferentes, por exemplo: peso (kg) e estatura (cm) de pré-escolares, largura (cm) e peso (g) de folhas, etc. Deve-se pressupor distribuição normal para os dados.

10.16.1 – EXEMPLO

Um estudo foi efetuado em uma maternidade, registrando-se o peso e o comprimento de 40 recém-nascidos do sexo feminino, obtendo-se os dados inseridos no *grid* geral, mostrados, em parte, na Figura 10.29.

H_0 : a variabilidade inerente ao peso é a mesma variabilidade inerente ao comprimento das crianças recém-nascidas do sexo feminino;

H_1 : a variabilidade inerente ao peso não é a mesma da variabilidade inerente ao comprimento das crianças recém-nascidas do sexo feminino;

Nível de decisão: $\alpha = 0.05$.

10.16.2 – INSTRUÇÕES

- Clicar em *Coeficiente de Variação*;
- Selecionar as colunas do *grid* geral;
- Clicar em *Executar*, para obter os resultados.

Figura 10.29
Visão parcial dos dados do Exemplo 10.16.1.

	- 1 - Peso (g)	- 2 - Comprimento (cm)
1	3200	50
2	3000	49
3	3100	53
4	2950	51
5	2900	48
6	3300	47
7	3125	50

Figura 10.30
Coeficientes de Variação: Resultados do Exemplo 10.16.1.

Resultados	- 1 - Peso	- 2 - Comprimento
Tamanho das amostras	40	40
Médias	3095.5000	50.5250
Desvios padrões	178.9062	2.0753
Coefs. de variação	0.0578	0.0411
Z =	2.1070	---
p =	0.0351	---

Os Coeficientes de Variação do peso ($CV = 0.0578$) e do comprimento ($CV = 0.0411$) das crianças recém-nascidas do sexo feminino mostram maior variabilidade intrínseca da primeira variável, com um *p-valor* significativo (0.0351). Rejeita-se, portanto, a hipótese de nulidade.

10.17 – TESTE DE POISSON

De uma população com distribuição de *Poisson*, pode-se testar, através de dois valores amostrais (Y_1 e Y_2), se eles são oriundos do mesmo universo.

10.17.1 – EXEMPLO (hipotético)

Duas capivaras foram capturadas de modo randômico, sendo uma do sexo masculino e outra do sexo feminino. Foi contado o número de coleópteros que parasitavam cada animal, os quais apresentavam distribuição de *Poisson*. No animal do sexo masculino observaram-se 26 parasitas, enquanto o registrado na capivara fêmea foi igual a 12. Testar a hipótese de que a quantidade de parasitas encontrados, em termos médios, está relacionada com o sexo dos animais capturados.

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

Nível de decisão: alfa bilateral = 0.05.

10.17.2 – INSTRUÇÕES

- Clicar o *mouse* em *Teste de Poisson*;
- Preencher as *caixas de texto* com os valores da primeira e da segunda amostra;
- Clicar em *Executar* para resultados.

Figura 10.31
Teste de
Poisson
(Exemplo
10.17.1).

Unidades na 1ª amostra

Unidades na 2ª amostra

Z = 2.2880

p (unilateral) = 0.0111

p (bilateral) = 0.0221

O valor de Z obtido é estatisticamente significativo ao nível alfa previamente fixado ($p = 0.0221$), rejeitando-se a hipótese de nulidade. O número de parasitas encontrados em cada animal é diferente, podendo-se admitir que a parasitose é mais intensa nas capivaras do sexo masculino.

CAPÍTULO 11

DUAS AMOSTRAS RELACIONADAS

11.1 – INTRODUÇÃO

É um procedimento onde dois conjuntos são investigados, o segundo constituído dos mesmos elementos do primeiro, formando-se pares de unidades onde cada indivíduo contribui com dois escores, ou seja, funciona como seu próprio controle. São, pois, amostras relacionadas, pareadas ou dependentes, tomadas para testar-se a média das diferenças *antes* e *depois* da condição objeto da pesquisa.

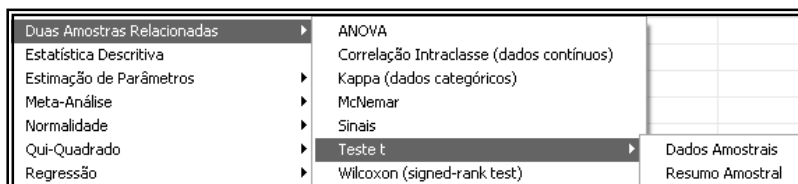
Os testes estatísticos para duas amostras relacionadas classificam-se em:

- a) *Testes paramétricos*:
 - i – Teste *t* de *Student*: dados amostrais;
 - ii – Teste *t* de *Student*: resumo amostral;
 - iii – ANOVA.
- b) *Testes não-paramétricos*:
 - i – Correlação Intraclass;
 - ii – *Kappa*;
 - iii – *McNemar*;
 - iv – Sinais;
 - v – *Wilcoxon (Signed-Rank Test)*.

11.1.1 – INSTRUÇÕES

- a) Preencher duas colunas do *grid* geral com os dados: *antes* e *depois*, com exceção do *Teste de McNemar*;
- b) Dirigir o *mouse* para o *menu* superior e clicar em *Estatística*;
- c) Apontar o *mouse* em *Dois Amostras Relacionadas*, para a escolha do teste:

Figura 11.1
Menu dos aplicativos para duas amostras relacionadas.



11.2 – TESTE *t* DE *STUDENT*: DADOS AMOSTRAIS

Teste paramétrico destinado a verificar se os dois grupos de escores dos mesmos indivíduos, cujas unidades foram retiradas ao acaso da população, não apresentam diferença em relação às médias. Os dados devem ser mensurados em escala intervalar ou de razões, sendo o teste empregado quando a variância da população das diferenças é desconhecida, com $n - 1$ graus de liberdade, sendo n o número de pares.

11.2.1 – EXEMPLO

Em um grupo de dez (10) pacientes, infectados pelo *P. vivax*, contou-se o número de plaquetas por mL do sangue periférico em duas ocasiões: no momento do diagnóstico e após o tratamento. Os dados foram inseridos no *grid* geral de entrada de dados.

Figura 11.2
Dados do
Exemplo
11.2.1.

	- 1 - Plaq.(antes)	- 2 - Plaq.(depois)	- 3 -
1	80300	109000	
2	217000	312000	
3	102000	138000	
4	101000	175000	
5	115000	179400	
6	132000	164220	
7	149600	207900	
8	158400	187500	
9	125000	178000	
10	122000	146900	

H_0 : a infecção não altera o número de plaquetas no sangue periférico: $\mu_D = 0$;

H_1 : a infecção altera o número de plaquetas no sangue: $\mu_D \neq 0$;

Nível de decisão: $\alpha = 0.01$.

11.2.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Teste t: dados amostrais*;
- Selecionar as colunas referentes às amostras do *grid* geral;
- Clicar em *Executar o Teste*, para resultados.

Figura 11.3
Resultados
do Exemplo
11.2.1.

	- 1 -	- 2 -
Indivíduos	10	10
Média	130230.0000	179792.0000
Desvio Padrão	38256.8527	54198.4069
Erro Padrão	12097.8791	17139.0411
Desv. Padrão da Diferença	23373.7031	---
Erro Padrão da Diferença	7391.4139	---
Média das diferenças	-49562.0000	---
(t)=	-6.7053	---
Graus de Liberdade	9	---
(p) unilateral =	< 0.0001	---
(p) bilateral =	< 0.0001	---
IC (95%)	-66281.3783 a -32842.6217	---
IC (99%)	-73584.0952 a -25539.9048	---

O valor de t calculado é igual a -6.7053, com $p < 0.0001$, ou seja, a diferença observada é estatisticamente muito significativa, rejeitando-se a hipótese de nulidade e aceitando-se a alternativa. Como o valor de t é negativo, deduz-se que a contagem plaquetária antes do tratamento era bem inferior àquela obtida após a terapêutica contra a infecção malárica, responsável pela diminuição desses elementos figurados do sangue periférico do homem.

11.3 – TESTE t - DUAS AMOSTRAS RELACIONADAS: RESUMO AMOSTRAL

11.3.1 – EXEMPLO

Um psicólogo acredita que a parte lateral do hipotálamo esteja envolvida no comportamento do apetite. Dez (10) ratos foram submetidos a uma cirurgia de implante de eletrodos: em cada animal foram implantados dois (2) eletrodos: um na parte lateral do hipotálamo e o outro em um área comprovadamente sem qualquer efeito sobre o apetite. Cada rato foi submetido à estimulação durante 30 minutos em cada eletrodo, com período de repouso entre os estímulos. No período de estimulação de cada uma das áreas foi determinada a quantidade de alimento ingerido por animal, com os seguintes resultados:

- Tamanho da amostra (n) = 10;
- Média após estímulo do hipotálamo: 17.3g
- Média após estímulo da área neutra: 11.89g
- Média das diferenças: 5.41g;
- Desvio padrão das diferenças: 3.31g.

$H_0: \mu_1 = \mu_0$;

$H_1: \mu_1 \neq \mu_0$;

Nível de decisão: $\alpha = 0.01$.

Figura 11.4
Dados
do Exemplo
11.3.1

Tamanho da Amostra (Número de Pares)	Média das Diferenças	Desvio Padrão das Diferenças
10	5.41	3.31

Cancelar Executar

11.3.2 – INSTRUÇÕES

- Selecionar no menu:
 - Duas Amostras Relacionadas;
 - Teste t: resumo amostral;
- Introduzir os dados no formulário (Figura 11.4) e clicar em *Executar*;

Figura 11.5
Resultados
do Exemplo
11.3.1.

Número de pares	10
Média das diferenças	5.4100
Desvio Padrão da diferença	3.3100
Erro Padrão da diferença	1.0467
t de Student	5.1686
Graus de liberdade	9
p(unilateral)	0.0003
p(bilateral)	0.0006
IC 95%	3.0423 a 7.7777
IC 99%	2.0082 a 8.8118

A estimulação elétrica do hipotálamo afeta o comportamento alimentar dos ratos ($p < 0.01$), rejeitando-se, portanto, a H_0 .

11.4 – ANOVA

Tem as mesmas características do teste *t* de Student para amostras pareadas, mas ao invés de utilizar a média das diferenças entre os pares, emprega a análise da variância entre os blocos.

11.4.1 – EXEMPLO

Determinaram-se os níveis sanguíneos de triglicerídios em doze (12) pessoas em duas ocasiões: a) em jejum e b) quatro (4) horas após o almoço. Testar:

H_0 : não há diferença entre os níveis de triglicerídios em jejum e quatro (4) horas após o almoço: $\mu_D = 0$;

H_1 : ocorre aumento dos níveis de triglicerídios após o almoço quando comparados com os valores obtidos em jejum: $\mu_D > 0$;

Nível de decisão: alfa = 0.01, unilateral.

Figura 11.6
Dados do
Exemplo
11.4.1.

	- 1 - Jejum	- 2 - Pós-prandial	- 3 -
1	145.000	183.000	
2	139.000	192.000	
3	176.000	210.000	
4	152.000	250.000	
5	181.000	240.000	
6	132.000	197.000	
7	138.000	203.000	
8	161.000	235.000	
9	123.000	170.000	
10	155.000	230.000	
11	158.000	245.000	
12	150.000	270.000	

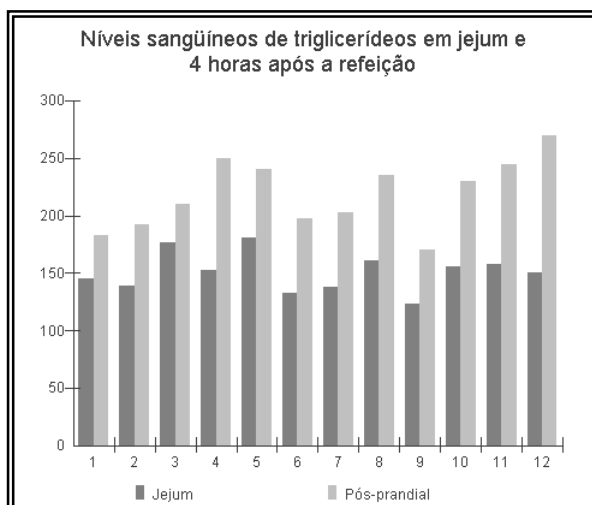
11.4.2 – INSTRUÇÕES

- Pressionar o *mouse* em ANOVA;
- Selecionar as colunas referentes às amostras do *grid* geral;
- Clicar em *Executar o Teste*, para resultados.

Figura 11.7
Resultados do
Exemplo
11.4.1.

Arquivo Editar Gráfico			
FONTES DE VARIAÇÃO	GL	SQ	QM
Tratamentos	1	27676.0417	27676.0417
Indivíduos	11	10052.4583	913.8598
Erro	11	3435.4583	312.3144
F (Tratamentos) =	88.6160		
(p) =	< 0.0001		
F (Indivíduos) =	2.9261		
(p) =	0.0444		

Figura 11.8
Gráfico de Colunas Justapostas do Exemplo 11.4.1.



O F obtido (88.6160) é altamente significativo ($p < 0.0001$), para 1,11 graus de liberdade, rejeitando-se a hipótese de nulidade e aceitando-se a alternativa. Os níveis de triglicerídios são maiores quatro (4) horas após o almoço do que os observados em jejum. Se o teste t fosse realizado, obter-se-ia um valor calculado igual à raiz quadrada de F . Pode-se, ainda, visualizar o gráfico desses dados, exibidos na Figura 11.8, bastando clicar na barra superior – *Gráfico* – da Figura 11.7, o qual ilustra a relação entre os níveis de triglicerídios em jejum e quatro (4) horas após refeição.

11.5 – CORRELAÇÃO INTRACLASSE: DADOS CONTÍNUOS

A correlação Intraclass é um aplicativo para testar a replicabilidade amostral para dados contínuos, não sendo tratadas as variáveis como X e Y mas sim como Replicação 1 e Replicação 2. É, desse modo, medida de replicabilidade amostral.

11.5.1 – EXEMPLO

Foram determinados os níveis de Testosterona Livre em seis (6) pacientes, tendo sido retiradas amostras sanguíneas de cada indivíduo, sendo cada amostra dividida em dois frascos distintos (A e B), os quais foram enviados, aleatoriamente, a dois laboratórios, tornando-se impossível distinguir para os técnicos laboratoriais a Replicação 1 da Replicação 2.

Figura 11.9
Dados do Exemplo 11.5.1.

	- 1 - Replic. 1	- 2 - Replic. 2
1	10.1	10.3
2	11.3	10.9
3	12.8	12.5
4	11.7	11.3
5	10.9	10.8
6	13.4	13.6

11.5.2 – INSTRUÇÕES

- Introduzir os dados no *grid* do *BioEstat* (Figura 11.9);
- Clicar em Duas Amostras Relacionadas e, a seguir, em *Replicabilidade*;
- Pressionar o *mouse* em *Correlação Intraclass*;
- Clicar em *Executar*, para Resultados;

Figura 11.10
Dados do
Exemplo
11.5.1.

Variação entre os grupos	Erro experimental
0.0214	0.0003
GL1	GL2
5	6
F	p-valor
68.3370	p = 0.0002
Correlação Intraclass	
0.9712	
IC 95% Lim. Inferior	IC 95% Lim. Superior
0.8388	0.9958
Conclusão segundo Fleiss, J. L. (1986)	
Replicabilidade Excelente	

A *Replicabilidade* foi considerada Excelente, com valor da Correlação Intraclass igual a 0.9712 e o *p-valor* = 0.002. Nos seis (6) pacientes, os níveis de Testosterona Livre em ambos os laboratórios podem ser considerados semelhantes, segundo resultados fornecidos pelos dois laboratórios.

11.6 – TESTE DE McNEMAR

Teste não-paramétrico para duas proporções, no qual procura-se avaliar o grau de discordância de dois tratamentos a que foram submetidos os mesmos indivíduos. Os dados devem ser mensurados a nível nominal ou ordinal e são dispostos em tabelas de contingência 2 x 2, conforme definido a seguir:

A (+ +)	B (+ -)
C (- +)	D (- -)

A forma mais usual de realizar a entrada de dados é introduzir nas caselas A e D os pares concordantes. As caselas (B e C) representam os pares discordantes de interesse no teste. O programa, entretanto, está disposto de maneira que pode haver inversão na introdução dos dados, a saber:

A (+ -)	B (+ +)
C (- -)	D (- +)

Desta forma, portanto, os pares concordantes podem ser inseridos nas caselas B e C ou nas caselas A e D, e o programa calcula a probabilidade desses pares em ambas as situações, devendo o pesquisador selecionar o resultado de acordo com o modo de

inserir os escores nas *caixas de texto*. Quando ao soma dos pares discordantes for menor que 20, devendo seus valores serem introduzidos nas caselas **B** e **C**, o programa calcula o *p-valor* pelo teste das Proporções Correlacionadas, ou seja, o Teste Exato.

11.6.1 – EXEMPLO 1

Mediu-se a pressão arterial sistólica em 620 indivíduos utilizando-se o esfigmomanômetro convencional na primeira medição e um modelo eletrônico na segunda. Considerou-se normotensa a pessoa cuja pressão sistólica era inferior a 150 mmHg. Os resultados foram estes:

- hipertenso com ambos medidores: 240 (+ +);
- hipertenso com modelo eletrônico e normotenso com o convencional: 100 (+ –);
- normotenso com o modelo eletrônico e hipertenso com o convencional: 60 (– +);
- normotenso com ambos aparelhos: 220 (– –).

H_0 : não há discordância nas proporções dos resultados da pressão arterial sistólica medida pelo esfigmomanômetro convencional e pelo modelo eletrônico;

H_1 : há discordância nas proporções dos resultados da pressão arterial sistólica medida pelo esfigmomanômetro convencional e pelo modelo eletrônico;

Nível de decisão: $\alpha = 0.01$.

11.6.2 – INSTRUÇÕES

- Pressionar o *mouse* em *McNemar*;
- Preencher as *caixas de texto* com os escores (Figura 11.11);
- Clicar em *Executar*, para resultados.

Figura 11.11
Teste de
McNemar
Exemplo
11.6.1.

Dados			
A	240	C	100
B	60	D	220
Total Linha	300	320	620
<input type="button" value="Executar"/> <input type="button" value="Sair"/>			
Resultados			
Qui-Quadrado, se (A/D) são pares discordantes		Qui-Quadrado, se (B/C) são pares discordantes	
0.7848		9.5063	
p (A/D) =		p (B/C) =	
0.3757		0.0020	

Os pares discordantes foram inseridos nas caselas B e C e o resultado do teste é altamente significativo ($p = 0.0020$), rejeitando-se a hipótese de nulidade e aceitando-se a alternativa. Há discordância nos resultados das medidas de pressão arterial

sistólica entre os modelos utilizados na investigação, tendo ocorrido maior número de pessoas consideradas hipertensas com o aparelho eletrônico.

11.6.3 – EXEMPLO 2 (Hipotético)

Determinou-se a presença ou a ausência de cárie dentária em **30** pessoas do sexo masculino, verificando-se, simultaneamente se esses indivíduos eram Secretores ou Não-Secretores. Os resultados foram estes:

- presença de cárie dentária e Secretor: **5** (+ +);
- presença de cárie dentária e Não-Secretor: **16** (+ -);
- ausência de cárie dentária e Secretor: **3** (- +);
- ausência de cárie dentária e Não-Secretor: **6** (- -).

Tendo em vista que o Total de Pares Discordantes é igual a dezenove (19) (16+3), sendo neste caso $B+C < 20$, o programa é ainda mais preciso, pois efetua o Teste de McNemar pelo Método Exato.

H_0 : não há discordância nas proporções dos portadores de cáries dentárias em relação ao Sistema Secretor;

H_1 : há discordância nas proporções dos portadores de cáries dentárias em relação ao Sistema Secretor;

Nível de decisão: $\alpha = 0.05$.

11.6.2 – INSTRUÇÕES

- Pressionar o mouse em *McNemar*;
- Preencher as *caixas de texto* com os escores (Figura 11.12), introduzindo os pares discordantes nas caselas B e C;
- Clicar em *Executar*, para resultados.

Figura 11.12
Teste de
McNemar
Exemplo
11.6.1.

Dados			
A	5	Total / Linha	
C	16		
B	3	Total / Linha	
D	6		
Total Coluna	8	22	30
<input type="button" value="Executar"/> <input type="button" value="Sair"/>			
Resultados			
Qui-Quadrado, se (A/D) são pares discordantes		Qui-Quadrado, se (B/C) são pares discordantes	
—		—	
p (A/D) =		p (B/C) =	
—		0.0044	
Calculado pelo Método Exato.			

Considerando-se o *p-valor* (0.0044), rejeita-se H_0 , concluindo-se que há associação entre a presença de cárie dentária e Sistema Secretor.

11.7 – TESTE DOS SINAIS

Teste não-paramétrico onde se define, em cada par, a diferença dos escores *antes* e *depois* do procedimento investigatório, baseando-se apenas no sentido das diferenças dos escores, registrados sob a forma de *sinais*: (+) e (-). Os pares com resultados iguais - considerados *empatados* -, são eliminados dos cálculos pelo programa. Quando o número de pares (não empatados) for maior que 30, o programa calcula a probabilidade pela curva normal $N(0,1)$, caso contrário, utiliza-se a distribuição binomial.

11.7.1 – EXEMPLO

Em vinte (20) voluntários testou-se o tempo de reação a determinado estímulo *antes* e *após* exercício físico prolongado, registrado em segundos.

H_0 : não há diferença no tempo de resposta ao estímulo utilizado *antes* e *após* o exercício físico prolongado: $p = q = \frac{1}{2}$;

H_1 : há diferença no tempo de resposta ao estímulo utilizado *antes* e *após* o exercício físico prolongado: $p \neq q$;

Nível de decisão: $\alpha = 0.05$, bilateral.

Figura 11.13
Visão parcial
dos Dados
do Exemplo
11.7.1.

	- 1 - Antes	- 2 - Após	- 3 -	
1	59.000	71.000		
2	56.000	63.000		
3	67.000	65.000		
4	62.000	68.000		
5	55.000	57.000		
6	53.000	63.000		
7	61.000	59.000		
8	58.000	62.000		

11.7.2 – INSTRUÇÕES

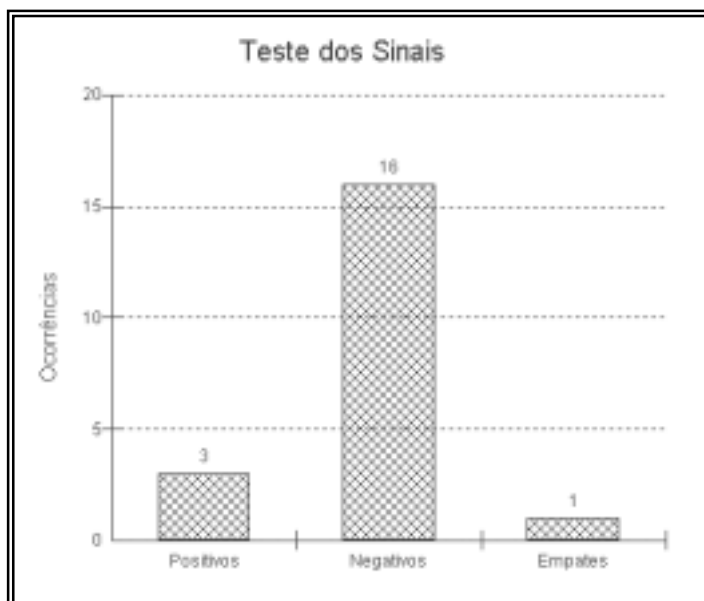
- Pressionar o *mouse* em *Sinais*;
- Selecionar as colunas do *grid* geral;
- Clicar em *Executar o Teste*, para resultados;
- Para obter o gráfico clique no menu *Gráfico*.

Figura 11.14
Resultados
do Exemplo
11.7.1.

Arquivo Editar Gráfico	
	Resultados
Tamanho da amostra (n) =	20
Número de sinais positivos =	3
Número de sinais negativos =	16
Número de escores empatados =	1
p (Binomial) =	0.0022
Poder do teste =	0.9970

O resultado é muito significativo ($p = 0.0022$) e o poder do teste é igual a 0.9970, rejeitando-se a hipótese de nulidade e aceitando-se a alternativa. O exercício físico prolongado aumenta o tempo de resposta ao estímulo utilizado no experimento.

Figura 11.15
Gráfico
do Exemplo
11.7.1.



11.8 – TESTE KAPPA

Prova não-paramétrica destinada a comparar as proporções da mesma variável mensurada a nível nominal em duas ocasiões diferentes. Testa-se a reprodutibilidade dos resultados, admitindo-se que haverá concordância nas proporções das respostas nos dois períodos considerados. Os dados são inseridos em *caixas de texto*, sob a forma de tabela de contingência 2 x 2.

11.8.1 – EXEMPLO

Foram enviados, em março de 2006, questionários a 400 pessoas, solicitando informações sobre o atendimento do posto de saúde localizado no bairro onde elas residiam. Um ano mais tarde o procedimento foi repetido para os mesmos indivíduos, obtendo-se 400 respostas para cada um dos dois períodos (400 em 2006 e 400 em 2007), conforme os dados contidos na tabela de contingência abaixo:

		ANO 2007	
		Satisfeito	Insatisfeito
ANO 2006	Satisfeito	180	19
	Insatisfeito	20	181

H_0 : Não há concordância entre as respostas nos dois períodos ($K = 0$);

H_1 : Há concordância entre as respostas nos dois períodos ($K \neq 0$);

Nível de decisão: alfa = 0.05 (unilateral).

11.8.2 – INSTRUÇÕES

- Pressionar o mouse em *Kappa*;
- Preencher as *caixas de texto* com os escores obtidos;
- Clicar em *Executar*, para resultados.

Figura 11.16
Teste
Kappa
Exemplo
11.8.1.

A	180	C	19	Total / Linha	199
B	20	D	181		201
Total / Coluna	200		200		400
<input type="button" value="Executar"/> <input type="button" value="Sair"/>					
Resultados					
Concordância observada =		Concordância esperada =			
0.9025		0.5000			
Kappa =		Z (Kappa) =			
0.8050		16.1002			
Conclusão da Replicabilidade		p (unilateral) =			
Excelente		< 0.0001			
Rosner B. (2006)					

O valor de *Kappa* (*K*) obtido foi igual a 0.8050, se aproximando da unidade, e denotando Excelente replicabilidade. O *p*-valor obtido (< 0.0001) é muito significativo, rejeitando-se a hipótese de nulidade, de não haver concordâncias de respostas nos dois períodos com os mesmos indivíduos. A escala de replicabilidade adotada foi a de Bernard Rosner em *Fundamental of Biostatistics* (2006).

11.9 – TESTE DE WILCOXON (SIGNED-RANK TEST)

O teste de *Wilcoxon* – *teste T* – é uma prova não-paramétrica destinada a comparar dados pareados de uma amostra obtidos em ocasiões distintas: *antes* e *após*. Equivale ao teste *t* de *Student* para dados relacionados, mas aplicado a dados mensurados pelo menos a nível ordinal, baseando-se no sentido e na magnitude das diferenças entre os pares amostrais.

11.9.1 – EXEMPLO

Culturas de linfócitos de nove (9) indivíduos foram investigadas, determinando-se os índices mitóticos *antes* e *após* exposição à luz fluorescente branca durante 60 minutos.

Figura 11.17
Dados do
Exemplo
11.9.1.

	- 1 - Antes	- 2 - Após	- 3 -
1	0.750	0.200	
2	0.670	0.150	
3	0.820	0.130	
4	0.800	0.250	
5	1.000	0.270	
6	0.870	0.300	
7	0.900	0.220	
8	0.900	0.250	
9	0.770	0.170	

H_0 : não há diferença entre os índices mitóticos das células de culturas observadas *antes* e *após* exposição à luz fluorescente branca;

H_1 : há diferença entre os índices mitóticos das células de culturas observadas *antes* e *após* exposição à luz fluorescente branca;

Nível de decisão: $\alpha = 0.01$.

11.9.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Wilcoxon*;
- Selecionar as colunas do *grid* geral;
- Clicar em *Executar o Teste*, para resultados:

Figura 11.18
Resultados
do Exemplo
11.9.1.

	Resultados Cols. 1 e 2
T =	0
Número de pares =	98
Z =	2.6656
p-valor (unilateral) =	0.0038
p-valor (bilateral) =	0.0077

O T obtido é altamente significativo ($p = 0.0077$), rejeitando-se a hipótese de nulidade e aceitando-se a alternativa. Os índices mitóticos são menores após a aplicação nas culturas de luz fluorescente branca.

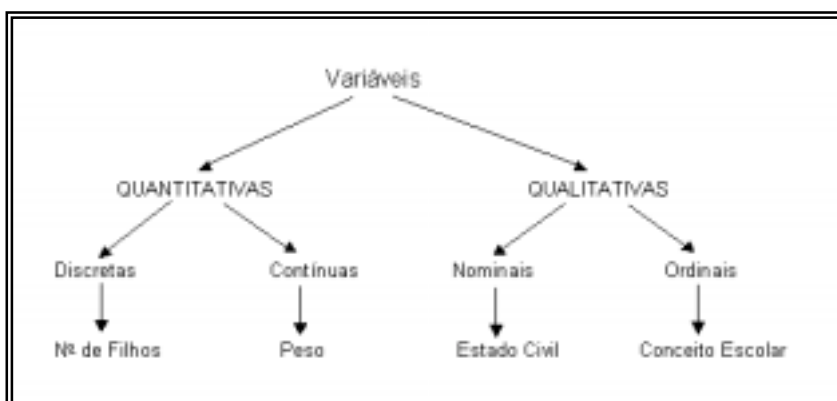
CAPÍTULO 12

ESTATÍSTICA DESCRITIVA

12.1 – INTRODUÇÃO

A função descritiva da estatística consiste em sumarizar os dados coletados – da amostra ou da população –, ordenando-os e classificando-os, para torná-los de fácil entendimento, conforme o tipo da variável que está em estudo.

Figura 12.1
Classificação
das variáveis
e exemplos
de cada tipo.



O programa calcula o seguinte:

I – Variáveis Quantitativas

- a) Número de escores (N);
- b) Valor mínimo;
- c) Valor máximo;
- d) Amplitude total (Ar);
- e) Mediana (Md);
- f) Desvio Interquartilico;
- g) Primeiro e Terceiro Quartis;
- h) Média aritmética (μ ou \bar{X});
- i) Variância (σ^2 ou s^2);
- j) Desvio padrão (σ ou s);
- k) Erro padrão ($\sigma_{\bar{x}}$ ou $s_{\bar{x}}$);
- l) Coeficiente de variação (CV);
- m) Coeficiente de assimetria (g_1);
- n) Coeficiente de curtose (g_2).

II – Variáveis Qualitativas

- a) Índice de *Shannon-Wiener*;
- b) Índice de *Simpson*.

12.2 - VARIÁVEIS QUANTITATIVAS

As três medidas de tendência central mais usadas em medicina e biologia são a média, a mediana e a moda. Todas as três são usadas para dados quantitativos, entretanto a mediana e a moda também podem ser usadas para dados ordinais.

12.2.1 – EXEMPLO 1

Um pesquisador coletou três (3) amostras, cuja variável investigada era a pressão arterial sistólica:

- a) Amostra A: 12 estudantes do sexo feminino;
- b) Amostra B: 9 alunos do sexo masculino;
- c) Amostra C: 10 docentes.

Figura 12.1

Dados do
Exemplo
12.2.1.

	- 1 - Amostra A	- 2 - Amostra B	- 3 - Amostra C
1	110		121
2	115	117	125
3	116	120	119
4	109	121	129
5	110	113	125
6	117	116	126
7	116	117	127
8	118	115	128
9	118	119	122
10	119	120	125
11	115		
12	114		

12.2.2 – EXEMPLO 2

Em uma cultura de bactérias, o número aumentou de 1431, no final do 1º dia, para 3250, 9610, 26800 e 63300 ao término do 2º, 3º, 4º e 5º dias, respectivamente. Descrever a amostra calculando a média geométrica.

Figura 12.2

Dados do
Exemplo
12.2.2.

	- 1 - ** Tempo	- 2 - Bactérias	- 3 -
1	1º dia	1431	
2	2º dia	3250	
3	3º dia	9610	
4	4º dia	26800	
5	5º dia	63300	
6			

12.2.3 – INSTRUÇÕES

- Introduzir os dados no *grid* geral do *BioEstat*;
- Dirigir o *mouse* para o *menu* superior e clicar em *Estatística*;
- Clicar o *mouse* em *Estatística Descritiva* e, a seguir, em *Dados Quantitativos*;
- Selecionar as colunas do *grid* (1, 2, 3, ... , n);
- Pressionar em *Executar o Teste*, para Resultados.

Figura 12.3
Resultados do
Exemplo
12.2.1.

	- 1 - Amostra A	- 2 - Amostra B	- 3 - Amostra C
N =	12	9	10
Mínimo	109.0000	113.0000	119.0000
Máximo	119.0000	121.0000	129.0000
Amplitude Total	10.0000	8.0000	10.0000
Mediana	115.5000	117.0000	125.0000
Primeiro Quartil (25%)	113.0000	116.0000	122.7500
Terceiro Quartil (75%)	117.2500	120.0000	126.7500
Desvio Interquartilico	4.2500	4.0000	4.0000
Média Aritmética	114.7500	117.5556	124.7000
Variância	11.4773	7.0278	10.0111
Desvio Padrão	3.3878	2.6510	3.1640
Erro Padrão	0.9780	0.8837	1.0008
Coefficiente de Variação	2.95%	2.26%	2.54%
Assimetria	-0.6849	-0.3644	-0.5509
Curtose	-0.8137	-0.7973	-0.4467

A tabela acima contém os resultados das três (3) amostras coletadas: medidas de tendência central, de dispersão e de forma.

Figura 12.4
Resultados
do Exemplo
12.2.2.

	- 2 - Bactérias
N =	5
Mínimo	1431.0000
Máximo	63300.0000
Amplitude Total	61869.0000
Mediana	9610.0000
Primeiro Quartil (25%)	3250.0000
Terceiro Quartil (75%)	26800.0000
Desvio Interquartilico	23550.0000
Média Aritmética	20878.2000
Variância	66265e+04
Desvio Padrão	25741.9705
Erro Padrão	11512.1592
Coefficiente de Variação	123.30%
Assimetria	1.5215
Curtose	1.9508
N (média geométrica) =	5
Média geométrica	9461.4343
Variância (geom.)	2.7768
Desvio padrão (geom.)	4.6343

A média geométrica é a medida de tendência central que mais se ajusta aos dados amostrais, sendo duas vezes menor que a aritmética e estando mais próxima da mediana.

12.3 - VARIÁVEIS QUALITATIVAS

Quando a variável é qualitativa, as observações são medidas numa escala nominal ou categórica. Logo, medidas como a média e o desvio padrão não se aplicam neste caso. Esses dados podem ser medidos de outras formas como: contagens, proporções, razões, taxas e índices.

12.3.1 – ÍNDICE DE SHANNON - WIENER: k amostras

O Índice de *Shannon-Wiener* estima a diversidade de variáveis categóricas em uma população, avaliando os aspectos da *riqueza* e *equitabilidade*, os quais dizem respeito ao número de categorias da variável em questão e às proporções de cada uma destas, respectivamente. A uma população com maior número de categorias que outra é atribuída maior riqueza, e a uma população com mais homogeneidade nas proporções de suas categorias é atribuída maior equitabilidade. O índice poderá ser efetuado para k amostras.

12.3.1.1 – EXEMPLO

Foram coletadas amostras de ninhos de andorinhas em localidades distintas de dois Municípios: A e B. Os dados obtidos estão contidos na Figura 12.5.

Figura 12.5
Dados do
Exemplo
12.3.1.1.

- 1 - Ninhos (A)	- 2 - Ninhos (B)
2	3
1	2
59	3
1	2
2	2
	3

12.3.1.2 – INSTRUÇÕES

- Introduzir os dados no *grid* geral do *BioEstat*;
- Dirigir o *mouse* para o *menu* superior e clicar em *Estatística*;
- Clicar o *mouse* em *Estatística Descritiva* e, a seguir, em *Dados Qualitativos*;
- Clicar em *Índice de Shannon-Wiener* e selecionar as colunas do *grid* (1, 2, ..., n);
- Pressionar em *Executar o Teste*, para Resultados.

Figura 12.6
Resultados do
Exemplo
12.3.1.1.

	- 1 -	- 2 -
Tamanho da Amostra	65	15
Número de Categorias	5	6
Índice de Shannon-Wiener	0.1870	0.7694
Máxima diversidade	0.6990	0.7782
Homogeneidade	0.2675	0.9888
Heterogeneidade	0.7325	0.0112

Os resultados evidenciam menor homogeneidade dos dados coletados no Município A (coluna 1), cujo valor é igual a 0.2675, em contraste com o índice dos dados do Município B (coluna 2), cujo valor obtido foi é igual a 0.9888.

12.3.2 – ÍNDICE DE SIMPSON: k amostras

Semelhante ao anterior, mede também a diversidade, quando os dados são categóricos.

12.3.2.1 – EXEMPLO

Em duas ilhas, foi contado o número de espécies de mamíferos encontrados em cada uma. Os dados estão resumidos na Figura 12.7. Estimar o Índice de *Simpson* para essas duas amostras.

Figura 12.7
Dados do
Exemplo
12.3.2.1.

- 1 - Ilha A	- 2 - Ilha B
443	708
332	221
322	144
298	81

12.3.2.2 – INSTRUÇÕES

- Introduzir os dados no *grid* geral do *BioEstat*;
- Dirigir o *mouse* para o *menu* superior e clicar em *Estatística*;
- Clicar o *mouse* em *Estatística Descritiva* e, a seguir, em *Dados Qualitativos*;
- Clicar em *Índice de Simpson* e selecionar as colunas do *grid* (1, 2, ... , n);
- Pressionar em *Executar o Teste*, para Resultados.

Figura 12.8
Resultados do
Exemplo
12.3.2.1.

	- 1 -	- 2 -
Espécies investigadas	4	4
Total de animais	1395	1154
Índice de Simpson	0.7436	0.5664
p 1	0.3176	0.6135
p 2	0.2380	0.1915
p 3	0.2308	0.1248
p 4	0.2136	0.0702

O Índice de *Simpson* foi maior na amostra da Ilha A (0.7436), enquanto o da ilha B atingiu 0.5664. Os resultados mostram, ainda, as proporções (p_i) de cada amostra.

CAPÍTULO 13

ESTIMAÇÃO DE PARÂMETROS

13.1 – INTRODUÇÃO

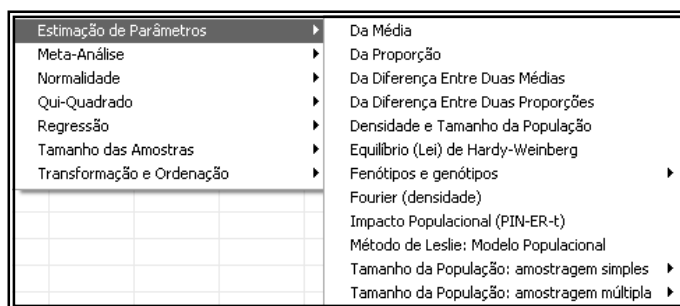
A estimação de parâmetros é parte da inferência estatística através da qual, a partir dos dados amostrais, se inferem os valores numéricos da população investigada. O *BioEstat* apresenta os seguintes programas de estimação paramétrica:

- a) Da Média;
- b) Da Proporção;
- c) Da Diferença entre Duas Médias;
- d) Da Diferença entre Duas Proporções;
- e) Da Densidade e Tamanho da População;
- f) Equilíbrio de *Hardy-Weinberg*;
- g) Fenótipos e Genótipos;
- h) *Fourier* (densidade);
- i) Impacto Populacional (PIN-ER-t);
- j) Método de *Leslie*: Modelo Populacional
- k) Tamanho da População (amostragem simples, método direto);
- l) Tamanho da População (amostragem simples, método inverso);
- m) Tamanho da População (amostragem múltipla, população fechada);
- n) Tamanho da População (amostragem múltipla, população aberta).

13.1.1 – INSTRUÇÕES

- a) Dirigir o mouse para o menu superior e clicar em *Estatística*;
- b) Apontar o mouse em *Estimação de Parâmetros*, como demonstra a Fig. 13.1:

Figura 13.1
Menu de
Estimação de
Parâmetros.



13.2 – ESTIMAÇÃO DA MÉDIA

A estimação da média populacional (μ) é efetuada de duas maneiras:

- a) Estimação de Ponto;
- b) Estimação de Intervalos de Confiança: 95% e 99%.

O procedimento consiste em retirar amostra aleatória da população e computar a média aritmética cujo valor, com base no Teorema do Limite Central, constituirá a

estimação de *ponto* da média do universo objeto da pesquisa. O desvio padrão será o da população (σ) ou, se inexistir, o do valor amostral (s). Os intervalos de confiança, para amostras com desvio padrão paramétrico conhecido, têm por base a distribuição de probabilidades da curva normal, ou seja, os escores padronizados z . Em caso contrário, quando se utilizar o desvio padrão amostral, o programa calcula os intervalos de confiança com suporte na distribuição t de *Student*.

13.2.1 – EXEMPLO

Determinaram-se os níveis de creatinina do sangue periférico de 35 estudantes universitários, escolhidos de modo randômico e cujas idades variavam de 18 a 23 anos, obtendo-se, como média amostral, 1.041 mg/dL. O desvio padrão da população é conhecido, cujo valor é igual a 0.2729 mg/dL. Desse modo, há:

- Média amostral = 1.041 mg/dL;
- Desvio padrão paramétrico = 0.2729 mg/dL;

Efetuar as estimativas de ponto e de intervalos de confiança de 95% e 99%.

13.2.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Estimação da Média*;
- Preencher as *caixas de texto*;
- Clicar o botão correspondente ao desvio padrão da população (σ);
- Clicar em *Executar*, para obter resultados.

Figura 13.2
Estimação de
ponto e de
intervalos da
média
(Exemplo
13.2.1).

A imagem mostra a interface de um programa de computador intitulado "Estimação da Média". No topo, há uma barra de menu com "Imprimir". Abaixo, a seção "Dados" contém três campos de entrada com botões de seta para cima e para baixo: "Tamanho da amostra:" com o valor 35, "Média:" com o valor 1.041, e "Desvio Padrão:" com o valor .2729. À direita desses campos estão os botões "Executar" e "Sair". Abaixo do campo de desvio padrão, há duas opções de seleção: "da população (σ)" (selecionada com um botão de rádio) e "da amostra (s)". A seção "Resultados" apresenta a "Estimativa de Ponto" como 1.0410. Abaixo disso, são mostrados dois intervalos de confiança: "IC 95%" com valores 0.9506 e 1.1314, e "IC 99%" com valores 0.9220 e 1.1600. Os intervalos são representados como $0.9506 \leq \mu \leq 1.1314$ e $0.9220 \leq \mu \leq 1.1600$.

Há 95% de chance de que a média paramétrica seja um valor situado no espaço compreendido entre 0.9506 e 1.1314, e de 99%, entre 0.9220 e 1.1600.

13.3 – ESTIMAÇÃO DA PROPORÇÃO

A estimação de uma proporção é efetuada para populações binomiais, onde a proporção paramétrica – π – é uma fração do universo que possui determinada modalidade ou categoria de interesse do pesquisador. A proporção da amostra aleatória é uma *estatística* que estima a *proporção paramétrica*, representando a estimativa de ponto da proporção populacional, assim definida: $p = k / n$, sendo k o número de eventos observados da categoria de interesse e n o tamanho amostral. Os cálculos dos intervalos de confiança são efetuados com base na distribuição normal, desde que o valor npq seja igual ou maior que 5, aspecto definido no programa do *BioEstat*.

13.3.1 – EXEMPLO

Um levantamento efetuado em hospital especializado em doenças neoplásicas revelou que em uma coorte de 12000 mulheres, cujas idades variavam entre 40 e 50 anos, o número de pacientes com câncer mamário foi igual a 420. O valor npq é igual a 40.53. Efetuar as estimativas de ponto e de intervalos de confiança de 95% e 99%.

13.3.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Estimação da Proporção*;
- Preencher as *caixas de texto*;
- Clicar em *Executar*, para resultados.

Figura 13.3
Estimação
da proporção
(Exemplo
13.3.1).

O programa 'Estimação da Proporção' apresenta uma interface com as seguintes seções:

- Dados:**
 - Tamanho da amostra: 12000
 - Eventos favoráveis: 420
 - Botões: Executar, Sair
- Resultados:**
 - Estimativa de Ponto: 0.035 (3.5%)
 - IC 95%: 0.032 (3.2%) ≤ π ≤ 0.038 (3.8%)
 - IC 99%: 0.031 (3.1%) ≤ π ≤ 0.039 (3.9%)

A estimativa de ponto é igual a 0.035 (3.5%), devendo a proporção paramétrica (π) ser um valor situado no intervalo de 0.032 a 0.038, com probabilidade de 95%, ou entre 0.031 e 0.039, com chance de 99%.

13.4 – ESTIMAÇÃO DA DIFERENÇA ENTRE DUAS MÉDIAS

Do mesmo modo que se infere a média de uma população, pode-se estimar a da diferença entre duas médias ($\mu_1 - \mu_2$). Quando se conhecem as variâncias paramétricas,

pressupõe-se uma distribuição normal da amostragem dessa diferença, independente do tamanho das amostras. Quando as variâncias das populações são desconhecidas, as variâncias amostrais são introduzidas e a distribuição da amostragem passa a ser a *t* de *Student*. Todas as informações concernentes à estimação da diferença entre duas médias, inclusive, quando for o caso, sobre o pressuposto da homogeneidade das variâncias amostrais, são inseridas diretamente nas *caixas de texto* do programa.

13.4.1 – EXEMPLO

Coletaram-se duas amostras referentes ao peso (g) de recém-nascidos: uma de população caucasóide e outra de universo ameríndio, com 32 e 27 recém-natos, respectivamente. Os dados obtidos foram os seguintes:

- Média da primeira amostra = 3300g;
- Variância da primeira amostra = 80200;
- Média da segunda amostra = 3050g;
- Variância da segunda amostra = 81000;

Em estudo prévio, as variâncias amostrais foram consideradas homogêneas. Efetuar as estimativas de ponto e de intervalos de confiança de 95% e 99% da diferença entre as duas médias.

13.4.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Diferença entre Duas Médias*;
- Preencher as *caixas de texto* com os dados;
- Clicar o botão correspondente a variâncias amostrais (s^2);
- Pressionar em *Executar* e, a seguir, em *Homogêneas*, para resultados.

Figura 13.4
Estimação da
diferença
entre duas
médias
(Exemplo
13.4.1).

Diferença entre duas Médias

Imprimir

Dados

Tamanho da amostra 1: 32 Tamanho da amostra 2: 27

Média 1: 3300 Média 2: 3050

Variância 1: 80200 Variância 2: 81000

☐ variâncias paramétricas (σ^2)

☒ variâncias amostrais (s^2)

Resultados

Estimativa de ponto da diferença: 250.0000

IC 95%: 101.3438 ≤ $\mu_1 - \mu_2$ ≤ 398.6562

IC 99%: 52.0488 ≤ $\mu_1 - \mu_2$ ≤ 447.9512

BioEstat 5.0

? As variâncias são homogêneas?

A estimativa de ponto da diferença entre as médias é igual a 250g, devendo o verdadeiro valor paramétrico dessa diferença das populações investigadas estar situado entre os limites dos intervalos de confiança, com erro de estimação de 0.05 ou 0.01, IC (95%) e IC (99%), respectivamente.

13.5 – ESTIMAÇÃO DA DIFERENÇA ENTRE DUAS PROPORÇÕES

Em algumas investigações há interesse em calcular a magnitude da diferença entre proporções de duas populações binomiais, desde que as amostras tenham valores npq iguais ou superiores a cinco (5) e que os escores p estejam próximos de 0.5. A distribuição de amostragem da diferença entre duas proporções binomiais ($p_1 - p_2$) se aproxima da curva normal, podendo-se, com base no Teorema do Limite Central, utilizar os escores z para o cálculo dos intervalos de confiança.

13.5.1 – EXEMPLO

Um candidato a Reitor de uma universidade obteve, em consulta prévia efetuada em dois Centros Profissionais da referida instituição de ensino, os seguintes resultados:

- Centro Profissional A: 430 eleitores e 312 votos favoráveis;
- Centro Profissional B: 400 eleitores e 243 votos favoráveis.

Efetuar as estimativas de ponto e de intervalos de confiança de 95% e 99% da diferença entre as duas proporções

13.5.2 – INSTRUÇÕES

- Pressionar o mouse em *Diferença entre Duas Proporções*;
- Preencher as *caixas de texto* com os dados;
- Clicar em *Executar*, para resultados.

Figura 13.5
Estimação da
diferença
entre duas
proporções
(Exemplo
13.5.1).

Estimación da diferença entre duas Proporções

Imprimir

Dados

Tamanho da amostra 1: 430

Tamanho da amostra 2: 400

Eventos favoráveis 1: 312

Eventos favoráveis 2: 243

Executar Sair

Resultados

Estimativa de ponto da diferença: 0.1181

IC 95%: 0.0543 ≤ $\pi_1 - \pi_2$ ≤ 0.1819

IC 99%: 0.0341 ≤ $\pi_1 - \pi_2$ ≤ 0.2020

A estimativa de ponto da diferença entre as duas proporções e os respectivos intervalos de confiança de 95% e 99% estão contidos na Figura 13.5.

13.6 – DENSIDADE E TAMANHO DA POPULAÇÃO

Em muitas pesquisas há interesse em determinar o tamanho do universo investigado como, por exemplo, o número de indivíduos acometidos de determinada doença epidêmica, o total de indivíduos residentes em determinada área geográfica, a quantidade de animais de uma espécie ameaçada de extinção, etc.

O *BioEstat* estima o número de indivíduos da área geográfica definida, ou seja, a densidade populacional e, a seguir, procede a estimação da quantidade de indivíduos em todo o espaço objeto da pesquisa.

13.6.1 – EXEMPLO

Em uma região geográfica de 400 ares plantados com seringueiras, selecionaram-se aleatoriamente 15 ares, examinando-se em cada um o número de plantas infectadas por fungos, contando-se, ao final, 35 árvores parasitadas. Estimar a densidade por are e o total de plantas infectadas.

13.6.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Densidade e tamanho da população*;
- Preencher as *caixas de texto*;
- Clicar em *Executar*, para resultados.

Figura 13.6
Densidade e
tamanho da
população
(Exemplo
13.6.1).

Densidade e Tamanho da População	
Imprimir	
Unidade da área objeto da investigação: m2, acre, hectare, etc.	are
Tamanho total da área objeto da investigação (em are)	400
Tamanho de cada quadrilátero amostral (em are)	1
Número de quadriláteros amostrais randomizados	15
Número de indivíduos observados em todos os quadriláteros amostrais	35
RESULTADOS: Densidade estimada: 2.33 Indivíduos p/ are Erro de estimação: 0.79 I.C. 95%: $1.54 \leq \mu \leq 3.12$ População estimada: 933 indivíduos. Erro de estimação: 315.52 I.C. 95%: $617 \leq \mu \leq 1249$	
<input type="button" value="Executar"/> <input type="button" value="Cancelar"/>	

A densidade de seringueiras infectadas é de 2.33 por are e a população de árvores com fungo em todos os 400 ares de terra atinge 933 unidades. Os resultados assinalam os erros de estimação para o intervalo de confiança de 95%.

13.7 – EQUILÍBRIO DE HARDY-WEINBERG

A estimação das frequências genotípicas e das fenotípicas esperadas podem ser calculadas, determinando-se se há aderência entre os valores observados e aqueles esperados, estes com base na determinação da frequência gênica.

O teste do Qui-Quadrado efetuado com base nos dados fenotípicos mostra, em termos probabilísticos, se a população está ou não em *Equilíbrio de Hardy-Weinberg*, sendo o grau de liberdade igual ao número de alelos dominantes menos um (1). O procedimento é bastante semelhante ao utilizado no item 13.8 deste Capítulo sobre *Fenótipos e Genótipos*.

O *BioEstat* apresenta os seguintes exemplos:

- Sistema ABO;
- Sistema Haptoglobinas;
- Sistema MN

Vamos apresentar um exemplo – Sistema MN –, pois os procedimentos para os demais sistemas são similares.

13.7.1 – EXEMPLO

Coletou-se amostra de 500 indivíduos de uma etnia de origem oriental e seus descendentes, determinando-se os respectivos grupos sanguíneos do Sistema MN, com os dados contidos na Figura 13.7. Determinar se a população de onde a amostra foi retirada está em equilíbrio de *Hardy-Weinberg*.

13.7.2 – INSTRUÇÕES

- Apontar o mouse em *Lei de Hardy-Weinberg*;
- Preencher as *caixas de texto* (Fig. 13.7) como os dados dos grupos M, MN e N;
- Pressionar em *Executar*, para resultados.

Figura 13.7
Dados do
Exemplo
13.7.1.

Sistema MN			
Fenótipo M	Fenótipo MN	Fenótipo N	
40	128	32	<input type="button" value="Executar"/>

Figura 13.8
Resultados
do Exemplo
13.7.1.

Systema MN	Fenótipo M	Fenótipo M N	Fenótipo N
Observado =	40	128	32
Esperado =	54.0800	99.8400	46.0800
Qui-Quadrado =	15.9106		
Graus de liberdade =	1		
p valor =	0.0001		
	Alelo M	Alelo N	
Estimação =	0.5200	0.4800	

O Qui-Quadrado é igual a 15.9106, sendo o *p-valor* = 0.0001, para um (1) grau de liberdade. A população não está em Equilíbrio de *Hardy-Weinberg*.

13.8 – FENÓTIPOS E GENÓTIPOS

A estimação das frequências fenotípicas e genotípicas, calculadas pelo método da *máxima-verossimilhança*, compreende os seguintes sistemas:

- | | |
|----------------------|-----------------|
| a) Sistema Secretor; | e) Sistema ESD; |
| b) Sistema ABO; | f) Sistema Gm; |
| c) Sistema A1A2BO; | g) Sistema Rh. |
| d) Sistema MN; | |

Tendo em vista que os procedimentos são similares, será tomado um dos sistemas para exemplificar a estimação paramétrica.

13.8.1 – EXEMPLO

Em uma população urbana brasileira retirou-se amostra aleatória de 200 pessoas de ambos os sexos, determinando-se os grupos sanguíneos do sistema ABO, tendo sido observado:

Figura 13.9
Fenótipos e genótipos: Sistema ABO (Exemplo 13.8.1).

Valores observados para os fenótipos		Estimativas iniciais para os alelos	
	Valor		Estimativa
A	70	A	.25
B	26	B	.11
AB	12		
O	92		

Valores Inteiros		Valores Decimais	
Executar		Cancelar	

Estimar os fenótipos e genótipos do sistema sanguíneo investigado.

13.8.2 – INSTRUÇÕES

- Apontar o mouse em *Fenótipos e Genótipos*;
- Clicar em *Sistema ABO*;
- Preencher o *grid* específico com os dados amostrais;
- Clicar em *Executar*, para resultados.

Figura 13.10
Resultados do Exemplo 13.10.1.

Sistema ABO	Fenótipo A	Fenótipo B	Fenótipo AB	Fenótipo O	Total
No. observado =	70	26	12	92	200
No. esperado =	72.4563	28.6339	9.1807	89.7292	200.0000
Prop. observada =	0.3500	0.1300	0.0600	0.4600	1.0000
Prop. esperada =	0.3623	0.1432	0.0459	0.4486	1.0000
Qui-quadrado =	0.0833	0.2423	0.8658	0.0575	1.2488
(p) =	0.2638	---	---	---	---
Graus de liberdade =	1	---	---	---	---
	Alelo: A	Alelo: B	Alelo: O		
Estimação inicial =	0.2500	0.1100	0.6400	---	---
Estimação final =	0.2307	0.0995	0.669810	---	---
Variância =	0.000509	0.000236	0.000631	---	---
Desvio padrão =	0.022564	0.015367	0.025129	---	---

A diferença entre os fenótipos observados e os esperados não é significativa ($p = 0.2638$). A estimação final dos alelos está bem próxima da calculada inicialmente, sendo a diferença entre o cálculo inicial e o final dos alelos A e B inferior a um desvio padrão, enquanto do alelo O é de apenas 1.19 desvio padrão.

13.9 – FOURIER (densidade)

O programa *Fourier*, do *BioEstat*, é destinado ao cálculo dos estimadores da densidade de populações animais. A largura da transecção é determinada pela curva mais apropriada de uma série de probabilidades baseada na distribuição das distâncias de observações.

13.9.1 – EXEMPLO

Em levantamento efetuado em certa reserva, no sentido de determinar a densidade de uma espécie de macacos, a distância perpendicular máxima foi igual a 50 metros e a inventariada atingiu 2000 metros, tendo sido observados doze (12) grupos de indivíduos nesse espaço. Estimar a densidade pelo método de *Fourier*.

13.9.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Fourier*;
- Preencher as *caixas de texto* e o *grid* específico;
- Clicar em *Executar*, para resultados.

Figura 13.11
Análise da
série de
Fourier
(Exemplo
13.9.1).

Distância perpendicular máxima (em metros)	Grupo e/ou indivíduo	Distância perpendicular
50	1	0
	2	0
	3	5
	4	10
Distância total inventariada (em metros)	5	10
2000	6	20
	7	20
	8	25
	9	30
Número total de grupos e/ou indivíduos observados	10	30
12	11	40
	12	50

Executar Cancelar

A estimativa é de 88.0 grupos de indivíduos por km^2 . Os intervalos de confiança e as medidas de variação estão contidos nos resultados.

Figura 13.12
Resultados
do Exemplo
13.9.1.

Número de termos usados =	1
Densidade (grupos e/ou indivíduos/ km^2) =	87.601
Desvio padrão =	35.384
Intervalo de confiança 95% =	18.247 <= u <= 156.954
Intervalo de confiança 99% =	0.000 <= u <= 178.893
Distância perpendicular máxima (metros) =	50
Número total de grupos e/ou indivíduos =	12
Distância total inventariada (metros) =	2000

13.10 – IMPACTO POPULACIONAL (PIN-ER-t)

Medidas epidemiológicas relacionadas com o fator de risco, tais como o Risco Relativo (RR) e o *Odds Ratio* (OR) que podem indicar a intensidade da associação entre exposição e doença, não esclarecem, todavia, os benefícios que poderiam advir para a população se a exposição ao fator de risco fosse modificada ou eliminada. Uma medida descrita recentemente (PIN-ER-t) proporciona essa quantificação populacional, denominada de Número de Impacto Populacional.

13.10.1 – EXEMPLO (hipotético)

Em estudo piloto em pessoas com idade superior a 56 anos, verificou-se que o Risco Relativo (RR), decorrente da associação entre Obesidade e Acidente Coronariano (AC), foi igual a **1.4**. Foram considerados obesos os indivíduos de ambos os sexos que apresentaram Índice de Massa Corporal (IMC) acima de 28 kg/m². A **Prevalência** de Obesidade nesse grupo foi mensurada em **0.3** (30%) e a **Incidência** de Acidente Coronariano em torno de **0.015** (1.5%).

Qual será o resultado da eliminação do fator de risco Obesidade (IMC > 28) em uma população de 300 mil habitantes com mais de 56 anos?

13.10.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Impacto Populacional (PIN-ER-t)*;
- Preencher as *caixas de texto*, conforme as indicações;
- Clicar em *Calcular o PIN-ER-t*, para resultados.

Figura 13.13
Cálculo do
Número de
Impacto
Populacional
(Exemplo
13.10.1).

Impacto Populacional (PIN-ER-t)	
Obesidade	Fator de risco
Acidente Coronariano	Doença associada ao Fator de risco
Acima de 56 anos	Tempo para cálculo da Incidência
0.015	Incidência da Doença
300000	Tamanho da População
0.3	Prevalência do Fator de Risco
1.4	Risco Relativo da associação entre o Fator de Risco e a Doença
Resultado	
PIN-ER-t = 482 IC(95%): 289 a 646	
*PIN: Population Impact Number	
Calcular o PIN-ER-t	Ver Exemplo Cancelar

Conforme o resultado do teste, o número de casos de Acidentes Coronarianos que poderiam ser prevenidos nessa população, acima de 57 anos, se o 'fator de risco' Obesidade fosse eliminado, seria de 482, com IC 95%: 289 a 646.

13.11 – MODELO DE LESLIE: MATRIZ DE CRESCIMENTO POPULACIONAL

Trata-se de um procedimento, desenvolvido por *P. H. Leslie*, destinado a descrever o crescimento de populações fechadas à migração e onde somente um sexo, geralmente o feminino, é considerado nos cálculos, determinando-se as mudanças numéricas dos organismos estudados em determinados períodos de tempo.

Do ponto de vista matemático, o método de *Leslie* baseia-se na construção de uma *matriz quadrada*. As linhas representam classes etárias ou estádios de vida reprodutiva. As colunas representam períodos de tempo. Para a construção a Matriz, certas informações devem ser conhecidas da população que está sendo investigada:

n_x : número de indivíduos de cada *classe* etária ou estágio;

f_x : taxa de fecundidade;

s_x : proporção de sobrevivência;

t : tempo.

13.11.1 – EXEMPLO (Hipotético)

Um estudo cujo objetivo era estimar o crescimento populacional do pirarucu (*Arapaima gigas*) foi realizado em um ambiente controlado. O experimento foi realizado a partir de uma colônia com 20 machos e 20 fêmeas. Num estudo piloto foi observado que a idade máxima que o pirarucu atinge é 13 anos e que nos primeiros cinco (5) anos de vida o peixe ainda não tem capacidade reprodutiva.

Será utilizado o método de *Leslie* para estimar o crescimento populacional em um período de cinco (5) anos. Os dados são apresentados na tabela abaixo.

Figura 13.14
Modelo de
entrada de
Dados do
Exemplo
13.11.1.

- 1 -	- 2 -	- 3 -	- 4 -	- 5 -
** Faixas etárias	n_x	f_x	s_x	
Zero a 5 anos	5.000	0.000	0.130	
6 a 8 anos	5.000	4.000	0.470	
9 a 10 anos	5.000	5.000	0.630	
11 a 12 anos	5.000	6.000		

A última célula da coluna ' s_x '
não deve ser preenchida.

13.11.2 – INSTRUÇÕES

- Apontar o mouse em *Método de Leslie: Modelo Populacional*;
- Selecione as colunas desejadas e clique em *Executar Estatística*;
- Quando surgir o diálogo “Método de *Leslie*”, digite o “Número de passos a serem calculados”, neste exemplo são cinco (5) e clique em OK.

Figura 13.15
Resultados
do Exemplo
13.11.1.

	n_x t_0	n_x t_1	n_x t_2	n_x t_3	n_x t_4	n_x t_5
Faixa etária 1	5.00	75.00	33.25	49.41	41.36	53.17
2	5.00	0.65	9.75	4.32	6.42	5.38
3	5.00	2.35	0.31	4.58	2.03	3.02
4	5.00	3.15	1.48	0.19	2.89	1.28
Total	20.00	81.15	44.79	58.51	52.70	62.85

Os resultados mostram que a população mantida em ambiente controlado terá aumento de 20 para 62 indivíduos cinco (5) anos, representando um aumento de 314%. A maioria dos indivíduos, entretanto, estará concentrada na primeira faixa etária (entre zero e cinco anos).

13.12 – TAMANHO DA POPULAÇÃO (amostragem simples, método direto)

A estimativa do tamanho de uma população – N – é aplicada na área animal pelo método direto a partir dos dados amostrais e referentes a uma determinada área geográfica. Este método compreende duas etapas: na primeira capturam-se t animais, os quais são marcados e devolvidos ao meio ambiente, e na segunda, alguns dias mais tarde, capturam-se n indivíduos, contando-se os marcados e os não-marcados. Com esses dados pode-se estimar o tamanho da população (N), processo que poderá repetir-se, periodicamente, avaliando-se o aumento ou a diminuição do universo pesquisado.

13.12.1 – EXEMPLO

Em uma reserva florestal de tamanho bem definido capturaram-se 250 capivaras, as quais foram marcadas. Transcorridos 20 dias, repetiu-se a operação de recaptura obtendo-se, ao final, 180 animais, dos quais 59 estavam marcados. Estimar o tamanho da população da espécie estudada.

Figura 13.16
Tamanho da população: método direto (Exemplo 13.12.1).

A janela de software, intitulada "Estimación da População (amostragem direta)", possui uma barra de título com um ícone de impressora e o texto "Imprimir". O corpo da janela é dividido em seções. À esquerda, há três campos de entrada de texto com rótulos: "No. de indivíduos capturados e marcados na 1a. amostra" (valor 250), "No. de indivíduos capturados na 2a. amostra (marcados e não marcados)" (valor 180) e "No. de indivíduos marcados e capturados na 2a. amostra" (valor 59). À direita desses campos, há dois botões: "Executar" e "Cancelar". Abaixo, uma seção intitulada "RESULTADOS:" exibe "População estimada: 763 indivíduos" e "Erro de estimação: 162.82". No rodapé, há uma expressão matemática para o intervalo de confiança de 95%: $600 \leq \mu \leq 926$, com "I.C. 95%" centralizado acima.

13.12.2 – INSTRUÇÕES

- Pressionar o mouse em *Tamanho da população (amostragem simples, método direto)*;
- Preencher as *caixas de texto*;
- Clicar em *Executar*, para resultados.

Os resultados mostram que o tamanho estimado da população é de 763 capivaras, podendo o parâmetro ser um valor situado entre 600 e 926 animais, conforme o cálculo do intervalo de confiança de 95%.

13.13 – TAMANHO DA POPULAÇÃO (amostragem simples, método inverso)

A metodologia é semelhante à do procedimento direto para o cálculo do tamanho da população (N), com a seguinte diferença: na primeira etapa capturam-se t animais, os quais são marcados e devolvidos ao meio ambiente, e, na segunda, algum tempo mais tarde, recapturam-se tantos espécimes até que nesta amostra tenham sido apreendidos n animais marcados e previamente determinados.

13.13.1 – EXEMPLO

Foram capturados e marcados 250 mergulhões e devolvidos, a seguir, ao respectivo habitat. Decorridas duas semanas, programou-se a recaptura de 33 aves marcadas, sendo que o total da amostra para atingir esse objetivo foi de 120 indivíduos. Estimar o tamanho da população objeto da pesquisa.

13.13.2 – INSTRUÇÕES

- Pressionar o mouse em *Tamanho da população (amostragem simples, método inverso)*;
- Preencher as *caixas de texto*;
- Clicar em *Executar*, para resultados.

Figura 13.17
Tamanho da população: método inverso (Exemplo 13.13.1).

O tamanho da população de mergulhões foi estimada em 909 indivíduos, podendo o verdadeiro parâmetro N ser um valor situado entre 643 e 1175 aves, definido pelo intervalo de confiança de 95%.

13.14 – TAMANHO DA POPULAÇÃO (amostragem múltipla, população fechada)

O modelo para este teste é o de *Schumacher* e *Eschmeyer*, considerando-se 'população fechada' aquele universo que mantém constante seu tamanho durante o período de investigação. O procedimento desses autores baseia-se na regressão linear, com ajustes pelo método dos mínimos quadrados.

Os dados devem ser introduzidos no *grid* geral do *BioEstat*, obedecendo aos seguintes critérios:

n_i = tamanho de cada amostra;

m_i = número de indivíduos marcados em cada amostra;

M_i = número de indivíduos marcados na população exatamente antes da i -ésima amostra ser coletada.

13.14.1 – EXEMPLO (hipotético)

Uma população de uma espécie de anuros foi investigada durante seis (6) dias consecutivos para calcular o tamanho desse universo. Os dados abaixo indicam a maneira de introduzir os dados no *grid* geral. Estimar o tamanho da população e os intervalos de confiança de 95% e 99%. Sendo a amostragem múltipla (6 amostras) e a população considerada fechada, o método escolhido foi o de *Schumacher* e *Eschmeyer*.

Figura 13.18
Dados
do Exemplo
13.14.1.

	- 1 - **	- 2 - Amostra 1	- 3 - Amostra 2	- 4 - Amostra 3	- 5 - Amostra 4	- 6 - Amostra 5	- 7 - Amostra 6	-
1	n_i	33.000	50.000	39.000	75.000	80.000	90.000	
2	m_i	0.000	16.000	28.000	41.000	45.000	70.000	

13.14.2 – INSTRUÇÕES

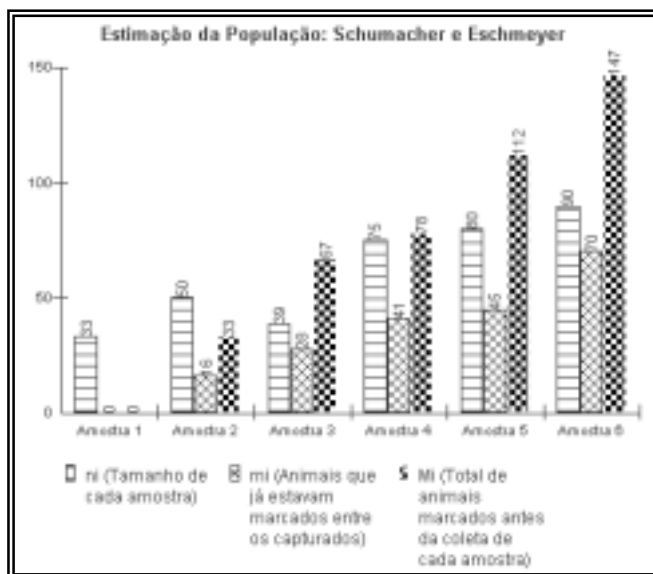
- Cada linha deve ter a seguinte disposição:
Na linha 1: n_i , número de animais capturados em cada amostra (marcados e não marcados);
Na linha 2: m_i , número de animais que já estavam marcados entre os capturados na amostra;
Todos os animais capturados em cada amostra (n_i) são marcados, exceto os que já apresentavam a marca, portanto, quantificados na linha (m_i);
- Introduzir os dados no *grid* como mostra a Figura 13.18;
- Pressionar o *mouse* em *Tamanho da população (amostragem múltipla, método de Schumacher e Eschmeyer)*;
- Selecionar as colunas do *grid* a partir da coluna 2, sendo neste exemplo de dois (2) a sete (7), totalizando as 6 amostras e clicar em *Executar*, para resultados.

Figura 13.19
Resultados
do Exemplo
13.14.1.

Arquivo Editar Gráfico	
	Resultados
Número de amostras	6
População estimada	174
Variancia	1.7072
t(alfa=0.05), (GL= 4)	2.7760
t(alfa=0.01), (GL= 4)	4.6040
IC (95%)	131 a 259
IC (99%)	112 a 384

Para obter o gráfico desse Exemplo pressionar o *mouse* em *Gráfico* da barra superior da Figura 13.19.

Figura 13.20
Gráfico
do Exemplo
13.14.1.



13.15 – TAMANHO DA POPULAÇÃO (amostragem múltipla, população aberta)

O modelo para este teste é o de *Parker*, considerando-se “população aberta” aquele universo que altera seu tamanho – por migrações, nascimentos e óbitos –, durante o período de investigação. O procedimento desse autor baseia-se na regressão linear, com ajustes pelo método dos mínimos quadrados. De acordo com o autor, o *intercepto* constitui uma estimativa do quociente entre o número de indivíduos marcados na população e o tamanho do universo.

Os dados devem ser introduzidos no *grid* geral do *BioEstat*, obedecendo aos seguintes critérios:

NACMD = número de animais inicialmente capturados, marcados e devolvidos ao respectivo meio ambiente;

períodos = número de períodos de recaptura ($i = 1, 2, 3, \dots, p$);

n_i = número de indivíduos capturados em cada amostra;

m_i = número de indivíduos marcados e recapturados em cada amostra;

13.15.1 – EXEMPLO (hipotético)

Na região de Mamirauá, Estado do Amazonas, amostra de 3250 peixes de determinada espécie foram capturados, marcados e devolvidos ao meio ambiente. Posteriormente, efetuaram-se em cinco (5) períodos distintos e consecutivos novas capturas, contando-se, em cada apreensão, o número de animais capturados (marcados e não marcados). Os dados da Figura 13.21 indicam o procedimento da introdução dos dados no *grid* geral. Estimar o tamanho da população e os intervalos de confiança de 95% e 99%. Tratando-se de amostragem múltipla e a população sendo considerada aberta, o método escolhido foi o de *Parker*.

Obs.: Somente foram marcados os 3250 indivíduos iniciais, nenhum outro animal foi marcado durante as capturas posteriores.

13.15.2 – INSTRUÇÕES

- Introduzir os dados no *grid* como mostra a Figura 13.21;
- Pressionar o *mouse* em *Tamanho da População* (*amostragem múltipla, método de Parker*);
- Selecionar as colunas do *grid*: 1, 2 e 3, neste exemplo;
- Clicar em *Executar* e preencher a *caixa de texto* com o número de espécimes inicialmente capturados (*NACMD*), marcados e devolvidos ao meio ambiente, o qual somou 3250 unidades;
- Pressionar em *OK*, para resultados.

Figura 13.21
Dados do
Exemplo
13.15.1.

	- 1 - Período	- 2 - ni	- 3 - mi	- 4 -
1	1	12456	50	
2	2	16221	28	
3	3	19004	29	
4	4	18957	15	
5	5	21388	17	

Figura 13.22
Espécimes
inicialmente
capturados.
(Exemplo
13.15.1).

Entrada de dados (NACMD)

Número de Animais Capturados, Marcados e Devolvidos

OK

Cancel

3250

Figura 13.23
Resultados
do
Exemplo
13.15.1.

	Resultados
Número de amostras =	5
Intercepto =	0.00398
Coef. de regressão =	-0.00074
População estimada =	816066
Desvio padrão a =	0.00076
Desvio padrão b =	0.00023
IC (95%) =	508090 a 2071994
IC (99%) =	386273 a 7243500

O número estimado de animais da espécie em questão e os respectivos intervalos de confiança estão contidos nos resultados.

CAPÍTULO 14

META-ANÁLISE

14.1 – INTRODUÇÃO

Define-se Meta-Análise como um procedimento destinado a examinar, de modo simultâneo, várias investigações, diversas pesquisas sobre um mesmo tópico. Esse tipo de estudo vem sendo largamente utilizado na esfera médica, para, através de numerosos dados de vários trabalhos sobre determinado assunto, aumentar o nível de confiança nas inferências estatísticas, com objetivos terapêuticos e outros. Alguns estudos, como se sabe, sobre determinada matéria, são concordantes, mas podem também ser contraditórios, o que aumenta a necessidade de análises conjuntas para que se possa concluir com maior segurança.

Embora o conceito de Meta-Análise seja oriundo de uma escola de metodologia desenvolvida por *Glass* em 1976, estudos genéticos sobre genótipos e fenótipos já abrangiam esse conceito, conforme descrito por *Sokal e Rohlf* (*Biometry*, 1969), sob o título "*Replicated tests of goodness of fit*" (página 575), ou seja, a combinação de várias amostras com informações sobre o mesmo assunto, para obter estimativas mais precisas das proporções paramétricas.

É importante distinguir os efeitos observados entre os diversos estudos examinados em conjunto:

- 1 - quando diferem somente em decorrência do erro amostral, ou seja, o verdadeiro efeito é o mesmo em cada um dos estudos, na mesma direção e não devido a diferenças sistemáticas em cada trabalho, diz-se que as amostras são homogêneas e as diferenças são consideradas efeitos randômicos ou aleatórios;
- 2 - outras vezes os efeitos observados ultrapassam o esperado pelo erro amostral, existindo diferenças reais entre os estudos examinados. Quando isso ocorre, denominando-se o efeito, neste caso, de efeito fixo, o qual pode acontecer na mesma direção dos tratamentos efetuados - favorável ou não - ou em diferentes direções - benéfico e adverso, sendo as amostras consideradas heterogêneas.

Alguns autores aplicam em tabelas 2x2 o teste do Qui-Quadrado de heterogeneidade. Quando o teste é significativo, escolhem testes de efeitos aleatórios, como o de *DerSimonian-Laird*; caso contrário, os de efeitos fixos, sendo o mais conhecido o de *Mantel-Haenszel*.

Existem várias discussões, contudo, sobre a escolha do teste, mas é preciso lembrar ser fundamental preencher, pelo menos, estas recomendações:

- a) as amostras devem ser aleatórias, representativas da população em estudo;
- b) os experimentos devem ser controlados, escolhendo-se os indivíduos de modo randômico para cada grupo. Estes procedimentos são designados na literatura inglesa como RCT (*Randomized Controlled Trials*);

- c) os critérios de tratamento devem ser semelhantes entre os estudos submetidos à meta-análise;
- d) rigorosa escolha dos trabalhos publicados, podendo-se incluir pesquisas ainda não editadas, desde que atendam os requisitos antes referidos.

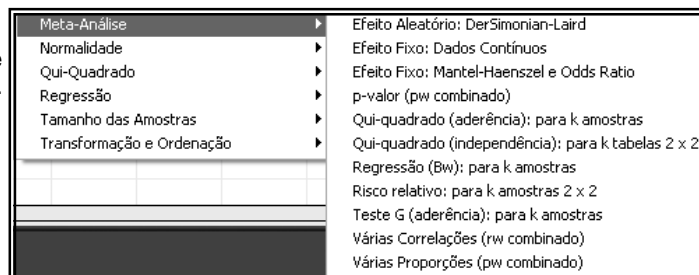
O BioEstat apresenta os seguintes aplicativos para Meta-Análise:

- a) Efeito Aleatório: *DerSimonian-Laird*: para k tabelas 2×2 ;
- b) Efeito Fixo: Dados Contínuos
- c) Efeito Fixo: *Mantel-Haenszel* e *Odds Ratio*: para k tabelas 2×2 ;
- d) p -valor (pw , combinado);
- e) Qui-Quadrado (aderência): para k amostras;
- f) Qui-Quadrado (independência): para k tabelas 2×2 ;
- g) Regressão (Bw): para k amostras;
- h) Risco relativo: para k tabelas 2×2 ;
- i) Teste G (aderência): para k amostras;
- j) Várias Correlações (rw combinado);
- k) Várias Proporções (pw combinado).

14.1.1 – INSTRUÇÕES

- a) Dirigir o *mouse* para o menu superior e clicar em *Estatística*;
- b) Apontar o *mouse* em *Meta-Análise*, conforme mostra a Figura 14.1;
- c) Introduzir os dados no *grid* geral ou nas *caixas de texto*, conforme o teste.

Figura 14.1
Menu de
Meta-Análise.



14.2 – EFEITO ALEATÓRIO: *DerSimonian-Laird*

Este teste é recomendado quando o Qui-Quadrado da heterogeneidade é significativo, aplicável em k tabelas de contingência 2×2 .

14.2.1 – EXEMPLO (dados hipotéticos)

Foram efetuados estudos sobre doença grave ocorrida em pessoas adultas, sendo, em todos esses estudos, um grupo submetido ao tratamento (A) e o outro ao tratamento tradicional (B). As amostras foram aleatórias e os indivíduos de cada grupo escolhidos pelo processo de randomização. Os resultados obtidos, após o período de tratamento previamente estabelecido, foram os seguintes:

Tabelas (Estudos)	Grupo A		Grupo B		Total
	Sucessos (a)	Insucessos (b)	Sucessos (c)	Insucessos (d)	
1	181	29	157	52	413
2	214	69	108	37	428
3	118	36	78	39	271
4	100	3	40	4	147
5	1	31	32	4	68
6	217	60	192	81	550
7	163	40	151	53	407
8	109	21	107	25	262
9	906	110	901	112	2029
10	345	80	113	26	564
11	212	30	201	50	493
12	33	17	38	12	100
13	24	16	28	20	88
14	30	1	47	10	88
15	100	1	29	3	133
16	250	173	244	177	844
17	180	29	165	32	406
18	307	41	318	47	713
19	100	6	97	7	210
20	1113	37	1080	48	2234
21	190	40	200	27	457
22	92	16	50	2	160
23	132	10	130	14	286
24	4270	268	4267	247	9052
25	369	48	355	61	833
26	123	1	97	2	223
27	293	18	300	11	622
28	1840	67	1830	72	3809
29	2000	43	1990	42	4075
30	6540	33	1661	3	8237

H_0 : a proporção de casos favoráveis é igual em ambos tratamentos;

H_1 : a proporção de casos favoráveis é maior com o tratamento A;

Nível de decisão: $\alpha = 0.05$.

14.2.2 – INSTRUÇÕES

- Apontar e pressionar o mouse em *Efeito Aleatório: DerSimonian-Laird*;
- Preencher as *caixas de texto* com os dados do Exemplo 14.2.1, Estudo 1;
- Clicar em *Nova tabela* e introduzir os dados do Estudo 2, e assim, sucessivamente, até o Estudo 30;
- Após introduzir os últimos dados, clicar em *Executar*, para resultados.

Figura 14.2
Resultados
do Exemplo
14.2.1.

Meta-análise: Efeito Aleatório

Novo Deletar Salvar Gráfico Análise Imprimir

Arquivo atual: Ex 14_2_1.ERND

Arquivos disponíveis: Ex 14_2_1.ERND

Entrada de Dados

	Sucessos	Insucessos
Amostra 1	6540	33
Amostra 2	1661	3

TABELAS 30 / 30

Nova tabela

Excluir

Limpar

Resultados

"Odds ratio" (combinado): 1.1486

Erro padrão: 0.0631

IC 95% (odds): 1.0149 ≤ μ ≤ 1.2999

Z: 2.1935

p-valor: 0.0283

Teste de Heterogeneidade

Qui-quadrado: 47.1320

Graus de Liberdade: 29

p-valor: 0.0181

O *Odds Ratio* (1.1486) é significativo ($p\text{-valor} = 0.0285$) e contido no Intervalo de Confiança 95% (1.0149 a 1.2999), cujo limite inferior está acima da unidade (valor 1), rejeitando-se, portanto, a hipótese de nulidade. O teste de heterogeneidade foi significativo ($p\text{-valor} = 0.0181$), o que sugere o teste escolhido - *DerSimonian-Laird*. O tratamento A, em conclusão, é superior ao B. O número total de indivíduos cuidados com os tratamentos referidos é de 38202, conferindo maior relevância aos resultados.

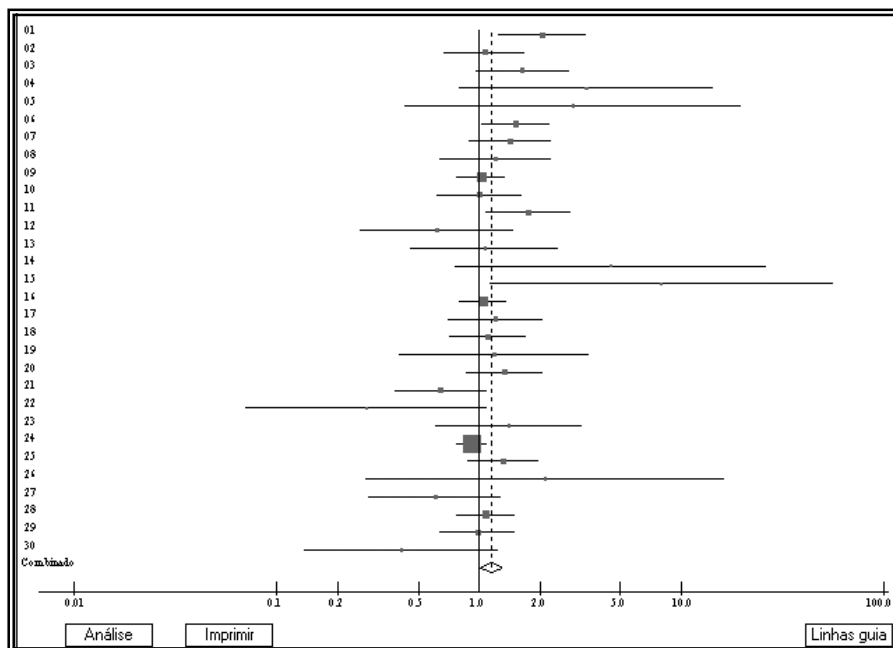
Está também disponível a análise dos dados de cada amostra: *Odds Ratio*, Intervalos de Confiança (95%) e o Peso (inverso da variância); para obter a tabela abaixo basta pressionar com o mouse em *Análise* na barra da Figura 14.2.

Figura 14.3
Visão parcial
da *Análise* e da
Meta-Análise
do Exemplo
14.2.1.

Amostra	Odds	IC Inf.	IC Sup.	Peso
01	2.051	1.245	3.378	15.431
02	1.067	0.674	1.689	18.204
03	1.634	0.959	2.783	13.533
...
26	2.111	0.274	16.262	0.922
27	0.607	0.286	1.290	6.768
28	1.080	0.770	1.514	33.673
29	0.982	0.640	1.505	21.044
30	0.411	0.137	1.238	3.161
Combinado	1.149	1.015	1.300	

Para obter o gráfico *Forest Plot* pressionar o *mouse* em *Gráfico* na barra da Figura 14.2, para resultado.

Figura 14.4
Gráfico do
Exemplo
14.2.1



O *Odds Ratio* (*w*) combinado está representado na parte inferior do gráfico, sob a forma de um pequeno losango, cujo limite inferior é maior que a unidade (1.015).

14.3 – EFEITO FIXO: DADOS CONTÍNUOS

Em determinadas situações dispomos de vários estudos sob a forma de dados contínuos, medidos em escala intervalar ou de razões, onde cada estudo compara duas amostras independentes. Essas pesquisas podem ser reunidas, desde que se tratem do mesmo tópico.

14.3.1 – EXEMPLO (dados hipotéticos)

Duas drogas foram utilizadas no tratamento de indivíduos com taxas elevadas de colesterolemia, cujas idades variavam entre 40 a 60 anos. A finalidade da pesquisa era testar o 'risco de aterosclerose' em dez (10) estudos independentes, do ponto de vista do *Coefficiente Total/HDL*. Dos dados coletados, registrou-se o tamanho das amostras, os valores médios dos coeficientes obtidos e os respectivos desvios padrões, dispostos no *grid* geral (Fig. 14.5), como segue:

- a) N (A): tamanho das amostras com a droga A;
- b) Média (A): média dos *Coefficientes* das amostras (tratamento A);
- c) DP (A): desvio padrão das amostras (tratamento A);

- d) N (B): tamanho das amostras com a droga B;
- e) Média (B): média dos *Coefficientes* das amostras (tratamento B);
- f) DP (B): desvio padrão das amostras (tratamento B);

Figura 14.5
Dados
do Exemplo
14.3.1.

	- 1 - N (A)	- 2 - Média (A)	- 3 - DP (A)	- 4 - N (B)	- 5 - Média (B)	- 6 - DP (B)	- 7 -
1	102.000	6.000	4.300	100.000	6.700	4.900	
2	170.000	4.800	4.600	149.000	5.060	5.360	
3	139.000	2.000	2.600	142.000	2.530	3.300	
4	180.000	2.690	2.400	180.000	3.210	2.450	
5	175.000	6.050	4.900	170.000	5.800	5.200	
6	752.000	4.690	5.400	740.000	4.750	5.310	
7	201.000	10.000	8.090	200.000	10.800	7.890	
8	1148.000	2.810	3.020	1100.000	3.000	3.200	
9	670.000	3.900	4.860	670.000	4.400	5.400	
10	255.000	4.680	5.200	250.000	4.700	5.120	

H_0 : não há diferença entre as médias dos *Coefficientes Total/HDL* entre os tratamentos A e B: $\mu_1 = \mu_0$;

H_1 : há diferença entre as médias dos *Coefficientes Total/HDL* entre os tratamentos A e B: $\mu_1 \neq \mu_0$;

Nível de decisão: alfa = 0.05

14.3.2 – INSTRUÇÕES

- a) Inserir os dados no *grid* geral, conforme mostra a Figura 14.5, obedecendo rigorosamente à ordem e à disposição dos dados nas colunas desse *grid*;
- b) Apontar e pressionar o mouse em *Efeito Fixo: Dados Contínuos*;
- c) Selecionar as seis (6) colunas do *grid* geral e *Executar*, para resultados.

Figura 14.6
Resultados
do Exemplo
14.3.1.

Arquivo Editar Gráfico Análise	
Resultados	
Número de estudos =	10
Média das diferenças (MD w) =	-0.2725
Erro padrão (EP w) =	0.0904
Z (w) =	3.0143
p-valor (w) =	0.0026
IC 95% =	-0.4497 a -0.0962

Os resultados apresentam diferença muito significativa ($p\text{-valor} = 0.0026$) entre os dois tratamentos, podendo-se dizer que a média dos *Coefficientes Total/HDL* com a droga A é inferior aos obtidos com a droga B (Média das diferenças = -0.2725). Foram testados 3792 indivíduos no grupo A e 3701, no B, totalizando 7493 pessoas.

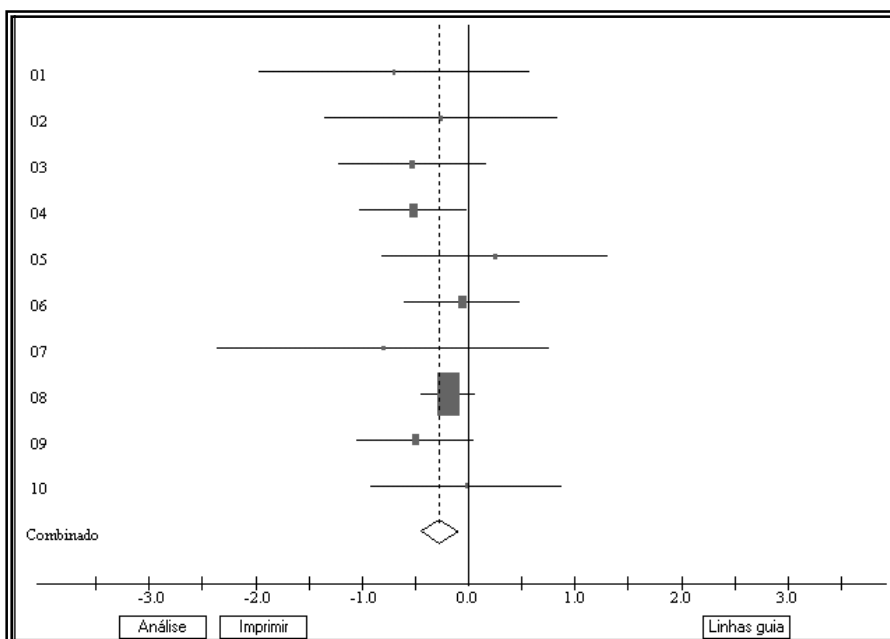
Se desejar os dados de cada amostra (Média das Diferenças e respectivo Intervalo de Confiança de 95%), pressionar com o *mouse* em *Análise* (Figura 14.6).

Figura 14.7
Análise e
Meta-Análise
do Exemplo
14.3.1

Amostra	Med. Dif.	IC Inf.	IC Sup.	Peso
01	-0.700	-1.971	0.571	2.379
02	-0.260	-1.353	0.833	3.216
03	-0.530	-1.226	0.166	7.939
04	-0.520	-1.021	-0.019	15.303
05	0.250	-0.816	1.316	3.381
06	-0.060	-0.604	0.484	13.004
07	-0.800	-2.364	0.764	1.570
08	-0.190	-0.447	0.067	58.102
09	-0.500	-1.050	0.050	12.694
10	-0.020	-0.920	0.880	4.740
Combinado	-0.273	-0.450	-0.096	

Para obter a Figura 14.8: pressionar o mouse em *Gráfico* na barra da Figura 14.6.

Figura 14.8
Gráfico do
Exemplo
14.3.1



A Média Geral das Diferenças (MD_w) e os respectivos Intervalos de Confiança estão representados na parte inferior do gráfico, sob a forma de um pequeno losango cujo limite superior é menor que a unidade (-0.096).

14.4 – EFEITO FIXO: MANTEL-HAENSZEL E ODDS RATIO

Este teste é também utilizado em análise de sobrevivência, mas é indicado para avaliação dos chamados efeitos fixos, utilizando-se k tabelas 2×2 . Calcula-se, simultaneamente, o valor do *Odds Ratio* e o respectivo intervalo de confiança.

14.4.1 – EXEMPLO (Dados hipotéticos)

Estudos realizados durante cinco (5) anos em diferentes centros populacionais, sobre a ocorrência de câncer pulmonar em fumantes e não fumantes, revelaram o seguinte:

BELFAST	Com câncer	Sem câncer
Fumantes	12	598
Não Fumantes	2	613

LISBOA	Com câncer	Sem câncer
Fumantes	14	734
Não Fumantes	3	758

LONDRES	Com câncer	Sem câncer
Fumantes	14	709
Não Fumantes	5	742

MADRI	Com câncer	Sem câncer
Fumantes	17	684
Não Fumantes	4	650

BERLIM	Com câncer	Sem câncer
Fumantes	22	918
Não Fumantes	3	832

H_0 : a prevalência de câncer pulmonar (*Efeito Fixo*) em indivíduos fumantes é igual a da observada em pessoas não fumantes: $\mu_1 = \mu_0$;

H_1 : a prevalência de câncer pulmonar em indivíduos fumantes é diferente da observada em pessoas não fumantes: $\mu_1 \neq \mu_0$;

Nível de decisão: $\alpha = 0.01$.

14.4.2 – INSTRUÇÕES

- Apontar e pressionar o mouse em *Efeito Fixo: Mantel-Haenszel e Odds Ratio*;
- Preencher as *caixas de texto* com os dados da primeira tabela 2 x 2;
- Clicar em *Nova tabela* e introduzir os dados da segunda tabela e assim, sucessivamente, até a última;
- Clicar em *Executar*, para resultados.

Figura 14.9
Resultados
do Exemplo
14.4.1.

Meta-Análise: Qui-quadrado, Mantel-Haenszel

Novo Deletar Salvar Gráfico Análise Imprimir

Arquivo atual: **Ex 14.4.1.EFIX**

Arquivos disponíveis:
Ex 14.4.1.EFIX
Ex_4_8_1.EFIX

Entrada de Dados

	Sucessos	Insucessos
Amostra 1	22	918
Amostra 2	3	832

TABELAS 5 / 5
Nova tabela
Excluir
Limpar

Resultados

Qui-Quadrado	p-valor	
37.1521	< 0.0001	
Graus de liberdade	"Odds-Ratio"(combinado)	NNH (Combinado)
1	4.5821	61
IC 95% (odds)		
2.7052	$\leq \mu \leq$	7.7610

Executar Cancelar

A meta-análise, efetuada em amostras coletadas por vários investigadores em cinco (5) cidades do continente europeu, revelou o seguinte: a diferença expressa em termos de Qui-quadrado é estatisticamente muito significativa ($p < 0.0001$), ou seja, a proporção de indivíduos fumantes com câncer pulmonar é maior que a de não fumantes com essa neoplasia. O *Odds Ratio* indica que essa proporção é quatro (4) vezes e meia maior entre fumantes, com intervalo de confiança (95%) de 2.71 a 7.76. O total de indivíduos testados nessas cinco (5) amostras foi igual a 7334, o que confere maior confiabilidade nos resultados obtidos. O NNH obtido evidencia que em cada grupo de 61 fumantes deve ocorrer um (1) novo caso de câncer pulmonar no período de cinco (5) anos.

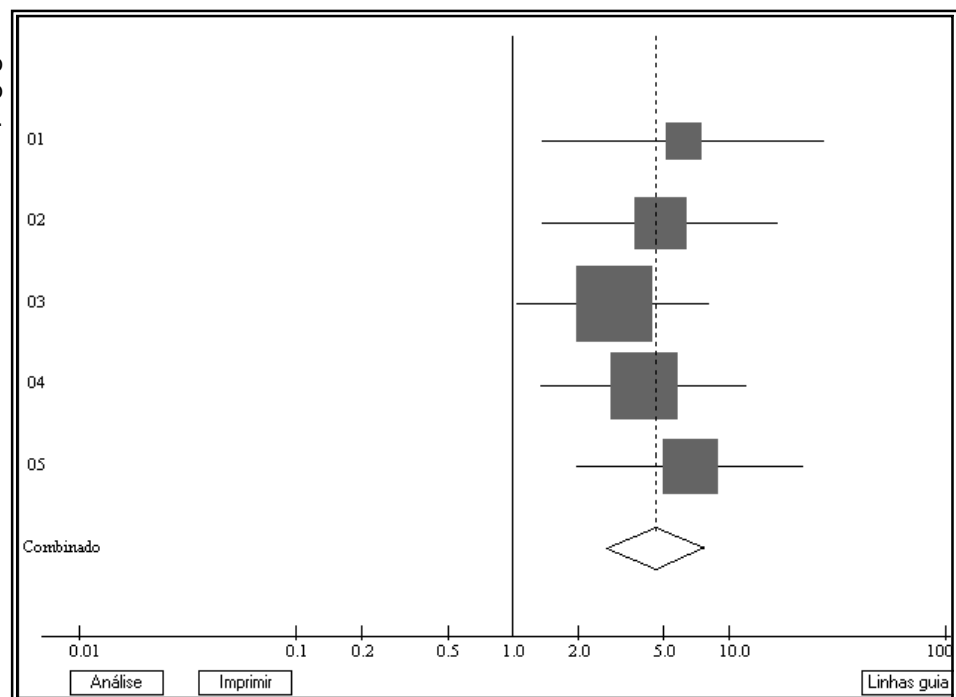
O procedimento para se obter a *Análise dos Dados* é o mesmo dos exemplos anteriores, clicando em *Análise* na barra da Figura 14.9.

Figura 14.10
Análise e
Meta-Análise
do Exemplo
14.4.1.

Meta Análise - Dados ilustrados no Forest Plot				
Amostra	Odds	IC Inf.	IC Sup.	Peso
01	6.151	1.371	27.598	1.705
02	4.819	1.379	16.839	2.454
03	2.930	1.050	8.178	3.647
04	4.039	1.352	12.066	3.207
05	6.646	1.982	22.287	2.624
Combinado	4.582	2.705	7.761	

Pressionar o mouse em Gráfico na barra da Figura 14.9, para obter o resultado.

Figura 14.11
Gráfico do
Exemplo
14.4.1.



O teste de *Mantel-Haenszel* é altamente significativo e o *Odds Ratio* combinado, com o respectivo Intervalo de Confiança (95%), está representado pelo losango na parte inferior do gráfico.

14.5 – P-VALOR (*pw combinado*)

Em alguns estudos, sobre o mesmo tópico, dispõe-se apenas das probabilidades de erro na conclusão de cada pesquisa de acordo com a hipótese formulada. Há, assim, o *p-valor* de cada trabalho, sendo possível, nesses casos, reuni-los para *Meta-Análise* obtendo-se, ao final, um *p-valor* que representa o conjunto de dados submetidos ao teste. O *p-valor* de cada estudo é transformado em *logaritmo neperiano*, aplicando-se, ao final, o teste do Qui-quadrado para obtenção do valor combinado ($w = pooled$).

14.5.1 – EXEMPLO

Coletaram-se 15 estudos e as respectivas probabilidades (*p-valor*) de testes efetuados em amostras aleatórias, com o pressuposto de distribuição normal das populações de onde eram oriundas e o conhecimento das variâncias paramétricas.

Foi aplicado o teste Z em cada um dos estudos efetuados. Os dados referidos estão parcialmente dispostos na Fig. 14.12.

H_0 : as probabilidades dos testes submetidos à *Meta-Análise* não apresentam *p-valor* combinado (w) significativo;

H_1 : as probabilidades dos testes submetidos à *Meta-Análise* apresentam *p-valor* combinado (w) significativo;

Nível de decisão: alfa = 0.01.

Figura 14.12
Dados
parciais do
Exemplo
14.5.1.

	- 1 -	- 2 -	-
** Estudo	p-valor		
1	1	0.0500	
2	2	0.0300	
3	3	0.0700	
4	4	0.3000	
5	5	0.0600	
6	6	0.0400	
7	7	0.0500	

14.5.2 – INSTRUÇÕES

- Inserir os dados no *grid* geral conforme mostra a Figura 14.12;
- Apontar e pressionar o *mouse* em *p-valor*;
- Selecionar a coluna 2 e clicar em *Executar*, para resultados.

Figura 14.13
Resultados
do Exemplo
14.5.1.

Número de Estudos	15
Qui-quadrado (w)	86.6554
Graus de liberdade	30
p-valor (w)	< 0.0001

O Qui-quadrado com 30 graus de liberdade foi igual a 86.6554, cujo *p-valor* (w) é altamente significativo ($p < 0.0001$).

14.6 – QUI-QUADRADO (aderência): PARA k AMOSTRAS

Indicado para comparar várias amostras em testes de aderência para verificar se são ou não homogêneas.

14.6.1 – EXEMPLO

Coletaram-se quatro (4) amostras sobre sazonalidade de nascimentos de crianças cujos pais residiam em localidades rurais ou urbanas do Médio Solimões:

Categoria A	Categoria B
Nascimentos de maio a outubro (tempo menos chuvoso e de vazante do rio)	Nascimentos de novembro a abril (período chuvoso e de enchente do rio)

H_0 : Os desvios entre o Observado e o Esperado nas categorias de cada amostra não ocorrem na mesma direção, sendo as mesmas heterogêneas;

H_1 : Os desvios entre o Observado e o Esperado nas categorias de cada amostra ocorrem na mesma direção, sendo as mesmas homogêneas;

Nível de decisão: alfa = 0.05.

14.6.2 – INSTRUÇÕES

- Apontar o mouse em *Qui-quadrado Aderência: k amostras*;
- Preencher as *caixas de texto* com o número de amostras e de categorias;
- Clicar em *Ativar entrada de dados* e, em seguida, preencher o *grid específico* com os dados;

Figura 14.14
Dados do
Exemplo
14.6.1.

- Clicar em *Executar*, para resultados.

Figura 14.15
Dados do
Exemplo
14.6.1.

χ^2	Amostra 1 Observado	Amostra 1 Esperado	Amostra 2 Observado	Amostra 2 Esperado	Amostra 3 Observado	Amostra 3 Esperado	Amostra 4 Observado	Amostra 4 Esperado
Categoria A	474	487	523	532	277	303	266	283
Categoria B	500	487	541	532	329	303	300	283

O total de nascidos no período A atingiu 1540, enquanto no período B somou 1670.

Figura 14.16
Resultados
do Exemplo
14.6.1.

	Amostra 1	Amostra 2	Amostra 3	Amostra 4
Qui-Quadrado	0.6940	0.3045	4.4620	2.0424
Graus de Liberdade	1	1	1	1
(p)=	0.4048	0.5811	0.0347	0.1530
Qui-Quadrado Total	7.5030	---	---	---
Graus de Liberdade	4	---	---	---
(p)=	0.1116	---	---	---
Homogeneidade	5.2648	---	---	---
Graus de Liberdade	1	---	---	---
(p)=	0.0218	---	---	---
Heterogeneidade	2.2382	---	---	---
Graus de Liberdade	3	---	---	---
(p)=	0.5245	---	---	---

As amostras são homogêneas ($p = 0.0218$), ou seja, ocorreu, em todas as amostras, maior número de nascimentos no período chuvoso e de enchente do rio (Categoria B). Não há evidência de heterogeneidade das amostras ($p = 0.5245$).

14.7 – QUI-QUADRADO (independência): PARA k TABELAS 2 x 2

Da mesma forma, neste teste procura-se verificar se as amostras foram retiradas da mesma população, isto é, se são homogêneas ou não.

14.7.1 – EXEMPLO

Em oito localidades endêmicas de malária efetuou-se pesquisa de controle de casos (hospitalares), registrando-se os indivíduos que eram admitidos com manifestações sugestivas da doença, oriundos da zona urbana ou da área rural, confirmando-se ou

não o diagnóstico *a posteriori*. As proporções de casos positivos e negativos de cada município foram estas:

ZONA	Município A		Município B		Município C		Município D	
	POSITIVO	NEGATIVO	POSITIVO	NEGATIVO	POSITIVO	NEGATIVO	POSITIVO	NEGATIVO
URBANA	35	50	48	56	17	18	30	52
RURAL	70	55	74	60	29	20	73	39

ZONA	Município E		Município F		Município G		Município H	
	POSITIVO	NEGATIVO	POSITIVO	NEGATIVO	POSITIVO	NEGATIVO	POSITIVO	NEGATIVO
URBANA	33	52	43	66	31	55	31	59
RURAL	75	50	78	47	78	40	74	38

Os totais gerais foram estes:

- Zona Urbana: testes positivos 268 casos;
- Zona Urbana: testes negativos 408 casos;
- Zona Rural: testes positivos 551 casos;
- Zona Rural: testes negativos 349 casos;
- Total geral = $268 + 408 + 551 + 349 = 1576$ casos

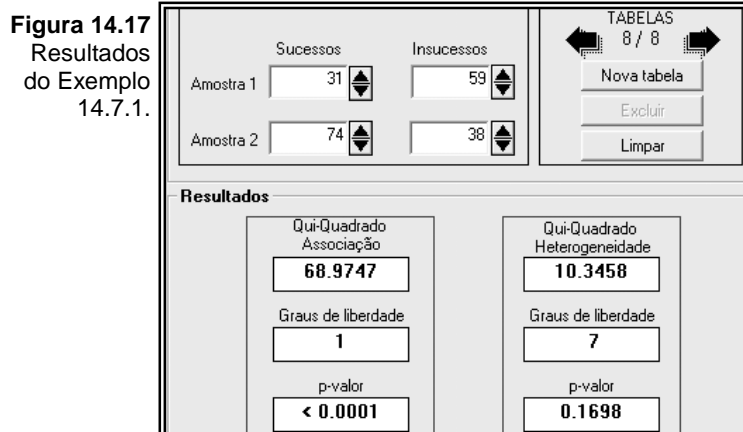
H_0 : a incidência da malária não está relacionada com a área de residência dos pacientes;

H_1 : a incidência da malária ocorre com mais frequência em pessoas residentes na zona rural, havendo, portanto, associação;

Nível de decisão: $\alpha = 0.01$.

14.7.2 – INSTRUÇÕES

- Apontar o mouse em *Qui-quadrado (independência)*: para k tabelas 2×2 , preenchendo as caixas de texto com os valores das tabelas;
- Após introduzir a última tabela (município H, nesse exemplo), clicar em *Executar*, para resultados.



O Qui-Quadrado de associação é muito significativo ($p < 0.0001$), rejeitando-se a hipótese de nulidade e aceitando-se a alternativa. A incidência de malária é maior nas zonas rurais do que nas urbanas nos oito municípios investigados, compreendendo 1576 pessoas. As amostras não são heterogêneas.

14.8 – REGRESSÃO (B_w): PARA k AMOSTRAS

Trata-se de verificar se várias amostras, no que diz respeito aos coeficientes de regressão linear e aos interceptos, provieram do mesmo universo. Trata-se de uma combinação das análises de variância e de regressão, conhecida como Análise de Covariância ou ANCOVA. O programa determina, ainda, o coeficiente de regressão (b) e o intercepto (a) de cada amostra, bem como o coeficiente de regressão conjunto, também denominado *pooled regression*, simbolizado por B_w . A homogeneidade dos coeficientes de regressão e dos interceptos é determinada pela análise de variância, constatando-se ou não se eles são oriundos da mesma população, ou seja, se são homogêneos ou heterogêneos. As amostras podem ser de diferentes tamanhos.

14.8.1 – EXEMPLO

Registrou-se a pressão arterial diastólica (P.A.D.) de pessoas do sexo feminino, cujas idades variavam de 39 a 73 anos, selecionando-se, ao acaso, três amostras obtidas em diferentes cidades, cada uma constituída de oito (8) indivíduos. O objetivo do estudo é verificar se a relação entre a *idade* e a P.A.D. é homogênea, para as três amostras levantadas. As hipóteses formuladas foram as seguintes:

- H_0 : Os coeficientes de regressão são iguais, pois as amostras foram retiradas da mesma população: $b_1 = b_2 = b_3$;
 H_1 : Os coeficientes de regressão são desiguais, pois as amostras foram retiradas de populações diferentes: pelo menos um $b_i \neq b_j$;
 H_0 : Os interceptos são iguais, pois as amostras foram retiradas da mesma população: $a_1 = a_2 = a_3$;
 H_1 : Os interceptos são desiguais, pois as amostras foram retiradas de populações diferentes: pelo menos um $a_i \neq a_j$;
 Nível de decisão: alfa = 0.05.

Figura 14.18
Dados do
Exemplo
14.8.1.

	- 1 - Y 1	- 2 - X 1	- 3 - Y 2	- 4 - X 2	- 5 - Y 3	- 6 - X 3
1	61.000	82.000	61.000	83.000	44.000	68.000
2	65.000	84.000	64.000	85.000	62.000	84.000
3	54.000	75.000	53.000	71.000	63.000	85.000
4	43.000	70.000	43.000	69.000	51.000	72.000
5	39.000	68.000	65.000	76.000	64.000	77.000
6	66.000	75.000	67.000	84.000	66.000	85.000
7	67.000	85.000	71.000	91.000	70.000	92.000
8	72.000	90.000	69.000	89.000	73.000	91.000

Note que neste teste a ordem de seleção das colunas requer que para cada par (Y,X) a variável dependente (Y) deve preceder a variável independente (X).

14.8.2 – INSTRUÇÕES

- Preencher o *grid* geral com os escores amostrais, introduzindo na primeira coluna os valores da variável P.A.D. (Y_1) e na seguinte, os valores da variável *Idade* (X_1) e, de modo sucessivo, para as demais amostras;
- Clicar em *Regressão (Bw)*: para k amostras e selecionar as colunas do *grid* geral;
- Clicar em *Executar*, para resultados.

Figura 14.19
Resultados
do Exemplo
14.8.1.

	Resultados	
Número de amostras	3	---
F (Regressão)	0.7463	---
Graus de liberdade	2 , 18	---
(p) =	0.5063	---
F (Interceptos)	0.0661	---
Graus de liberdade	2 , 20	---
(p) =	0.9359	---
AMOSTRA 1: Intercepto (a)	-49.4185	---
Coef. regressão (b)	1.3710	(p) = 0.0034
AMOSTRA 2: Intercepto (a)	-20.7776	---
Coef. regressão (b)	1.0173	(p) = 0.0040
AMOSTRA 3: Intercepto (a)	-23.5215	---
Coef. regressão (b)	1.0415	(p) = 0.0011
Bw (Pooled Regression)	1.1327	---
F (Bw)	80.1078	---
Graus de liberdade	1 , 20	---
(p) =	< 0.0001	---

A análise comparativa dos coeficientes de regressão exibe um valor de F (Regressão) não-significativo, pois $p = 0.5063$, para 2,18 graus de liberdade, aceitando-se por isso a hipótese de nulidade. O mesmo pode-se afirmar da análise comparativa dos interceptos, cujo valor de F (Interceptos) não é significativo, sendo $p = 0.9359$, com 2,20 graus de liberdade. Assim, os coeficientes de regressão das amostras (b_1 , b_2 e b_3) são homogêneos, o mesmo acontecendo com os interceptos (a_1 , a_2 e a_3). Conclui-se que as amostras são homogêneas. O coeficiente de regressão conjunto - *pooled regression* - é igual a 1.1327 (Bw), F (Bw) igual a 80.11, cujo p -valor é menor que 0.0001, muito significativo. A Figura 14.19 exibe, ainda, os valores do intercepto (a) e do coeficiente de regressão (b) de cada amostra.

14.9 – RISCO RELATIVO (w): PARA k TABELAS 2×2

Da mesma forma que se calcula o "*Odds Ratio*" (Teste de *Mantel-Haenszel*) em tabelas 2×2 , pode-se verificar o *Risco Relativo* (RR), como se demonstra a seguir.

14.9.1 – EXEMPLO

O exemplo 14.4.1 e as respectivas tabelas 2×2 , com as seguintes hipóteses:

H_0 : o risco relativo de câncer pulmonar em indivíduos fumantes é igual ao observado em pessoas não fumantes;

H_1 : o risco relativo de câncer pulmonar em indivíduos fumantes é maior que o observado em pessoas não fumantes;

Nível de decisão: $\alpha = 0.01$.

14.9.2 – INSTRUÇÕES

- Apontar e pressionar o mouse em *Risco Relativo*: para k amostras 2×2 ;
- Preencher as caixas de texto com os dados da primeira tabela 2×2 ;
- Clicar em *Nova tabela* e introduzir os dados da segunda tabela e assim, sucessivamente, até a última tabela (quinta, neste exemplo);
- Após introduzir a última tabela, clicar em *Executar*, para resultados.

Figura 14.20
Resultados
do Exemplo
14.9.1.

O *Risco Relativo* de câncer pulmonar é cerca de quatro (4) vezes maior em fumantes do que em não fumantes, sendo a diferença muito significativa ($p < 0.0001$). As amostras são homogêneas, ou seja, na mesma direção, como pode ser visto no resultado do teste do Qui-quadrado de homogeneidade = 35.26 e $p < 0.0001$.

O procedimento para se obter a *Análise dos Dados* é o mesmo dos exemplos anteriores, clicando-se em *Análise* na barra da Figura 14.20.

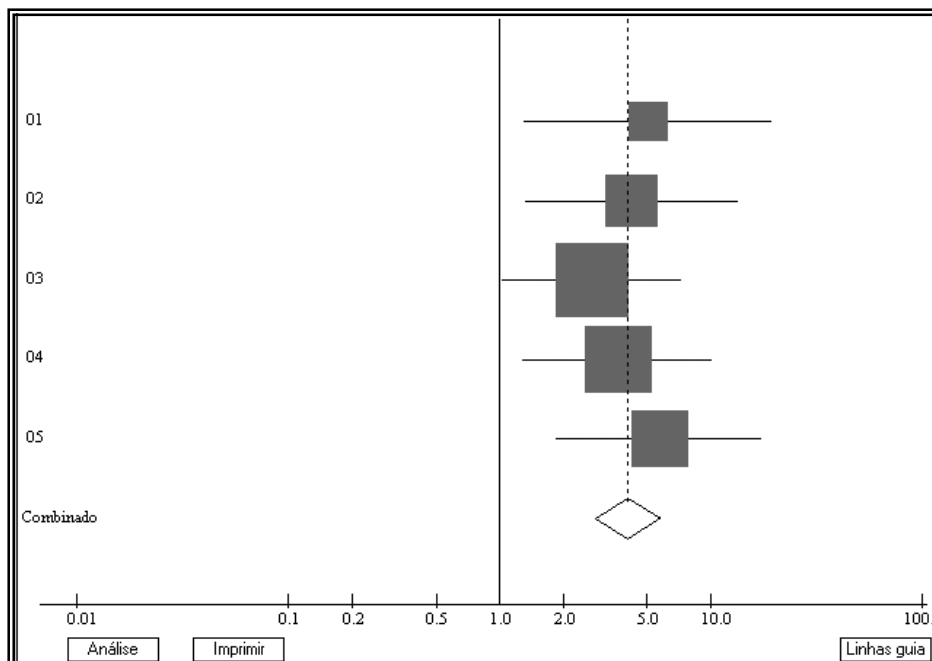
O NNH obtido evidencia em cada grupo de 61 fumantes deve ocorrer um (1) novo caso de câncer pulmonar no período de cinco (5) anos.

Figura 14.21
Análise e
Meta-Análise
do Exemplo
14.9.1

Amostra	Risco Relativo	IC Inf.	IC Sup.	Peso
01	5.041	1.302	19.509	2.098
02	4.215	1.317	13.484	2.841
03	2.724	1.026	7.230	4.031
04	3.629	1.295	10.169	3.618
05	5.711	1.859	17.545	3.050
Combinado	4.053	2.829	5.799	

Pressionar o mouse em *Gráfico* na barra da Figura 14.20, para resultado:

Figura 14.22
Gráfico do
Exemplo
14.9.1



O *Risco Relativo combinado* é altamente significativo e está representado pelo losango na parte inferior do gráfico, com os respectivos intervalos de confiança (95%), e cujo limite inferior é maior que a unidade ($2.829 > 1$).

14.10 – TESTE G (aderência): PARA k AMOSTRAS

O teste G de aderência é semelhante ao teste do Qui-Quadrado de aderência para várias amostras, sendo que os valores esperados em cada categoria estão dispostos em somente uma coluna, em números decimais.

14.10.1 – EXEMPLO

O mesmo exemplo do Qui-Quadrado (Exemplo 14.6.1), para o teste de Homogeneidade (aderência) com quatro (4) amostras, sobre sazonalidade de nascimentos de crianças cujos pais residiam em localidades rurais ou urbanas do Médio Amazonas, estando os dados assim dispostos:

Categoria A: nascimentos de maio a outubro;

Categoria B: nascimentos de novembro a abril.

H_0 : os desvios entre os valores observados e os esperados em cada categoria não ocorrem na mesma direção: as amostras são heterogêneas;

H_1 : os desvios entre os valores observados e os esperados em cada categoria ocorrem nas mesmas direções: as amostras são homogêneas;

Nível de decisão: $\alpha = 0.05$.

14.10.2 – INSTRUÇÕES

- Apontar o mouse em *Teste G (aderência)*: para k amostras, surgindo a Figura 14.23;
- Preencher as *caixas de texto* com o número de amostras e o de categorias, 4 e 2, respectivamente, no presente caso;
- Clique em *Ativar entrada de dados*;

Figura 14.23
Quantidade de amostras e de categorias.

- Preencher, a seguir, com os dados amostrais, as *caixas de texto* conforme demonstrado na Figura 14.24, clicando em *Executar*, para resultados.

Figura 14.24
Entrada de dados do Teste G.

G Teste: Homogeneidade						
Novo Deletar Salvar Imprimir Executar						
Exemplo 15 7 1.G Exemplo_14_10_1		Amostra 1 Observado	Amostra 2 Observado	Amostra 3 Observado	Amostra 4 Observado	Esperado (em decimais)
Categoria A		474	523	277	266	0.5
Categoria B		500	541	329	300	0.5

Figura 14.25
Resultados do exemplo 14.10.1.

	Amostra 1	Amostra 2	Amostra 3	Amostra 4
G Amostral =	0.6941	0.3045	4.4675	2.0436
GL =	1	1	1	1
(p) =	0.4048	0.5811	0.0345	0.1528
G Total =	7.5098	---	---	---
GL =	4	---	---	---
(p) =	0.1113	---	---	---
G Homogeneidade =	5.2662	---	---	---
GL =	1	---	---	---
(p) =	0.0217	---	---	---
G Heterogeneidade =	2.2436	---	---	---
GL =	3	---	---	---
(p) =	0.5234	---	---	---

Os resultados mostram que as amostras são homogêneas, com valores observados e esperados na mesma direção, sendo o teste significativo com p -valor igual a 0.0217. Aceita-se a hipótese de alternativa, rejeitando-se a hipótese alternativa. Os resultados são semelhantes aos obtidos no teste de Homogeneidade (aderência) pelo Qui-Quadrado, sendo também o teste G Heterogeneidade não significativo ($p = 0.5234$).

14.11 – VÁRIAS CORRELAÇÕES (r_w combinado)

Destinado a testar vários coeficientes de correlação linear (r), no sentido de verificar se foram retirados da mesma população, ou seja, se as amostras de onde provieram apresentam homogeneidade.

14.11.1 – EXEMPLO

Pesquisa realizada sobre a associação linear entre peso e estatura de recém-nascidos foi efetuada em três (3) maternidades, com a finalidade de verificar se as amostras eram homogêneas. Cada amostra, independentemente do tamanho, tem um valor de r correspondente ao grau de associação entre as duas variáveis analisadas.

$H_0: r_1 = r_2 = r_3;$

$H_1:$ pelo menos $r_i \neq r_k;$

Nível de decisão: $\alpha = 0.05.$

14.11.2 – INSTRUÇÕES

- Pressionar o mouse em *Várias correlações: homogeneidade*;
- Preencher a tela seguinte (Figura 14.26) com o número de amostras a testar, sendo três (3) neste exemplo, pressionando-se, então, em *Montar Grid*;

Figura 14.26
Número de amostras.
(Exemplo 14.11.1).

Figura 14.27
Dados do Exemplo 14.11.1.

	Tamanho (n)	r (Pearson)
Amostra 1	25	0.51
Amostra 2	30	0.58
Amostra 3	34	0.88

- Preencher então o *grid* específico com o Tamanho de cada amostra e o respectivo valor de r (Figura 14.27) e clicar em *Executar* para resultados.

Figura 14.28
Resultados do Exemplo 14.11.1.

	Resultado	
Número de amostras =	3	
Qui-Quadrado =	11.0325	
Graus de liberdade =	2	
p-valor(Qui-quadrado)=	0.0040	
Qui-quadrado ajustado =	10.9874	
p-valor(ajustado)=	0.0041	
rw (combinado)=	0.7218	
Comparações (amostras)	q	p-valor
r (1 e 2)	0.4911	ns
r (1 e 3)	4.1246	< 0.01
r (2 e 3)	3.8321	< 0.05
Valor crítico (0.05) =	3.314	---
Valor crítico (0.01) =	4.120	---

O p -valor ajustado (0.0041) do teste com todas as amostras é muito significativo, rejeitando-se a hipótese de nulidade: as amostras são heterogêneas. Quando os valores de r são comparados entre si, constata-se que há diferença significativa entre os coeficientes das amostras 2 e 3 ($p < 0.05$) e muito significativa entre aqueles representados pelas amostras 1 e 3 ($p < 0.01$). Observa-se, ainda, que não há diferença significativa entre os coeficientes das amostras 1 e 2 (ns).

14.12 – VÁRIAS PROPORÇÕES (pw combinado)

Em determinados estudos há interesse em comparar várias proporções no sentido de determinar se elas provieram ou não do mesmo universo, podendo ser efetuada a pesquisa por este modelo de procedimento.

14.12.1 – EXEMPLO

Procurou-se avaliar se os resultados da contagem de linfócitos na fórmula leucocitária do sangue periférico eram influenciados pelos tipos de aparelhos eletrônicos atualmente utilizados pelos laboratórios clínicos. Retirou-se amostra sangüínea de um indivíduo considerado clinicamente normal e efetuou-se o procedimento laboratorial em quatro (4) aparelhos de marcas diferentes: A, B, C e D. Os resultados foram estes:

Figura 14.29
Dados do
Exemplo
14.12.1.

	- 1 -	- 2 -	- 3 -	- 4 -
	** Aparelhos	n (leucócitos)	x (linfócitos)	
1	A	200.000	40.000	
2	B	150.000	51.000	
3	C	100.000	28.000	
4	D	140.000	35.000	

H_0 : Não há diferença nas proporções de linfócitos obtidas através de quatro (4) aparelhos diferentes: $p_1 = p_2 = p_3 = p_4$;

H_1 : Há diferença nas proporções de linfócitos obtidas nos aparelhos testados: pelo menos duas proporções são desiguais; $p_i \neq p_j$;

Nível de decisão: $\alpha = 0.05$.

14.12.2 – INSTRUÇÕES

- Clicar o mouse em *Várias Proporções: (pw combinado)*;
- Selecionar as colunas do *grid* geral referentes aos dados das colunas 2 e 3;
- Clicar em *Executar*, para resultados.

Figura 14.30
Resultados
(Exemplo
14.12.1).

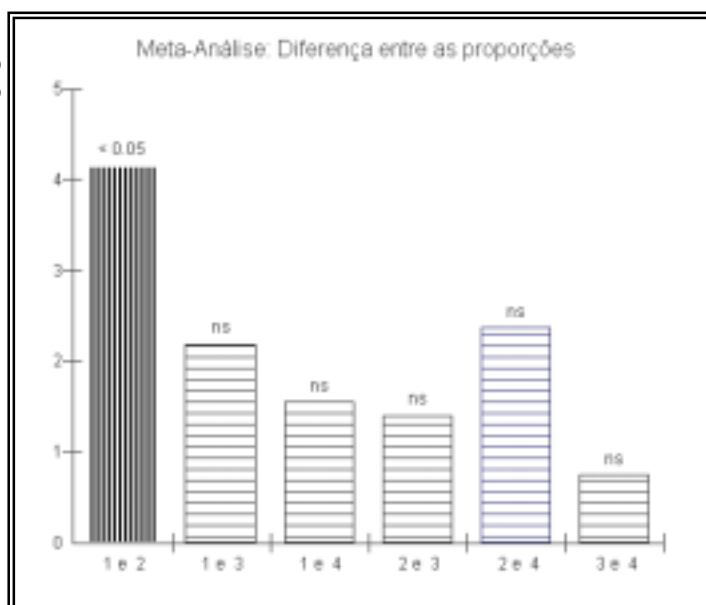
Editar Gráfico		
	Resultado	
Pw =	0.2610	---
Qw (1-Pw) =	0.7390	---
Qui-Quadrado =	8.9866	---
Graus de liberdade =	3	---
p-valor (w) =	0.0295	---
Comparações:	q	p-valor
Proporções 1 e 2	4.1466	< 0.05
Proporções 1 e 3	2.1866	ns
Proporções 1 e 4	1.5455	ns
Proporções 2 e 3	1.3971	ns
Proporções 2 e 4	2.3627	ns
Proporções 3 e 4	0.7440	ns
Valor crítico (0.05) =	3.633	---
Valor crítico (0.01) =	4.403	---

O Qui-Quadrado obtido é igual a 8.9866 ($gl = 3$) e o p -valor (w) = 0.0295, rejeitando-se a hipótese de nulidade. Observa-se que as proporções 1 (A) e 2 (B) apresentam

diferença estatisticamente significativa ($p < 0.05$) e nas demais comparações os resultados não são significativos (*ns*).

Pressionar o mouse em *Gráfico* na barra da Figura 14.30.

Figura 14.31
Gráfico do
Exemplo
14.12.1.



Pelo teste efetuado, no qual o *pw-valor* é estatisticamente significativo, fica evidente que as amostras 1 (A) e 2 (B) apresentam diferenças do ponto de vista estatístico, com erro de conclusão inferior a 5% ($p < 0.05$).

CAPÍTULO 15

NORMALIDADE

15.1 – INTRODUÇÃO

Uma das pressuposições dos testes estatísticos paramétricos diz respeito à distribuição normal das variáveis nas populações. Quando se retira, portanto, uma amostra para esses modelos de testes, deve-se supor que as unidades do universo em questão apresentem distribuição normal.

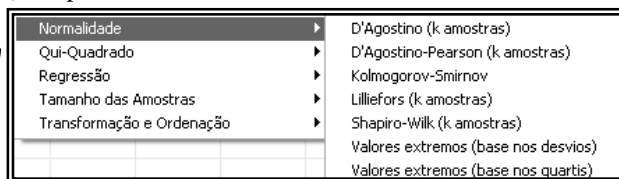
Vários testes têm sido descritos para a identificação da normalidade das variáveis. O *BioEstat* apresenta os seguintes:

- Teste *D'Agostino*, para $n \geq 10$ e k amostras;
- Teste *D'Agostino-Pearson*, para $n \geq 20$ e k amostras;
- Teste de *Kolmogorov-Smirnov*, para uma amostra;
- Teste de *Lilliefors*, para k amostras;
- Teste de *Shapiro-Wilk*, para: $2 < n < 51$ e k amostras;
- Teste para Valores Extremos (base nos desvios);
- Teste para Valores Extremos (base nos quartis).

15.1.1 – INSTRUÇÕES

- Introduzir os dados no *grid* padrão do *BioEstat*;
- Dirigir o *mouse* para o *menu* superior e clicar em *Estatística*;
- Apontar o *mouse* em *Normalidade*, conforme mostra a Figura 15.1.

Figura 15.1
Menu
Normalidade.



15.2 – TESTE *D'AGOSTINO* (k amostras)

Trata-se de um teste para pequenas amostras, mas cujo valor de n deve ser igual ou superior a dez (10) unidades. Os resultados são comparados com os valores críticos previamente estabelecidos. A hipótese de nulidade, portanto de normalidade, deve ser rejeitada quando o valor calculado é menor ou igual que o valor crítico mínimo ou igual ou maior que o valor crítico máximo. É recomendável, ainda, que os resultados sejam fornecidos em cinco (5) decimais, uma vez que o valor estatístico do *desvio* (D) pode ser muito pequeno. Pode ser efetuado para várias (k) amostras simultaneamente.

15.2.1 – EXEMPLO

Foram determinados os níveis de glicemia (mg/dL) de quinze (15) crianças escolares, obtendo-se os valores constantes no *grid* padrão, Figura 15.2.

H_0 : a distribuição da variável apresenta normalidade;

H_1 : a distribuição da variável não atende o pressuposto de normalidade;

Nível de decisão: $\alpha = 0.05$.

Figura 15.2
Dados do
Exemplo
15.2.1.

	- 1 - Glicemia
1	75.000
2	80.000
3	81.000
4	76.000
5	79.000
6	75.000
7	74.000
8	77.000
9	76.000
10	79.000

15.2.2 – INSTRUÇÕES

- Pressionar o *mouse* em *D'Agostino (k amostras)*;
- Selecionar a coluna correspondente à amostra do *grid* geral;
- Clicar em *Executar*, para resultados.

Figura 15.3
Resultados
do teste
D'Agostino
(Exemplo
15.2.1.)

Resultados	Col 1
Tamanho da amostra =	15
D (Desvio) =	0.2812
Valores críticos 5%	0.2578 a 0.2859
Valores críticos 1%	0.2469 a 0.2866
p =	ns

O *D* obtido é igual a 0.2812, o qual está contido nos limites dos valores críticos tabelados ao nível alfa de 5%, traduzindo um $p > 0.05$. A variável em questão apresenta distribuição normal.

15.3 – TESTE *D'AGOSTINO-PEARSON (k amostras)*

Este teste baseia-se nas medidas de simetria (g_1) e curtose (g_2) e o *p-valor* é calculado pelo Qui-Quadrado com dois (2) graus de liberdade. Deve ser utilizado para amostras iguais ou maiores que 20 unidades, com uma ou *k* amostras.

15.3.1 – EXEMPLO

O peso (*g*) de sementes de uma espécie de leguminosa foi calculado de amostra aleatória de 30 unidades, cujos valores estão inseridos no *grid* geral.

H_0 : a variável peso das sementes da leguminosa é normalmente distribuída;

H_1 : a variável peso das sementes da leguminosa não é normalmente distribuída;

Nível de decisão: alfa = 0.05.

15.3.2 – INSTRUÇÕES

- Pressionar o *mouse* em *D'Agostino-Pearson (k amostras)*;
- Selecionar a coluna do *grid* geral correspondente à amostra;
- Clicar em *Executar*, para resultados.

Figura 15.4
Dados do
Exemplo
15.3.1
e resultados
do teste.

	- 1 -
	Peso (g)
1	1.000
2	1.230
3	0.980
4	1.100
5	0.990
6	1.210
7	1.170
8	1.300
9	1.050
10	1.080
11	0.970
12	1.000

D'Agostino-Pearson	
Arquivo	Editar Gráfico
Resultados - 1 -	
Tamanho da amostra =	30
G1 =	0.3138
G2 =	-1.3038
Zg1 =	0.9237
Zg2 =	1.4764
K2 =	3.0329
Graus de liberdade =	2
p =	0.2195

O p -valor é igual a 0.2195, aceitando-se a hipótese de nulidade. A variável estudada apresenta distribuição normal.

15.4 – TESTE DE KOLMOGOROV-SMIRNOV

Este teste já foi descrito em detalhes no Capítulo 9, referente a testes para uma amostra. Assim sendo, o exemplo e os resultados podem ser vistos nesse capítulo e o teste compreende:

- Introdução dos dados amostrais no *grid* geral;
- Pressionar o *mouse* em *Kolmogorov-Smirnov*;
- Selecionar a coluna do *grid* referente aos dados amostrais;
- Clicar em *Executar*;
- Preencher as *caixas de texto* com os parâmetros do universo em estudo (média e desvio padrão);
- Clicar em *OK*, para resultados.

15.5 – TESTE DE LILLIEFORS (k amostras)

O teste de *Lilliefors* já está descrito no Capítulo 9, para uma amostra. Efetuar o mesmo teste, mas com cinco (5) amostras de modo simultâneo.

15.5.1 – EXEMPLO

Foram coletadas cinco (5) amostras, cada uma representando diferentes populações. Testar a normalidade dessas amostras pelo *Teste de Lilliefors*.

H_0 : as amostras apresentam normalidade em suas distribuições;

H_1 : as amostras não apresentam normalidade em suas distribuições;

Nível de decisão: $\alpha = 0.05$.

15.5.2 – INSTRUÇÕES

- Introduzir os dados de cada amostra no *grid* geral;
- Pressionar o *mouse* em *Lilliefors (k amostras)*;
- Selecionar as colunas do *grid* referentes aos dados amostrais;
- Clicar em *Executar*, para resultados.

Figura 15.5
Dados das
amostras do
Exemplo
15.5.1.

- 1 - Amostra A	- 2 - Amostra B	- 3 - Amostra C	- 4 - Amostra D	- 5 - Amostra E
158.0	245.0	31.6	18.5	20.5
154.0	240.0	30.4	17.9	19.6
153.0	240.0	31.0	18.4	32.0
153.0	236.0	30.9	17.7	20.2
155.0	243.0	31.5	18.6	20.3
162.0	247.0	32.0	19.0	20.9
157.0	238.0	31.9	18.4	31.0
156.0	239.0	31.8	18.6	20.2
164.0	248.0	32.7	19.1	21.1
158.0	238.0	31.0	18.8	
158.0	240.0		18.6	
160.0			18.6	

Figura 15.6
Resultados
das
Amostras
do Exemplo
15.5.1.

	- 1 - Amostra A	- 2 - Amostra B	- 3 - Amostra C	- 4 - Amostra D	- 5 - Amostra E
Tamanho da amostra =	12	11	10	12	9
Desvio máximo =	0.1733	0.2634	0.1638	0.2185	0.4180
Valor crítico (0.05) =	0.2420	0.2490	0.2580	0.2420	0.2710
Valor crítico (0.01) =	0.2750	0.2840	0.2940	0.2750	0.3110
p(valor)	ns	< 0.05	ns	ns	< 0.01

O teste de *Lilliefors* mostra anormalidade nas amostras 2 e 5, cujos valores de p correspondem a < 0.05 e < 0.01 , respectivamente. As demais amostras não apresentam diferenças significativas ($p > 0.05$), podendo-se considerá-las como representantes de populações com normalidade de distribuição.

15.6 – TESTE DE SHAPIRO–WILK (k amostras)

O teste de *Shapiro-Wilk* – teste W – é obtido dividindo-se o quadrado da combinação linear apropriada dos valores ordenados da amostra pela variância simétrica estimada. O *BioEstat* calcula este teste para amostras: $2 < n < 51$. O teste pode ser efetuado para k amostras de modo simultâneo.

15.6.1 – EXEMPLO (hipotético)

Foram escolhidas de modo aleatório 42 laranjas, cujos pesos estão inseridos no *grid* geral, Figura 15.7. As hipóteses formuladas foram as seguintes:

H_0 : a variável peso das laranjas é normalmente distribuída;

H_1 : a variável peso das laranjas não apresenta normalidade em sua distribuição;

Nível de decisão: $\alpha = 0.05$.

Figura 15.7
Visão parcial
dos dados do
Exemplo
15.6.1.

	- 1 - Laranja (g)	- 2 -
1	27.0	
2	23.0	
3	20.6	
4	12.0	
5	32.0	
6	22.0	

15.6.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Shapiro-Wilk* (*k amostras*);
- Selecionar a coluna correspondente à amostra do *grid* geral;
- Clicar em *Executar*, para resultados.

Figura 15.8
Resultados
do Teste de
*Shapiro-
Wilk* do
Exemplo
15.6.1.

Teste de Shapiro-Wilk	
Arquivo Editar Gráfico	
- 1 -	
Resultados	
Tamanho da amostra =	42
Média =	19.6595
Desvio padrão =	5.7468
W =	0.9249
p =	0.0137

O *p*-valor obtido é igual a 0.0137, rejeitando-se a hipótese de nulidade.

15.7 – TESTE DE VALORES EXTREMOS, modelo com base nos desvios (OUTLIERS)

A determinação dos chamados *valores extremos* ou *outliers* pode ser desejada, uma vez que podem desempenhar importante papel nas conclusões das pesquisas. Um valor extremo pode ser introduzido em um conjunto de observações em decorrência de falha na coleta de dados, de erro no registro durante a coleta, na introdução equivocada desse escore no momento da execução dos cálculos ou, ainda, do fato de que o valor extremo obtido está correto. Torna-se, por conseguinte, importante sua identificação, devendo-se tomar, conforme o caso, as medidas adequadas.

O *BioEstat* identifica os *valores extremos* contidos em uma só amostra, sem considerar, contudo, sua origem. De acordo com a experiência de *Rosner* (2006), o programa determina a quantidade de possíveis *outliers* com base na equação $\min(\{n/10\}, 5)$, onde *n* é o tamanho da amostra e cinco (5) é a quantidade máxima de valores extremos que podem ser detectados em uma distribuição normal.

15.7.1 – EXEMPLO (baseado nos desvios)

Foi efetuada coleta de 50 peixes de uma determinada espécie em um lago criatório, cujos escores estão no *grid* geral. Parte da amostra está contida na Figura 15.9.

H_0 : Não há valores extremos na amostra obtida;

H_1 : Há valores extremos na amostra obtida;

Nível de decisão: $\alpha = 0.05$.

Figura 15.9
Visão parcial
dos dados
do Exemplo
15.7.1.

		- 1 -	- 2 -
		Peixes (kg)	
1		28.0	
2		23.0	
3		20.6	
4		13.2	
5		32.0	
6		24.0	

15.7.2 – INSTRUÇÕES

- Preencher o *grid* geral com os dados amostrais;
- Pressionar o *mouse* em *Valores Extremos (base nos desvios)*;
- Selecionar a coluna correspondente aos dados;
- Clicar em *Executar*;

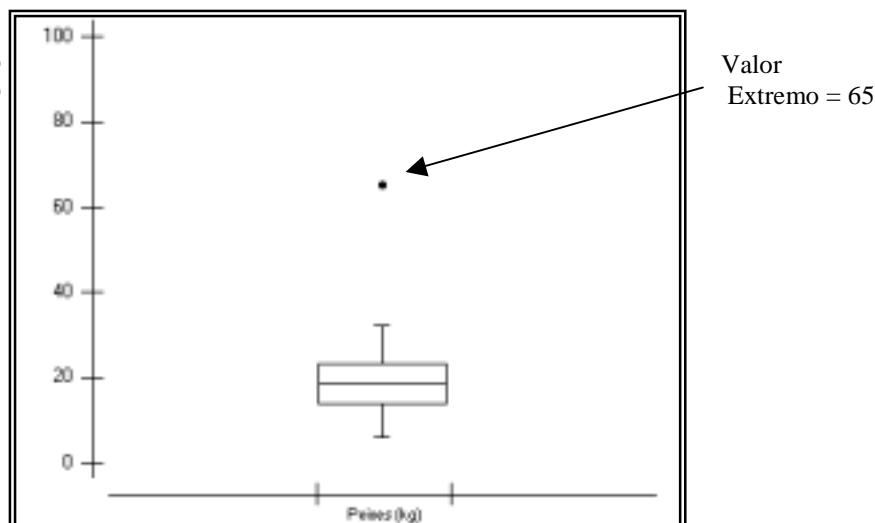
Figura 15.10

Resultados
do Teste de
Valores
Extremos,
com
base nos
desvios
(*Outliers*).

	Col. 1	Col. 1	Col. 1	Col. 1	Col. 1
Valor(es) extremo(s) =	65.000	32.000	6.000	28.000	11.000
Tamanho da amostra =	50	49	48	47	46
Média =	19.4180	18.4878	18.2063	18.4660	18.2587
Desvio padrão =	8.1193	4.8091	4.4333	4.0955	3.8836
ESD =	5.6140	2.8097	2.7533	2.3279	1.8691
ESD (0.05) =	3.1300	3.1228	3.1152	3.1072	3.0988
(p) =	< 0.05	ns	ns	ns	ns

O valor 65 kg é considerado um valor extremo com $p < 0.05$. Se esse escore não for excluído da amostra, o peso médio dos peixes será de 19.418 kg e, com sua exclusão, passa a ser igual a 18.4878 kg. Há, portanto, diferença de 0.93 kg entre esses valores médios. Deve-se, assim, verificar a origem do valor extremo referido. Como neste exemplo, $n = 50$, logo $n/10$ é igual a 5, então o programa admitiu a possibilidade de existirem até 5 *outliers*, como demonstrado na Figura 15.10.

Pressionando-se em *Gráfico* na barra da Figura 15.10 obtém-se a seguinte imagem:

Figura 15.11
Gráfico do
Exemplo
15.7.1.

O gráfico (*Box-Plot* – desvios) confirma o *Valor Extremo* detectado pelo teste.

15.8 – TESTE DE VALORES EXTREMOS, modelo com base nos quartis (OUTLIERS)

A determinação dos *valores extremos* pode ser calculada com base nos quartis, onde os *outliers* serão:

- a) Os escores menores que o limite inferior, com base no primeiro quartil ajustado;
- b) Os escores maiores que o limite superior, baseado no terceiro quartil ajustado.

O *BioEstat* identifica, como se verá em seguimento, os *valores extremos* contidos em uma só amostra, sem considerar, contudo, sua origem.

15.8.1 – EXEMPLO (baseado nos quartis)

O exemplo é o mesmo descrito no item 15.7.1, Figura 15.9.

H_0 : Não há valores extremos na amostra obtida;

H_1 : Há valores extremos na amostra obtida;

Nível de decisão: $\alpha = 0.05$.

15.8.2 – INSTRUÇÕES

- e) Preencher o *grid* geral com os dados amostrais;
- f) Pressionar o *mouse* em *Valores Extremos* (base nos quartis);
- g) Selecionar a coluna correspondente aos dados;
- h) Clicar em *Executar*.

Figura 15.12
Resultados
do Exemplo
15.8.1,
baseado
nos quartis.

	- 1 -	
Tamanho da amostra =	50	
Mediana =	19.2000	
Limite inferior =	6.0000	
Extremos abaixo =	0 escores	
Limite superior =	30.0000	
Extremos acima =	2 escores	65.0000 32.0000

Os escores 65 e 32 são considerados *valores extremos*, eis que se encontram acima do limite superior. Pelo método anterior, com base nos desvios, apenas o escore 65 foi considerado *outlier*.

CAPÍTULO 16

QUI-QUADRADO

16.1 – INTRODUÇÃO

O teste do Qui-Quadrado é um dos mais empregados em pesquisas biomédicas, sendo aplicado para dados mensurados em escala nominal ou ordinal.

Os testes de aderência e de independência do Qui-Quadrado e o teste G já foram exibidos nos capítulos relativos a uma e duas amostras independentes. Por sua vez, os testes envolvendo várias amostras sobre o mesmo assunto, tanto de aderência como o de n tabelas de contingência, estão dispostos agora no Capítulo 14 sobre Meta-Análise.

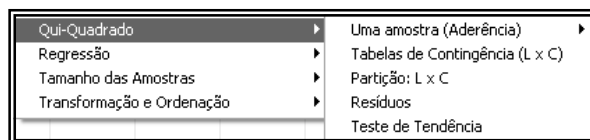
Este Capítulo, embora abranja os testes de Aderência e de Independência, inclui exemplos apenas dos aplicativos de Partição e de Tendência:

- a) Uma amostra (Aderência): Qui-Quadrado, Capítulo 9;
- b) Duas amostras independentes: $l \times c$, Capítulo 10;
- c) Partição: $l \times c$;
- d) Resíduos do Qui-Quadrado;
- e) Teste de Tendência.

16.1.1 – INSTRUÇÕES

- a) Dirigir o mouse para o menu superior e clicar em *Estatística*;
- b) Apontar o mouse em *Qui-Quadrado*, quando surgirá a figura abaixo:

Figura 16.1
Menu do Qui-Quadrado.



16.2 – PARTIÇÃO

O teste de independência do Qui-Quadrado é destinado para duas amostras dispostas em tabelas de contingência $l \times c$: 2×3 , 3×3 , 3×4 etc. Nessas tabelas, quando o resultado é significativo ($p\text{-valor} \leq 0.05$), não se pode saber com precisão em qual dos conjuntos testados ocorreu a associação estatística. A partição dessas tabelas de contingência em outras de tamanho 2×2 , cada uma com um grau de liberdade, permite esclarecer o problema. Inicia-se a partilha pelo canto superior esquerdo da tabela geral, combinando-se linhas e colunas até abranger o todo.

16.2.1 – EXEMPLO

Efetuiu-se levantamento de 250 granjas de trigo, com a finalidade de determinar a influência da lagarta sobre a produção do cereal. As granjas foram classificadas em quatro (4) tipos, de acordo com a intensidade da infestação – leve, moderada, alta, muito alta –, registrando-se em cada uma o resultado da colheita – satisfatória e não-satisfatória. Os dados foram inseridos no *grid* geral.

Figura 16.2
Dados do
Exemplo
16.2.1.

	- 1 - ** Colheita	- 2 - Leve	- 3 - Moderada	- 4 - Alta	- 5 - Muito alta
1	Satisfatória	94.000	62.000	31.000	15.000
2	Insatisfatória	15.000	15.000	17.000	11.000

H_0 : A produção de trigo independe do grau de infestação do cereal pela lagarta;
 H_1 : A produção de trigo depende do grau de infestação do cereal pela lagarta;
 Nível de decisão: $\alpha = 0.01$.

16.2.2 – INSTRUÇÕES

- Clicar o *mouse* em *Partição: L x C*;
- Selecionar as colunas do *grid* geral;
- Clicar em *Executar o Teste*, para obter os resultados.

Figura 16.3
Resultados
do Exemplo
16.2.1.

Teste do Qui-Quadrado: Partição				
Arquivo Editar				
	Lin : Col	Qui- Quadrado	GL	(p)
Partição 1	2 : 2	0.8516	1	0.3561
Partição 2	2 : 3	8.1896	1	0.0042
Partição 3	2 : 4	6.6674	1	0.0098
Total =	2 x 4	15.7086	3	0.0013

Pelos resultados pode-se concluir que as colheitas com infestação leve e moderada não apresentam diferença estatística significativa ($p = 0.3561$). Quando comparadas, entretanto, com às infestações alta e muito alta, partições 2:3 (2 linhas e 3 colunas) e 2:4 (duas linhas e 4 colunas), verifica-se que as discrepâncias são muito significativas (p -valor = 0.0042 e p -valor = 0.0098, respectivamente). Se fosse efetuado o teste de independência, o Qui-Quadrado seria altamente significativo (p -valor = 0.0013, $\chi^2 = 15.7086$), mas não informaria as proporções de interesse no caso: colheitas *Satisfatórias* e *Insatisfatórias*.

16.3 – ANÁLISE DE RESÍDUOS DO QUI-QUADRADO

O *Teste de Resíduos* é realizado em tabelas de contingência $L \times C$, o qual colabora na interpretação do Qui-Quadrado, eis que determina a importância probabilística de cada uma das caselas, complementando com mais especificidade o Teste de Partição, pois é possível comparar os resultados de cada valor obtido com a probabilidade padronizada da curva normal, uma vez que o denominador da equação final dos *Resíduos* é comparável ao *erro padrão*. Desse modo, o teste do Qui-Quadrado adquire

uma nova dimensão, informando a importância de cada valor original para o resultado dessa estatística não-paramétrica.

16.3.1 – EXEMPLO (hipotético)

Em um hospital público, localizado em Pauxís-PA, foi realizado um estudo para dimensionar o impacto da fibromialgia entre os profissionais de enfermagem. Os indivíduos que preencheram os critérios estabelecidos neste estudo foram distribuídos aleatoriamente em dois grupos:

- **Grupo A:** 84 pacientes com Lombalgia;
- **Grupo B:** 58 pacientes sem Lombalgia (Grupo Controle).

A ocorrência da doença em cada faixa etária foi registrada conforme a tabela abaixo.

Figura 16.4
Dados do
Ex 16.3.1.

	- 1 - ** Faixa Etária	- 2 - Lombalgia	- 3 - Controle
1	30 a 34 anos	8	9
2	35 a 39 anos	9	10
3	40 a 44 anos	11	13
4	45 a 49 anos	22	16
5	50 a 54 anos	34	10

H_0 : Os valores das caselas, após ajustamento, não apresentam diferença significativa;

H_1 : Os valores das caselas, após ajustamento, apresentam diferença significativa;

Nível de decisão: $\alpha = 0.05$.

16.3.2 – INSTRUÇÕES

- a) Clicar o *mouse* em *Resíduos*;
- b) Selecionar as colunas do *grid* geral;
- c) Clicar em *Executar o Teste*, para obter os resultados.

Figura 16.5
Resultados
do Exemplo
16.3.1.

Resultados	Col 2	Col 3
Linha 1	-1.0814	1.0814
Linha 2	-1.1230	1.1230
Linha 3	-1.4565	1.4565
Linha 4	-0.1847	0.1847
Linha 5	2.9430	-2.9430
Nível Alfa 0.05	1.96	
Nível Alfa 0.01	2.576	

Os valores apresentados na Figura 16.5 expressam a contribuição de cada casela para a respectiva avaliação probabilística, demonstrando que os valores da linha 5 foram mais expressivos nos resultados obtidos, revelando dados probabilísticos inferiores a 0.01. Quando se calcula o Qui-Quadrado de Partição observa-se a confirmação dos resíduos obtidos.

16.4 – TESTE DE TENDÊNCIA

O *Teste de Tendência* é realizado em k amostras, cada uma com duas categorias (A e B), em tabelas $2 \times k$. Cada amostra reflete determinados períodos de tempo, os quais devem ser espaçados de maneira regular. O objetivo é determinar o aumento ou diminuição do evento *sucesso* nas amostras, ou seja, se há alguma *Tendência* – *crescente* ou *decrecente* – ou se inexistente qualquer associação dos eventos com os períodos investigados.

16.4.1 – EXEMPLO

Em pesquisa ambulatorial, procurou-se estabelecer a tendência da hipertrofia prostática em relação à idade.

Os dados obtidos foram os seguintes:

Hipertrofia de próstata	IDADE EM ANOS			
	54 – 59	60 – 64	65 – 69	70 – 74
SIM	120	135	162	190
NÃO	784	800	950	950

H_0 : Não há tendência de maior número de indivíduos com hipertrofia prostática com o aumento da idade;

H_1 : Há tendência de maior número de indivíduos com hipertrofia prostática com o aumento da idade;

Nível de decisão: alfa = 0.05.

16.4.2 – INSTRUÇÕES

- Clicar o mouse em *Teste de Tendência*;
- Preencher o *grid* específico com os dados;

Figura 16.6
Dados do Exemplo 16.4.1.

	Grupo 1	Grupo 2	Grupo 3	Grupo 4
Casos positivos	120	135	162	190
Casos controles	784	800	950	950

Clicar em *Executar*, para obter os resultados.

Figura 16.7
Resultados do Exemplo 16.4.1.

	Resultados
A =	52.8443
Tendência crescente	A > 0
Qui-Quadrado =	4.3651
Graus de liberdade =	1
(p) =	0.0367

A tendência é aumentar o número de casos de hipertrofia prostática à medida que o indivíduo acresce a idade. O teste é significativo ($p = 0.0367$), indicando tendência crescente pelo valor positivo de A (52.8443). Rejeita-se a hipótese de nulidade e aceita-se a alternativa. A seção 22.14.4, do capítulo Fórmulas Estatísticas, define a variável A . Quando $A > 0$, a tendência é crescente; e se $A < 0$, a tendência é decrescente.

CAPÍTULO 17

ANÁLISE DE REGRESSÃO

17.1 – INTRODUÇÃO

No estudo de *correlação* procura-se verificar a magnitude e o sentido da associação que possa existir entre duas variáveis, sem haver qualquer grau de dependência de uma em relação à outra. No teste de *regressão*, ao contrário, a finalidade é determinar a *dependência* de uma variável em relação à chamada variável *independente* ou preditora.

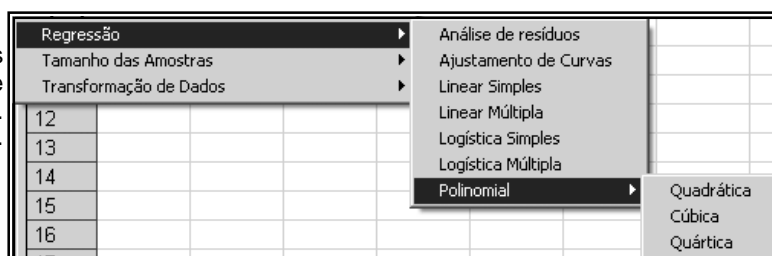
O *BioEstat* contém os seguintes programas de regressão:

- Regressão linear simples;
- Regressão linear múltipla;
- Ajustamento de curvas;
- Análise de Resíduos da Regressão;
- Regressão logística simples;
- Regressão logística múltipla;
- Regressões polinomiais;
- Regressões passo-a-passo (*Stepwise regression*);
- Regressão: comparação de duas regressões lineares

17.1.1 – INSTRUÇÕES

- Dirigir o *mouse* para o *menu* superior e clicar em *Estatística*;
- Apontar o *mouse* em *Regressão*, quando surgirá a figura abaixo:

Figura 17.1
Menu dos
Aplicativos de
Regressão.



17.2 – REGRESSÃO LINEAR SIMPLES

É um dos modelos de regressão de largo emprego nas pesquisas biomédicas, no qual procura-se prever o valor de uma variável dependente (Y) a partir de uma variável independente (X) ou preditora, pressupondo que elas assumam modelo linear.

17.2.1 – EXEMPLO

Determinou-se a pressão arterial diastólica (P.A.D.) de 20 indivíduos do sexo masculino, registrando-se a idade, em anos, de cada pessoa. Os dados obtidos estão inseridos no *grid* geral.

H_0 : a pressão arterial diastólica dos homens não sofre alterações com a idade: $\beta = 0$;
 H_1 : a pressão arterial diastólica dos homens é alterada pela idade: $\beta \neq 0$;
 Nível de decisão: $\alpha = 0.05$.

Figura 17.2

Dados do
Exemplo
17.2.1
(visão
parcial).

	- 1 - Y (P.A.D.)	- 2 - X (Idade)	- 3 -
1	80.000	35.000	
2	70.000	37.000	
3	76.000	45.000	
4	80.000	53.000	
5	80.000	67.000	

17.2.2 – INSTRUÇÕES

- Preencher o *grid* geral com os escores amostrais, representando a primeira coluna os valores da variável Y (dependente) e a segunda os escores da variável X (independente);
- Clicar em *Linear Simples* e selecionar as colunas do *grid* geral;
- Clicar em *Executar o Teste*, para resultados.

Figura 17.3

Resultados
do Exemplo
17.2.1.

Teste de Regressão Linear			
Arquivo Editar Gráfico Estimar Y			
Fontes de variação	GL	SQ	QM
Regressão	1	347.6388	347.6388
Erro	18	385.5612	21.4201
Total	19	733.2000	---
F (regressão) =	16.2296	p = 0.0011	
Variável dependente =	Coluna 1		
Variável independente =	Coluna 2		
Média (\bar{X}) =	54.0000		
Média (\bar{Y}) =	78.8000		
Coef. de Determinação (R^2) =	0.4741		
R^2 (ajustado) =	0.4449		
Coefficiente de Correlação =	0.6886		
Intercepto (a) =	60.2501	t = 12.7664	p < 0.0001
Coef. de Regressão (b) =	0.3435	t = 4.0286	p = 0.0008
IC 95% (a)	50.335 a 70.166		
IC 95% (b)	0.164 a 0.523		
Equação	$Y' = a + bX$		

O valor de F (regressão) é muito significativo ($p = 0.0011$), rejeitando-se a hipótese de nulidade e aceitando-se a alternativa: a pressão arterial diastólica acresce à medida que aumenta a idade do homem. Consta-se, neste exemplo, pelo coeficiente de determinação ajustado (R^2), que 44.49% da variável dependente é explicada pela variável preditiva, devendo outros fatores atuar como preditores do aumento da P.A.D.

17.2.3 – ESTIMAÇÃO DE Y

Para obter o valor estimado da Pressão Arterial Diastólica, pressione a opção “Estimar Y” que está visível no topo da Figura 17.3. Neste exemplo definimos o valor da variável X em 50 (anos). O valor estimado foi 77.4251 conforme ilustrado na figura abaixo. Para o cálculo da estimação está implícito o uso do modelo $Y' = a + bX$ referente ao Exemplo 17.2.1.

Figura 17.4
Resultado da
Estimação
de Y do
Exemplo
17.2.1.

17.3 – REGRESSÃO LINEAR MÚLTIPLA

A regressão múltipla linear é semelhante à regressão linear simples, mas a quantidade de variáveis independentes ou preditoras é de duas ou mais.

17.3.1 – EXEMPLO

Foi efetuada pesquisa em doze (12) indivíduos do sexo masculino, com idades variando de 30 a 55 anos, determinando-se: Y: creatininemia em mg/dL; X₁: idade em anos; X₂: peso em kg; X₃: estatura em cm.

Figura 17.5
Visão parcial
dos dados
do Exemplo
17.3.1.

	- 1 - Creatinina	- 2 - Idade	- 3 - Peso	- 4 - Estatura
1	0.870	30.000	62.000	163.000
2	1.000	34.000	73.000	167.000
3	1.150	38.000	72.000	166.000
4	1.310	41.000	80.000	181.000
5	1.610	44.000	70.000	166.000
6	1.630	45.000	69.000	168.000
7	1.220	46.000	84.000	174.000

H₀: o nível sérico da creatinina não é dependente da idade e/ou do peso e/ou da estatura: $\beta_1 = 0, \beta_2 = 0, \beta_3 = 0$;

H₁: o nível sérico da creatinina é dependente da idade e/ou do peso e/ou da estatura: pelo menos um $\beta_i \neq 0$;

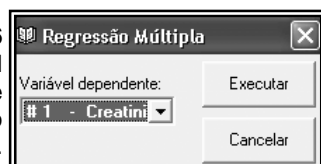
Nível de decisão: alfa 0.05.

17.3.2 – INSTRUÇÕES

- Clicar o mouse em *Linear Múltipla*;
- Selecionar as colunas do *grid* geral;
- Clicar em *Executar Estatística*, então surgirá uma *caixa de texto* para que seja informado em qual das colunas está a variável dependente Y que, neste exemplo, foi a coluna 1 (Creatinina);

Figura 17.6

A variável
Dependente
do Exemplo
17.3.1.



- Clicar em *Executar*, para resultados.

Figura 17.7

Resultados
do Exemplo
17.3.1.

Arquivo Editar Estimar Y			
Fontes de variação	GL	SQ	QM
Regressão	3	0.7121	0.2374
Erro	8	0.4244	0.0531
Total	11	1.1366	---
F Regressão (3, 8) =	4.4744	---	---
(p) =	0.0399	---	---
Coef. de determinação múltipla (R^2_{yy}) =	0.6266	---	---
F R^2_{yy} (3, 8) =	7.5505	---	---
(p) =	0.0105	---	---
R^2_{yy} ajustado =	0.4865	---	---
Coef. de correlação múltipla (R_{yy}) =	0.7916	---	---
Intercepto (a) =	1.4957	t = 0.5717	p = 0.5832
Coef. parcial de regressão (b_1) =	0.0223	t = 2.3102	p = 0.0496
Coef. parcial de regressão (b_2) =	0.0222	t = 1.7992	p = 0.1096
Coef. parcial de regressão (b_3) =	-0.0164	t = -0.8444	p = 0.4230
Comparação: Coef. (b_1) e Coef. (b_2) =	---	t = 0.0037	p = 0.9971
Comparação: Coef. (b_1) e Coef. (b_3) =	---	t = 1.7458	p = 0.1189
Comparação: Coef. (b_2) e Coef. (b_3) =	---	t = 1.2850	p = 0.2347
Variável dependente =	Coluna 1	---	---
$Y' = a + b_1X_1 + b_2X_2 + \dots + b_iX_i$			

O valor de F (regressão) é significativo ($p = 0.0399$), rejeitando-se a hipótese de nulidade e aceitando-se a alternativa: pelo menos uma das variáveis independentes – X_1 , X_2 ou X_3 – influencia o nível sérico de creatinina. Entre os coeficientes parciais de regressão somente o b_1 é estatisticamente significativo ($t = 2.3102$, $p = 0.0496$), sendo, portanto, X_1 a variável que mais explica a variação de Y . O nível sérico de creatinina (Y) deve aumentar 0.0223 mg/dL para cada acréscimo de uma unidade de X_1 (cada ano de idade, a partir de 39 e até 55).

Dados sobre os coeficientes de determinação (R^2) e de correlação múltiplos (R_{yy}) estão contidos nos resultados, assinalando-se, ainda, a comparação entre os coeficientes parciais de regressão.

Assinale-se, por fim, que o modelo linear para predição de Y , baseado nas variáveis X_1 , X_2 e X_3 , é representado pela seguinte equação:

$$\hat{Y} = 1.4957 + 0.0223X_1 + 0.0222X_2 + (-0.0164X_3).$$

17.3.3 – ESTIMAÇÃO DE Y

Para obter um valor estimado da variável Y , pressione a opção “Estimar Y” que está visível no topo da Figura 17.7. Neste exemplo fixamos os valores da variável X_1 em 54 (anos), da variável X_2 em 80 (peso) e da variável X_3 em 174 (estatura). O valor obtido da estimação da variável Y (creatinina) é igual a 1.6223, conforme ilustrado na figura abaixo.

Figura 17.8
Estimação
de Y do
Exemplo
17.3.1.

X	Digite
X 1	54
X 2	80
X 3	174

$\hat{Y} = 1.4957 +$
 $(0.0223 \times 54) +$
 $(0.0222 \times 80) +$
 (-0.0164×174)

 $\hat{Y} = 1.6223$

Estimar Y Cancela

17.4 – AJUSTAMENTO DE CURVAS

Em algumas pesquisas, nas quais se procura analisar a relação de dependência entre duas variáveis, o modelo de distribuição pode não ser *linear*, eis que os dados podem se ajustar a uma distribuição *exponencial*, *logarítmica* ou *geométrica*. Torna-se mais fácil efetuar a regressão das variáveis em questão de maneira simultânea, verificando-se, *pelo coeficiente de determinação*, qual a curva que mais se ajusta aos escores coletados.

17.4.1 – EXEMPLO

Supondo-se que, em determinado experimento, os escores amostrais foram submetidos ao teste de regressão. Os dados sobre as variáveis X e Y , independente e dependente, respectivamente, estão inseridos no *grid* geral. Efetuar o ajustamento.

Figura 17.9
Dados do
Exemplo
17.4.1.

	- 1 - X	- 2 - Y	- 3 -
1	1.000	0.500	
2	3.000	0.700	
3	6.000	1.000	
4	10.000	1.200	
5	18.000	1.400	
6	33.000	1.700	
7	76.000	1.900	

17.4.2 – INSTRUÇÕES

- Preencher o *grid* geral com os escores amostrais, representando:
 - Na primeira coluna os valores da variável *X* (independente);
 - Na segunda, os escores da variável *Y* (dependente);
- Clicar em *Ajustamento de Curvas* e selecionar as colunas do *grid* geral;
- Clicar em *Executar Estatística*, para obter os resultados.

Figura 17.10
Resultados
do Exemplo
17.4.1.

Ajustamento de Curvas				
Arquivo Editar Gráfico				
	Regressão Linear	Regressão Exponencial	Regressão Logarítmica	Regressão Geométrica
Tamanho da amostra =	7	7	7	7
Intercepto (a) =	0.8544	0.8190	0.4153	0.5247
Coef. regressão (b) =	0.0165	0.0139	0.3452	0.3241
Coef. determinação (R ²) =	73.73%	58.88%	98.48%	97.48%
Média (X) =	21.0000	21.0000	9.7079	9.7079
Média (Y) =	1.2000	1.0960	1.2000	1.0960
Var. independente =	Coluna 1	Coluna 1	Coluna 1	Coluna 1
Var. dependente =	Coluna 2	Coluna 2	Coluna 2	Coluna 2
Equação =	$Y' = a + bX$	$Y' = a * e^{(bX)}$	$Y' = a + b * \ln(X)$	$Y' = a * X^b$
Graus de liberdade =	5	5	5	5
(p) =	0.0133	0.0440	p < .00001	p < .00001

Considerando os valores dos Coeficientes de Determinação (R^2), constata-se que as regressões logarítmica e geométrica, 98% e 97%, respectivamente, são as curvas que mais se ajustam aos dados obtidos no experimento. Em ordem decrescente, o ajustamento seria o seguinte:

- Logarítmica**, com $R^2 = 98.48\%$;
- Geométrica**, com $R^2 = 97.48\%$;
- Linear**, com $R^2 = 73.73\%$;
- Exponencial**, com $R^2 = 58.88\%$.

Assim, a estimação da variável dependente pela preditora deve obedecer ao modelo logarítmico:

$$\hat{Y} = 0.42 + 0.35 \times \ln(X).$$

Assim, para estimar o valor *Y* quando *X* = 8, teremos:

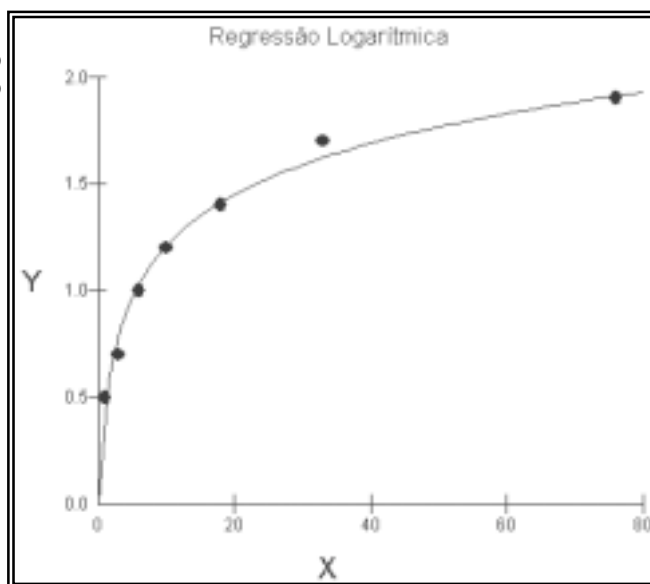
$$\hat{Y} = 0.4153 + 0.3452 \times \ln(8)$$

$$\hat{Y} = 0.4153 + 0.3452 \times 2.07944$$

$$\hat{Y} = 1.1331$$

Para obter a curva da função $\hat{Y} = 0.42 + 0.35 \times \ln(X)$, pressione em *Gráfico* na barra da Figura 17.10. A qual confirma que o melhor ajuste é representado por uma curva logarítmica, conforme indicado pelo coeficiente de determinação (r^2) de 98%.

Figura 17.11
Gráfico
do Exemplo
17.4.1.



17.5 – ANÁLISE DE RESÍDUOS

Na análise de regressão, a predição da variável dependente Y é efetuada diretamente através do escore da variável independente X e dos coeficientes respectivos, pela equação: $\hat{Y} = a + bX$. No entanto, de modo indireto, pode-se examinar a variável dependente pela avaliação dos *resíduos*, ou seja, dos escores e_i , os quais são obtidos pela conhecida diferença: $e_i = (y_i - \hat{y}_i)$, que estimam os valores paramétricos ε_i . A média dos resíduos é sempre igual a zero, obtendo-se a variância pela equação da soma dos quadrados e respectivo quadrado médio.

Essa análise pode ser útil para verificar os seguintes pressupostos:

- homogeneidade das variâncias dos valores de Y ;
- linearidade dos valores observados;
- normalidade dos escores de Y ;
- presença de valores extremos (*outliers*).

O exame dos resíduos é efetuado, em geral, com diagramas de dispersão. Desse modo, após a análise residual, é conveniente efetuar com os valores obtidos os referidos gráficos.

17.5.1 – EXEMPLO 1 (hipotético)

Suponha-se que foi determinada a pressão arterial sistólica (P.A.S.) de doze (12) crianças de ambos os sexos consideradas normais, cujas idades estavam compreendidas entre um (1) e seis (6) anos (incluindo meses de vida). Os escores estão inseridos no *grid* geral.

Figura 17.12
Dados do
Exemplo
17.5.1
(visão
parcial).

	- 1 - Idade	- 2 - P.A.S.	-
1	1.600	83.000	
2	2.400	116.000	
3	3.000	100.000	
4	4.700	130.000	
5	5.000	115.000	
6	6.000	126.000	
7	5.200	122.000	
8	3.800	118.000	
9	2.000	108.000	
10	4.100	125.000	
11	5.000	172.000	
12	1.000	80.000	

H_0 : os escores da P.A.S. são homogêneos (não apresentam valores extremos);

H_1 : os escores da P.A.S. não são homogêneos (apresentam valores extremos);

Nível de significância: $\alpha = 0.05$.

17.5.2 – INSTRUÇÕES

- Preencher o *grid* geral com os escores amostrais, introduzindo na primeira coluna os valores da variável X (independente) e na seguinte os valores da variável Y (dependente);
- Orientar o *mouse* para *Regressão* e clicar em *Análise de Resíduos*;
- Selecionar as colunas correspondentes às variáveis X e Y e clicar em *Executar*, para resultados.

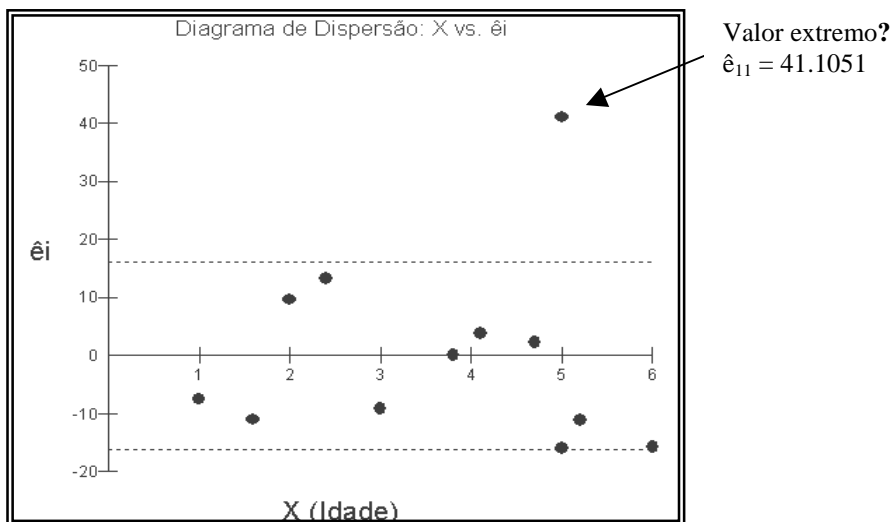
Figura 17.13
Resultados
do Exemplo
17.5.1.

Arquivo Editar Gráfico					
Caso	Y	Y (esperado)	Resíduo	Val. Esperado Padrão	Resíduo Padrão
1	83.0000	94.0114	-11.0114	-1.2663	-0.6505
2	116.0000	102.6899	13.3101	-0.7721	0.7863
3	100.0000	109.1988	-9.1988	-0.4015	-0.5434
4	130.0000	127.6405	2.3595	0.6486	0.1394
5	115.0000	130.8949	-15.8949	0.8339	-0.9390
6	126.0000	141.7430	-15.7430	1.4516	-0.9300
7	122.0000	133.0645	-11.0645	0.9574	-0.6536
8	118.0000	117.8772	0.1228	0.0927	0.0073
9	108.0000	98.3507	9.6493	-1.0192	0.5700
10	125.0000	121.1316	3.8684	0.2780	0.2285
11	172.0000	130.8949	41.1051	0.8339	2.4283
12	80.0000	87.5026	-7.5026	-1.6369	-0.4432
Mínimo	80.0000	87.5026	-15.8949	-1.6369	-0.9390
Máximo	172.0000	141.7430	41.1051	1.4516	2.4283
Média	116.2500	116.2500	0.0000	0.0000	0.0000
Mediana	117.0000	119.5044	-3.6899	0.1853	-0.2180

- Clique na opção “Gráfico” do menu superior para obter o diagrama dos resíduos.

Pelo gráfico de dispersão entre os valores da variável X e o dos resíduos (\hat{e}_i), contido no diagrama da Figura 17.14, observa-se:

Figura 17.14
Gráfico de dispersão do Exemplo 17.5.1.



A linha contínua horizontal representa a *média dos resíduos* valores de Y , a qual é igual zero (0). As linhas tracejadas correspondem a $\pm 1\sigma$ em torno da média residual. Nota-se que há muita diferença entre o valor observado $Y_{11} = 172.0000$ e o esperado $\hat{Y}_{11} = 130.8949$ (Figura 17.13), cujo resíduo é igual a 41.1051, o qual está inserido na Figura 17.14 (ver seta indicativa). Testando-se o conjunto dos resíduos, no sentido de verificar se esse escore pode ser considerado um valor extremo, obtém-se:

Figura 17.15
Valores Extremos do Exemplo 17.5.1.

	Col. 1
Valor(es) extremo(s) =	41.105
Tamanho da amostra =	12
Média =	0.0000
Desvio padrão =	16.1401
ESD =	2.5468
ESD (0.05) =	2.4100
(p) =	< 0.05

Constata-se, pelos dados da Figura 17.15, que o resíduo 41.1051 foi considerado um valor extremo ($p < 0.05$). Esse resíduo corresponde à P.A.S. igual a 172 mmHg, de uma criança de cinco (5) anos de idade. Considerando-se essa faixa etária, deve-se rever a medição referida, uma vez que pode ter ocorrido erro do observador, equívoco no registro, problemas com a aferição do aparelho medidor ou, ainda, hipertensão arterial verdadeira.

17.5.3 – EXEMPLO 2

Pode-se, também, testar os valores dos resíduos para verificar a *normalidade* desses escores. Para isso, escolheu-se o teste de *Shapiro-Wilk* com os resíduos já calculados, contidos na Figura 17.13 e reproduzidos, de modo isolado, na Figura 17.16:

Figura 17.16
Dados do
Exemplo
17.5.3.

	- 1 -	- 2 -
1	-11.0114	
2	13.3101	
3	-9.1988	
4	2.3595	
5	-15.8949	
6	-15.7430	
7	-11.0600	
8	.1228	
9	9.6493	
10	3.8684	
11	41.1051	
12	-7.5026	

H_0 : os resíduos dos escores da pressão arterial sistólica apresentam normalidade;

H_1 : os resíduos dos escores da pressão arterial sistólica não apresentam normalidade;

Nível de significância: $\alpha = 0.05$.

17.5.4 – INSTRUÇÕES

- Preencher o *grid* geral com os escores residuais, como se verifica na Figura 17.16;
- Clicar em *Normalidade* e pressionar em *Shapiro-Wilk*;
- Selecionar a coluna do *grid* correspondente aos valores dos resíduos;
- Pressionar em *Executar*, para resultados.

Figura 17.17
Resultados
do Exemplo
17.5.3.

Tamanho da amostra =	12
Média =	0.0000
Desvio padrão =	16.1401
W =	0.8511
p =	0.0424

O teste de *Shapiro-Wilk* (W) é significativo com p -valor igual a 0.0424, confirmando o resultado do teste de valores extremos (*outliers*). Os resíduos não apresentam distribuição normal, rejeitando-se a hipótese de nulidade.

17.6 – REGRESSÃO LOGÍSTICA SIMPLES

A regressão logística simples testa uma variável dependente Y e uma variável independente, sendo a variável Y binária, ou seja, assume valores 1 (sucessos) e valores 0 (insucessos). Por exemplo, óbito e sobrevivente, curado e não-curado, viciado em drogas e não-viciado, fumante e não-fumante, e assim por diante. A variável independente pode ser binária ou contínua.

17.6.1 – EXEMPLO (hipotético)

Uma investigação foi efetuada no sentido de verificar a incidência de câncer cervical (Y) em relação com a história clínica de cervicite crônica (X). A classificação estabelecida foi a seguinte:

Variável Y: neoplasia cervical presente = 1; ausente = 0;

Variável X: antecedentes de cervicite crônica presente = 1; ausente = 0.

O número total de indivíduos investigados foi igual a 20, conforme mostra, parcialmente, a Figura 17.18:

Figura 17.18
Dados
parciais do
Exemplo
17.6.1.

	- 1 - Neoplasia	- 2 - Cervicite
1	1.000	1.000
2	.000	.000
3	1.000	.000
4	.000	.000
5	1.000	1.000
6	1.000	1.000
7	1.000	1.000
8	.000	.000
9	.000	.000

$H_0: \beta = 0$;

$H_1: \beta \neq 0$;

Nível de decisão: $\alpha = 0.05$.

17.6.2 – INSTRUÇÕES

- Clicar o mouse em *Regressão Logística Simples*;
- Selecionar as colunas do *grid* geral, começando pela coluna 1, a qual será considerada como a variável dependente (Y);
- Clicar em *Executar o Teste*, para resultados.

Figura 17.19
Resultados
do Exemplo
17.6.1.

Regressão Logística (Logit)

No. de 0 (valor zero) 11.0000 55.00 %
No. de 1 (valor um) 9.0000 45.00 %
Obs: Estes valores referem-se à variável dependente Y

Número de Variáveis Independentes: 1
Número de Casos: 20
Função Objeto: Máxima Verossimilhança
Valor Final: 9.9829
Estimação de Y: ☒ Exibir

-2*log(verossimilhança)
Para este modelo: 19.9658
Só Intercepto: 27.5256
Qui-Quadrado: 7.5598
G. Liberdade: 1
(p): 0.0060

Logit Pi = -1.5041 + (2.7568 X1)

Variáveis	Coefficiente	Erro padrão	Z	p-valor	odds ratio	IC 95%
Intercepto	-1.5041	0.7817	---	---	---	---
X1	2.7568	1.1198	2.4619	0.0138	15.7500	1.754 a 141.410

A equação de regressão, assinalada na Figura 17.19, é a seguinte:

$$\text{Logit } P_i = -1.5041 + 2.7568X_i.$$

Tendo em vista o *p*-valor (0.0138) do coeficiente *b*, rejeita-se a hipótese de nulidade. A chance de neoplasia cervical é de cerca de 16 vezes maior se a mulher tiver antecedentes de cervicite crônica (*odds ratio* = 15.75).

Com base na equação acima referida, as probabilidades seriam assim calculadas:

Para $X = 1$ (antecedentes de cervicite crônica):

$$\text{Logit } P = -1.5041 + 2.7568 \times 1 = 1.2527$$

$$P = 1/(1 + \text{EXP}(-1.2527)) = \mathbf{0.78 = 78\%}.$$

Para $X = 0$ (sem antecedentes de cervicite crônica):

$$\text{Logit } P = -1.5041 + 2.7568 \times 0 = -1.5041$$

$$P = 1/(1 + \text{EXP}(1.5041)) = \mathbf{0.18 = 18\%}.$$

Pode-se obter essas probabilidades diretamente do *BioEstat*, pressionando-se o botão *Exibir*, assinalado pela seta (Figura 17.19), preenchendo-se as *caixas de texto* da Figura 17.20 com **1** (com cervicite) ou **0** (sem cervicite). Pressionar o *mouse* em *Estimar Y*, para resultados.

Figura 17.20
Resultados
do Exemplo
17.6.1.

The figure shows two side-by-side screenshots of the BioEstat software interface. Both windows have a 'Valor de X 1' input field and a 'Logit Pi' output field. The left window shows 'Valor de X 1' as 1, 'Logit Pi' as 1.2527, and 'P(Y)' as 0.7778 = 77.78%. The right window shows 'Valor de X 1' as 0, 'Logit Pi' as -1.5041, and 'P(Y)' as 0.1818 = 18.18%. Both windows have 'Estimar Y' and 'Cancelar' buttons at the bottom.

Os valores estimados de *Y* são os mesmos anteriores: 78% e 18%, para $X_1 = 1$ e $X_1 = 0$, respectivamente.

17.7 – REGRESSÃO LOGÍSTICA MÚLTIPLA

A regressão logística múltipla testa uma variável dependente *Y* e duas ou mais variáveis independentes, sendo a variável *Y* binária, ou seja, assume valores 1 (sucessos) e valores 0 (insucessos). As variáveis independentes são geralmente binárias, mas uma delas pode ser contínua.

17.7.1 – EXEMPLO (hipotético)

Uma investigação verificou que a incidência de câncer pulmonar (*Y*) em relação ao fumo (X_1) e ao sexo dos indivíduos (X_2). O número total de indivíduos investigados foi igual a 20, sendo os cinco (5) primeiros contidos na Fig. 17.21.

Figura 17.21
Dados
parciais do
Exemplo
17.7.1

	- 1 - Câncer pulmonar Y	- 2 - Sexo X1	- 3 - Fumante X2
1	1	1	1
2	0	0	0
3	1	0	1
4	0	0	0
5	1	1	0

DEFINIÇÕES:

Y: Câncer presente = 1; Ausente = 0;

X₁: Sexos: Masculino = 1; Feminino = 0;

X₂: Fumante = 1; Não Fumante = 0

$H_0: \beta_1 = 0; \beta_2 = 0;$
 $H_1: \beta_1 \neq 0; \beta_2 \neq 0;$
 Nível de decisão: alfa 0.05.

17.7.2 – INSTRUÇÕES

- Clicar o mouse em *Regressão Logística Múltipla*;
- Selecionar as colunas do *grid* geral, começando com a coluna 1, a qual será considerada a variável dependente (Y);
- Clicar em *Executar o Teste*.

Figura 17.22
 Resultados
 do Exemplo
 17.7.1.

Regressão Logística (Logit)

No. de 0 (valor zero) 11.0000 55.00 % Obs: Estes valores referem-se à variável dependente Y
 No. de 1 (valor um) 9.0000 45.00 %

Número de Variáveis Independentes: 2 Número de Casos: 20 Função Objeto: Máxima Verossimilhança Valor Final: 6.9313 Estimação de Y: ☒ Exibir

-2*log(verossimilhança)
 Para este modelo Só Intercepto Qui-Quadrado G. Liberdade (p)
 13.8626 27.5256 13.6629 2 0.0011

Logit $P_i = -2.9892 + (2.1813 X_1) + (3.1450 X_2)$

Variáveis	Coefficiente	Erro padrão	Z	p-valor	odds ratio	IC 95%
Intercepto	-2.9892	1.3509
X 1	2.1813	1.3905	1.5687	0.1167	8.8574	0.580 a 135.177
X 2	3.1450	1.4321	2.1961	0.0281	23.2208	1.402 a 384.532

A chance (*odds ratio*) de neoplasia pulmonar é cerca de 23 vezes se o indivíduo é fumante. Os valores probabilísticos exatos são estimados (P_i) pela equação:

$$\text{Logit } P_i = -2.9892 + 2.1813 X_1 + 3.1450 X_2$$

Com base na equação acima referida, as probabilidades seriam assim calculadas:

- Para $X_1 = 0$ (não-fumante) e $X_2 = 0$ (sexo feminino):
 $\text{Logit } P_1 = -2.9892 + 2.1813 \times 0 + 3.1450 \times 0 = -2.9892$
 $P_1 = 1/(1 + \text{EXP}(-2.9892)) = 0.0479 = 4.79\%;$
- Para $X_1 = 1$ (fumante) e $X_2 = 0$ (sexo feminino):
 $\text{Logit } P_1 = -2.9892 + 2.1813 \times 1 + 3.1450 \times 0 = 0.8079$
 $P_1 = 1/(1 + \text{EXP}(0.8079)) = 0.31 = 31\%;$
- Para $X_1 = 0$ (não-fumante) e $X_2 = 1$ (sexo masculino):
 $\text{Logit } P_1 = -2.9892 + 2.1813 \times 0 + 3.1450 \times 1 = 0.1558$
 $P_1 = 1/(1 + \text{EXP}(0.1558)) = 0.54 = 54\%;$
- Para $X_1 = 1$ (fumante) e $X_2 = 1$ (sexo masculino):
 $\text{Logit } P_1 = -2.9892 + 2.1813 \times 1 + 3.1450 \times 1 = 2.3371$
 $P_1 = 1/(1 + \text{EXP}(2.3371)) = 0.91 = 91\%.$

Pode-se obter essas probabilidades diretamente do *BioEstat*, pressionando-se o botão *Exibir*, assinalado pela seta (Figura 17.22), preenchendo-se as *caixas de texto* da Figura 17.18 com $X_1 = 1$ (fumante) ou 0 (não-fumante.) e $X_2 = 1$ (sexo masculino) ou 0 (sexo feminino). Pressione *Estimar Y*, para resultados.

Figura 17.23
Resultados
do Exemplo
17.7.1.

The figure shows two side-by-side screenshots of the BioEstat software interface. Each screenshot displays a table for input variables and a box for the calculated probability P(Y).

Left Screenshot:

Valor de X 1	0
Valor de X 2	0

Logit Pi: -2.9892

$$p = \frac{1}{1 + e^{-(\alpha + \beta_1 x_1 + \dots + \beta_k x_k)}}$$

P(Y): 0.0479 = 4.79%

Buttons: Estimar Y, Cancelar

Right Screenshot:

Valor de X 1	1
Valor de X 2	0

Logit Pi: -0.8079

$$p = \frac{1}{1 + e^{-(\alpha + \beta_1 x_1 + \dots + \beta_k x_k)}}$$

P(Y): 0.3083 = 30.83%

Buttons: Estimar Y, Cancelar

17.8 – REGRESSÃO POLINOMIAL

O ajustamento de regressões polinomiais pode ser efetuado através do emprego dos conhecidos *coeficientes dos polinômios ortogonais*, os quais facilitam os cálculos quando os escores da variável independente apresentam intervalos iguais. O uso da computação permite calcular essas regressões sem auxílio dos referidos coeficientes, como será observado em seguimento. O *BioEstat* executa as seguintes regressões polinomiais:

- Quadrática;
- Cúbica;
- Quártica.

A escolha de uma dessas Regressões – Quadrática, Cúbica ou Quártica - depende da disposição dos dados coletados. Assim, no Exemplo dado a seguir – Regressão Quártica: Exemplo 17.8.1. – verifica-se que a produção de milho (variável Y) aumenta a medida que aumenta a quantidade de fosfato na adubação (variável X). A partir, porém, de determinada quantidade deste produto agrícola, há uma tendência de diminuição na produção de milho. Então a curva desses valores Y e X não será linear e sim parabólica, como se verifica no gráfico da Figura 11.27. Daí a escolha da Regressão Quadrática. Pelas Curvas de outros Exemplos pode-se escolher outro modelo de Regressão Polinomial.

17.8.1 – EXEMPLO (Regressão quadrática)

Em um experimento de emprego de fosfato (X) e produção de milho (Y), os resultados obtidos estão inseridos no *grid* geral (Figura 17.24). Efetuar a regressão quadrática para esses dados:

$$H_0: \beta_1 = 0; \beta_2 = 0;$$

$$H_1: \beta_1 \neq 0; \beta_2 \neq 0.$$

$$\text{Nível alfa} = 0.05$$

Figura 17.24
Dados do
Exemplo
17.8.1.

	- 1 - Y	- 2 - X	- 3 -
1	65.000	.000	
2	94.000	80.000	
3	118.000	160.000	
4	112.000	240.000	
5	109.000	320.000	

17.8.2 - INSTRUÇÕES

- Pressionar o *mouse* em *Polinomial*;
- Escolher uma das regressões polinomiais, que neste exemplo foi a *Quadrática*;
- Selecionar as colunas do *grid*, começando com a que representa a variável *Y*;
- Clicar em *Executar Estatística*, para resultados.

Figura 17.25
Resultados
do Exemplo
17.8.1.

Fontes de variação	GL	SQ	QM
Regressão	2	1754.7429	877.3714
Erro	2	54.4571	27.2286
Total	22	1809.2000	---
F Regressão (2, 2) =	32.2225	---	---
(p) =	0.0281	---	---
Coef. de determinação múltipla (R ² _{yy}) =	0.9699	---	---
F R ² _{yy} (2, 2) =	96.6674	---	---
(p) =	0.0085	---	---
R ² _{yy} ajustado =	0.9398	---	---
Coef. de correlação múltipla (R _{yy}) =	0.9848	---	---
Intercepto (a) =	64.9714	t = 13.2301	p = 0.0057
Coef. parcial de regressão (b ₁) =	0.4682	t = 6.4389	p = 0.0233
Coef. parcial de regressão (b ₂) =	-0.0010	t = -4.8145	p = 0.0405
Comparação: Coef. (b ₁) e Coef. (b ₂) =	---	t = 6.4348	p = 0.0233
Variável dependente =	Coluna 1	---	---
Y' = a + (b ₁)X + (b ₂)X ²			

O *F* calculado é significativo, como também os coeficientes *b*₁ e *b*₂.

17.8.3 – ESTIMAÇÃO DE Y

Para obter um valor estimado da variável *Y*, pressione a opção 'Estimar Y' que está visível no topo da Figura 17.25. Neste exemplo estamos supondo que o valor da variável *X*₁ é 120 (fosfato). O valor obtido da estimação da variável *Y* (produção de milho) é igual a 106.7554, conforme ilustrado na figura seguinte.

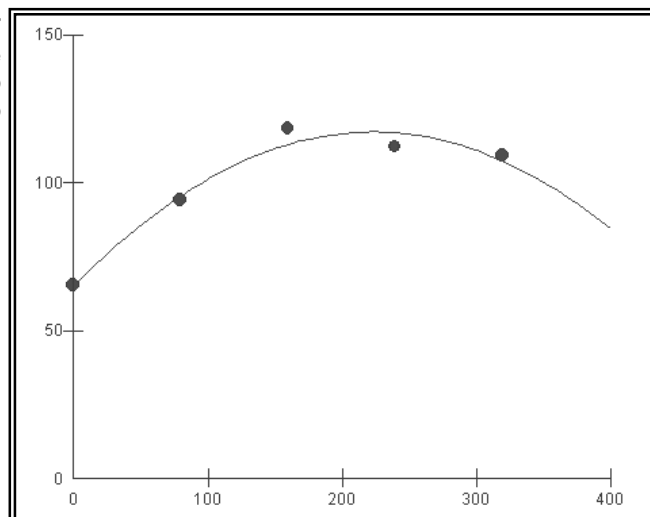
Figura 17.26
Estimação.
Exemplo
17.8.1.

Valor de	Digite	
	120	Y' = 64.9714 + (0.4682 x 120 ¹ + (-0.0010 x 120 ²) ----- Y' = 106.7554

17.8.4 – GRÁFICO DO EXEMPLO 17.8.1

Pode-se obter o gráfico em curva dos dados do Exemplo 17.8 basta selecionar a opção 'Gráfico' no menu mostrado na Figura 17.25.

Figura 17.27
Gráfico de
linha do
Exemplo
17.8.1.



Obs.: O BioEstat dispõe também de exemplos para as regressões Cúbica (Ex 17.08.2) e Quártica (Ex 17.08.3), com os respectivos gráficos.

17.9 – SELEÇÃO DE REGRESSORES (STEPWISE REGRESSION)

Esta análise possibilita a escolha das variáveis preditoras que se relacionam com mais intensidade com a variável dependente, aspecto importante na Regressão Linear Múltipla. Uma variável é considerada tão importante para o modelo quanto for capaz de influenciar de maneira significativa a variável dependente (Y). O método de execução pode ser de duas maneiras:

- a) Método Progressivo (*forward*), a cada passo é incluída uma variável.
- b) Método Regressivo (*backward*), a cada passo é excluída uma variável.

A inclusão ou exclusão de variáveis depende de três fatores: Coeficiente de Determinação, *p-valor* e Erro Quadrático Médio.

17.9.1 – EXEMPLO: MÉTODO PROGRESSIVO (FORWARD)

Em um estudo pesquisadores procuraram determinar a influência de algumas variáveis sobre a Pressão Arterial Sistólica (P.A.S.). Os dados, para este exemplo hipotético, estão representados na Figura 17.28 pelas seguintes variáveis:

- a) Pressão Arterial Sistólica (P.A.S.), medida em mm/Hg, representando a variável dependente;
- b) Idade, em anos completos;
- c) Peso, em quilogramas;
- d) Tempo de diagnóstico da doença, em anos;
- e) Nível de estresse, medido em escala de 1 a 100.

H_0 : as variáveis independentes não têm influência sobre a variável dependente;
 H_1 : as variáveis independentes têm influência sobre a variável dependente;
Nível alfa = 0.05.

Figura 17.28
Visão parcial
do Exemplo
17.9.1.

- 1 - Y= PAS	- 2 - Idade	- 3 - Peso	- 4 - Duração	- 5 - Estresse
105	48	86.000	5.2	32
115	50	95.000	3.9	13
116	49	95.300	8.3	9
117	51	94.700	5.9	89
112	51	90.000	7.1	94
121	47	99.700	9.4	11
121	48	99.800	2.5	43
110	47	91.100	6.2	9
110	49	89.200	7.1	61
115	48	92.700	5.6	34
114	47	94.400	5.0	90

17.9.2 – INSTRUÇÕES

- Pressionar o mouse em Regressão Linear Passo a Passo (Stepwise);
- Selecionar as colunas, começando com a que representa a variável Y;
- Clicar em Executar Estatística, quando surgir a Figura 17.29;
- Clique no botão “Progressivo >>” para obter a regressões inicial.
- Para executar um novo passo da análise clique no botão “Próximo Passo”.

Figura 17.29
Escolha do
modelo de
stepwise do
Exemplo
17.9.1.

Escolha o sentido das regressões

Progressivo >>

<< Regressivo

Sair

Figura 17.30
Resultados
do Exemplo
17.9.1.

Var. Dependente (Y): Coluna 1	R	R2	Variação R2	GL	F	p-valor	QM Erro	Coluna Incluída
Col: 3	0.9570	91.58%	91.58%	1,22	239.2314	0.0000	1.5899	3
Cols: 3, 2,	0.9737	94.81%	3.23%	2,21	191.8975	0.0000	1.0260	2
Cols: 3, 2, 5,	0.9753	95.11%	0.30%	3,20	129.7693	0.0000	1.0147	5
Cols: 3, 2, 5, 4,	0.9754	95.13%	0.02%	4,19	92.8691	0.0000	1.0637	4
Próximo Passo								

Pela análise dos resultados verifica-se que a variável da Coluna 3, representando o Peso dos pacientes, foi responsável por 91.58% do Coeficiente de Determinação (R^2), enquanto as demais variáveis – idade, duração da doença e estresse – acresceram apenas 3.55% a esse Coeficiente, totalizando, ao final 95.13%. Se observarmos o QM do Erro, constata-se que o modelo que melhor representa o fenômeno é composto pelas variáveis Peso (3), Idade (2) e Estresse (5), cujo Quadrado Médio do Erro (1.0147) é o menor entre as simulações realizadas pelo programa. Nos dados deste exemplo, considerando-se quatro casas decimais, o *p*-valor mostra-se estável, indicando que em quaisquer das combinações acima haverá rejeição da hipótese de nulidade. O Coeficiente *F* aponta para a variável Peso (3) como a que obteve o valor mais elevado (239.2314).

Observação: Aconselha-se, na análise do indicador de melhor desempenho do modelo de regressão linear, verificar conjuntamente os valores de R^2 e do Quadrado Médio do Erro (QM do Erro).

17.9.3 – EXEMPLO: MÉTODO REGRESSIVO (BACKWARD)

Trata-se do mesmo Exemplo anterior (Exemplo 17.9.1), efetuando-se, agora, a Regressão passo a passo no sentido inverso (*backward*), ou seja, excluindo-se gradativamente cada uma das variáveis independentes.

17.9.4 – INSTRUÇÕES

- Pressionar o *mouse* em Regressão Linear Passo a Passo (*Stepwise*);
- Selecionar as colunas, começando com a que representa a variável *Y*;
- Clicar em *Executar o Teste*;
- Ao surgir o diálogo 'Escolha o sentido das regressões', clicando no botão 'Regressivo >>' para obter a regressão inicial.
- Para executar um novo passo da análise clique no botão 'Próximo Passo'.

Figura 17.31
Visão parcial
do resultado
do Exemplo
17.9.1

Var. Dependente (Y): Coluna 1	R	R2	Varição R2	GL	F	p-valor	QM Erro	Coluna Eliminada
Cols: 2, 3, 4, 5,	0.9754	95.13%	95.13%	4,22	92.8691	0.0000	1.0637	---
Cols: 2, 3, 5,	0.9753	95.11%	-0.02%	3,22	129.7693	0.0000	1.0147	4
Cols: 2, 3,	0.9737	94.81%	-0.30%	2,22	191.8975	0.0000	1.0260	5
Col: 3	0.9570	91.58%	-3.23%	1,22	239.2314	0.0000	1.5899	2
Próximo Passo								

O método *stepwise* regressivo preconiza que o modelo inicial seja composto por todas as variáveis independentes. Neste caso o Coeficiente de Determinação (R^2) foi responsável por 95.13% da variável dependente (P.A.S.). A cada passo é eliminada uma das variáveis independentes, fica bem claro que a de maior importância neste Exemplo é a variável da Coluna 3 (Peso), as demais exercem menor influência. Notar que a variável da Coluna 2 (Idade) é responsável por apenas 3.23% da determinação da variável dependente. Observa-se que, nas quatro simulações realizadas neste exemplo o *p*-valor é muito significativo, não

podendo ser considerado, neste exemplo, um indicador para a eliminação das variáveis em questão.

17.10 – COMPARAÇÃO ENTRE DUAS REGRESSÕES

Este teste procura determinar se os valores do Coeficiente de Regressão (b) e do intercepto (a) de duas regressões provieram da mesma população:

17.10.1 – EXEMPLO (hipotético)

Os dados obtidos em uma amostragem foram divididos em dois extratos:

- a) Grupo A, indivíduos com idade entre 50 e 60 anos;
- b) Grupo B, indivíduos com idade entre 15 e 19 anos;

O objetivo do estudo é determinar se a influência da idade sobre a Pressão Arterial é idêntica em ambos os grupos.

$H_0: b_1 = b_2;$

$H_1: b_1 \neq b_2;$

Nível alfa = 0.05.

H_0 : Os interceptos são semelhantes ($a_1 = a_2$);

H_1 : Os interceptos são diferentes ($a_1 \neq a_2$);

Nível alfa = 0.05.

Figura 17.32
Visão parcial
do Exemplo
17.10.1.

- 1 -	- 2 -	- 3 -	- 4 -	- 5 -	- 6 -
Y (PAD) A	X (Idade) A		Y (PAD) B	X (Idade) B	
62	80		44	29	
65	70		37	30	
89	76		53	27	
93	80		51	25	
85	80		41	23	
83	87		42	22	
80	82		50	24	
86	80		62	21	
84	82		44	20	
82	84		62	25	
91	75		63	26	
84	70		51	22	
80	70		64	19	
89	75		65	18	
97	85		64	17	

17.10.2 – INSTRUÇÕES

- a) Pressionar o mouse em *Comparação entre Duas Regressões*;
- b) Selecionar as quatro colunas do *grid*, correspondentes às variáveis Y e X da primeira e da segunda regressões;
- c) Clicar em *Executar o Teste*, para resultados.

Figura 17.33
Resultados
do Exemplo
17.10.1.

Resultados	Regressão A	Regressão B
Tamanho das Amostras	16	16
Médias X	79.1250	23.0000
Médias Y	84.2500	53.4375
Coefficientes de Regressão	0.6081	-1.3636
Interceptos	36.1368	84.8011
Teste t, Coefs. Regressão	2.7623	---
gl	28	---
p-valor	0.0100	---
Teste t, Interceptos	1.3872	---
gl	29	---
p-valor	0.1759	---

Percebe-se que o Teste t dos Coeficientes de Regressão (0.6081 e -1.3636, respectivamente) é muito significativo ($p\text{-valor} = 0.0100$), sendo, portanto, dissimilares. Os Interceptos, por sua vez, não evidenciam diferenças, eis que o $p\text{-valor}$ do Teste T para esses Coeficientes não é significativo do ponto de vista estatístico (0.1759).

CAPÍTULO 18

ESTATÍSTICA CIRCULAR

18.1 – INTRODUÇÃO

A estatística circular (ou direcional) é aplicada a dados que estão dispostos ao redor de uma circunferência. Exemplos de dados deste tipo são diversos, e podem incluir as direções de vôo 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.

O *BioEstat* oferece os seguintes recursos para trabalhar com dados circulares:

- a) Estatística Descritiva:
 - i – Tamanho da amostra;
 - ii – Média dos cossenos;
 - iii – Média dos senos;
 - iv – Média angular;
 - v – Comprimento do vetor Médio;
 - vi – Variância angular;
 - vii – Desvio angular;
 - viii – Intervalo de confiança (95%);
- b) Uma Amostra: Teste de *Rayleigh*;
- c) Uma Amostra: Teste V (*Rayleigh*);
- d) Duas Amostras Independentes (*Watson-Williams*);
- e) Três Amostras Independentes (*Watson-Williams*);
- f) Duas Amostras Relacionadas: Paramétrico;
- g) Correlação Circular.

18.1.1 – INSTRUÇÕES

- a) Introduzir os dados no *grid* padrão do *BioEstat*. Os ângulos devem ser inseridos em graus;
- b) Dirigir o *mouse* para o *menu* superior e clicar em *Estatísticas*;
- c) Apontar o *mouse* em *Estatísticas Circulares*, conforme demonstra a Figura 18.1:

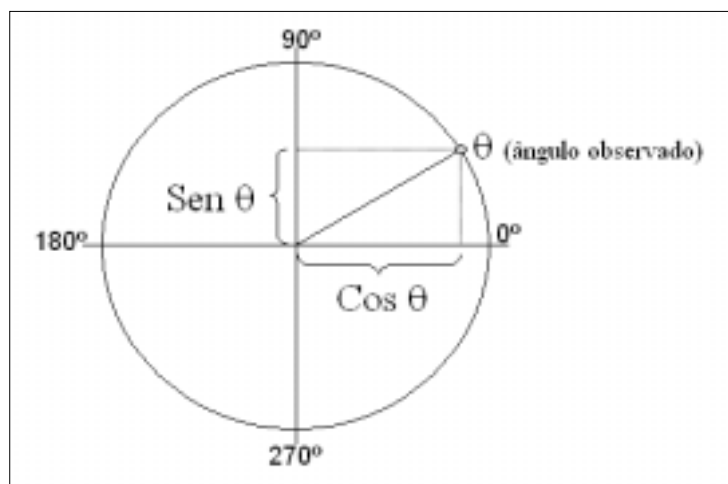
Figura 18.1
Menu de
Estatísticas
Circulares.

Estatísticas Circulares	▶	Estatística Descritiva
Distribuição de Probabilidades	▶	Uma Amostra: Teste de Rayleigh
Uma Amostra	▶	Uma Amostra: Teste V (Rayleigh)
Duas Amostras Independentes	▶	Duas Amostras Independentes (Watson-Williams)
Duas Amostras Relacionadas	▶	Três Amostras: ANOVA (Watson-Williams)
Estatística Descritiva	▶	Duas Amostras Relacionadas: Paramétrico
Estimação de Parâmetros	▶	Correlação Circular

18.2 – ESTATÍSTICA DESCRITIVA CIRCULAR

Diferentemente da estatística descritiva para dados lineares, aqui cada escore é interpretado como um vetor unitário. As projeções desse vetor nos eixos Y e X são o seno e o cosseno do ângulo, respectivamente (ver figura 18.2). Isto é importante pois o seno e o cosseno do ângulo é que são computados para calcular a média angular. A utilização da média aritmética não é recomendada para a análise de dados circulares. Em muitos casos, ela gera resultados errados facilmente percebidos até pelo observador leigo (ver exemplo 18.2.1).

Figura 18.2
Projeções do
ângulo
observado
nos eixos
vertical (seno)
e horizontal
(cosseno).



As medidas de dispersão adotadas no *BioEstat* são a variância angular e o desvio angular. A variância angular é sempre um número entre zero e dois (2), e o desvio angular varia no intervalo de zero a 81.03. Valores mais altos podem indicar maior dispersão ou distribuição circular uniforme.

18.2.1 – EXEMPLO

Um pesquisador estudou um ninho contendo nove (9) filhotes de pássaros. Ao anotar a direção do primeiro vôo de cada animal, observou os seguintes ângulos:

Figura 18.3
Dados do
Exemplo
18.2.1.

	- 1 -	- 2 -
	Ângulo do vôo	
1		6.0
2		354.0
3		18.0
4		280.0
5		90.0
6		260.0
7		355.0
8		22.0
9		35.0

Tentaremos caracterizar a amostra através da estatística descritiva.

18.2.2 – INSTRUÇÕES

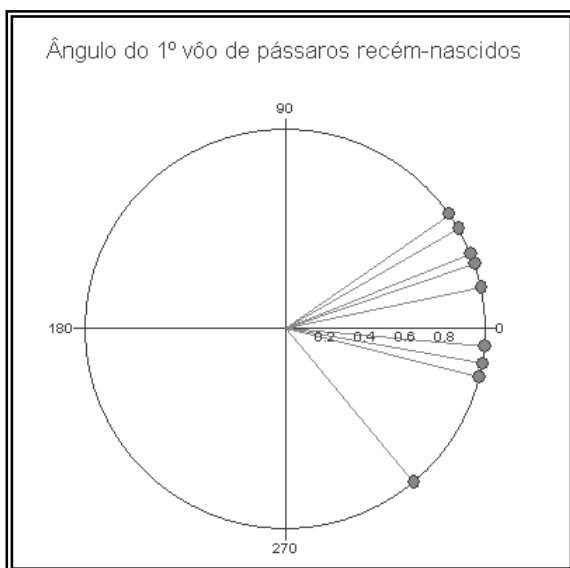
- Clicar em *Estatística Descritiva*;
- Na janela *Seleção de Amostras* escolha a coluna desejada, neste caso representado pela Coluna 1, e clique no botão *Executar Estatística*.

Figura 18.4
Resultado
da estatística
descritiva
circular
Ex 18.2.1.

Circular Descritiva	
Arquivo Editar Gráfico	
- 1 -	
Tamanho da amostra	9
Média dos cossenos	0.6314
Média dos senos	0.0223
Média angular	2.0200
Comprimento do Vetor Médio (R)	0.6318
Variância angular	0.7364
Desvio angular	49.1679
IC 95%	-41.776 a 45.816

A estatística descritiva apontou a média angular próxima de dois (2) graus. Note que a média aritmética seria 164. Destacamos, assim, a necessidade de métodos apropriados para lidar com dados de natureza circular. O Gráfico pode ser acionado para emitir a figura abaixo.

Figura 18.5
Gráfico do
Ex 18.2.1.



18.3 – UMA AMOSTRA: TESTE DE RAYLEIGH

O teste de *Rayleigh* é usado para verificar se uma população está distribuída uniformemente em torno da circunferência. A hipótese nula declara que a população apresenta uma distribuição circular uniforme. A hipótese alternativa, por sua vez, afirma que a população não é uniformemente distribuída em torno da circunferência e, portanto, existe uma direção preferida.

18.3.1 – EXEMPLO

Em seguimento ao estudo do Exemplo 18.2.1, o pesquisador observou outro ninho, contendo doze (12) filhotes de pássaros. O objetivo era comprovar a existência de orientação direcional logo na primeira tentativa de vôo.

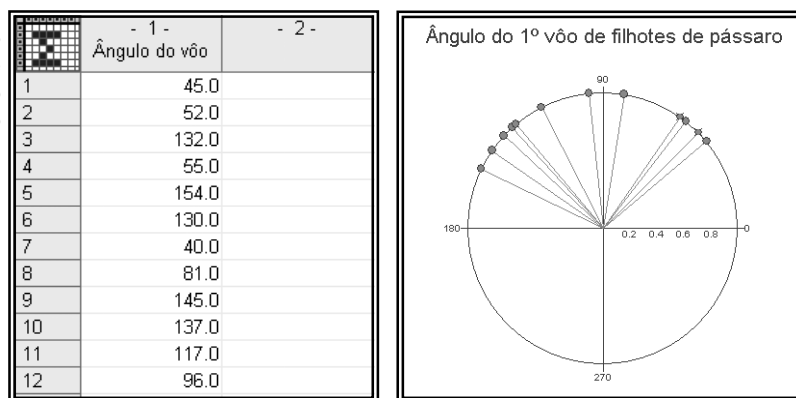
H_0 : A população está uniformemente distribuída ao redor do círculo;

H_1 : A população não está distribuída uniformemente ao redor do círculo;

Nível de decisão: $\alpha = 0.05$.

Verificaram-se os seguintes ângulos:

Figura 18.6
Dados e
Gráfico do
Exemplo
18.3.1.

**18.3.2 – INSTRUÇÕES**

- Selecionar *Uma Amostra: Teste de Rayleigh*;
- Na janela *Seleção de Amostras* escolha a coluna desejada;
- Clicar em *Executar Estatística*, para obter os resultados.

Figura 18.7
Resultado do
Exemplo
18.3.1.

Tamanho da amostra =	12
r =	0.7631
R de Rayleigh =	9.1572
Z de Rayleigh =	6.9879
u =	4.344
p-valor =	< 0.01
Média Angular =	99.4336

O *p-valor* obtido foi menor que o nível de significância (0.05). Rejeita-se, portanto, a hipótese nula e conclui-se que há uma orientação direcional no primeiro vôo desses pássaros. A média angular para esta amostra foi igual a 99.4 graus.

18.4 – UMA AMOSTRA: TESTE V (RAYLEIGH)

É usado para testar um valor paramétrico de média angular contra a hipótese de uniformidade em torno da circunferência. As hipóteses nula e alternativa são como segue:

H_0 : A população está distribuída de maneira uniforme ao redor da circunferência.

H_1 : A população não está distribuída de maneira uniforme ao redor da circunferência, e a média da população é igual ao valor esperado.

Obs.: quando a hipótese nula é aceita, não se deve descartar a possibilidade de que uma outra média angular, diferente do valor esperado, de fato exista.

18.4.1 – EXEMPLO

Antes de depositar seus ovos, dez tartarugas foram soltas num ponto exatamente a leste de seu local de desova. Ao serem recapturadas, algum tempo depois, mediu-se o ângulo formado entre a posição inicial e a final. A expectativa do estudo era de que os quelônios nadassem diretamente para oeste, realizando retorno num ângulo de 180 graus.

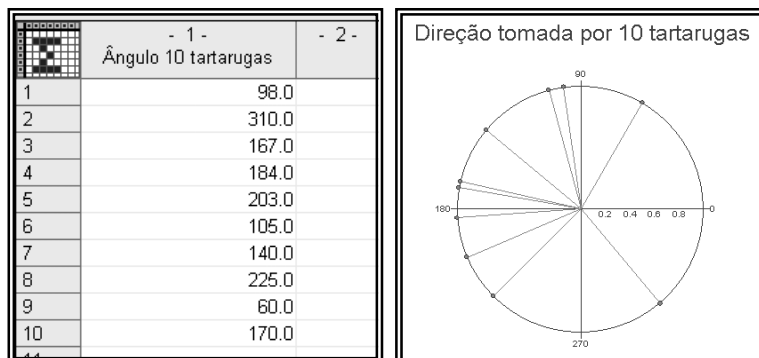
H_0 : As tartarugas nadaram em direções de distribuição aleatória;

H_1 : As tartarugas nadaram, no retorno, num ângulo de 180 graus;

Nível de decisão: $\alpha = 0.05$

A coleta de dados revelou os seguintes ângulos:

Figura 18.8
Dados e
Gráfico do
Exemplo
18.4.1.



18.4.2 – INSTRUÇÕES

- Selecionar o menu o *Estatística Circular*;
- Clicar sobre a opção *Uma Amostra: Teste V (Rayleigh)*;
- Na janela *Seleção de Amostras*, selecione a coluna que contém os dados e clique no botão *Executar Estatística*;
- Ao surgir a janela *Teste V de Rayleigh*, digite o ângulo esperado e confirme com um clique no botão *OK* para que seja exibido o resultado do teste. Neste exemplo o ângulo médio esperado é de 180 graus.

Figura 18.9
Entrada de dados do ângulo médio esperado.

Figura 18.10
Dados e Gráfico do Exemplo 18.4.1.

Tamanho da amostra =	10
$r =$	0.4994
R de Rayleigh =	4.9936
Ângulo Médio Esperado =	180.0
Ângulo Médio Observado =	157.3
Teste Estatístico V =	4.6056
$u =$	2.0597
Valor Crítico =	1.648
p-valor =	< 0.05

O *p-valor* foi inferior ao nível de significância pré-estabelecido. Conclui-se que a média angular da amostra é diferente do da população ($157.3 \neq 180$).

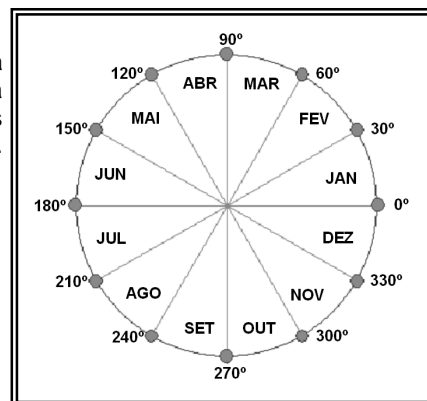
18.5 – DUAS AMOSTRAS INDEPENDENTES (WATSON-WILLIAMS)

Prova paramétrica para comparar as médias de duas amostras circulares independentes. A hipótese nula afirma que as populações apresentam médias iguais, enquanto a hipótese alternativa assegura que elas são diferentes. Este teste assume que as duas populações têm distribuição circular normal (de *von Mises*) e dispersões angulares iguais.

18.5.1 – EXEMPLO

Em um estudo observacional, as visitas de abelhas às flores de uma planta foram sendo registradas ao longo de dois anos. O pesquisador desejava saber se havia sincronização entre o pico de atividade das abelhas e o pico anual de desabrochamento das flores. Antes de executar o teste de *Watson-Williams*, verificamos, pelo teste de *Rayleigh*, que havia sazonalidade tanto na atividade das abelhas quanto no desabrochamento das flores. Os dias e meses do ano foram convertidos em graus conforme mostra a próxima figura:

Figura 18.11
Divisão da circunferência em meses (Ex 18.5.1).



H_0 : O pico de atividade das abelhas é sincronizado com o desabrochar das flores;
 H_1 : O pico de atividade das abelhas não é sincronizado com o desabrochar das flores;

Nível de decisão: $\alpha = 0.05$.

Figura 18.12
Visão parcial
dos dados
(Ex 18.5.1).

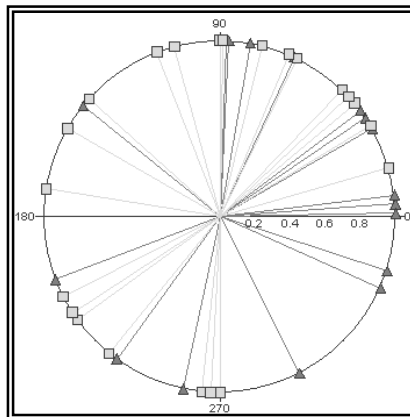
	- 1 - Abelhas	- 2 - Flores	- 3 -
1	171	76	
2	30	31	
3	105	105	
4	201	171	
5	270	138	
6	34	216	
7	258	46	
8	4	46	
9	7	16	
10	88	111	
11	111	270	
12	65	213	
13	65	264	
14	90	64	
15	37	90	
16	141	43	

18.5.2 – INSTRUÇÕES

- Clicar o mouse em *Duas Amostras Independentes (Watson-Williams)*;
- Na janela *Seleção de Amostras*, escolha as duas amostras que serão comparadas e clique no botão *Executar Estatística*;

Figura 18.13
Resultados
do Exemplo
18.5.1.

Arquivo Editar Gráfico	
	- 1 -
Tamanho das amostras	28 e 25
R1	11.1866
R2	7.2319
rw	0.3475
R	17.7391
K	1.5048
F (Watson-Williams)	1.5077
GL Tratamentos	1
GL Erro Experimental	51
p-valor	0.2230
Média Paramétrica Estimada	62.96



Com o p -valor maior que 0.05, aceitamos a hipótese nula. Conclui-se que há, de fato, uma sincronização entre a atividade das abelhas e a época do ano em que as flores desabrocham em maior quantidade.

18.6 – TRÊS AMOSTRAS INDEPENDENTES (WATSON-WILLIAMS)

Esta é uma generalização do teste de *Watson-Williams* em que devem ser inseridas três amostras circulares independentes. A estatística F é empregada para determinar o *p*-valor.

18.6.1 – EXEMPLO

Em um experimento, três grupos de pombos-correio foram soltos a quatro (4) km de seu viveiro habitual. O primeiro grupo consistiu de pombos anósmicos (sem olfato), e os pombos do segundo grupo tiveram seus olhos vendados. Um terceiro grupo, de animais hígidos e sem vendas, serviu de controle.

Após serem todos soltos, anotaram-se as direções dos vôos. O ângulo zero marcou a direção em que deveriam seguir (em linha reta) para chegar ao viveiro. O bloqueio da visão ou do olfato afetou a direção de vôo dos pombos?

Figura 18.14
Dados do
exemplo
18.6.1.

	- 1 - Grupo A	- 2 - Grupo B	- 3 - Grupo Controle
1	33.0	52.0	51.0
2	20.0	23.0	8.0
3	337.0	21.0	3.0
4	352.0	9.0	3.0
5	6.0	3.0	1.0
6	7.0	1.0	340.0
7	341.0	353.0	350.0
8	10.0	343.0	355.0
9	9.0	324.0	349.0
10	331.0	340.0	343.0
11	27.0		2.0
12			5.0

H_0 : A média da direção dos vôos é igual nas três (3) populações: $\mu_1 = \mu_2 = \mu_3$;

H_1 : A média da direção dos vôos não é igual nas três (3) populações: $\mu_i \neq \mu_k$.

Nível de decisão: alfa = 0.05.

18.6.2 – INSTRUÇÕES

- Selecionar a opção: Três Amostras Independentes (*Watson-Williams*);
- Na janela *Seleção de Amostras*, selecionar três amostras e clicar em *Executar Estatística*.

Figura 18.15
Resultado do
exemplo
18.6.1.

Tamanho das amostras	11, 10 e 12
R1	10.3748
R2	9.1631
R3	11.4711
rw	0.9397
R	31.0020
K	1.0436
F (Watson-Williams)	0.0549
GL Tratamentos	2
GL Erro Experimental	30
p-valor	0.9465
Média Paramétrica Estimada	1.94

O p -valor obtido (0.9465) aponta para uma diferença não-significativa entre as médias circulares dos três grupos de pombos-correios. Assim, aceitamos a hipótese nula e concluímos que nem a olfação nem a visão foram essenciais para o retorno das aves ao pombal.

18.7 – DUAS AMOSTRAS RELACIONADAS

Neste teste paramétrico, comparam-se as medidas de um mesmo grupo de indivíduos em dois momentos distintos. As amostras devem ser compostas por dados circulares pareados e em graus.

18.7.1 – EXEMPLO

Com o objetivo de verificar a eficácia de um protocolo fisioterapêutico, um grupo de doze (12) pacientes foi submetido a testes antes e depois de dez (10) sessões do tratamento. A variável avaliada foi a flexão do joelho, medida em graus.

Figura 18.16
Dados do exemplo 18.7.1.

	- 1 - Antes	- 2 - Depois
1	35.000	42.000
2	42.000	42.000
3	68.000	80.000
4	28.000	90.000
5	41.000	40.000
6	38.000	110.000
7	46.000	87.000
8	60.000	120.000
9	46.000	46.000
10	38.000	60.000
11	54.000	72.000
12	60.000	95.000

H_0 : Não houve alteração na média angular após o tratamento.

H_1 : Houve alteração na média angular após o tratamento.

Nível de decisão: alfa = 0.01.

18.7.2 – INSTRUÇÕES

- Clicar em *Duas Amostras Relacionadas: Paramétrico*;
- Na janela *Seleção de Amostras*, escolha as duas colunas correspondentes aos valores Antes e Depois. Clique em *Executar Estatística*;

Figura 18.17
Resultados do Exemplo 18.7.1.

	- 1 -
Número de pares	12
Diferença média dos Senos	0.1497
Diferença média dos Cossenos	-0.4229
Estatística F	6.3659
p-valor	0.0164

O p -valor foi muito significativo (< 0.0001), rejeitando-se H_0 . O tratamento aumentou a média angular de flexão dos joelhos dos indivíduos submetidos ao tratamento.

18.8 – CORRELAÇÃO CIRCULAR

Este teste paramétrico avalia a associação entre duas variáveis circulares independentes. Usamos a notação r_{aa} para identificar o *Coefficiente de Correlação*, que mede o grau de associação entre as variáveis. O r_{aa} pode atingir qualquer valor real entre -1 e 1 , e o sinal deste coeficiente determina o sentido da correlação. O valor $r_{aa}=0$ indica ausência de correlação, e quando r_{aa} for igual a 1 ou -1 tem-se uma correlação perfeita. Por último, alertamos que nenhum valor de r_{aa} pode ser usado como indicativo de relação *causa-e-efeito*.

18.8.1 – EXEMPLO

Desejamos avaliar a relação entre a direção do vento e a direção de vôo de um grupo de 20 pássaros. Foram coletados os seguintes dados:

Figura 18.18
Visão Parcial
do Dados
do Exemplo
18.8.1.

	- 1 - Direção do Vento	- 2 - Direção do Vôo
1	26.0	98.00
2	321.0	51.00
3	10.0	24.00
4	207.0	310.00
5	49.0	300.00
6	30.0	90.00

$$H_0: r_{aa}=0;$$

$$H_1: r_{aa}\neq 0;$$

Nível de decisão: $\alpha = 0.05$.

18.8.2 – INSTRUÇÕES

- Clicar o mouse em *Regressão Linear*;
- Na janela *Seleção de Amostras*, escolha as duas colunas correspondentes aos valores *Direção do Vento* e *Ângulo de Vôo*;

Figura 18.19
Resultado do
Exemplo
18.8.1.

Número de pares	20
Coefficiente de Correlação	0.0768
IC 95%	0.0654 a 0.0881
Estatística t	0.3268
GL	18
p-valor	0.7476

- Pressionar em *Executa Estatística*, para resultados.

Com um *p-valor* maior que o nível alfa de 0.05, aceitamos a hipótese nula e concluímos que não houve correlação entre a direção do vento e a direção de vôo dos pássaros.

CAPÍTULO 19

TAMANHO DAS AMOSTRAS

19.1 – INTRODUÇÃO

O tamanho amostral está relacionado com o grau de variação dos dados, a magnitude dos efeitos a serem detectados e o poder do teste desejado. Cumpre salientar que:

- a) quanto maior a *variância*, maior deverá ser o tamanho da amostra, e vice-versa;
- b) quanto menor o nível *alfa*, maior deverá ser o tamanho da amostra;
- c) quanto maior o *poder* do teste, maior deverá ser o tamanho da amostra;
- d) quanto maior a diferença entre as médias a serem comparadas, menor poderá ser o tamanho da amostra.

É importante obter informações de outros trabalhos sobre o assunto objeto da investigação, pois é possível coletar informações sobre *parâmetros* e/ou *estatísticas* da média e da variância, a fim de permitir, com mais precisão, determinar o tamanho amostral. Pode-se, ainda, efetuar um teste-piloto, cujo tamanho da amostra ficará a critério do pesquisador, obtendo-se *estatísticas* com definições prévias do nível alfa e do poder do teste que irão proporcionar os dados necessários para o cálculo do tamanho amostral.

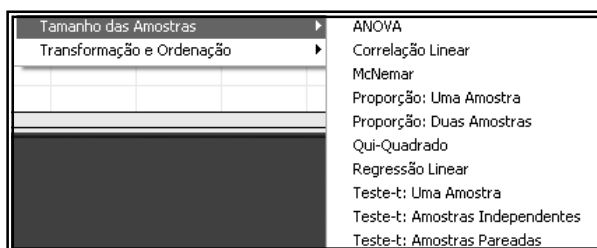
O *BioEstat* calcula o tamanho da amostra para os seguintes testes:

- i – Análise da variância;
- ii – Correlação linear;
- iii – Teste de McNemar;
- iv – Proporção: uma amostra;
- v – Proporção: duas amostras;
- vi – Qui-Quadrado: 2 x 2;
- vii – Regressão Linear;
- viii – Teste *t*: uma amostra;
- ix – Teste *t*: duas amostras independentes;
- x – Teste *t*: duas amostras pareadas.

19.1.1 – INSTRUÇÕES

- a) Dirigir o *mouse* para o *menu* superior e clicar em *Estatística*;
- b) Apontar o *mouse* em *Tamanho das Amostras*, quando surgirá a figura abaixo:

Figura 19.1
Menu de
Tamanho
das
Amostras.



19.2 – ANÁLISE DE VARIÂNCIA

O programa é adequado para o cálculo do tamanho amostral da ANOVA, um critério, abrangendo de três (3) a oito (8) tratamentos e permitindo a escolha do poder do teste, com a recomendação de que não deve ser inferior a 0.80 a fim de não aumentar o erro tipo II (β). A escolha do nível alfa apresenta duas alternativas: 0.05 e 0.01.

19.2.1 – EXEMPLO

Determinar o tamanho da amostra para um teste de análise da variância, um critério, em que se dispõe de cinco (5) tratamentos, considerando o nível alfa de 0.05, o poder do teste de 0.80 e tendo em vista as seguintes informações obtidas em experimento piloto: diferença mínima entre as médias dos tratamentos = 39.48; desvio padrão do erro experimental = 26.18.

19.2.2 – INSTRUÇÕES

- Clicar o *mouse* em ANOVA;
- Preencher as *caixas de texto* com os dados;
- Pressionar em *Alfa 0.05* ou *Alfa 0.01*, para resultados.

Figura 19.2
Tamanho da amostra: ANOVA.

A janela de diálogo 'Tamanho da amostra: ANOVA' contém as seguintes informações:

Entrada de dados	
Diferença mínima entre as médias dos tratamentos:	39.48
Desvio padrão do erro:	26.18
Número de tratamentos (3 a 8):	5
Poder do teste:	0.8
<input checked="" type="radio"/> Alfa 0.05 <input type="radio"/> Alfa 0.01	
Resultado	
Número de repetições por tratamento:	12
<input type="button" value="Cancelar"/>	

19.3 – CORRELAÇÃO LINEAR

O tamanho amostral na correlação linear – teste de *Pearson* – depende do conhecimento prévio do valor de r obtido em testes anteriores, escolhendo-se o poder do teste e o nível alfa desejado.

19.3.1 – EXEMPLO

Uma pesquisa-piloto revelou que o coeficiente de correlação linear entre duas variáveis era de 0.73. Calcular o tamanho da amostra, considerando o nível alfa de 0.05 e o poder do teste de 0.85.

19.3.2 – INSTRUÇÕES

- Clicar o mouse em *Correlação Linear*;
- Preencher as *caixas de texto* com os dados;
- Clicar em *Executar*, para resultados.

Figura 19.3
Tamanho da amostra: Correlação linear.

19.4 – TESTE DE McNEMAR

O teste de *McNemar*, um caso especial de teste binomial, é efetuado em tabelas de contingência 2 x 2, onde se calcula a probabilidade dos chamados *pares discordantes*: *a* e *d* ou *b* e *c*. Os dados a serem introduzidos nas *caixas de texto* para o cálculo amostral apresentam a seguinte configuração:

- Proporção, em decimais, de todos os pares concordantes: com o tratamento *A* e com o tratamento *B*;
- Proporção, em decimais, dos pares discordantes com um dos tratamentos;
- poder do teste e o nível alfa, informando, ainda, se o teste é unilateral ou bilateral.

19.4.1 – EXEMPLO

Supondo-se que mulheres com neoplasias mamárias, cujas idades variavam de 45 a 60 anos, foram submetidas a dois tipos de tratamento: *A* e *B*. Inicialmente, fez-se o pareamento das pacientes em dois grupos, de acordo com a idade, o tempo de doença, o grupo étnico e a situação sócio-econômica. Ao final de um período de 4 anos, verificou-se o seguinte:

- em 87% de todos os pares amostrais o tratamento foi concordante, seja com o esquema terapêutico *A*, quer com o *B*;
- dos pares discordantes, 75% ocorreram com o tratamento *A*.

Calcular o tamanho da amostra ao nível alfa de 0.05, poder igual a 0.85 e teste unilateral.

19.4.2 – INSTRUÇÕES

- Clicar o mouse em *McNemar*;
- Preencher as *caixas de texto* com os dados;
- Pressionar em *Teste Unilateral* ou *Teste Bilateral*, para resultados.

Figura 19.4
Tamanho da amostra: McNemar.

A janela 'Tamanho da amostra: McNemar' contém as seguintes informações:

- Entrada de dados:**
 - Proporção de todos os pares concordantes com ambos os tratamentos (em decimais): 0.87
 - Proporção de pares discordantes com um dos tratamentos (em decimais): 0.75
 - Poder: 0.85
 - Alfa: 0.05
 - Teste: ☒ Teste Unilateral, ☐ Teste Bilateral
- Resultados:**
 - Tamanho: 198 pares
 - Tamanho: 396 indivíduos
- Botão: Cancelar

19.5 – PROPORÇÃO: UMA AMOSTRA

Em diversas ocasiões o pesquisador deseja realizar teste binomial para uma amostra, considerando as seguintes informações: o poder do teste, o nível alfa de significância, o direcionamento da H_1 e o conhecimento das proporções paramétrica e amostral, esta última obtida em pesquisa-piloto.

19.5.1 – EXEMPLO

Quanto indivíduos seriam necessários para um teste bilateral, sabendo-se que a proporção do universo é igual a 0.01 e a da amostra, obtida em pesquisa-piloto, igual a 0.04? Dados adicionais: poder do teste 0.80 e nível alfa 0.05.

19.5.2 – INSTRUÇÕES

- Clicar o mouse em *Proporção: uma amostra*;
- Preencher as *caixas de texto* com os dados;
- Pressionar em *Teste Unilateral* ou *Teste Bilateral*, para resultados.

Figura 19.5
Tamanho da amostra: Uma Proporção.

Proporção da amostra	0.04
Proporção da população	0.01
Poder do teste	0.80
Nível alfa	0.05
<input type="radio"/> Teste Unilateral <input checked="" type="radio"/> Teste Bilateral	
Resultado	
Tamanho da amostra	144

19.6 – PROPORÇÃO: DUAS AMOSTRAS

O teste para duas proporções binomiais, diferente do McNemar, é para amostras independentes. Deve-se informar as proporções estimadas de p de cada amostra, a relação entre elas (1:1, 1:2, 1:½, etc.), o poder do teste, o nível alfa e se o teste é bilateral ou unilateral.

19.6.1 – EXEMPLO

Suponha-se que a incidência da rubéola na população infantil não vacinada é de 0.30, e que, após campanha vacinal, a incidência diminuiu para 0.15. Determinar o tamanho da amostra necessária para testar a diferença entre as duas proporções, ao nível alfa de 0.05, poder de 0.80 e teste unilateral. Nota-se que houve decréscimo na incidência da doença após imunização.

19.6.2 – INSTRUÇÕES

- Clicar o mouse em *Proporção: duas amostras*;
- Preencher as *caixas de texto* com os dados;
- Pressionar, neste caso, em *Teste Unilateral*, para resultados.

Figura 19.6
Tamanho da amostra: Duas Proporções.

Proporção estimada da 1a. amostra	0.30
Proporção estimada da 2a. amostra	0.15
Relação entre as amostras	Amostra 1 : Amostra 2 1 : 1
Poder do teste	0.80
Nível alfa	0.05
<input checked="" type="radio"/> Teste Unilateral <input type="radio"/> Teste Bilateral	
Resultados	
Tamanho da 1a. amostra	95
Tamanho da 2a. amostra	95

19.7 – QUI-QUADRADO

O tamanho da amostra para testes do Qui-Quadrado atende testes de aderência e de independência, estes dispostos em tabela de contingência 2 x 2. As informações sobre os escores obtidos em teste-piloto – Qui-quadrado e graus de liberdade - são introduzidas nas *caixas de texto*. Escolher o nível alfa - 0.05 ou 0.01 – e o poder do teste.

19.7.1 – EXEMPLO

Por suposição, em determinado experimento piloto, obtiveram-se os dados abaixo:

- a) Qui-Quadrado = 4.756;
- b) Graus de liberdade = 1

Calcular o tamanho da amostra (n) para um teste ao nível alfa de 0.05 e poder do teste igual a 0.95.

19.7.2 – INSTRUÇÕES

- a) Clicar o *mouse* em *Qui-Quadrado*;
- b) Preencher as *caixas de texto* com os dados;
- c) Selecionar *Alfa 0.05 e Poder do teste* de 0.95;
- d) Pressionar *Executar*, para resultados.

Figura 19.7
Tamanho da amostra: Qui-Quadrado.

Tamanho da amostra

Imprimir

Valor do Qui-quadrado estimado da amostra 4.756

Graus de liberdade 1

Nível alfa ☒ 0.05 ☐ 0.01

Poder do teste 0.95

Cancelar Executar

Tamanho da amostra 273

19.8 – REGRESSÃO LINEAR

O cálculo é baseado em informações sobre estatísticas da Regressão Linear obtidas em teste piloto. As estatísticas necessárias para calcular o tamanho da amostra são as seguintes:

- Coefficiente de regressão (b);*
- Desvio padrão da variável X;*
- Desvio padrão da variável Y.*

19.8.1 – EXEMPLO

Os dados preliminares da pesquisa piloto mostraram os seguintes valores:

b (Coeficiente de regressão) = 4.8;

s_x (Desvio padrão da variável X) = 7;

s_y (Desvio padrão da variável Y) = 80.

Calcular o tamanho da mostra para um teste de Regressão Linear, considerando o valor $\alpha = 0.05$ e o poder de 0.80.

19.8.2 – INSTRUÇÕES

- Clicar o mouse em *Regressão Linear*;
- Preencher as *caixas de texto* com os dados do Exemplo 19.8.1;
- Pressionar em *Executar*, para resultados.

Figura 19.8
Tamanho da amostra: Regressão Linear.

19.9 – TESTE t : UMA AMOSTRA

O tamanho da amostra é calculado com base nas informações sobre o parâmetro da população e de estatísticas obtidas em teste-piloto. O poder, o nível de significância e a determinação prévia de teste bilateral ou unilateral constituem elementos que definem o tamanho amostral.

19.9.1 – EXEMPLO

Uma amostra retirada de uma população indígena mostrou que o nível de colesterol no sangue periférico dessas pessoas era em média igual a 159 mg/dL e o desvio

padrão de 47.5 mg/dL. Tendo em vista que a média de colesterolemia na população brasileira é de 178 mg/dL, pergunta-se: qual o tamanho da amostra para confirmar essa diferença ao nível alfa de 0.01, com poder de 0.85, em teste unilateral ?

19.9.2 – INSTRUÇÕES

- Clicar o mouse em *Teste t: uma amostra*;
- Preencher as *caixas de texto* com os dados;
- Pressionar em *Teste Unilateral* ou *Teste Bilateral*, para resultados.

Figura 19.9
Tamanho da amostra: Teste t para Uma Amostra.

19.10 – TESTE t : DUAS AMOSTRAS INDEPENDENTES

O cálculo é baseado em informações sobre as médias amostrais e os respectivos desvios padrões, fixando-se o tamanho da segunda amostra em relação à primeira e determinando-se o poder, o nível alfa e o direcionamento do teste, unilateral ou bilateral. O tamanho da segunda amostra poderá ser igual ao da primeira (1:1) ou diferente (1:2), (1:1.5), (1:0.5), etc.

19.10.1 – EXEMPLO

Dados preliminares pressupõem que o peso de recém-nascidos de famílias de baixo nível socioeconômico (BNSE) é diferente do das crianças filhas de casais com elevado nível socioeconômico (ENSE). As estatísticas mostraram o seguinte:

	Média de peso	Desvio padrão
BNSE	2950g	468g
ENSE	3240g	470g

Calcular o tamanho das amostras para testar essa possível diferença, em teste bilateral, ao nível alfa de 0.05 e poder de 0.90, devendo a segunda amostra ser do mesmo tamanho da primeira (relação 1:1).

19.10.2 – INSTRUÇÕES

- Clicar o mouse em *Teste t: duas amostras independentes*;
- Preencher as *caixas de texto* com os dados;
- Pressionar em *Teste Unilateral* ou *Teste Bilateral*, para os resultados.

Figura 19.10
Tamanho da amostra: Teste t Duas Amostras Independentes.

Entrada de dados		
	1a. amostra	2a. amostra
Média	2950	3240
Desvio padrão	468	470
Relação entre as amostras	Amostra 1: 1	Amostra 2: 1
Poder do teste	0.90	
Nível alfa	0.05	
<input type="radio"/> Teste Unilateral <input checked="" type="radio"/> Teste Bilateral		
Resultados		
Tamanho da 1a. amostra	55	
Tamanho da 2a. amostra	55	
Cancelar		

19.11 – TESTE t: AMOSTRAS PAREADAS

No caso de duas amostras pareadas os dados são coletados antes e após determinado experimento, coletando-se as informações dos mesmos indivíduos nos dois períodos considerados. O cálculo prevê o nível alfa, o poder e, ainda, se o teste é unilateral ou bilateral.

19.11.1 – EXEMPLO

Determinou-se a hemoglobina no início da infecção malárica e oito (8) dias após a cura. Os dados coletados revelaram os seguintes valores:

- Antes do tratamento: média = 9.5g/dL;
- Oito (8) dias após a cura: média = 10.7g/dL;
- Desvio padrão da diferença entre as médias = 2.5g/dL.

Calcular o tamanho da amostra necessária para verificar se a diferença é estatisticamente significativa ao nível alfa de 0.05, poder de 0.85 e teste unilateral.

19.11.2 – INSTRUÇÕES

- Clicar o *mouse* em *Teste t: duas amostras pareadas*;
- Preencher as *caixas de texto* com os dados;
- Pressionar em *Teste Unilateral* ou *Teste Bilateral*, para resultados.

Figura 19.11
Tamanho da
amostra:
Teste t Duas
Amostras
Pareadas.

Tamanho: t teste duas amostras ...

Imprimir

Entrada de dados

Diferença entre as médias (antes depois)	1.20
Desvio padrão da diferença	2.50
Poder do teste	0.85
Nível alfa	0.05

☒ **Teste Unilateral** ☐ Teste Bilateral

Resultado

Tamanho da amostra	33
--------------------	----

Cancelar

CAPÍTULO 20

TRANSFORMAÇÃO DE DADOS

20.1 – INTRODUÇÃO

As transformações são procedimentos estatísticos de mudanças de escalas, com a finalidade de obter a normalidade da distribuição dos escores e a estabilização da variância.

Na distribuição normal os escores amostrais são transformados de modo linear em *z-escores*, cuja média é igual a zero (0) e o desvio padrão igual a uma unidade (1). Outras transformações não-lineares podem ser efetuadas para atender ao modelo de distribuição normal, sobretudo quando os escores brutos apresentam acentuada assimetria ou quando médias muito grandes são acompanhadas por variâncias também muito elevadas, com perda da normalidade, da homogeneidade e da homocedasticidade.

O *BioEstat* apresenta os seguintes programas de transformações:

- a) Box e Cox;
- b) Exponencial;
- c) Logaritmo natural (neperiano);
- d) Logaritmo decimal;
- e) Percentual;
- f) Quadrado;
- g) Raiz quadrada;
- h) Ordinal (*rank*);
- i) Recíproca.

Algumas indicações são oferecidas nos textos de Estatística sobre a escolha do modelo de transformação dos escores brutos coletados.

A transformação logarítmica, por exemplo, é indicada quando os valores numa distribuição linear são multiplicativos, como crescimento bacteriano, ou quando a distribuição é muito assimétrica, sendo indicada também em dados de pH. Esta transformação não admite, obviamente, valores negativos ou nulos. Quando o escore for zero, o programa adiciona um (1) a todos os valores.

A transformação em raiz quadrada é sugerida em variáveis referentes a medidas de superfície ou em variáveis discretas – número de leucócitos por mm^3 , número de filhos por casal, etc. Emprega-se, também, quando o modelo de distribuição dos escores brutos é o de *Poisson*, no qual a média e a variância têm o mesmo valor (np). Com essa transformação a variância e a média tornam-se independentes. Quando o valor da variável for zero, o programa acrescenta 0.5 a todos os valores.

A transformação ordinal (*rank*) é designada para distribuições muito assimétricas, sendo largamente usada nos testes não-paramétricos – correlação de *Spearman*, análise de variância de *Kruskal-Wallis*, etc.

Pode-se efetuar a transformação ordinal (*rank*) pela ordem numérica ou alfabética, de modo crescente ou decrescente. Algumas vezes há necessidade de ordenar um conjunto de colunas, sendo a primeira àquela que conduzirá os respectivos valores das demais.

A transformação recíproca é empregada em pesquisas cujos escores envolvem diluições, como ocorre em titulações sorológicas. Quando houver zeros entre os dados, não deverá ser usada essa transformação.

Box e Cox desenvolveram um procedimento a fim de estimar a melhor transformação para obter a normalidade de um conjunto de escores. Trata-se de um modelo de função de máxima-verossimilhança, podendo o usuário, por esse método, verificar qual a transformação *sugerida* para seus dados.

Recomenda-se, todavia, ponderar as vantagens e as desvantagens da transformação de dados. Quando ocorrer dúvida, é preferível manter os escores originais e utilizar testes não-paramétricos.

Figura 20.1
Menu de transformação.



20.1.1 – EXEMPLO 1

Determinaram-se as *estatísticas* dos dados inseridos no *grid*, verificando-se que a variância é cerca de 25 vezes maior que a média, sendo a curva assimétrica positiva. A transformação dos escores foi efetuada escolhendo-se, a título ilustrativo, a logarítmica (natural), a raiz quadrada e a recíproca, cujos valores estão contidos na Figura 20.2. Os dados originais, contidos na coluna 1 do *grid* geral, foram ainda submetidos ao procedimento de *Box e Cox*, cujo resultado está contido na Figura 20.4.

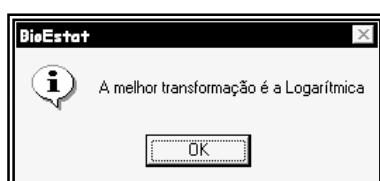
Figura 20.2
Dados do Exemplo 20.1.1.

	- 1 - Amostra	- 2 - Ln(x)	- 3 - Sqr(x)	- 4 - 1.0/(x)	- 5 -
1	25.000	3.219	5.000	0.040	
2	0.250	-1.386	0.500	4.000	
3	6.250	1.833	2.500	0.160	
4	27.250	3.305	5.220	0.037	
5	70.560	4.256	8.400	0.014	
6	46.240	3.834	6.800	0.022	
7	1.690	0.525	1.300	0.592	
8	8.410	2.129	2.900	0.119	
9	0.750	-0.288	0.866	1.333	
10	22.850	3.129	4.780	0.044	

20.1.2 – INSTRUÇÕES

- Preencher o *grid* geral com os escores amostrais;
- Dirigir o *mouse* para o *menu* superior e clicar em *Estatística*;
- Apontar o *mouse* em *Transformações*;
- Clicar em *Ln*, selecionar a coluna correspondente aos valores originais;
- O *BioEstat* informa a coluna a ser preenchida com os valores transformados;
- Repetir os passos *d* e *e*, para as demais transformações.
- Para o procedimento *Box e Cox*, efetuar *a*, *b* e *c*, pressionando a seguir nessa opção, o que foi efetuado com os dados amostrais (Figura 20.2), tendo sido sugerida a *transformação logarítmica*.

Figura 20.3
Box e Cox.



Observa-se, em relação ao *Coefficiente variância/média*, o seguinte:

Figura 20.4
Estatística
descritiva dos
dados
originais e
dos transfor-
mados.

	- 1 - Amostra	- 2 - Ln(x)	- 3 - Sqr(x)	- 4 - 1.0/(x)
Tamanho da amostra =	10	10	10	10
Mínimo	0.2500	-1.3863	0.5000	0.0142
Máximo	70.5600	4.2565	8.4000	4.0000
Amplitude Total	70.3100	5.6428	7.9000	3.9858
Mediana	15.6300	2.6292	3.8401	0.0813
Primeiro Quartil (25%)	2.8300	0.8517	1.6000	0.0375
Terceiro Quartil (75%)	26.6875	3.2835	5.1651	0.4838
Desvio Interquartilico	23.8575	2.4318	3.5651	0.4463
Média Aritmética	20.9250	2.0556	3.8266	0.6360
Variância	526.8057	3.5343	6.9799	1.5698
Desvio Padrão	22.9522	1.8800	2.6419	1.2529
Erro Padrão	7.2581	0.5945	0.8355	0.3962
Coefficiente de Variação	109.69%	91.46%	69.04%	196.99%
Assimetria (g1)	1.2583	-0.7566	0.3397	2.6166
Curtose (g2)	1.1958	-0.5819	-0.9077	7.0942
Média Harmônica =	1.5723	65.6893	1.8065	0.0478
N (média harmônica) =	10	10	10	10
Média Geométrica =	7.8115	2.4031	2.7949	0.1280
N (média geométrica) =	10	8	10	10
Variância (geom.) =	4.6410	1.2200	1.4678	4.6410
Desvio Padrão (geom.) =	6.5533	1.9673	2.5599	6.5533

- Dados originais: 25.2;
- Transformação logarítmica: 1.7;
- Transformação raiz quadrada: 1.8;
- Transformação recíproca: 2.5.

Esses *Coefficientes* e os valores da *assimetria* e *curtose* sugerem que a melhor transformação para os dados originais é a *logarítmica*, confirmando a indicação do procedimento *Box e Cox*.

20.1.3 – EXEMPLO 2 (hipotético)

Foram inseridos no *grid* geral os dados referentes a diversos municípios de um estado brasileiro e respectivos valores da população e da área municipal em quilômetros quadrados.

Figura 20.5
Dados do
Exemplo
20.1.3.

	- 1 - Município	- 2 - População	- 3 - Área (k2)	- 4 -
1	Salgado	120000.0	320.0	
2	Santa Maria	139000.0	411.0	
3	Abaeté	75000.0	166.0	
4	Pauxis	12000.0	135.0	
5	Mocorongo	200000.0	178.0	
6	Fivela	58000.0	290.0	
7	Serra Alegre	44200.0	140.0	

20.1.4 – INSTRUÇÕES

- Preencher o *grid* geral com os municípios selecionados e respectivos valores das populações e áreas;
- Dirigir o *mouse* para o *menu* superior e clicar em *Estatística*;
- Apontar o *mouse* em *Transformações* e clicar em *ordinal*, pressionando em ordem alfabética ou numérica, a qual, neste exemplo, é *alfabética*;
- Selecionar as colunas do *grid* de interesse, começando com a coluna que irá comandar as demais, a qual, neste caso é a coluna 1 (Município);
- Escolher entre ordem crescente ou decrescente, e que, neste caso, foi *crescente*.

Figura 20.6
Rank
Alfabético.

	- 1 - Município	- 2 - População	- 3 - Área (k2)	- 4 - "Rank Alfabético"	- 5 - População	- 6 - Área (km2)
1	Salgado	120000.0	320.0	Abaeté	75000.0	166.0
2	Santa Maria	139000.0	411.0	Fivela	58000.0	290.0
3	Abaeté	75000.0	166.0	Mocorongo	200000.0	178.0
4	Pauxis	12000.0	135.0	Pauxis	12000.0	135.0
5	Mocorongo	200000.0	178.0	Salgado	120000.0	320.0
6	Fivela	58000.0	290.0	Santa Maria	139000.0	411.0
7	Serra Alegre	44200.0	140.0	Serra Alegre	44200.0	140.0

Verifica-se que os Municípios estão ordenados em ordem alfabética e que os valores de cada Município – População (coluna 5) e Área (Coluna 6) - estão de acordo com os números registrados nas Coluna 2 e 3, respectivamente.

CAPÍTULO 21

GRÁFICOS E TABELA

21.1 – INTRODUÇÃO

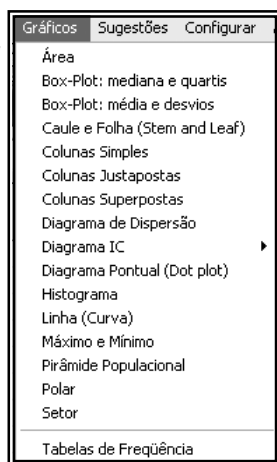
Os gráficos constituem importantes instrumentos de comunicação rápida, clara e efetiva, poupando, sobretudo, tempo e esforço na visualização de dados resumidos, permitindo fixar uma imagem duradoura, daí o largo uso dessas representações em trabalhos estatísticos. As tabelas de frequência permitem, também, condensar os escores de modo a torná-los de fácil compreensão. O *BioEstat* apresenta os seguintes gráficos e tabelas de frequência:

- Área;
- Box Plot: mediana e quartis;
- Box Plot: média e desvios;
- Caule e Folha (*Stem and Leaf*);
- Colunas;
- Diagrama de Dispersão;
- Diagramas IC;
- Diagrama Pontual (*Dot Plot*);
- Histograma;
- Linha (Curva);
- Máximo e Mínimo;
- Pirâmide Populacional;
- Polar;
- Setor;
- Tabelas de Frequência.

21.1.1 – INSTRUÇÕES

Dirigir o *mouse* para *menu* e clicar em *Gráficos*, quando surgirá a figura abaixo;

Figura 21.1
Menu
Gráficos.



- Pressionar com o *mouse* o gráfico escolhido ou *Tabelas de Frequência*;
- Selecionar a(s) coluna(s) do *grid*, conforme o caso, e pressionar em *Executar*, para resultados;
- Configurar cada gráfico de acordo com a orientação contida no item 21.16.

21.2 – GRÁFICO DE ÁREA

Destina-se a comparar as áreas abrangidas pelos escores de duas ou mais amostras, embora possa ser utilizado para apenas um conjunto de dados.

21.2.1 – EXEMPLO

Tomou-se a pressão arterial de quatorze (14) pessoas do sexo feminino, cujas idades variavam de 30 a 40 anos. Os dados obtidos em condições basais estão inseridos no *grid* padrão e representam: PAD-F: pressão arterial *diastólica*, sexo feminino; PAS-F: pressão arterial *sistólica*, sexo feminino.

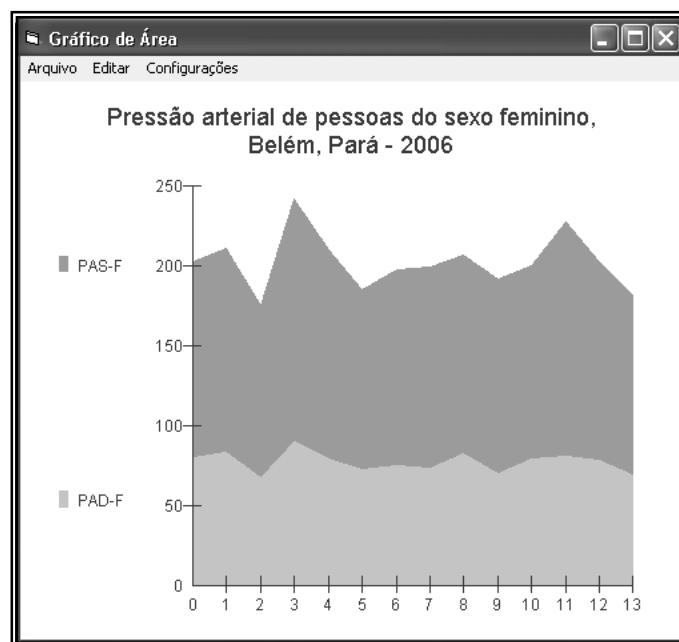
Figura 21.2
Dados
do Exemplo
21.2.1
(Visão parcial).

	- 1 - PAD-F	- 2 - PAS-F	- 3 -
1	81.000	122.000	
2	84.000	127.000	
3	68.000	107.000	
4	91.000	151.000	
5	80.000	130.000	
6	73.000	112.000	
7	76.000	122.000	
8	74.000	125.000	
9	83.000	124.000	
10	71.000	121.000	
11	80.000	120.000	

21.2.2 – INSTRUÇÕES

- Clicar em *Área*;
- Selecionar as colunas do *grid* e pressionar em *Executar*.

Figura 21.3
Gráfico de
área do
Exemplo
21.2.1.



Verifica-se que a pressão arterial diastólica (cinza-escuro), embora menor, acompanha, de modo aproximadamente paralelo e em cada indivíduo, o valor da pressão sistólica (cinza-claro). Valores menos elevados da pressão sistólica correspondem a valores mais baixos da pressão diastólica, e vice-versa. A correlação linear de *Pearson* mostra forte associação entre as duas variáveis, com valor de $r = 0.813$, sendo muito significativa ($p = 0.0004$).

21.3 – GRÁFICO BOX-PLOT (mediana e quartis)

Trata-se de gráfico muito utilizado nas áreas de ciências biológicas e médicas, mostrando, no *box*, a mediana, o primeiro e o terceiro quartis. Exibe, ainda, o menor e o maior escores através do limite inferior e superior de linhas retas verticais, que se originam do primeiro e terceiro quartis, respectivamente.

21.3.1 – EXEMPLO

Foram tomadas as pressões arteriais sistólica e diastólica de 40 pessoas de ambos os sexos, no período de vigília e sono. Os valores estão inseridos no *grid* padrão e parcialmente representados na figura abaixo:

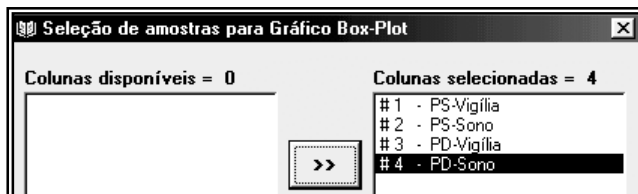
Figura 21.4
Dados do
Exemplo
21.3.1
(Visão parcial).

	- 1 - PS-Vigília	- 2 - PS-Sono	- 3 - PD-Vigília	- 4 - PD-Sono	- 5 -
1	122.000	101.000	75.000	61.000	
2	113.000	103.000	72.000	64.000	
3	133.000	112.000	86.000	67.000	
4	129.000	117.000	73.000	69.000	
5	117.000	104.000	74.000	69.000	
6	122.000	104.000	84.000	73.000	
7	120.000	93.000	79.000	62.000	
8	98.000	88.000	64.000	51.000	
9	109.000	98.000	71.000	65.000	
10	114.000	105.000	76.000	69.000	
11	109.000	100.000	77.000	68.000	
12	136.000	132.000	78.000	84.000	
13	114.000	104.000	70.000	65.000	

21.3.2 – INSTRUÇÕES

- Clicar em *Box Plot: mediana e quartis*;
- Selecionar as colunas do *grid* e pressionar em *Executar*, para resultados.

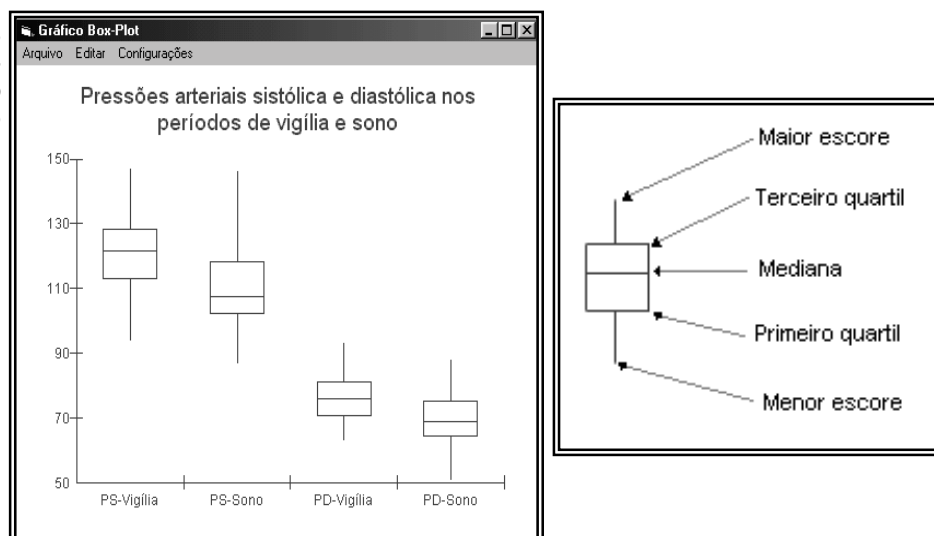
Figura 21.5
Colunas
selecionadas
para o
Exemplo
21.3.1.



O gráfico compara valores das pressões arteriais sistólicas nos períodos de vigília e de repouso noturno (sono), o mesmo ocorrendo com as diastólicas. As posições e as

amplitudes dos *boxes* e, ainda, o comprimento das linhas verticais evidenciam com clareza as diferenças entre as distribuições das variáveis referidas.

Figura 21.6
Gráfico *Box-Plot* do
Exemplo
21.3.1.



21.4 – GRÁFICO BOX-PLOT (média e desvios)

Semelhante ao gráfico anterior, mostrando, todavia, no *box*, a média e o desvio ou o erro padrão. Exibe, ainda, o menor e o maior escore através do limite inferior e superior de linhas retas verticais. Pode-se verificar, também, a presença de valores extremos (*outliers*), como ocorreu no exemplo abaixo, representados através de pequenos círculos acima ou abaixo das linhas verticais do gráfico. Com o *mouse*, pode-se visualizar, na parte superior do gráfico, o valor de cada uma dessas estatísticas ou dos escores amostrais.

21.4.1 – EXEMPLO

Foi efetuada coleta de 50 peixes de uma determinada espécie em um lago criatório, cujos escores estão inseridos no *grid* geral. Parte da amostra está contida na Fig. 21.7.

Figura 21.7
Dados do
Exemplo
21.4.1
(Visão
parcial).

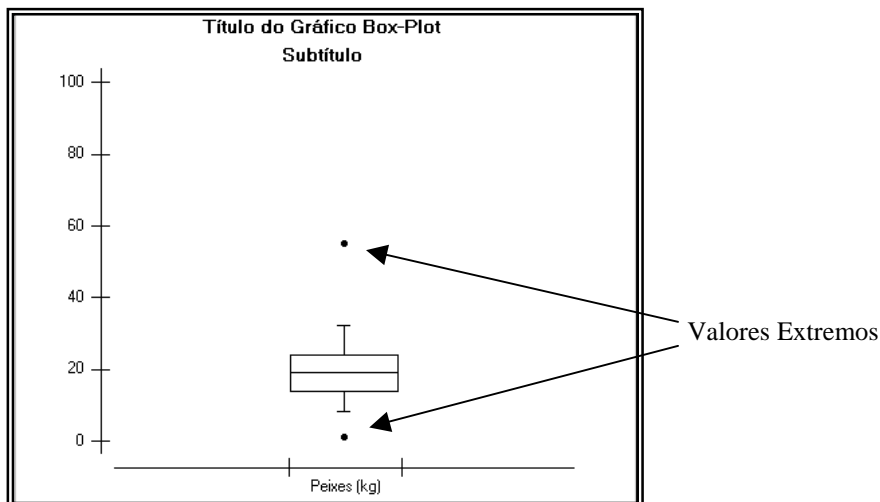
	- 1 - Peixes (kg)
1	28.000
2	23.000
3	20.600
4	13.200
5	32.000
6	24.000
7	18.700
8	1.000
9	19.400
10	18.400
11	18.600
12	21.400

21.4.2 – INSTRUÇÕES

- Clicar em *Box-Plot: média e desvios*, selecionar as coluna do *grid* e pressionar em *Executar*;
- Escolher um dos botões de opção: Caixa (*Box*) = **desvio padrão** e Suíças (*Whisker*) = **Máximo e Mínimo**, pressionando *Aplicar* para obter o gráfico.

Figura 21.8
Botões de opção do Exemplo 21.4.1.

Figura 21.9
Gráfico *Box-Plot* do Exemplo 21.4.1.



As linhas horizontais de cada *box* representam, agora, a média aritmética (linha central) e ± 1 desvio padrão (linhas superior e inferior).

21.5 – GRÁFICO CAULE E FOLHA (*STEM AND LEAF*)

Algumas vezes é difícil agrupar certos escores em tabelas e representá-los graficamente, perdendo-se, nesse propósito, o valor real de cada escore. O diagrama “caule e folha” contorna esse problema, mostrando, ainda, a disposição dos dados em forma de curva, evidenciando aspectos da simetria e da curtose.

Para sua construção, o programa separa o componente “caule” e respectiva “folha”, sendo o primeiro representado por todos os dígitos de cada escore, menos o dígito da direita, o qual constitui a “folha”. A precisão da “folha” pode ser escolhida pelo usuário. Exemplos:

<u>Escore</u>	<u>Caule</u>	<u>Folha</u>	<u>Precisão da Folha</u>
523	52	3	Unidade
46.74	46	7	Centésimo
115.8	115	8	Décimo
523	5	2	Dezena

O programa executa todos esses passos de acordo com as informações inseridas nas *caixas de textos*, conforme se verificará em seguimento.

21.5.1 – EXEMPLO

A pressão arterial sistólica de 87 pessoas do sexo masculino, cujas idades variavam de 20 a 45 anos, foram tomadas em condições basais. Os valores estão inseridos no *grid* geral, parcialmente mostrados na Figura 21.10.

Figura 21.10
Dados do
Exemplo
21.5.1
(Visão
parcial).

	- 1 - P A S	- 2 -
1	117.000	
2	111.000	
3	134.000	
4	133.000	
5	112.000	
6	118.000	
7	113.000	
8	96.000	

Estatísticas do Exemplo 21.5.1

Número de escores	= 87
Mediana	= 119
Média aritmética	= 119.8
Coef. de assimetria	= 0.29
Coef. de curtose	= 0.05

21.5.2 – INSTRUÇÕES

- Clicar em *Caule e Folha (Stem and Leaf)*;
- Selecionar a coluna do *grid* e pressionar o *mouse* em *Executar*;
- O formulário, mostrado na Figura 21.11 (lado direito), indica o maior e o menor escore e a amplitude dos dados; deve-se, com um clique do *mouse*, indicar o *nível de precisão da folha*, que neste caso foi a **unidade**, e o *número de linhas por classe*, o qual, neste exemplo, foram escolhidas duas (**2**) linhas;
- Pressionar em *Exibir Diagrama*, para resultados (Figura 21.11).

Figura 21.11
Gráfico de
Caule e Folha
do Exemplo
21.5.1.

Diagrama Caule e Folha		
Arquivo Editar		
	CAULE	FOLHA
Escore = 87		
1	9	3
3	9	6 6 7
4	10	2 3 3 4
8	10	7 8 8 8 9 9 9 9
14	11	0 0 1 1 2 2 2 2 2 3 3 3 3
15	11	5 5 5 5 5 6 6 7 7 8 8 8 9 9 9
14	12	0 0 0 0 1 1 2 2 2 2 3 3 3 3
10	12	5 6 7 7 7 8 8 9 9 9
8	13	0 0 0 0 2 3 4 4
4	13	5 5 6 6
3	14	1 3 4
2	14	5 7
1	15	2

A figura abaixo ilustra a configuração que precede a exibição do diagrama de Caule e Folha.

Note a importância da escolha adequada do item *Nível de Precisão da Folha*.

Informe o nível de precisão da folha

Coluna: 1 - P A S

Menor escore: 93
Maior escore: 152
Amplitude: 59

Nível de precisão da folha:

☐ Dez milésimos (0,0001) ☒ Unidade

☐ Um milésimo (0,001) ☐ Dezena

☐ Um centésimo (0,01) ☐ Centena

☐ Um décimo (0,1) ☐ Milhar

☐ Milhar x 10 ☐ Milhar x 100

Linhas por classe:

☐ 1 linha ☒ 2 linhas ☐ 5 linhas

Exibir Diagrama

O diagrama *Caule e Folha* apresenta uma distribuição aproximadamente simétrica ($g_1 = 0.29$) e mesocúrtica ($g_2 = 0.05$), com todos os escores amostrais, sendo a maior concentração de valores em torno da média e da mediana, 119.8 e 119, respectivamente.

21.6 – GRÁFICO DE COLUNAS

Destina-se a comparar escores de diferentes categorias representadas sob a forma de retângulos separados, mas da mesma largura, com alturas proporcionais às grandezas que representam. Os dados devem ser mensurados em escala nominal, ordinal ou numérica de valores discretos. O *BioEstat* apresenta os seguintes modelos para este tipo de gráfico:

- Colunas simples;
- Colunas justapostas;
- Colunas superpostas.

Em qualquer dos casos, as colunas podem ser apresentadas sob a forma vertical ou horizontal.

21.6.1 – EXEMPLO

Em um hospital foram efetuadas cirurgias classificadas em diferentes especialidades e realizadas no mês de março de 1999. Os dados estão contidos no *grid* geral:

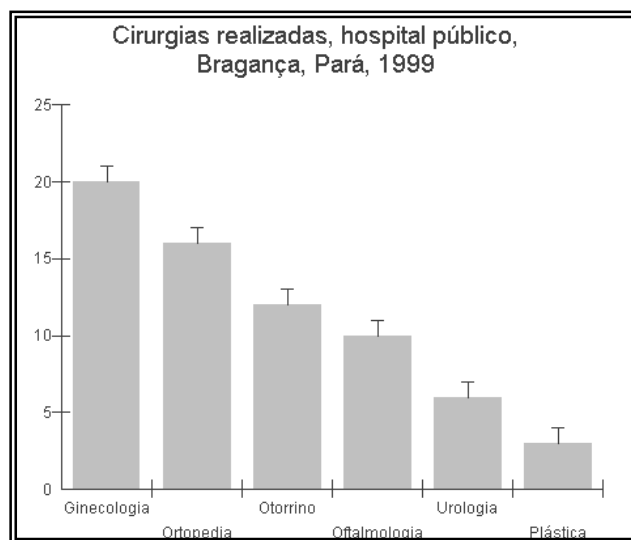
Figura 21.12
Dados do
Exemplo
21.6.1.

	- 1 - Cirurgias	- 2 -	- 3 -
1	Ginecologia	20.0	
2	Ortopedia	16.0	
3	Otorrino	12.0	
4	Oftalmologia	10.0	
5	Urologia	6.0	
6	Plástica	3.0	

21.6.2 – INSTRUÇÕES

- Clicar o *mouse* em *Colunas* e pressionar em *Simples*;
- Selecionar as duas colunas do *grid* e pressionar o *mouse* em *Executar*.

Figura 21.13
Gráfico de
Colunas
do Exemplo
21.6.1
contendo
barras de
erro.



Para visualizar as **Barras de Erro**, selecione na janela de **Configurações** a opção **Outras** e escolha o tipo de **Barra de Erro** conforme disponível na parte inferior da janela.

21.6.3

EXEMPLO (colunas justapostas, dados hipotéticos)

Foi efetuada investigação destinada a verificar a incidência de dengue de acordo com a escolaridade. Os dados estão inseridos no *grid* geral, conforme mostra a Figura 21.14.

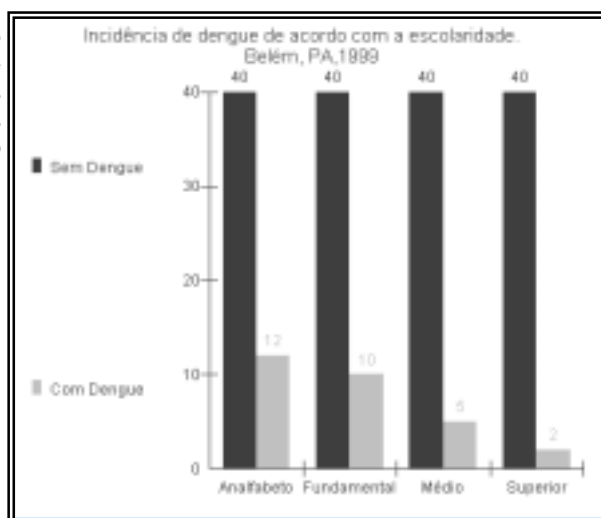
Figura 21.14
Dados do
Exemplo
21.6.3.

	- 1 -	- 2 -	- 3 -
	Escolaridade	Sem Dengue	Com Dengue
1	Analfabeto	40.000	12.000
2	Fundamental	40.000	10.000
3	Médio	40.000	5.000
4	Superior	40.000	2.000

21.6.4 – INSTRUÇÕES

- Clicar o *mouse* em *Colunas* e pressionar em *Justapostas*;
- Selecionar as três colunas do *grid* e pressionar em *Executar*.

Figura 21.15
Gráfico de
Colunas
justapostas
do Exemplo
21.6.3.

**21.6.5 – EXEMPLO (colunas superpostas, dados hipotéticos)**

A mesma investigação do exemplo anterior, mas com o tamanho da amostra diferente entre as diversas escolaridades. O modelo escolhido foi o de colunas superpostas, sendo os dados das colunas apresentados em forma percentual.

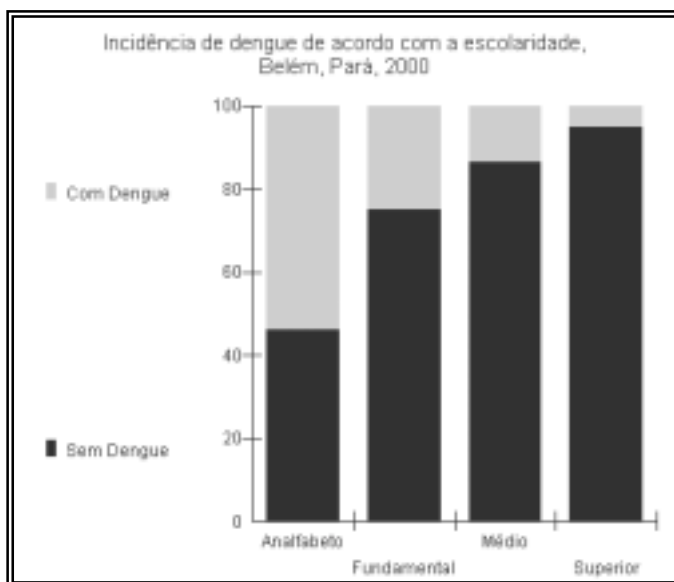
Figura 21.16
Dados do
Exemplo
21.6.5.

	- 1 -	- 2 -	- 3 -
	Escolaridade	Sem Dengue	Com Dengue
1	Analfabeto	46.000	54.000
2	Fundamental	45.000	15.000
3	Médio	52.000	8.000
4	Superior	38.000	2.000

21.6.6 – INSTRUÇÕES

- Clicar o mouse em *Colunas* e pressionar em *Superpostas*;
- Selecionar as três colunas do *grid* e pressionar em *Executar*;
- Após o aparecimento do gráfico, pressionar em *Configurações* e escolher a opção *Outras*;
- Clicar em *Percentuais*.

Figura 21.17
Gráfico de Colunas superpostas percentuais do Exemplo 21.6.5.



21.7 – DIAGRAMA DE DISPERSÃO

Gráfico utilizado para duas variáveis mensuradas de maneira simultânea – X e Y , como ocorre nos testes de Correlação e de Regressão.

21.7.1 – EXEMPLO

Peso e comprimento de folhas de gramínea coletadas em Ipixuna, Estado do Pará. Os escores estão inseridos no *grid* geral (Figura 21.18).

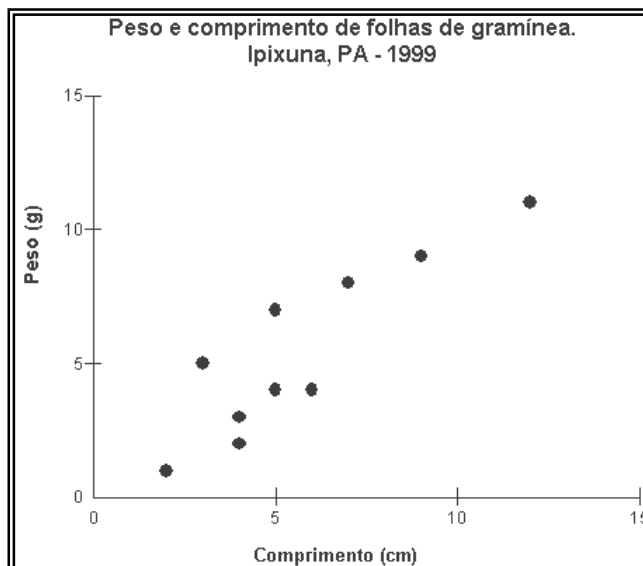
Figura 21.18
Dados do Exemplo 21.7.1.

	- 1 - Peso (g)	- 2 - Compr. (cm)	- 3 -
1	2.0	4.0	
2	5.0	3.0	
3	4.0	5.0	
4	8.0	7.0	
5	11.0	12.0	
6	9.0	9.0	
7	7.0	5.0	
8	4.0	6.0	
9	3.0	4.0	
10	1.0	2.0	

21.7.2 – INSTRUÇÕES

- Clicar em *Diagrama de Dispersão*;
- Selecionar as colunas do *grid* e pressionar o *mouse* em *Executar*.

Figura 21.19
Diagrama de
Dispersão do
Exemplo
21.7.1.



A figura abaixo ilustra o resultado do teste de Correlação Linear do Exemplo 21.7.1.

	Colunas 1 e 2
N (pares) =	10
r (Pearson) =	0.8878
IC 95% =	0.59 a 0.97
IC 99% =	0.41 a 0.98
R2 =	0.7882
t =	5.4561
GL =	8
(p) =	0.0006

Ao lado do gráfico de dispersão estão os resultados do teste de Correlação Linear, mostrando o coeficiente r muito significativo, com $p\text{-valor} = 0.0006$. Quando este gráfico expressa os escores da regressão linear, a reta representativa da equação dos mínimos quadrados pode ser ativada, conforme explicitado no item 21.16.

21.8 – DIAGRAMA IC (95% e 99%)

Este gráfico é aplicado quando em uma amostra há relação entre duas variáveis: uma categórica (nominal ou ordinal) e outra numérica (discreta ou contínua), sobretudo quando a categórica é, no mínimo, dicotômica (sexo masculino e feminino vs. idade, doentes e não doentes vs. níveis de colesterol).

O diagrama descreve graficamente a estimação da média e o respectivo intervalo de confiança (95% ou 99%) da variável numérica de cada categoria. Permite, ainda, comparar as médias e intervalos de confiança das categorias com a média geral e seu intervalo de confiança, o qual abrange todos os valores.

Trata-se de uma estimação de parâmetros de cada categoria, com base na *distribuição Z* (1.96 e 2.58) quando as amostras forem maiores que 30; e na *distribuição t*, se o tamanho amostral for igual ou inferior a 30 unidades.

Para este diagrama, o *BioEstat* estabelece o limite de dez (10) categorias.

21.8.1. – EXEMPLO 1

Foi efetuada avaliação da cardiotoxicidade do antimoníato de meglumina no tratamento da leishmaniose tegumentar americana. Um grupo de quinze (15) pacientes, escolhidos de maneira aleatória, recebeu a dose de 15 mg/kg/dia dessa droga durante um período de 20 dias. Foram efetuados eletrocardiogramas

convencionais de cada paciente, antes e logo após o tratamento, determinando-se o **Intervalo PR** de cada traçado. Critérios idênticos foram efetuados com outros quinze (15) pacientes mudando-se apenas a dose do medicamento para 20 mg/kg/dia. O total de eletrocardiogramas atingiu 60 unidades, com a seguinte distribuição:

- 15 Eletrocardiogramas antes do tratamento com dosagem de 15/mg/kg/dia;
- 15 Eletrocardiogramas após o tratamento com dosagem de 15/mg/kg/dia;
- 15 Eletrocardiogramas antes do tratamento com dosagem de 20/mg/kg/dia;
- 15 Eletrocardiogramas após o tratamento com dosagem de 20/mg/kg/dia;

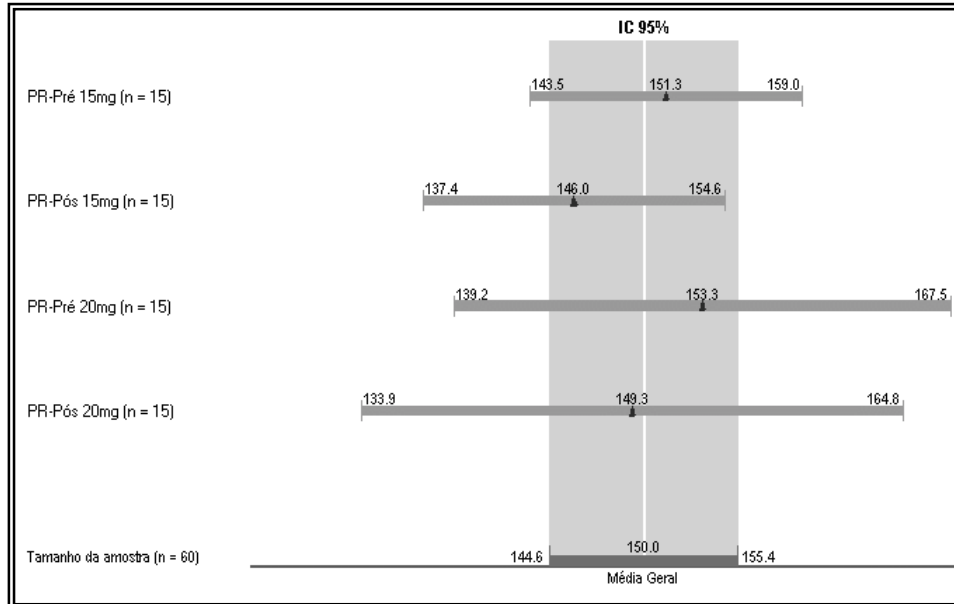
Figura 21.20
Visão parcial
dos dados
do
Exemplo
24.8.1.

	- 1 - Categoria	- 2 - Intervalo PR	- 3 - Categorias	- 4 - Legenda
1	1	159	1	PR-Pré 15mg
2	1	170	2	PR-Pós 15mg
3	1	130	3	PR-Pré 20mg
4	1	160	4	PR-Pós 20mg

21.8.2 – INSTRUÇÕES

- Introduzir os dados da seguinte forma:
Primeira coluna: Contém as categorias. Neste exemplo os valores são de um (1) a quatro (4), correspondentes às dosagens de 15 e 20 mg/kg/dia, antes e após os tratamentos.
Segunda coluna: Os valores da variável numérica. Neste caso é o valor do **Intervalo PR**, em cada situação indicada na primeira coluna.
Terceira e quarta colunas: indicam o número e o nome das categorias.
- Clicar em *Gráficos* e, a seguir, em *Diagrama IC 95%* e selecionar as quatro (4) colunas;
- Pressionar em *Executar* para obter o diagrama abaixo.

Figura 21.21
Diagrama
IC 95%
do
Exemplo
21.8.1.



Verifica-se que as médias dos Intervalos PR das amostras antes dos tratamentos foram ligeiramente superiores à média geral (150.0), ocorrendo o oposto após a medicação. Por sua vez, o IC (95%) da média geral foi o menor de todos os ICs, em decorrência do tamanho amostral (60), o qual reuniu todas as amostras.

21.8.3 – EXEMPLO 2

Em indivíduos infectados pelo *H. pylori* na mucosa gástrica, a avaliação das idades de 299 destes indivíduos apresentou a seguinte distribuição: 103 assintomáticos; 100 com gastrite; 65 com úlcera gástrica ou duodenal e 31 com câncer do estômago.

Figura 21.22
Visão parcial
dos dados
do
Exemplo
21.8.3.

	- 1 - Grupo	- 2 - Idade	- 3 - Grupo	- 4 - Legenda	- 5 -
1	4.000	44.000	1.000	Controle	
2	1.000	60.000	2.000	Gastrite	
3	3.000	37.000	3.000	Úlcera	
4	2.000	63.000	4.000	Câncer	
5	1.000	53.000			
6	3.000	30.000			
7	4.000	50.000			
8	2.000	35.000			
9	4.000	70.000			
10	1.000	56.000			
11	3.000	71.000			
12	2.000	58.000			
13	1.000	42.000			
14	4.000	42.000			
15	1.000	44.000			
16	1.000	36.000			
17	1.000	23.000			

21.8.4 – INSTRUÇÕES

- a) Introduzir os dados da seguinte forma:

Primeira coluna: Contém as categorias. Neste exemplo os valores são de 1 a 4 e correspondem às categorias:

- 1) assintomáticos (controles);
- 2) com gastrite;
- 3) com úlcera gástrica ou duodenal;
- 4) com câncer do estômago.

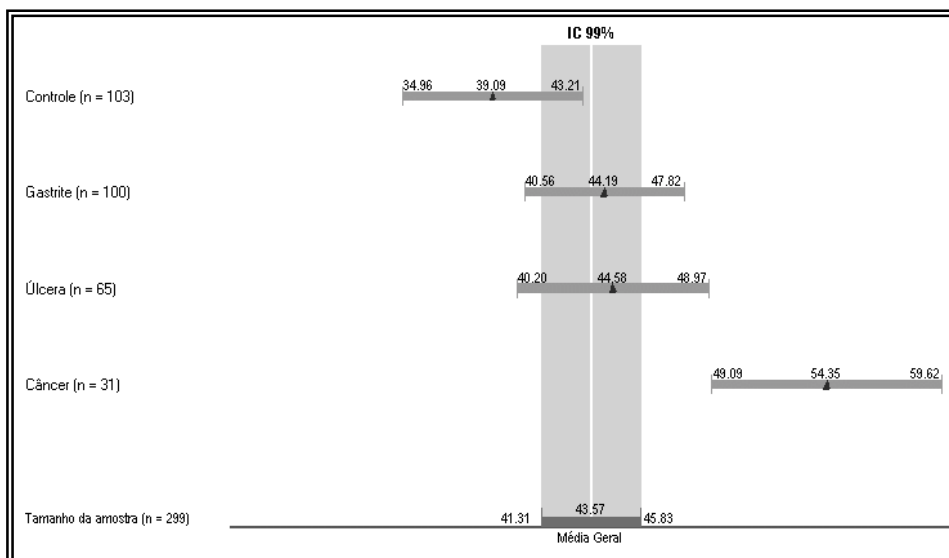
Obs.: Não há necessidade de ordenação prévia desses valores, conforme se verifica na Figura 21.23. O programa computacional efetua os respectivos “ranks” numéricos, sendo essa uma das vantagens deste aplicativo.

Segunda coluna: Variável numérica. Esta coluna contém o valor das idades de cada grupo: controles, gastrite, etc.

Terceira e quarta colunas: indicam o número e o nome das categorias.

- b) Clicar em *Gráficos* e, a seguir, em *Diagrama IC 95%* e selecionar as quatro (4) colunas;
- c) Pressionar em *Executar* para obter o diagrama abaixo.

Figura 21.23
Diagrama
IC 95%
do
Exemplo
24.8.3.



Percebe-se, nitidamente, o deslocamento da média e do IC dos acometidos de câncer gástrico. Neste grupo a média etária foi igual a 54.35 anos e o limite inferior (49.09) do respectivo IC foi maior que o limite superior do IC da média geral (45.83).

A média geral (43.57), por sua vez, foi menor que a média dos acometidos de gastrite (44.19) ou de úlcera gástrica ou duodenal (44.58), mas superou a dos indivíduos controles (39.09).

O Intervalo de Confiança (95%) da média geral, como seria o esperado, em face do tamanho amostral ($n = 299$), foi o menor (41.31 a 45.83), ou seja, de 4.52 anos.

21.9 – DIAGRAMA PONTUAL (DOT PLOT)

Este diagrama descreve graficamente duas ou mais amostras (até sete), permitindo a visualização pontual dos escores em sentido vertical e as respectivas comparações.

21.9.1 – EXEMPLO

O peso em gramas de sementes de duas variedades (A e B) da mesma espécie foram coletados de modo aleatório, e esses valores foram inseridos no *grid* do *BioEstat*, conforme visão parcial contida na Figura 21.24.

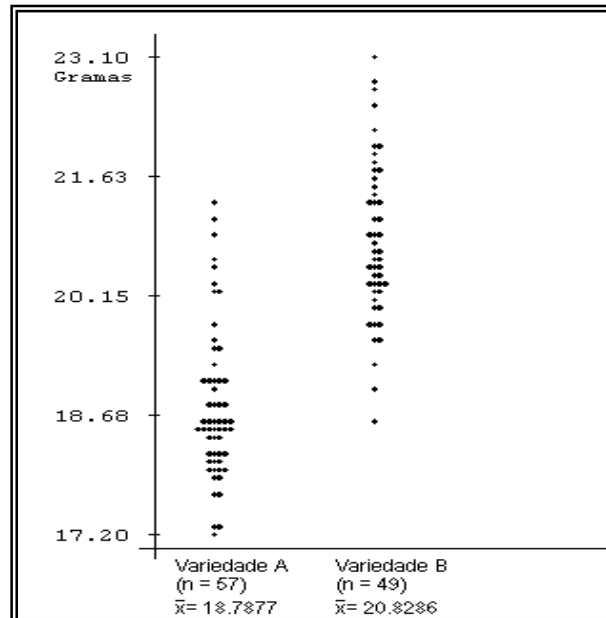
Figura 21.24
Visão parcial
dos dados
do
Exemplo
21.9.1.

	- 1 - Variedade A	- 2 - Variedade B	- 3 -
1	18.5	20.5	
2	17.9	19.6	
3	18.4	20.6	
4	17.7	20.2	
5	18.6	20.3	
6	19.0	20.9	
7	18.4	21.3	

21.9.2 – INSTRUÇÕES

- a) Clicar em *Gráficos* e, a seguir, em *Diagrama Pontual*;
 b) Pressionar em *Executar* para Resultados.

Figura 21.25
 Diagrama
 Pontual
 do
 Exemplo
 21.9.1.



Observa-se que o peso das sementes da Variedade A é inferior ao da Variedade B.

21.10 – HISTOGRAMA

O histograma é uma 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.

O programa solicita que o usuário informe se os dados estão ou não agrupados em classes e, também, o número desejado.

Para obter melhor representação gráfica, deve-se evitar que o número de classes seja muito pequeno ou muito grande, pois a representação do histograma, nesses casos, não atenderá as finalidades esperadas.

21.10.1 – EXEMPLO

Foi tomada a pressão arterial diastólica (P.A.D.) de 45 pessoas do sexo feminino, cujas idades variavam de 25 a 40 anos. Os dados estão inseridos no *grid* geral e parcialmente reproduzidos na Figura 21.26.

Figura 21.26

Visão parcial dos dados do Exemplo 21.10.1.

	- 1 - P A D	- 2 -	- 3 -
1	75.000		
2	72.000		
3	86.000		
4	73.000		
5	74.000		
6	84.000		

Dados do Exemplo 21.10.1	
Número de escores	= 45
Mínimo	= 63
Máximo	= 93
Média aritmética	= 77,067
Desvio Padrão	= 8,0068
Mediana	= 76

21.10.2 – INSTRUÇÕES

- Pressionar o *mouse* em *Histograma*;
- Selecionar a coluna do *grid* geral e clicar em *Executar*;
- Preencher o formulário *Especificação das Classes* (Fig. 21.27):
 - Tipo de Limites de Classes: *inferior fechado* e *superior aberto* ($\text{[} | \text{)}$);
 - Mantenha o conteúdo da caixa de texto *Número de Classe* (6);
 - Ajuste os *Limites Superior* e *Inferior*, 96 e 60, respectivamente;
 - O *Limite Inferior* não poderá ser **maior** que o *Limite Inferior da Amostra*, o qual corresponde ao menor valor amostral, 63 neste caso;
 - O *Limite Superior* não poderá ser **menor** que o *Limite Superior da Amostra*, correspondente ao maior valor amostral, 93 neste exemplo;
- Pressionar o botão *Confirmar* para obter o *Histograma*.

Figura 21.27

Histograma do Exemplo 21.10.1.

Especificação das Classes

Escolha o tipo de limites dos intervalos das classes

☒ $\text{[} | \text{)}$
☐ $\text{[} | \text{]}$
☐ $\text{[} | \text{)}$

Número de Classes:
 Intervalo de Classe:
 Casas decimais:

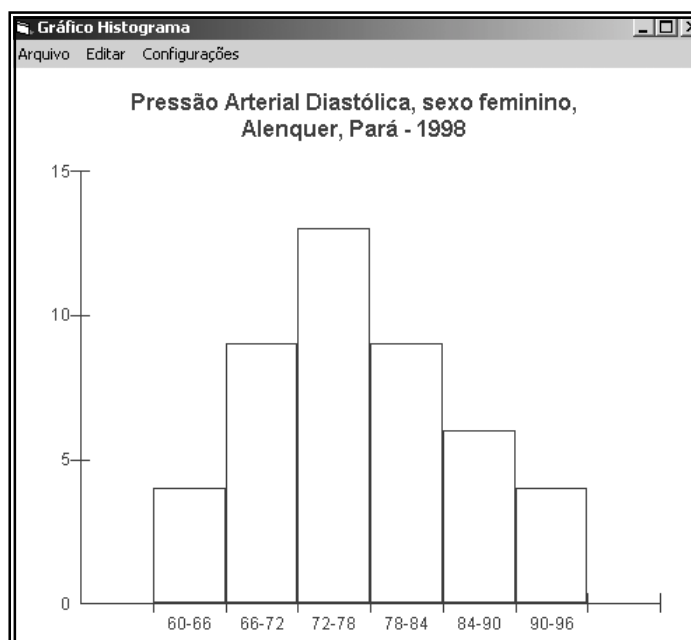
Limite Superior Ajustado:
 Amplitude Total:
 Limite Inferior Ajustado:

Limite Superior da Amostra:
 Tamanho da amostra:
 Limite Inferior da Amostra:

Obs: configuração para histograma e tabela de frequência.

Figura 21.28

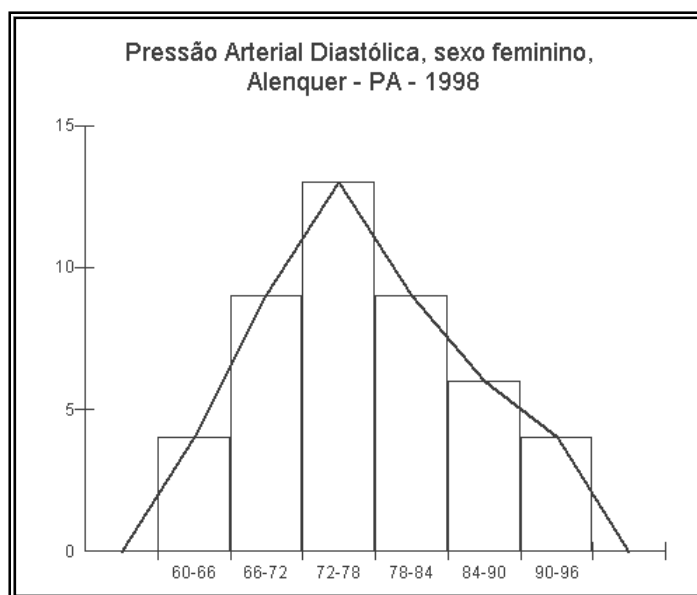
O *BioEstat* sugere o número de classes e os limites superior e inferior do histograma do Exemplo 21.10.1.



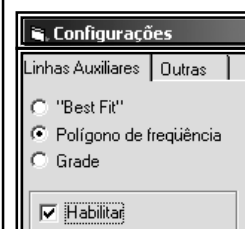
O histograma pode ser associado ao *polígono de freqüência*, como é observado na Figura 21.29.

Figura 21.29

Histograma com polígono de freqüência do Exemplo 21.10.1.



A figura abaixo ilustra a configuração para exibição do polígono de freqüência.



21.11 – GRÁFICO DE LINHA (CURVA)

Utiliza-se este gráfico sobretudo na representação de séries temporais, considerando-se a disposição dos dados nos eixos das abscissas (X) e das ordenadas (Y).

21.11.1 – EXEMPLO (hipotético)

Populações dos países A e B (em milhões de habitantes) determinadas a cada dez (10) anos, de acordo com a recomendação das Nações Unidas, obtendo-se os dados que estão inseridos no *grid* geral. Os escores da primeira coluna correspondem ao eixo das abscissas (X) e os das demais colunas, ao eixo das ordenadas (Y).

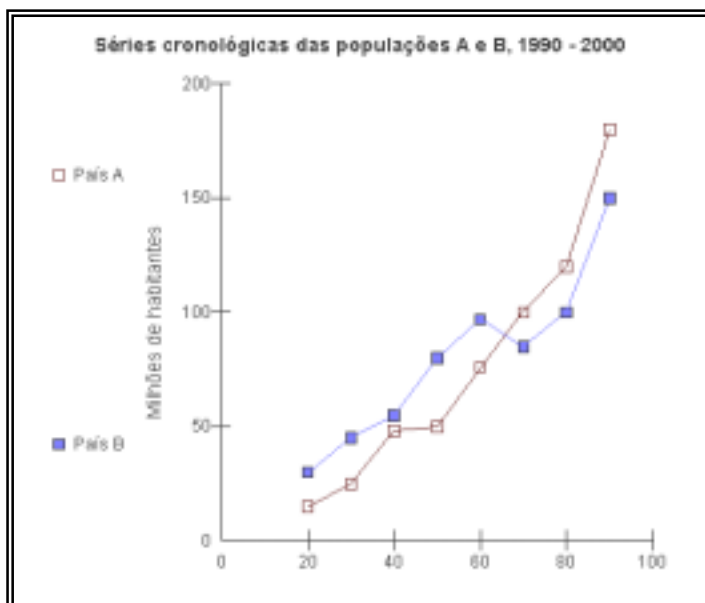
Figura 21.30
Dados do
Exemplo
21.11.1.

	- 1 - Ano	- 2 - País A	- 3 - País B
1	20	15	30
2	30	25	45
3	40	48	55
4	50	50	80
5	60	76	97
6	70	100	85
7	80	120	100
8	90	180	150

21.11.2 – INSTRUÇÕES

- Clicar em *Linha (Curva)*;
- Selecionar as colunas do *grid* e pressionar em *Executar*.

Figura 21.31
Gráfico de
linha do
Exemplo
21.11.1.



21.12 – GRÁFICO DE MÁXIMO E MÍNIMO

Permite comparar a amplitude e os valores máximo e mínimo de cada amostra.

21.12.1 – EXEMPLO

A pressão arterial – sistólica e diastólica – foi tomada de 25 pessoas do sexo feminino e de dez (10) indivíduos do sexo masculino, respectivamente. Os dados estão inseridos no *grid* geral, com as indicações da Legenda:

Figura 21.32
Dados
parciais do
Exemplo
21.12.1.

	- 1 - PMS-F	- 2 - PMS-M	- 3 - PMD-F	- 4 - PMD-M
1	119.000	118.000	63.000	73.000
2	102.000	135.000	63.000	72.000
3	128.000	142.000	62.000	88.000
4	120.000	126.000	64.000	72.000
5	118.000	141.000	65.000	86.000
6	116.000	89.000	74.000	48.000
7	154.000	124.000	84.000	70.000
8	121.000	119.000	65.000	72.000
9	115.000	135.000	68.000	71.000
10	135.000	125.000	66.000	72.000
11	111.000		71.000	
12	146.000		75.000	
13	124.000		67.000	

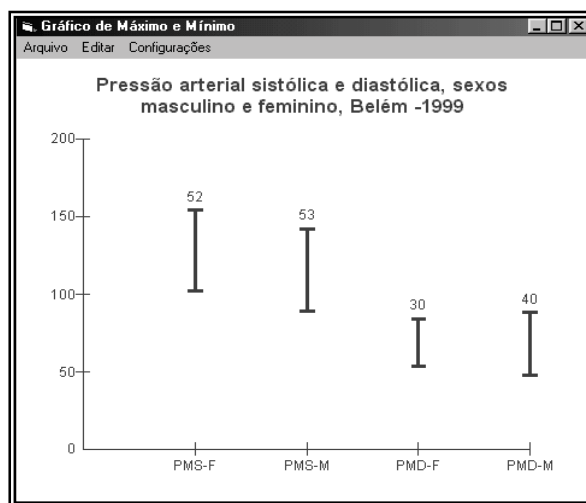
Legenda:

- 1: PMS-F → pressão máxima sistólica, sexo feminino;
 2: PMS-M → pressão máxima sistólica, sexo masculino;
 3: PMD-F → pressão máxima diastólica, sexo feminino;
 4: PMD-M → pressão máxima diastólica, sexo masculino.

21.12.2 – INSTRUÇÕES

- Clicar em *Máximo e Mínimo*;
- Selecionar as colunas do *grid* e pressionar em *Executar*.

Figura 21.33
Gráfico de
Máximo e
Mínimo do
Exemplo
21.12.1.



As *amplitudes*, tanto sistólicas como diastólicas, assinaladas na parte superior de cada barra vertical, foram maiores nos indivíduos do sexo masculino, o que demonstra maior variação dessas variáveis entre os homens. Já o valor *máximo* da pressão sistólica foi maior entre as mulheres, enquanto o da diastólica, observou-se nas pessoas do sexo masculino.

21.13 – PIRÂMIDE POPULACIONAL

A maioria dos *gráficos estatísticos* refere-se às amostras. Pode-se, contudo, representar Populações através de *Pirâmides*, as quais podem representar grupos etários, proporções sexuais, locais de residências, estados civis e outras variáveis.

21.13.1 – EXEMPLO 1

A população do Estado do Pará (IBGE) no ano 2000 foi tabulada em relação aos grupos etários e respectivos sexos.

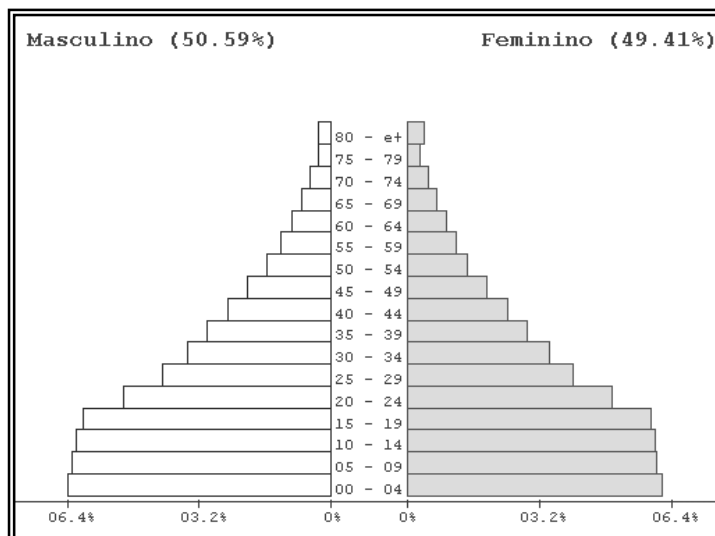
Figura 21.34
Visão parcial
dos dados
do
Exemplo
21.13.1.

	- 1 - Limite inferior	- 2 - Limite superior	- 3 - Masculino	- 4 - Feminino
1	0	4	396051	381704
2	5	9	389664	374270
3	10	14	382467	372500
4	15	19	371238	365028
5	20	24	312171	306983
6	25	29	252511	249916
7	30	34	214968	212604
8	35	39	185800	180571

21.13.2 – INSTRUÇÕES

- Inserir no *grid* as informações obtidas do IBGE;
- Pressionar em *Gráficos e*, a seguir, em *Pirâmide Populacional*;
- Clicar em Executar para Resultados.

Figura 21.35
Pirâmide
populacional
do
Exemplo
21.13.1.



Há ligeiro predomínio de indivíduos do sexo masculino nos primeiros anos de vida e do sexo feminino após os 60 anos de vida.

21.13.3 – EXEMPLO 2

A população do Estado do Pará (IBGE) no ano 1970 foi tabulada em relação aos grupos etários e locais de residência, urbana e rural.

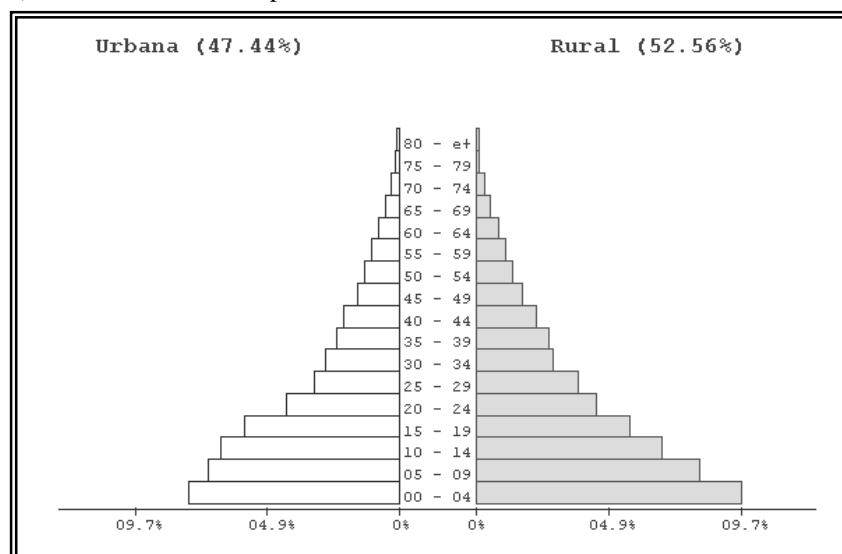
Figura 21.36
parcial dos
dados do
Exemplo
24.13.3.

- 1 - Limite inferior	- 2 - Limite superior	- 3 - Urbana	- 4 - Rural
0	4	167347	210151
5	9	151718	177549
10	14	141683	147341
15	19	123151	122358
20	24	90239	95795
25	29	68277	81272
30	34	58562	60988
35	39	50429	57823
40	44	44196	48076
45	49	33607	36454
50	54	27893	29490
55	59	22580	23304
60	64	16855	18527
65	69	11765	11669
70	74	7274	7064

24.13.4 – INSTRUÇÕES

- Inserir no *grid* as informações obtidas do IBGE;
- Pressionar em *Gráficos* e, a seguir, em *Pirâmide Populacional*;
- Clicar em Executar para Resultados.

Figura 21.37
Pirâmide
populacional
do
Exemplo
24.13.3.



Enquanto no ano 2000 a proporção de residentes na área rural era de 33.45%, no ano de 1970 a população dessa mesma área chegava a 52.56%.

21.13.5 – EXEMPLO 3

A população do Estado do Pará (IBGE) no ano 2000 foi tabulada em relação aos grupos etários e locais de residência, urbana e rural.

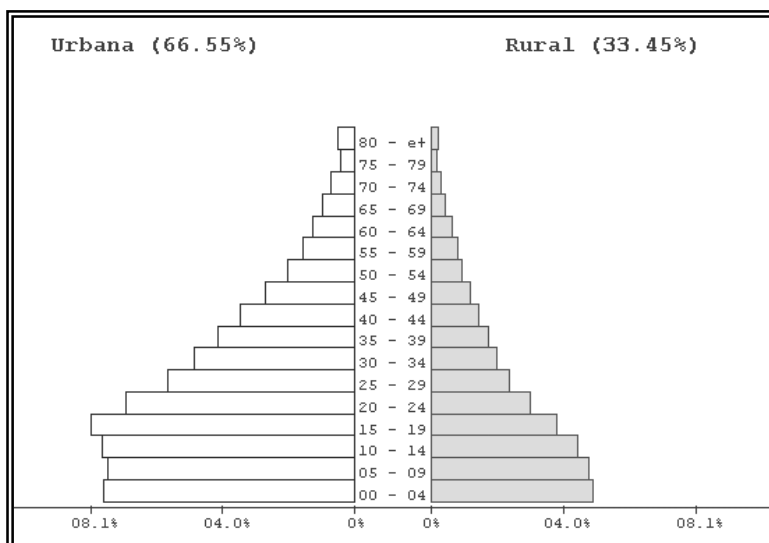
Figura 21.38
Visão parcial
dos dados
do
Exemplo
24.13.5.

	- 1 - Limite inferior	- 2 - Limite superior	- 3 - Urbana	- 4 - Rural
1	0	4	473137	304618
2	5	9	465551	298383
3	10	14	476952	278015
4	15	19	498808	237458
5	20	24	432249	186905
6	25	29	353224	149203
7	30	34	302864	124708
8	35	39	258785	107586
9	40	44	216308	89335
10	45	49	170030	74886
11	50	54	128083	59191
12	55	59	98381	51085
13	60	64	79479	39330
14	65	69	61272	28108
15	70	74	44728	18455

24.13.6 – INSTRUÇÕES

- Inserir no *grid* as informações obtidas do IBGE;
- Pressionar em *Gráficos* e, a seguir, em *Pirâmide Populacional*;
- Clicar em Executar para Resultados.

Figura 21.39
Pirâmide
populacional
do
Exemplo
21.13.5.



Há predominância de moradores na área urbana quando comparados com os da área rural, denotando uma migração para as cidades em todo o Estado do Pará.

21.13.7 – EXEMPLO 4

A população do Estado do Pará foi tabulada em relação aos grupos etários, sexos e estado civil (solteiros e outros). Os dados são hipotéticos, apenas para demonstrar a possibilidade da introdução de três (3) variáveis.

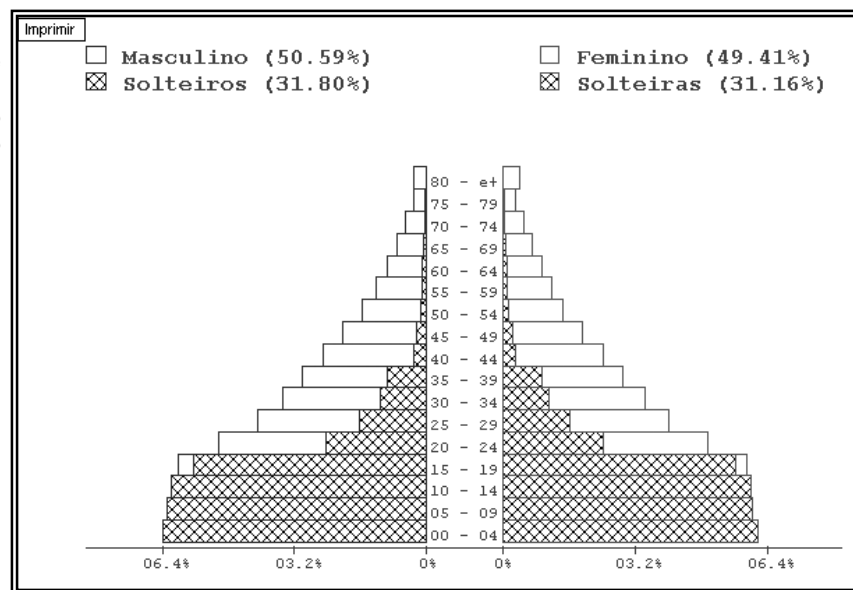
Figura 21.40
Visão parcial
dos dados
do
Exemplo
21.13.7.

- 1 -	- 2 -	- 3 -	- 4 -	- 5 -	- 6 -
Limite inferior	Limite superior	Masculino	Solteiros	Feminino	Solteiras
0	4	396051	396051	381704	381704
5	9	389664	389664	374270	374270
10	14	382467	382467	372500	372500
15	19	371238	350000	365028	350000
20	24	312171	150000	306983	150000
25	29	252511	100000	249916	100000
30	34	214968	70000	212604	70000
35	39	185800	60000	180571	60000
40	44	155518	20000	150125	20000
45	49	125771	15000	119145	15000
50	54	96586	10000	90688	10000

21.13.8 – INSTRUÇÕES

- Inserir no *grid* as informações obtidas do IBGE;
- Pressionar em *Gráficos* e, a seguir, em *Pirâmide Populacional*;
- Clicar em *Executar* para Resultados.

Figura 21.41
Pirâmide
populacional
do
Exemplo
21.13.7.



Esta pirâmide representa três variáveis: faixa etária, sexo, estado civil (solteiros e não solteiros).

21.14 – GRÁFICO POLAR

É recomendado para séries temporais de caráter cíclico, as quais representam determinados períodos.

21.14.1 – EXEMPLO (hipotético)

Foram registrados os limites das áreas endêmicas de malária e dengue durante doze (12) meses. As distâncias lineares registradas estão dimensionadas em quilômetros, em relação à sede do município. Os dados estão contidos na Fig. 21.42.

Figura 21.42
Dados do
Exemplo
21.14.1.

	- 1 -	- 2 -	- 3 -	-
	** Meses	Malária	Dengue	
1	Janeiro	19	5.0	
2	Fevereiro	17	7.0	
3	Março	20	6.0	
4	Abril	23	5.0	
5	Maio	18	8.0	
6	Junho	16	7.0	
7	Julho	17	10.0	
8	Agosto	18	8.0	
9	Setembro	21	6.0	
10	Outubro	15	21.0	
11	Novembro	18	5.0	
12	Dezembro	19	4.0	

21.14.2 – INSTRUÇÕES

- Introduzir os dados no *grid* geral do *BioEstat*;
- Apontar o *mouse* em *Gráficos*;
- Clicar em *Polar*;
- Selecionar as colunas do *grid* e pressionar o *mouse* em *Executar*.

Figura 21.43
Resultados
do Exemplo
21.14.1.



Apenas em um período (outubro) a ocorrência de dengue ultrapassou os limites geográficos dos casos de malária. A sede do município está representada pelo centro do gráfico (cruzamento das linhas das ordenadas com as das abscissas).

21.15 – GRÁFICO DE SETOR

Deve ser utilizado para dados nominais ou ordinais, quando o principal objetivo é o de comparar várias parcelas com o total. Divide-se um círculo em *setores*, cada um correspondendo, de forma proporcional, a uma das categorias da amostra.

21.15.1 – EXEMPLO (hipotético)

Dados da vacinação efetuada em crianças de zero (0) a um (1) ano de idade, em um Posto de Saúde da cidade de Castanhal, Pará, no primeiro trimestre de 1999.

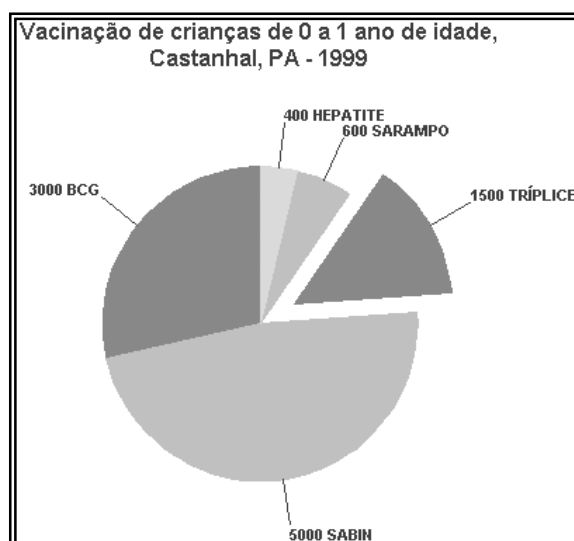
Figura 21.44
Dados do
Exemplo
21.15.1.

	- 1 - Vacinas	- 2 - Quantidade	- 3 -
1	BCG	3000.0	
2	SABIN	5000.0	
3	TRÍPLICE	1500.0	
4	SARAMPO	600.0	
5	HEPATITE	400.0	

21.15.2 – INSTRUÇÕES

- Clicar em *Setor*;
- Na janela do gráfico selecione *Configurações*;
- Na aba *Outras*, selecione entre *Absolutos* ou *Percentuais*;
- Selecionar as colunas do *grid* e pressionar o *mouse* em *Executar*.

Figura 21.45
Gráfico de
setor do
Exemplo
21.15.1.



Para destacar uma ou mais 'fatias' do gráfico de setor, pressione o menu **'Configurações'**, selecione a janela **'Fontes e Cores'**, clique no botão **'Elementos do gráfico'**, selecione o número do(s) elemento(s) desejado(s) e marque a caixa **'Destacar'**



É facilmente visualizado que a vacina Sabin e o BCG representaram as maiores proporções de vacinações efetuadas no período indicado. A fatia destacada corresponde à vacina tríplice, escolhida de forma aleatória.

21.16 – MANIPULAÇÃO DE GRÁFICOS

Os gráficos gerados pelo *BioEstat* podem ser alterados pelo usuário, impressos ou copiados para o *clipboard* (memória auxiliar do Microsoft Windows).

21.16.1 – IMPRIMIR GRÁFICOS

Após a exibição do gráfico, acione o menu *Arquivo*, *Imprimir* e escolha o tipo de impressão desejada.

Figura 21.46
Menu de
impressão
de gráficos.



21.16.2 – COPIAR E COLAR GRÁFICOS

Após a exibição do gráfico, acione o menu *Editar*, selecione a opção *Copiar*, para que a imagem do gráfico fique armazenada na memória, podendo, dessa forma, ser colada (aplicada, visualizada) em outros programas como o Microsoft Excel, Word ou PowerPoint, através do uso das teclas CTRL-V ou no acionamento do menu *Editar*, *Colar* daqueles aplicativos.

21.16.3 – CONFIGURAR APARÊNCIA DE GRÁFICOS

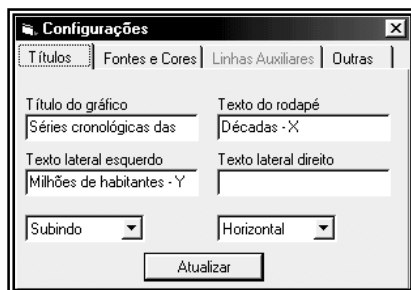
As características do gráfico podem ser alteradas pelo acionamento do menu *Configurações*. O formulário de Configurações contém quatro (4) grupos de características que podem ser manipuladas:

- a) Títulos;
- b) Fontes e Cores;
- c) Linhas Auxiliares;
- d) Outras.

21.16.3.1 - CONFIGURAR TÍTULOS

- a) Clique na aba *Títulos* para visualizar a Figura 21.47.
- b) Preencha as *caixas de texto* com os títulos desejados;
- c) Clique em *Atualizar*.

Figura 21.47
Configuração
de títulos da
Figura 21.31.



21.16.3.2 - CONFIGURAR FONTES E CORES

Clique na aba *Fontes e Cores* para visualizar a Figura 21.48.

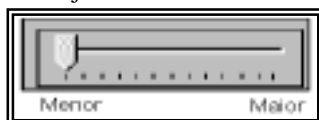
Figura 21.48
Configuração
de fontes e
cores.



Para configurar adequadamente o item *Fontes e Cores*, siga as instruções abaixo:

1. Escolha a abrangência da configuração, entre as seguintes opções:
 - a) Título;
 - b) Subtítulos;
 - c) Fundo (Cor de fundo);
 - d) Legendas;
 - e) Elementos do Gráfico;
 - f) Explodir.
2. Ajuste o tamanho dos caracteres utilizando a escala gráfica abaixo.

Figura 21.49
Tamanho dos
caracteres.



3. Escolha a aparência dos textos selecionados, entre as seguintes opções:
 - a) **Negrito**
 - b) *Itálico*
 - c) Sublinhado
4. Escolha a cor desejada clicando na grade colorida exibida na Figura 21.48.
5. Para alterar a cor de elementos como barras, linhas, pontos, etc.:
 - a) Clique no botão Elementos do Gráfico;
 - b) Selecione o número do elemento, na caixa de combinação;
 - c) Clique na cor desejada.

21.16.3.3 - CONFIGURAR LINHAS AUXILIARES

Com os dados do Exemplo 21.7.1 e respectivo gráfico, clique em *Linhas Auxiliares* para visualizar a Figura 21.50.

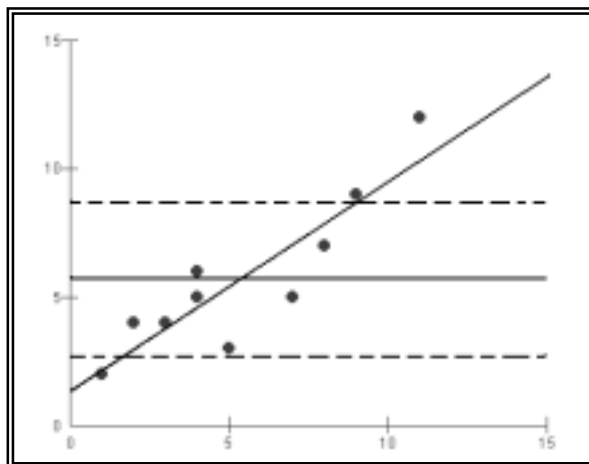
Figura 21.50
Configuração
de linhas
auxiliares.



Para configurar as linhas auxiliares:

- Escolha a linha desejada: Média, Desvio padrão, Máximo e Mínimo, Curva ajustada, *best Fit* ou Grade;
- Clique no botão *Habilitar* (localizado acima da grade de cores);
- Clique na cor desejada

Figura 21.51
Linhas
auxiliares.



As linhas tracejadas, Figura 21.51, correspondem a $\pm 1\sigma$ em torno da média de Y . A linha contínua horizontal representa a média aritmética da mesma variável e a linha contínua inclinada corresponde à reta de regressão ($\hat{Y} = a + bX$).

21.16.3.4 – OUTRAS (Dados, Dimensões, Eixos e Orientação)

De acordo com o modelo selecionado do gráfico, os dados poderão ser apresentados sob a forma percentual ou em termos absolutos, as figuras (setor) representadas em duas ou três dimensões, a intercessão entre os eixos cartesianos na origem (0,0) ou em um ponto variável determinado pelo *BioEstat*, e a orientação das colunas em barras verticais ou horizontais.

Para essas finalidades deve-se selecionar em *Configurações* a opção *Outras*, escolhendo uma das opções conforme o tipo de gráfico em questão, ou seja, *Dados*, *Dimensões*, *Eixos* ou *Orientação*.

21.17 – TABELAS DE FREQUÊNCIA

As tabelas de frequência constituem um modelo de representação tabular utilizado nas descrições de dados quantitativos discretos ou contínuos. Embora sua construção manual seja relativamente simples, o *BioEstat* apresenta esse modelo, com os intervalos de classes, os pontos médios e as frequências absoluta e percentual de cada classe, podendo, ainda, o usuário colocar o título, conforme o caso.

O número de classes fica a critério do usuário, mas o programa, considerando a amplitude total e o número de escores, sugere um número adequado, de acordo com a fórmula de *Sturges*.

21.17.1 – EXEMPLO (hipotético)

Foi determinada a quantidade de consumo de oxigênio, após exercício controlado, em 44 estudantes de uma Escola Superior de Educação Física, Pará., em março de 1998. Os dados estão inseridos no *grid* geral, Figura 21.52.

Figura 21.52
Visão parcial
do Exemplo
21.17.1

	- 1 - Oxigênio	- 2 -	- 3 -
1	46.000		
2	45.000		
3	63.000		
4	59.000		
5	42.000		

21.17.2 – INSTRUÇÕES





- Pressionar o mouse em *Tabelas de Frequência*;
- Selecionar a coluna do *grid* geral e clicar em *Executar*;
- Preencher o formulário *Especificação das Classes* (Fig. 21.53):
 - Tipo de Limites de Classes: *inferior fechado* e *superior fechado* ();
 - Aumente o conteúdo da caixa de texto *Número de Classes* para **7**;
 - Ajuste os *Limites Superior* e *Inferior*, **73** e **32**, respectivamente;
 - O *Limite Inferior* não poderá ser **maior** que o *Limite Inferior da Amostra*, o qual corresponde ao menor valor amostral, 33 neste caso;
 - O *Limite Superior* não poderá ser **menor** que o *Limite Superior da Amostra*, correspondente ao maior valor amostral, 70 neste exemplo;
- Pressionar o botão *Confirmar* para obter *Tabela de Frequência*.

Figura 21.53
Formulário
Especificação
das Classes.

Especificação das Classes

Escolha o tipo de limites dos intervalos das classes

☐  ☒  ☐ 

Número de Classes: Intervalo de Classe: Casas decimais:

Limite Superior Ajustado: Amplitude Total: Limite Inferior Ajustado:

Limite Superior da Amostra: Tamanho da amostra: Limite Inferior da Amostra:

Obs: configuração para histograma e tabela de frequência.

Figura 21.54
Tabela de
Frequência
do Exemplo
21.17.1.

Tabela de frequência			
Imprimir Precisão das classes Histograma			
Consumo de oxigênio, após exercício controlado, em 44 estudantes			
Escola de Educação Física, Pa., 1998			
Classes	Xi	Fi	Percentual
32 — 37	34.5	4	9.09 %
38 — 43	40.5	5	11.36 %
44 — 49	46.5	10	22.73 %
50 — 55	52.5	12	27.27 %
56 — 61	58.5	7	15.91 %
62 — 67	64.5	3	6.82 %
68 — 73	70.5	3	6.82 %
TOTAL		44	100 %

Se desejar a imagem desses dados, pressione o mouse em *Histograma* na barra superior da Fig. 21.54.

A figura 21.53 apresenta três opções de tipos de limites dos intervalos de classes:

- intervalos fechados à esquerda e abertos à direita, devendo-se, neste caso, aumentar o limite superior ajustado;
- intervalos fechados à direita e à esquerda;
- intervalos fechados à direita e abertos à esquerda, devendo-se, neste caso, diminuir o limite inferior ajustado.

21.18 – GRÁFICOS DE RESULTADOS

Além dos gráficos contidos neste capítulo, o *BioEstat* 5.0 disponibiliza outros diagramas adequados à interpretação de resultados específicos, os quais só podem ser obtidos através da opção *Gráfico* que está localizada na parte superior do *grid* de resultados do teste em questão.

21.19 – CONFIGURAÇÕES AVANÇADAS

Além das opções de configurações de gráficos disponíveis no menu Configurar, o *BioEstat* permite o acesso à interface de Configurações Avançadas. O acesso a este recurso está disponível quando uma janela de gráfico está sendo exibida pelo *BioEstat*. Para visualizar essa interface pressione o botão direito do mouse sobre o diagrama exibido.

CAPÍTULO 22

FÓRMULAS ESTATÍSTICAS

22.1 – INTRODUÇÃO

Neste capítulo estão dispostas as fórmulas dos testes estatísticos contidos no *BioEstat*.

22.2 – ANÁLISE MULTIVARIADA

22.2.1 – ANÁLISE DE CONGLOMERADOS

a) Distâncias: ver 22.2.5, 22.2.6 e 22.2.7

b) Distância *Chebyshev*: $d_{ij} = \max |X_{iv} - X_{jv}|$;

c) Distância de *Bray-Curtis*:

$$BC_{ij} = \sum \frac{|n_{ik} - n_{jk}|}{(n_{ik} + n_{jk})}$$

d) Ligação simples: $d_{(UV)W} = \min(d_{UW}, d_{VW})$;

e) Média do Grupo: $d_{(UV)W} = \frac{\sum_i \sum_k d_{ik}}{i \ k}$;

f) *Ward* (Variância Mínima): $ESS = \sum_{j=1}^N (x_j - \bar{x})(x_j - \bar{x})$;

g) *Gower* (Mediana): $S_{ij} = \frac{\sum_{v=1}^p S_{ijv}}{\sum_{v=1}^p W_{ijv}}$;

h) Centróide: $d_{(i,j)k} = \min |d_{ik}; d_{jk}|$.

22.2.2 – ANÁLISE DISCRIMINANTE

A combinação linear (Fisher) $\hat{y} = \hat{a}'x = [\bar{x}_1 - \bar{x}_2]' S_{pooled}^{-1} x$ é

é maximizada pela equação: $max = \frac{(\hat{a}'d)^2}{\hat{a}' S_{pooled} \hat{a}}$

$$s_y^2 = \frac{\sum_{j=1}^{n_1} (y_{1j} - \bar{y}_1)^2 + \sum_{j=1}^{n_2} (y_{2j} - \bar{y}_2)^2}{n_1 + n_2 - 2}, \text{ sendo}$$

$$y_{1j} = \hat{a}' x_{1j} \quad \text{e} \quad y_{2j} = \hat{a}' x_{2j}$$

22.2.3 – CORRELAÇÃO CANÔNICA

$$U_r = a_{r1}X_1 + a_{r2}X_2 + \dots + a_{rp}X_p$$

$$V_r = b_{r1}Y_1 + b_{r2}Y_2 + \dots + b_{rp}Y_p$$

$$Corr(U, V) = \frac{a' \sum_{12} b}{\sqrt{a' \sum_{11} a} \sqrt{b' \sum_{22} b}}$$

$$\phi_j^2 = -\{n - 0.5(pq + 1)\} \sum_{i=j+1}^r \log_e(1 - \lambda_i)$$

$$\text{Graus de liberdade} = (p - 1)(q - 1)$$

	$X_1 \ X_2 \dots X_p$	$Y_1 \ Y_2 \dots Y_p$
X_1	$p \times p \text{ matriz}$ A	$p \times q \text{ matriz}$ C
X_2		
.		
.		
X_p		
Y_1	$q \times p \text{ matriz}$ C	$q \times q \text{ matriz}$ B
Y_2		
.		
.		
Y_p		

22.2.4 – COMPONENTE PRINCIPAL

a) Matriz de covariância:

$$C = \begin{bmatrix} c_{11} & c_{12} & \Lambda & c_{1p} \\ c_{21} & c_{22} & \Lambda & c_{2p} \\ \mathbf{M} & \mathbf{M} & & \mathbf{M} \\ c_{p1} & c_{p2} & \Lambda & c_{pp} \end{bmatrix};$$

b) Autovalores:

$$\lambda_1 + \lambda_2 + \lambda_3 + \mathbf{K} + \lambda_p = c_{11} + c_{22} + c_{33} + \mathbf{K} + c_{pp};$$

c) Ordenação dos autovalores:

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \mathbf{K} \geq \lambda_p \geq 0;$$

d) Autovetores:

$$a_{i1}, a_{i2}, \mathbf{K}, a_{ip};$$

e) λ_i corresponde ao $i^{\text{ésimo}}$ componente principal:

$$Z_i = a_{i1}X_1 + a_{i2}X_2 + \mathbf{K} + a_{ip}X_p;$$

f) Variância:

$$\text{var}(Z_i) = \lambda_i.$$

22.2.5 – DISTÂNCIA MULTIVARIADA (Euclidiana)

i, j = indivíduos; p = número de variáveis;

$$d_{ij} = \sqrt{\left\{ \sum_{k=1}^p (x_{ik} - x_{jk})^2 \right\}}.$$

22.2.6 – DISTÂNCIA MULTIVARIADA (Penrose)

p = número de variáveis;

$$P_{ij} = \sum_{k=1}^p \frac{(\mu_{ki} - \mu_{kj})^2}{pV_k}.$$

22.2.7 – DISTÂNCIA MULTIVARIADA (Mahalanobis)

p = número de variáveis;

$$D_{ij}^2 = \sum_{r=1}^p \sum_{s=1}^p (\mu_{ri} - \mu_{rj}) v^{rs} (\mu_{si} - \mu_{sj}).$$

22.2.8 – TESTE DE HOTELLING

a) Estimação conjunta da matriz de covariância:

$$C = \frac{\{(n_1 - 1)C_1 + (n_2 - 1)C_2\}}{n_1 + n_2 - 2};$$

b) Teste de *Hotelling*:

$$T^2 = \frac{n_1 n_2 (\bar{x}_1 - \bar{x}_2)' C^{-1} (\bar{x}_1 - \bar{x}_2)}{n_1 + n_2};$$

$$F = (n_1 + n_2 - p - 1) T^2 / \{(n_1 + n_2 - 2)p\};$$

$$gl = (n_1 + n_2 - p - 1).$$

22.2.9 – TESTE DE BARTLETT

n = número de observações;

m = número de amostras;

p = número de variáveis;

$|T|$ = matriz da soma dos quadrados e produtos;

$|W|$ = matriz da soma dos quadrados dentre as amostras e dos produtos cruzados;

$$\phi = \left[n - 1 - \frac{1}{2}(p - m) \right] \ln[|T|/|W|].$$

22.2.10 – TESTE DE MANTEL

n = número de elementos da matriz;

M = elementos da primeira matriz;

E = elementos da segunda matriz;

$$Z = \Sigma \Sigma ME$$

$$r = \frac{n \Sigma Z - (\Sigma M)(\Sigma E)}{\sqrt{[n \Sigma M^2 - (\Sigma M)^2][n \Sigma E^2 - (\Sigma E)^2]}}$$

$$gl = n - 2$$

22.3 – ANÁLISE DE SOBREVIVÊNCIA**22.3.1 – ANÁLISE ATUARIAL**

$S(t)$ = função de risco, que indica a probabilidade de sobrevivência no tempo t , para cada $t > 0$;

$h(t)$ = risco no tempo t , determinado pela equação:

$$h(t) = \left[\frac{S(t) - S(t + \Delta t)}{\Delta t} \right] / S(t).$$

22.3.2 – KAPLAN-MEIER

$$\hat{S}(t) = \prod_{j=1}^k \left(\frac{n_j - d_j}{n_j} \right);$$

intervalos de $t_{(k)}$ a $t_{(k+1)}$, sendo $k = 1, 2, 3, \dots, r$;

$$\text{var}\{\ln \hat{S}(t)\} = \sum_{j=1}^k \text{var}\{\ln \hat{p}_j\}.$$

22.3.3 – LOG-RANK TEST

$$O = \sum_{i=1}^k a_i;$$

$$E = \sum_{i=1}^k \frac{(a_i + b_i)(a_i + c_i)}{N_i};$$

$$V = \sum_{i=1}^k \frac{(a_i + b_i)(c_i + d_i)(a_i + c_i)(b_i + d_i)}{N_i^2(N_i - 1)};$$

$$X_{LR}^2 = \frac{(|O - E| - 0.5)^2}{V}.$$

22.3.4 – COX-MANTEL TEST

$$U = r_2 - \sum_{i=1}^k m_i A_i;$$

$$I = \sum_{i=1}^k \frac{m_i(r_i - m_i)}{r_i - 1} A_i(1 - A_i); \quad Z = U\sqrt{I}.$$

22.3.5 – GEHAN (WILCOXON GENERALIZADO)

$$W = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} U_{ij};$$

$$\text{Var}(W) = \frac{n_1 n_2 \sum_{i=1}^{n_1+n_2} U_i^2}{(n_1 + n_2)(n_1 + n_2 - 1)}; \quad Z = W\sqrt{\text{Var}(W)}.$$

22.3.6 – COX: MODELO DE RISCO PROPORCIONAL

$$S(t) = e^{(a + b_1 X_1 + b_2 X_2 + \dots + b_k X_k)}$$

e = base dos logaritmos naturais (2.71828);

a = constante; b_k = coeficiente de regressão;

X_k = variável;

$$h(t) = -\ln S(t)$$

$$\left[\frac{h(t)}{h_0(t)} \right] = e^{(b_1 X_1 + b_2 X_2 + \dots + b_k X_k)}$$

$$\ln \left[\frac{h(t)}{h_0(t)} \right] = (b_1 X_1 + b_2 X_2 + \dots + b_k X_k)$$

22.4 – ANÁLISE DE VARIÂNCIA**22.4.1 – ANOVA (um critério)**

X_{ij} = observações individuais;

j = relativo às repetições (1, 2, 3... j ... r);

$X_{i.}$ = totais dos tratamentos;

i = tratamentos (1, 2, 3... i ... t);

rt = N de unidades experimentais;

$X_{..}$ = total geral; $\bar{X}_{i.} = \sum X_{ij} / r$;

SQ = soma dos quadrados:

$$SQ_{total} = \sum X_{ij}^2 - \frac{X_{..}^2}{rt}; \quad SQ_{tratamentos} = \frac{\sum X_{i.}^2}{r} - \frac{X_{..}^2}{rt};$$

$$SQ_{erro} = SQ_{total} - SQ_{tratamentos}; \quad gl_{tratamentos} = t - 1; \quad gl_{erro} = t(r - 1);$$

QM = quadrados médios (variâncias):

$$QM_{tratamentos} = SQ_{tratamentos} / gl_{tratamentos}; \quad QM_{erro} = SQ_{erro} / gl_{erro};$$

$$F_{tratamentos} = \frac{QM_{tratamentos}}{QM_{erro}}.$$

22.4.2 – ANOVA (dois critérios)

$$SQ_{blo cos} = \frac{\sum X_{.j}^2}{t} - \frac{X_{..}^2}{rt}; \quad SQ_{erro} = SQ_{total} - SQ_{tratamentos} - SQ_{blo cos};$$

$$gl_{blo cos} = r - 1; \quad gl_{erro} = (t - 1)(r - 1);$$

$$QM_{blo cos} = SQ_{blo cos} / gl_{blo cos}; \quad QM_{erro} = SQ_{erro} / gl_{erro};$$

$$F_{blo cos} = \frac{QM_{blo cos}}{QM_{erro}}.$$

22.4.3 – TESTE DE FRIEDMAN

N = número de linhas;

k = número de colunas;

R_j = soma dos postos da coluna j ;

$$X_r^2 = F_r = \left[\frac{12}{Nk(k+1)} \sum_{j=1}^k R_j^2 \right] - 3N(k+1).$$

22.4.4 – TESTE KRUSKAL-WALLIS

k = número de amostras ou grupos;

N = número total de escores;

n_j = número de escores da amostra j ;

R_j = somatório dos postos da amostra j ;

$$H = \left[\frac{12}{N(N+1)} \sum_{j=1}^k n_j \bar{R}_j^2 \right] - 3(N+1).$$

$$z = \frac{\bar{R}_i - \bar{R}_j}{\sqrt{\frac{N(N+1)}{12} \times \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}}$$

$$\text{alfa ajustado: } \alpha = \frac{\alpha}{k(k-1)}$$

22.4.5 – TESTE Q DE COCHRAN

$gl = k - 1$;

G_j = total de sucessos no posto j ;

L_i = total de sucessos na linha i ;

$$Q = \frac{(k-1) \left[k \sum_{j=1}^k G_j^2 - \left(\sum_{j=1}^k G_j \right)^2 \right]}{k \sum_{i=1}^N L_i - \sum_{i=1}^N L_i^2}.$$

22.5 – CORRELAÇÃO**22.5.1 – CONTINGÊNCIA C**

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}};$$

$$C = \sqrt{\frac{\chi^2}{N + \chi^2}};$$

N = soma de todas as categorias.

22.5.2 – CONCORDÂNCIA DE KENDALL

$$W = \frac{\sum \left(R_j - \frac{\sum R_j}{N} \right)^2}{\frac{1}{12} k^2 (N^3 - N)}; \quad r_{S_{av}} = \frac{kW - 1}{k - 1};$$

k = número de conjuntos de postos;

N = número de indivíduos.

22.5.3 – CORRELAÇÃO DE KENDALL

N = número de postos;

$$\tau = \frac{S}{\frac{1}{2} N(N-1)}; \quad \sigma_\tau = \sqrt{\frac{2(2N+5)}{9N(N-1)}};$$

$$z = \frac{\tau - \mu_\tau}{\sigma_\tau}.$$

22.5.4 – COEFICIENTE DE CORRELAÇÃO PARCIAL

N = número de trincas de variáveis (X, Y, Z);

$$r_{xy.z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{1 - r_{xz}^2} \sqrt{1 - r_{yz}^2}} \rightarrow \text{variável } z \text{ mantida constante};$$

$$gl = N - 3.$$

22.5.5 – COEFICIENTE DE CORRELAÇÃO LINEAR DE PEARSON

$$r = \frac{\sum XY - (\sum X)(\sum Y)/n}{\sqrt{\sum X^2 - (\sum X)^2/n} \sqrt{\sum Y^2 - (\sum Y)^2/n}};$$

n = número de pares (X, Y) ;

$$gl = n - 2;$$

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}.$$

22.5.6 – COEFICIENTE *Phi*

$$phi = r_{\phi} = \frac{ad - bc}{\sqrt{(a+c)(b+d)(a+b)(c+d)}}.$$

22.5.7 – COEFICIENTE DE CORRELAÇÃO DE SPEARMAN

N = número de pares; D = diferença entre postos;

$$r_s = 1 - \frac{6\sum D_i^2}{N(N^2 - 1)};$$

$$t = r_s \sqrt{\frac{N-2}{1-r_s^2}}.$$

22.6 – DISTRIBUIÇÃO DE PROBABILIDADES

22.6.1 – DISTRIBUIÇÃO BINOMIAL

n = tamanho da amostra;

$E(x) = \mu = np$ = média;

$V(x) = \sigma^2 = npq$ = variância;

$\sigma = \sqrt{npq}$ = desvio padrão;

$$P(x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}.$$

22.6.2 – DISTRIBUIÇÃO DE POISSON

n = tamanho da amostra;

$E(x) = \mu = np$ = média;

$V(x) = \sigma^2 = np$ = variância;

$$P(x) = \frac{\mu^x e^{-\mu}}{x!};$$

$e = 2.71828$ (constante, base dos logaritmos neperianos).

22.6.3 – DISTRIBUIÇÃO HIPERGEOMÉTRICA

n = tamanho da amostra;

N = tamanho da população;

k = número de elementos na população considerados “sucessos”;

$N - k$ = número de elementos na população considerados “insucessos”;

x = número de “sucessos” na amostra;

$$P(x) = \frac{\frac{k!}{x!(k-x)!} \times \frac{(N-k)!}{(n-x)!((N-k)-(n-x))!}}{\frac{N!}{n!(N-n)!}}$$

$$E(x) = \mu = n \frac{k}{N} = \text{média};$$

$$V(x) = \sigma^2 = n \left(\frac{k}{N} \right) \left(\frac{N-k}{N} \right) \left(\frac{N-n}{N-1} \right);$$

$$P(x) = \frac{\frac{k!}{x!(k-x)!} \times \frac{(N-k)!}{(n-x)!((N-k)-(n-x))!}}{\frac{N!}{n!(N-n)!}}.$$

22.6.4 – DISTRIBUIÇÃO EXPONENCIAL

$\mu = 1/\lambda$ = média; $\sigma = 1/\lambda$ = desvio padrão;

$$P(x \geq 0) = e^{-\lambda x}$$

$$P(x \geq 0) = e^{-\lambda x}, \text{ sendo } \lambda > 0.$$

22.6.5 – DISTRIBUIÇÃO NORMAL

$$\mu = \frac{\sum X}{N} = \text{média};$$

$$\sigma^2 = \frac{\sum X^2 - (\sum X)^2 / N}{N} = \text{variância};$$

$$\sigma = \sqrt{\sigma^2} = \text{desvio padrão};$$

$$Y = \frac{N}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2 / 2\sigma^2} = \text{equação da curva normal};$$

Y = frequência;

X = escore bruto;

$e = 2.71828$ (constante matemática, base dos logaritmos neperianos);

$\pi = 3.1416$ (constante matemática);

$z = \frac{X - \mu}{\sigma}$ = escores padronizados.

22.6.6 – CRIVO (Screening Test)

Tabela de contingência 2 x 2:

Teste	Doente	Não-doente	Total
+	a	b	$a + b$
-	c	d	$c + d$
Total	$a + c$	$b + d$	$a + b + c + d$

$\frac{a}{a + c} \times 100$ = sensibilidade; $\frac{d}{b + d} \times 100$ = especificidade;

$\frac{a}{a + b} \times 100$ = valor preditivo do teste positivo;

$\frac{d}{c + d} \times 100$ = valor preditivo do teste negativo;

$\frac{a + c}{a + b + c + d}$ = prevalência.

22.6.7 – CURVA ROC (Receiver Operating Characteristic curve)

NA = Número de pacientes

NN = Número de sadios

Ponto de Corte = 1 – especificidade, sensibilidade

$\text{Área} = \sum_{i=1}^k 0.5 * (\text{Sensibilidade}_i + \text{Sensibilidade}_{i-1}) * \text{Especificidade}_i$

$Q1 = \text{Área} / (2 - \text{Área}); \quad Q2 = (2 \times \text{Área}^2) / (1 + \text{Área})$

$Q3 = \text{Área} \times (1 - \text{Área}); \quad Q4 = (NA - 1) \times (Q1 - \text{Área}^2)$

$Q5 = (NN - 1) \times (Q2 - \text{Área}^2);$

Erro padrão = $\sqrt{(Q3 + Q4 + Q5) / (NA \times NN)}$

$d = \sqrt{(1 - \text{Sensibilidade})^2 + (1 - \text{Especificidade})^2}$

Eixo das ordenadas (Y) = Sensibilidade

Eixo das abscissas (X) = 1 – Especificidade (Falso-positivos)

22.6.8 – REGRA DE BAYES

$$\Pr(B_i|A) = \frac{\Pr(A|B_i) \times \Pr(B_i)}{\sum_{j=1}^k \Pr(A|B_j) \times \Pr(B_j)}$$

22.7 – UMA AMOSTRA**22.7.1 – KOLMOGOROV-SMIRNOV (aderência)**

k = número de postos; N = tamanho da amostra;

$Fo(X)$ = distribuição teórica acumulada sob H_0 ;

$Sn(X)$ = distribuição amostral acumulada;

D = máximo $|Fo(X) - Sn(X)|$ = desvio máximo;

$$D \geq \frac{1.36}{\sqrt{N}}, \text{ para } \alpha = 0.05; \quad D \geq \frac{1.63}{\sqrt{N}}, \text{ para } \alpha = 0.01.$$

22.7.2 – QUI-QUADRADO (aderência)

O = frequência observada; E = frequência esperada;

k = número de categorias;

$$\chi^2 = \sum \frac{(O - E)^2}{E};$$

$$\chi^2_{(Yates)} = \sum \frac{(|O - E| - 0.5)^2}{E};$$

$$gl = k - 1.$$

22.7.3 – TESTE G (aderência)

f_i = frequência observada; \hat{f}_i = frequência esperada;

k = número de categorias;

$$G = 2 \sum_{i=1}^k f_i \times \ln \left(\frac{f_i}{\hat{f}_i} \right); \quad gl = k - 1.$$

22.7.4 – TESTE BINOMIAL: UMA PROPORÇÃO

\hat{p} = proporção do evento amostral; p_0 = proporção do evento na população;

N = tamanho da amostra;

Quando $\hat{p} < p_0$: $p(\text{valor}) = 2\Phi[(\hat{p} - p_0)/\sqrt{(p_0q_0)/N}]$;

Quando $\hat{p} \geq p_0$: $p(\text{valor}) = 2\{1 - \Phi[(\hat{p} - p_0)/\sqrt{(p_0q_0)/N}]\}$.

22.7.5 – TESTE *t* (Student)

μ = média da população; n = tamanho da amostra;

$$\bar{X} = \frac{\sum X_i}{n}; \quad s_{\bar{x}} = s / \sqrt{n};$$

$$t = \frac{\bar{X} - \mu}{s_{\bar{x}}}; \quad gl = n - 1.$$

22.7.6 – TESTE Z

μ = média da população; n = tamanho da amostra;

$$\bar{X} = \frac{\sum X_i}{n}; \quad \sigma_{\bar{x}} = \sigma / \sqrt{n};$$

$$Z = \frac{\bar{X} - \mu}{\sigma_{\bar{x}}}.$$

22.7.7 – TESTE DE POISSON

$$p = 2 \times \left[1 - \sum_{k=0}^x \frac{e^{-\mu_0} \mu_0^k}{k!} \right] \quad \text{quando } x < \mu_0$$

$$p = 2 \times \left[\sum_{k=0}^x \frac{e^{-\mu_0} \mu_0^k}{k!} \right] \quad \text{quando } x \geq \mu_0.$$

22.8 – DUAS AMOSTRAS INDEPENDENTES

22.8.1 – TESTE EXATO DE FISHER

A = escore da casela A ; B = escore da casela B ;

C = escore da casela C ; D = escore da casela D ;

$N = A + B + C + D$;

$$p = \frac{(A+B)!(C+D)!(A+C)!(B+D)!}{N!A!B!C!D!}.$$

22.8.2 – TESTE DE KOLMOGOROV-SMIRNOV

k = número de postos; NA = tamanho da 1ª amostra; NB = tamanho da 2ª amostra;

Distribuição das frequências acumuladas:

$$S_{NA}(X) = k/NA; \quad S_{NB}(X) = k/NB;$$

D = máximo $|S_{NA}(X) - S_{NB}(X)|$ = desvio máximo;

$$\chi^2 = 4D^2 \frac{NA \times NB}{NA + NB}; \quad D \geq 1.36 \sqrt{\frac{NA + NB}{NA \times NB}}, \quad \text{para } \alpha = 0.05;$$

$$D \geq 1.63 \sqrt{\frac{NA + NB}{NA \times NB}}, \text{ para } \alpha = 0.01.$$

22.8.3 –TESTE DE MANN-WHITNEY

n_1 = tamanho da primeira amostra; n_2 = tamanho da segunda amostra;

R_1 = soma dos postos da primeira amostra;

R_2 = soma dos postos da segunda amostra;

$$U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1, \text{ ou } U = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2;$$

$$\mu_U = \frac{n_1 n_2}{2} = \text{média}; \quad \sigma_U = \sqrt{\frac{(n_1)(n_2)(n_1 + n_2 + 1)}{12}} = \text{desvio padrão};$$

$$z = \frac{U - \mu_U}{\sigma_U}; \quad p(\text{valor}) = 2 \times [1 - \Phi(z)].$$

22.8.4 –TESTE DA MEDIANA

MC = mediana combinada;

A = número de escores da amostra 1 acima da MC ;

B = número de escores da amostra 1 abaixo da MC ;

C = número de escores da amostra 2 acima da MC ;

D = número de escores da amostra 2 abaixo da MC ;

$$X^2 = \frac{N \left(|AD - BC| - \frac{N}{2} \right)^2}{(A + B)(C + D)(A + C)(B + D)}; \quad gl = 1.$$

22.8.5 – TAXA DE INCIDÊNCIA

$$E_1 = \frac{(a_1 + a_2)t_1}{t_1 + t_2};$$

$$V_1 = \frac{(a_1 + a_2)t_1 t_2}{(t_1 + t_2)^2};$$

a_1 e a_2 = número de eventos nos grupos 1 e 2;

t_1 e t_2 = número de pessoas nos tempos 1 e 2;

$$Z = \frac{a_1 - E_1 - 0.5}{\sqrt{V_1}}, \text{ quando } a_1 > E_1;$$

$$Z = \frac{a_1 - E_1 + 0.5}{\sqrt{V_1}}, \text{ quando } a_1 \leq E_1; \text{ Teste válido quando } V_1 \geq 5.$$

22.8.6 – ODDS RATIO (OR)

Tabela de contingência 2 x 2:

	Sucesso	Insucesso	Total
Amostra 1	a	b	n_1
Amostra 2	c	d	n_2
Total	n_a	n_b	N

$$p_1 = a/n_1; \quad q_1 = 1 - p_1;$$

$$p_2 = c/n_2; \quad q_2 = 1 - p_2;$$

$$OR = \frac{ad}{bc} = \frac{p_1 q_2}{p_2 q_1}; \quad X^2 = \sum \frac{(|O - E| - .5)^2}{E};$$

$$C_1 = \ln(OR) - 1.96 \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}};$$

$$C_2 = \ln(OR) + 1.96 \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}};$$

$$IC(95\%): 2.71828^{C_1} \leq OR \leq 2.71828^{C_2};$$

IC é válido somente quando:

$$n_1 p_1 q_1 \geq 5 \text{ e } n_2 p_2 q_2 \geq 5.$$

22.8.7 – QUI-QUADRADO

r = número de linhas;

k = número de colunas;

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}};$$

$$\chi^2_{Yates} = \sum_{i=1}^r \sum_{j=1}^k \frac{(|O_{ij} - E_{ij}| - 0.5)^2}{E_{ij}};$$

$$gl = (r - 1)(k - 1);$$

Correção de Yates: somente quando $gl = 1$.

22.8.8 – RISCO RELATIVO

Tabela de contingência 2 x 2:

	Sucesso	Insucesso	Total
Amostra 1	a	b	n_1
Amostra 2	c	d	n_2
Total	n_a	n_b	N

$$p_1 = a/n_1; \quad q_1 = 1 - p_1;$$

$$p_2 = c/n_2; \quad q_2 = 1 - p_2;$$

$$RR = \frac{a/n_1}{c/n_2} = \frac{p_1}{p_2};$$

$$X^2 = \sum \frac{(|O - E| - .5)^2}{E};$$

$$C_1 = \ln(RR) - 1.96 \sqrt{\frac{b}{an_1} + \frac{d}{cn_2}};$$

$$C_2 = \ln(RR) + 1.96 \sqrt{\frac{b}{an_1} + \frac{d}{cn_2}};$$

$$IC(95\%): 2.71828^{C_1} \leq RR \leq 2.71828^{C_2};$$

IC é válido somente quando:

$$n_1 \overline{p} \overline{q} \geq 5 \text{ e } n_2 \overline{p} \overline{q} \geq 5.$$

22.8.9 – TESTE BINOMIAL: DUAS PROPORÇÕES

$$\hat{p} = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2}, \quad \hat{q} = 1 - \hat{p};$$

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}(1/n_1 + 1/n_2)}};$$

$$p(\text{valor}) = 2[1 - \Phi(Z)], \text{ se } Z \geq 0;$$

$$p(\text{valor}) = 2\Phi(Z), \text{ se } Z < 0.$$

22.8.10 – TESTE t DE STUDENT

n_1 = tamanho da amostra 1; n_2 = tamanho da amostra 2;

$$\overline{X}_1 = \sum X_1 / n_1; \quad \overline{X}_2 = \sum X_2 / n_2;$$

$$s_{dif} = \sqrt{\left(\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \right) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)};$$

$$t = (\overline{X}_1 - \overline{X}_2) / s_{dif};$$

$$gl = n_1 + n_2 - 2.$$

22.8.11 – TESTE Z

n_1 = tamanho da amostra 1; n_2 = tamanho da amostra 2;

$$\overline{X}_1 = \sum X_1 / n_1; \quad \overline{X}_2 = \sum X_2 / n_2;$$

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\hat{\sigma}_{\bar{x}_1 - \bar{x}_2}};$$

$p\text{-valor} = 1 - \Phi(z) \rightarrow$ teste unilateral;

$p\text{-valor} = 2 \times (1 - \Phi(z)) \rightarrow$ teste bilateral

22.8.12 – COEFICIENTES DE VARIAÇÃO

$$Z = \frac{CV_1 - CV_2}{\sqrt{\left(\frac{V_p^2}{n_1 - 1} + \frac{V_p^2}{n_2 - 1}\right)(0.5 + V_p^2)}}; \quad V_p^2 = \frac{(n_1 - 1)CV_1 + (n_2 - 1)CV_2}{n_1 + n_2 - 2};$$

$p\text{-valor} = 2 \times (1 - \Phi(z)).$

22.8.13 – TESTE DE POISSON

$$z = \left| \sqrt{2X_1 + \frac{3}{4}} - \sqrt{2X_2 + \frac{3}{4}} \right|$$

22.9 – DUAS AMOSTRAS RELACIONADAS

22.9.1 – TESTE KAPPA

p_o = probabilidade de concordância observada;

p_e = probabilidade de concordância esperada;

$$K = \frac{p_o - p_e}{1 - p_e};$$

$$se(K) = \sqrt{\frac{1}{N(1 - p_e)^2} \times \left\{ p_e + p_e^2 - \sum_{i=1}^c [a_i b_i (a_i + b_i)] \right\}};$$

$$Z = \frac{K}{se(K)};$$

$p(\text{valor}) = 1 - \Phi(Z).$

22.9.2 – CORRELAÇÃO INTRACLASSE

Coefficiente

$$\rho_I = \sigma_A^2 / (\sigma_A^2 + \sigma_2)$$

Estimaco de ponto

$$\rho_I = [\hat{\sigma}_A^2 / (\sigma_A^2 + \hat{\sigma}^2), 0]$$

Interpretação de acordo com *J. L. Fleiss* (1986):

$\rho_1 < 0.04$ indica pouca reprodutibilidade;

$0.04 \leq \rho_1 < 0.75$ indica boa reprodutibilidade;

$\rho_1 \geq 0.75$ indica excelente reprodutibilidade.

22.9.3 – TESTE DE McNemar

N_A = total de pares discordantes tipo A;

N_D = total de pares discordantes;

Para $N_D \geq 20$:

$$X^2 = \left(\left| N_A - \frac{N_D}{2} \right| - \frac{1}{2} \right)^2 / \left(\frac{N_D}{4} \right);$$

$$p(\text{valor}) = \Pr(\chi_1^2 \geq X^2);$$

Para $N_D < 20$

$$p = 2 \times \sum_{k=0}^{n_A} \binom{n_D}{k} \left(\frac{1}{2} \right)^{n_D} \text{ se } n_A < n_D / 2$$

$$p = 2 \times \sum_{k=n_A}^{n_D} \binom{n_D}{k} \left(\frac{1}{2} \right)^{n_D} \text{ se } n_A > n_D / 2$$

$$p = 1 \text{ se } n_A = n_D / 2$$

22.9.4 – TESTE DOS SINAIS

$$H_0: P(+) = P(-) = \frac{1}{2};$$

$$H_1: P(+) \neq P(-);$$

$$P(k \leq x | n, p) = \sum_{k=0}^x \binom{n}{k} p^k q^{n-k};$$

$$Z = \frac{2x \pm 1 - n}{\sqrt{n}}.$$

22.9.5 – TESTE DE WILCOXON

T = soma dos postos do mesmo sinal;

$$\mu_T = \frac{N(N+1)}{4};$$

$$\sigma_T = \sqrt{\frac{N(N+1)(2N+1)}{24}};$$

$$Z(T) = \frac{T - \mu_T}{\sigma_T};$$

$$p(\text{valor}) = 2 \times [1 - \Phi(T)].$$

22.9.6 – TESTE t DE STUDENT

$$\bar{D} = \sum D / N;$$

$$s_D = \sqrt{\frac{\sum (D - \bar{D})^2}{N - 1}};$$

$$s_{\bar{D}} = \frac{s_D}{\sqrt{N}};$$

$$t = \frac{\bar{D}}{s_{\bar{D}}}; \quad gl = N - 1.$$

22.9.7 – ANOVA (dados pareados)

$a = 2$ (colunas); $b =$ número de itens;

\bar{Y}_{A1} = média da coluna 1; \bar{Y}_{A2} = média da coluna 2;

\bar{Y}_{Bi} = médias aritméticas de cada par de observações;

$$\bar{Y} = \frac{1}{ab} \sum_{i=1}^a \sum_{j=1}^b Y = \text{média geral};$$

$$SS_A = b \sum_{i=1}^a (\bar{Y}_A - \bar{Y})^2; \quad gl_A = 1; \quad QM_A = SS_A / gl_A;$$

$$SS_B = a \sum_{j=1}^b (\bar{Y}_B - \bar{Y})^2; \quad gl_B = b - 1; \quad QM_B = SS_B / gl_B;$$

$$SS_{erro} = \sum_{i=1}^a \sum_{j=1}^b (Y - \bar{Y}_A - \bar{Y}_B + \bar{Y})^2; \quad gl_{erro} = 1 \times (b - 1); \quad QM_{erro} = SS_{erro} / gl_{erro};$$

$$F_A = QM_A / QM_{erro}; \quad F_B = QM_B / QM_{erro}.$$

22.10 – ESTATÍSTICA DESCRITIVA

22.10.1 – AMPLITUDE TOTAL

At (valor máximo – valor mínimo) = amplitude total.

22.10.2 – MÉDIA ARITMÉTICA

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}.$$

22.10.3 – MEDIANA

a) quando o número de observações for ímpar:

$$Md = \text{valor da posição } \left(\frac{n+1}{2} \right);$$

b) quando o número de observações for par:

$$Md = \text{média dos valores das posições } \left(\frac{n}{2} \right) \text{ e } \left(\frac{n}{2} + 1 \right).$$

22.10.4 – VARIÂNCIA

$$s^2 = \frac{\sum (X_i - \bar{X})^2}{n-1} \text{ ou } s^2 = \frac{\sum X_i^2 - (\sum X_i)^2 / n}{n-1}.$$

22.10.5 – DESVIO-PADRÃO

$$s = \sqrt{s^2}.$$

22.10.6 – ERRO-PADRÃO

$$s_{\bar{X}} = \frac{s}{\sqrt{n}}.$$

22.10.7 – COEFICIENTE DE VARIAÇÃO (%)

$$CV = \frac{s}{\bar{X}} \times 100.$$

22.10.8 – MOMENTOS DE ORDEM r CENTRADOS NA MÉDIA

$r = 2, 3 \text{ e } 4$

$$m_2 = \sum (X_i - \bar{X})^2;$$

$$m_3 = \sum (X_i - \bar{X})^3;$$

$$m_4 = \sum (X_i - \bar{X})^4.$$

22.10.9 – COEFICIENTE DE ASSIMETRIA

$$g_1 = \frac{n \times m_3}{(n-1) \times (n-2) \times \sigma^3};$$

$g_1 = 0$: curva simétrica;

$g_1 > 0$: curva assimétrica positiva;

$g_1 < 0$: curva assimétrica negativa.

22.10.10 – COEFICIENTE DE CURTOSE

$$g_2 = \frac{[n \times (n+1) \times m_4 - 3 \times m_2 \times m_2 \times (n-1)]}{[(n-1) \times (n-2) \times (n-3) \times \sigma^4]};$$

$g_2 = 0$: curva normocúrtica;

$g_2 > 0$: curva leptocúrtica;

$g_2 < 0$: curva platicúrtica.

22.10.11 – MÉDIA GEOMÉTRICA

$$GM = \sqrt[n]{\prod_{i=1}^n X_i}, \text{ ou } \ln(G) = \frac{\sum \ln(X_i)}{n} \rightarrow GM = \text{antiln}(G).$$

22.10.12 – VARIÂNCIA GEOMÉTRICA

$$V_1 = \frac{\sum \ln(X_i)}{n}; \quad V_2 = \frac{\sum \ln(X_i) - V_1}{n-1};$$

$$Var_{geométrica} = \frac{\text{antiln}(V_2)}{\ln(10)}.$$

22.10.13 – DESVIO-PADRÃO GEOMÉTRICO

$$D_1 = \sqrt{V_2}; \quad Desvio-padrão_{geométrico} = \text{antiln}(D_1).$$

22.11 – ESTIMAÇÃO DE PARÂMETROS

22.11.1 – ESTIMAÇÃO DA MÉDIA

a) De ponto:

$$\bar{X} = \frac{\sum X_i}{n};$$

b) De intervalos:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}};$$

IC(95%):

para variância paramétrica desconhecida: $\bar{X} \pm t_{.05}(gl) s_{\bar{x}};$

para variância paramétrica conhecida: $\bar{X} \pm 1.96 \sigma_{\bar{x}};$

IC(99%):

para variância paramétrica desconhecida: $\bar{X} \pm t_{.01}(gl) s_{\bar{x}};$

para variância paramétrica conhecida: $\bar{X} \pm 2.58 \sigma_{\bar{x}}.$

22.11.2 – ESTIMAÇÃO DA PROPORÇÃO

k = número de eventos favoráveis; n = tamanho da amostra;

a) De ponto:

$$p = \frac{k}{n};$$

b) De intervalos:

$$\hat{\sigma}_p = \sqrt{\frac{p(1-p)}{n}};$$

IC(95%):

$$p \pm 1.96 \hat{\sigma}_p;$$

IC(99%):

$$p \pm 2.58 \hat{\sigma}_p.$$

22.11.3 – ESTIMAÇÃO DA DIFERENÇA ENTRE DUAS MÉDIAS

a) De ponto

$$\bar{X}_A = \frac{\sum X_i}{n_A}; \quad \bar{X}_B = \frac{\sum X_j}{n_B};$$

$$E = (\bar{X}_A - \bar{X}_B);$$

b) De intervalos:

$$\sigma_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \quad \text{ou} \quad s_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

IC(95%):

$$(\bar{X}_A - \bar{X}_B) \pm 1.96 \sigma_{(\bar{x}_1 - \bar{x}_2)} \quad \text{ou} \quad (\bar{X}_A - \bar{X}_B) \pm t_{.05}(gl) s_{(\bar{x}_1 - \bar{x}_2)}$$

IC(99%):

$$(\bar{X}_A - \bar{X}_B) \pm 2.58 \sigma_{(\bar{x}_1 - \bar{x}_2)} \text{ ou } (\bar{X}_A - \bar{X}_B) \pm t_{.01}(gl) s_{(\bar{x}_1 - \bar{x}_2)}$$

22.11.4 – ESTIMAÇÃO DA DIFERENÇA ENTRE DUAS PROPORÇÕES

a) De ponto:

p_1 = proporção do evento na amostra 1;

p_2 = proporção do evento na amostra 2;

$$E = (p_1 - p_2);$$

b) De intervalos:

$$\hat{\sigma}_{(p_1 - p_2)} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}};$$

$$IC(95\%): (p_1 - p_2) \pm 1.96 \hat{\sigma}_{(p_1 - p_2)};$$

$$IC(99\%): (p_1 - p_2) \pm 2.58 \hat{\sigma}_{(p_1 - p_2)}.$$

22.11.5 – DENSIDADE E TAMANHO DA POPULAÇÃO

A = número total de quadros da área investigada;

n = número de quadros selecionados de modo randômico;

a = tamanho de cada quadro;

$$\bar{m} = \frac{\sum m_i}{n} = \text{média de indivíduos por quadro};$$

$$\hat{\lambda} = \frac{\bar{m}}{a} = \text{estimador da densidade};$$

$$\hat{V}(\hat{\lambda}) = \frac{\hat{\lambda}}{an};$$

$$\hat{M} = \hat{\lambda}A = \text{população estimada};$$

$$IC(95\%): \hat{M} \pm 1.96 \sqrt{\hat{V}(\hat{M})};$$

$$IC(99\%): \hat{M} \pm 2.58 \sqrt{\hat{V}(\hat{M})}.$$

22.11.6 – FENÓTIPOS E GENÓTIPOS

Sistema ABO:

Fenótipo	Genótipo	Probabilidade	Número observado
<i>O</i>	<i>OO</i>	r^2	(<i>O</i>)
<i>A</i>	<i>AA+AO</i>	$p(p+2r)$	(<i>A</i>)
<i>B</i>	<i>BB+BO</i>	$q(q+2r)$	(<i>B</i>)
<i>AB</i>	<i>AB</i>	$2pq$	(<i>AB</i>)

Função logarítmica de máxima verossimilhança:

$$L = (O) 2 \ln r + (A) \{\ln p + \ln (p + 2r)\} + (B) \{\ln q + \ln (q + 2r)\} + (AB)(\ln p + \ln q).$$

Substituindo:

$$x = p/r; \quad y = q/r;$$

$$1 + x + y = 1/r;$$

Assim,

$$L = -2N \ln(1 + x + y) + \{(A) + (AB)\} \ln x + \{(B) + (AB)\} \ln y + (A) \ln (2 + x) + (B) \ln (2 + y).$$

Estimação inicial:

$$r' = \sqrt{[O]};$$

$$p' = 1 - \sqrt{[B] + [O]};$$

$$q' = 1 - \sqrt{[A] + [O]};$$

$$\text{Total} = 2 - \sqrt{[A] + [O]} - \sqrt{[B] + [O]} + \sqrt{[O]};$$

$$D - 1 - r' - p' - q' = \sqrt{[A] + [O]} + \sqrt{[B] + [O]} - \sqrt{[O]} - 1.$$

Equações de máxima verossimilhança:

$$2 = \frac{p'(2 - p')}{p} + \frac{(p' + D)(p' + D + 2r')}{p + 2r} = \frac{q'(2 - q')}{q} + \frac{(q' + D)(q' + D + 2r')}{q + 2r};$$

sendo:

$$r = (r' + \frac{1}{2}D) (1 + \frac{1}{2}D);$$

$$p = p' (1 + \frac{1}{2}D);$$

$$q = q' (1 + \frac{1}{2}D);$$

$$\text{Total} = (1 + \frac{1}{2}D) (r' + p' + q' + \frac{1}{2}D)$$

Sistema MN:

$$(m) = 2 (MM) + (MN);$$

$$(n) = (MN) + 2(NN);$$

Total = 2G, sendo G = tamanho da amostra;

(*MM*), (*MN*) e (*NN*) correspondem aos genótipos;

Proporções gênicas:

$$m = \{2(MM) + (MN)\}/2G;$$

$$n = \{2(NN) + (MN)\}/2G;$$

Número esperado de genótipos:

$$Gm^2, 2Gmn \text{ e } Gn^2$$

$$\text{Variância} = (mn)/2G$$

$$\text{Desvio padrão} = \sqrt{(mn)/2G}.$$

22.11.7 – **FOURIER (densidade)**

N = número total de indivíduos ou grupos observados;

X_i = distância perpendicular de cada indivíduo ou grupo observado;

W = distância perpendicular máxima;

L = distância total inventariada (em metros);

M = número de termos para análise (iteração);

k = 1, 2, 3, ..., *M*;

$$Z_k = \cos[(k \times 3.14159 \times X_i)/W];$$

$$S_k = \sum_{k=1}^M Z_k; \quad A_k = [2/(N \times W)] \times S_k;$$

$$U = \sum_{k=2}^{M+1} A_k; \quad H = (1/W) + U + A_1;$$

$$D = N \times H / (2 \times L) = \text{densidade};$$

$$V = [1/(N-1)] \times (1/W) \times [A_2 + (2/W) - A_1^2];$$

$$DA = (N \times H) / (2 \times L); \quad DB = (N/N^2) + (V/H^2);$$

$$DD = \sqrt{DA^2 \times DB} : \text{desvio padrão};$$

$$D \times 1000000 = \text{densidade (indivíduos ou grupos por km}^2\text{)};$$

$$IC(95\%): D \pm 1.96 \times DD; \quad IC(99\%): D \pm 2.58 \times DD.$$

22.11.8 – **TAMANHO DA POPULAÇÃO (amostragem simples, método direto)**

t = número de animais capturados, marcados e devolvidos ao meio ambiente (1ª captura);

n = número de animais capturados (2ª captura);

s = número de animais marcados entre os indivíduos da 2ª captura;

$$\hat{p} = \frac{s}{n} = \text{proporção de animais marcados na 2ª captura};$$

$$\hat{N} = \frac{nt}{s} = \text{população estimada}; \quad \hat{\sigma}_p = \sqrt{\frac{t^2 n (n-s)}{s^3}};$$

$$IC (95\%): \hat{N} \pm 1.96 \hat{\sigma}_p; \quad IC (99\%): \hat{N} \pm 2.58 \hat{\sigma}_p.$$

22.11.9 – TAMANHO DA POPULAÇÃO (amostragem simples, método inverso)

t = número de animais capturados, marcados e devolvidos ao meio ambiente (1ª captura);

s = número de animais marcados que devem ser capturados (2ª captura);

n = número de animais capturados entre marcados e não-marcados (2ª captura);

$$\hat{p} = \frac{s}{n} = \text{proporção de animais marcados (2ª captura)};$$

$$\hat{N} = \frac{nt}{s} = \text{população estimada}; \quad \hat{\sigma}_p = \sqrt{\frac{t^2 n (n-s)}{s^2 (s+1)}};$$

$$IC (95\%): \hat{N} \pm 1.96 \hat{\sigma}_p; \quad IC (99\%): \hat{N} \pm 2.58 \hat{\sigma}_p.$$

22.11.10 – TAMANHO DE POPULAÇÃO FECHADA (amostragem múltipla) Método de *Schumacher e Eschmeyer*

\hat{N} = tamanho da população (estimação);

s = número de amostras;

n_i = tamanho de cada amostra;

m_i = número de animais marcados em cada amostra;

M_i = número de animais marcados na população exatamente antes da i -ésima amostra ser tomada;

$$\hat{N} = \frac{\sum_{i=2}^s n_i M_i^2}{\sum_{i=2}^s m_i M_i}$$

$$(s-2) \hat{\sigma}^2 = \sum_{i=2}^s \frac{m_i^2}{n_i} - \frac{\left(\sum_{i=2}^s m_i M_i \right)^2}{\sum_{i=2}^s n_i M_i^2}$$

Intervalo de Confiança:

$$IC = \frac{\sum_{i=2}^s n_i M_i^2}{\sum_{i=2}^s m_i M_i \pm t_{s-2}(\alpha/2) \left(\tilde{\sigma}^2 \sum_{i=2}^s n_i M_i^2 \right)^{1/2}}$$

22.11.11 – TAMANHO DE POPULAÇÃO ABERTA (AMOSTRAGEM MÚLTIPLA) Método de *Parker*

$$\hat{N}_o = M_o / \hat{\beta}_o$$

$$IC = M_o / (\hat{\beta}_o \pm t_{n-2}(\alpha/2) \times \sigma \beta_o)$$

22.12 – META-ANÁLISE

22.12.1 – *DerSimonian-Laird*

$$w = \sum_{i=1}^k w_i k ; \quad s_w^2 = \frac{1}{k-1} \left(\sum_{i=1}^k w_i^2 - k \bar{w}^2 \right)$$

$$U = (k-1) \left(\bar{w} - \frac{s_i^2}{k \bar{w}} \right); \quad \hat{\tau}^2 = 0 \text{ if } Q \leq k-1$$

$$\hat{\tau} = (Q - (k-1)) / U \text{ if } Q > k-1$$

$$w_i^* = \frac{1}{\left[(1/w_i) + \hat{\tau}^2 \right]}; \quad \bar{T}_{.RND} = \sum_{i=1}^k w_i^* T_i / \sum_{i=1}^k w_i^*$$

$$\text{var} = (\bar{T}_{.RND}) = 1 / \sum_{i=1}^k w_i^*$$

$$\bar{T}_{.RND} - z_{\alpha/2} / \sqrt{\sum_{i=1}^k w_i^*} \leq \bar{\theta} \leq \bar{T}_{.RND} + z_{\alpha/2} / \sqrt{\sum_{i=1}^k w_i^*}$$

22.12.2 – *MANTEL-HAENSZEL*

Tabelas de Contingência 2 x 2:

Teste	Exposto	Não-exposto	Total
+	a_i	b_i	$a_i + b_i$
-	c_i	d_i	$c_i + d_i$
Total	$a_i + c_i$	$b_i + d_i$	n_i

$$E = \sum_{i=1}^k \frac{(a_i + b_i)(a_i + c_i)}{n_i};$$

$$V = \sum_{i=1}^k \frac{(a_i + b_i)(c_i + d_i)(a_i + c_i)(b_i + d_i)}{n_i^2(n_i - 1)};$$

$$X_{MH}^2 = \frac{(|O - E| - 0.5)^2}{V}; \quad gl = 1.$$

22.12.3 – P-VALOR (Fisher)

$$-2 \sum_{i=1}^k \log(p_i);$$

Comparar o valor crítico da distribuição χ^2 com $2k$ graus de liberdade.

22.12.4 – QUI-QUADRADO (aderência) para k amostras

$gl_a = k - 1$, para o qui-quadrado de cada amostra;

$gl_t = \sum gl_a$, para o qui-quadrado total;

$gl_{ho} = k - 1$, para o qui-quadrado de homogeneidade;

$gl_{he} = gl_t - gl_{ho}$, para o qui-quadrado de heterogeneidade.

22.12.5 – QUI-QUADRADO (independência): para k tabelas 2 x 2

	Doente	Não-doente	Total
Exposto	n_{11}	n_{12}	$n_{1\bullet}$
Não-exposto	n_{21}	n_{22}	$n_{2\bullet}$
Total	$n_{\bullet 1}$	$n_{\bullet 2}$	$n_{\bullet\bullet}$

$$s_i = \sqrt{1/n_{11} + 1/n_{12} + 1/n_{21} + 1/n_{22}};$$

$$X_A^2 = \frac{\left(\left| \sum_{i=1}^k n_{11}(i) - \sum_{i=1}^k n_{1\bullet}(i)n_{\bullet 1}(i)/n_{\bullet\bullet}(i) \right| - 1/2 \right)^2}{\sum_{i=1}^k \left(\frac{n_{1\bullet}(i)n_{2\bullet}(i)n_{\bullet 1}(i)n_{\bullet 2}(i)}{n_{\bullet\bullet}(i)^2(n_{\bullet\bullet}(i) - 1)} \right)};$$

$$X_H^2 = \sum \frac{a_i^2}{s_i^2} - \chi_A^2; \quad gl_A = 1; \quad gl_H = k - 1.$$

22.12.6 – REGRESSÃO (w): para n amostras

$$\Sigma x^2 = \Sigma X^2 - (\Sigma X)^2/n; \quad \Sigma y^2 = \Sigma Y^2 - (\Sigma Y)^2/n;$$

$$\Sigma xy = \Sigma XY - (\Sigma X)(\Sigma Y)/n;$$

$$b_w = \frac{\sum_{i=1}^n xy}{\sum_{i=1}^n x^2};$$

$$SS_1 = \sum_{i=1}^n (Y - Y')^2, \text{ com } gl_1 = k(n - 2);$$

$$SS_2 = \sum y^2 - \frac{(\sum xy)^2}{\sum x^2}, \text{ com } gl_2 = k(n - 1) - 1;$$

$$SS_3 = SS_2 - SS_1, \text{ com } gl_3 = k - 1;$$

Quando o tamanho das amostras são desiguais:

$$gl_1 \text{ de } SS_1 = \sum_{i=1}^k n_i - 2k; \quad gl_2 \text{ de } SS_2 = \sum_{i=1}^k n_i - k - 1; \quad F = \frac{SS_3 / gl_3}{SS_1 / gl_1}.$$

22.12.7 – RISCO RELATIVO (w): para k tabelas 2 x 2

$$RR_{MH} = \frac{\sum_{i=1}^k \left(\frac{b_i + d_i}{n_i} \right) a_i}{\sum_{i=1}^k \left(\frac{a_i + c_i}{n_i} \right) b_i}; \quad n_i = a_i + b_i + c_i + d_i$$

22.12.8 – VÁRIAS CORRELAÇÕES (rw combinado)

n_i = tamanho da amostra i ;

r_i = coeficiente de correlação linear da amostra i ;

$$z_i = \frac{1}{2} [\ln(1 + r_i) - \ln(1 - r_i)];$$

$$X^2 = \sum_{i=1}^k (n_i - 3) z_i^2 - \frac{\left[\sum_{i=1}^k (n_i - 3) z_i \right]^2}{\sum_{i=1}^k (n_i - 3)}; \quad gl = k - 1.$$

22.12.9 – VÁRIAS PROPORÇÕES (pw combinado)

n_i = tamanho de cada amostra i ; x_i = número de eventos de cada amostra i ;

$$\hat{p} = \frac{\sum_{i=1}^k x_i}{\sum_{i=1}^k n_i}; \quad \hat{q} = 1 - \hat{p}; \quad \chi^2 = \sum_{i=1}^k \frac{(x_i - n_i \hat{p})^2}{n_i \hat{p}(1 - \hat{p})}; \quad gl = k - 1;$$

$$p'_i = 0.5 \left[\arcsen \sqrt{X/(n+1)} + \arcsen \sqrt{(X+1)/(n+1)} \right];$$

$$\text{Desvio Padrão} = \sqrt{\frac{410.35}{n_i + 0.5} + \frac{410.35}{n_j + 0.5}}.$$

22.13 – NORMALIDADE**22.13.1 – D'AGOSTINO**

$$D = \frac{\sum_{i=1}^n (i - n + 1/2) x(i)}{\sqrt{n^3 \sum_{i=1}^n (x_i - \bar{x})^2}};$$

Valores críticos: tabelados.

22.13.2 – D'AGOSTINO-PEARSON

$$K^2 = Z_{g1}^2 + Z_{g2}^2$$

$$gl = 2$$

22.13.3 – KOLMOGOROV-SMIRNOV

k = número de postos; N = tamanho da amostra;

μ = média da população; σ = desvio padrão paramétrico;

$Fo(X)$ = distribuição teórica acumulada sob H_0 ;

$Sn(X)$ = distribuição amostral acumulada;

D = máximo $|Fo(X) - Sn(X)|$ = desvio máximo;

$$D \geq \frac{1.36}{\sqrt{N}}, \text{ para } \alpha = 0.05; \quad D \geq \frac{1.63}{\sqrt{N}}, \text{ para } \alpha = 0.01.$$

22.13.4 – LILLIEFORS

k = número de postos; N = tamanho da amostra;

Média e desvio padrão amostrais;

$Fo(X)$ = distribuição teórica acumulada sob H_0 ;

$Sn(X)$ = distribuição amostral acumulada;

$D = \text{máximo } |Fo(X) - Sn(X)| = \text{desvio máximo};$

$$D \geq \frac{0.886}{\sqrt{N}}, \text{ para } \alpha = 0.05;$$

$$D \geq \frac{1.031}{\sqrt{N}}, \text{ para } \alpha = 0.01.$$

22.13.5 – SHAPIRO-WILK

$$V = \text{cov}(x_i, x_j); \quad m' = (m_1 + m_2, \dots, m_n)$$

$$\bar{y} = \frac{\sum y_i}{n}; \quad \hat{\sigma} = \frac{m' V^{-1} y}{m' V^{-1} m};$$

$$R^2 = m' V^{-1} m; \quad C^2 = m' V^{-1} V^{-1} m;$$

$$S^2 = \sum_1^n (y - \bar{y})^2; \quad b = R^2 \hat{\sigma} / C$$

$$W = \frac{b^2}{C^2 S^2}$$

22.14 – QUI-QUADRADO

22.14.1 – AMOSTRAS INDEPENDENTES

$r = \text{número de linhas}; \quad k = \text{número de colunas};$

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}};$$

$$\chi^2_{Yates} = \sum_{i=1}^r \sum_{j=1}^k \frac{(|O_{ij} - E_{ij}| - 0.5)^2}{E_{ij}};$$

$gl = (r - 1)(k - 1); \quad \text{Correção de Yates: somente quando } gl = 1.$

22.14.2 – RESÍDUOS L x C

Resíduo Padronizado

Resíduo Ajustado

$$R_{\text{pad}} = \frac{O - E}{E}$$

$$R_{\text{ajus}} = \frac{R_{\text{pad}}}{\sqrt{\left(1 - \frac{TC}{TG}\right)\left(1 - \frac{TL}{TG}\right)}}$$

TC = Total da Coluna

TL = Total da Linha

TG = Total Geral

Valor crítico : 1.96, para alfa = 0.05

Valor crítico : 2.576, para alfa = 0.01

22.14.3 – PARTIÇÃO $R \times C$

Partições:

 R = linha; C = coluna;

Exemplo:

Tabela $3 \times 4 = 3R \times 4C$

N_{11}	n_{12}	n_{13}	n_{14}
N_{21}	n_{22}	n_{23}	n_{24}
N_{31}	n_{32}	n_{33}	n_{34}

Modelo para a partição:

A	B
C	D

1a. partição $\rightarrow 2R:2C$ $A = n_{11}$; $B = n_{12}$; $C = n_{21}$; $D = n_{22}$.**2a. partição** $\rightarrow 3R:2C$ $A = n_{11} + n_{21}$; $B = n_{12} + n_{22}$; $C = n_{31}$; $D = n_{32}$.**3a. partição** $\rightarrow 2R:3C$ $A = n_{11} + n_{12}$; $B = n_{13}$; $C = n_{21} + n_{22}$; $D = n_{23}$.**4a. partição** $\rightarrow 3R:3C$ $A = n_{11} + n_{12} + n_{21} + n_{22}$; $B = n_{13} + n_{23}$; $C = n_{31} + n_{32}$; $D = n_{33}$.**5a. partição** $\rightarrow 2R:4C$ $A = n_{11} + n_{12} + n_{13}$; $B = n_{14}$; $C = n_{21} + n_{22} + n_{23}$; $D = n_{24}$.**6a. partição** $\rightarrow 3R:4C$ $A = n_{11} + n_{12} + n_{13} + n_{21} + n_{22} + n_{23}$; $B = n_{14} + n_{24}$; $C = n_{31} + n_{32} + n_{33}$; $D = n_{34}$. gl (para cada partição) = 1.

22.14.4 – TENDÊNCIA: TABELAS DE CONTINGÊNCIA 2 x k

k = número de grupos (colunas); x_i = sucessos no grupo i ;
 n_i = número de unidades no grupo i ; $\hat{p}_i = x_i/n_i$ = proporção de sucessos no grupo i ;
 x = número total de sucessos em todos os grupos; n = número total de unidades;
 $\bar{p} = x/n$ = proporção geral de sucessos; $\bar{q} = 1 - \bar{p}$ = proporção geral de insucessos;
 S_i = escore da variável do grupo i ;

$$A = \left(\sum_{i=1}^k x_i S_i \right) - x \bar{S}; \quad B = \bar{p} \bar{q} \left[\left(\sum_{i=1}^k n_i S_i^2 \right) - \left(\sum_{i=1}^k n_i S_i \right)^2 / n \right];$$

Quando $A > 0$, tendência crescente;

Quando $A < 0$, tendência decrescente;

$X_1^2 = A^2/B$; $gl = 1$; Deve ser usado somente quando $n\bar{p}\bar{q} \geq 5$.

22.15 – REGRESSÃO

22.15.1 – AJUSTAMENTO DE CURVAS

Equações:

$$\hat{Y} = a + bX : \text{Linear};$$

$$\hat{Y} = ae^{bX} (a > 0) : \text{Exponencial};$$

$$\hat{Y} = a + b \ln(X) : \text{Logarítmica};$$

$$\hat{Y} = aX^b (a > 0) : \text{Geométrica};$$

Coefficientes:

$$b = \frac{\sum (XY) - (\sum X)(\sum Y)/n}{\sum X^2 - (\sum X)^2/n};$$

$$a = \bar{Y} - b\bar{X};$$

$$R^2 = \frac{a\sum Y + b\sum XY - (\sum Y)^2/n}{\sum Y^2 - (\sum Y)^2/n};$$

Variáveis:

Linear:

$$X_i = X_i; \quad Y_i = Y_i; \quad a = a.$$

Exponencial:

$$X_i = X_i; \quad Y_i = \ln(Y_i); \quad a = \ln(a).$$

Logarítmica:

$$X_i = \ln(X_i); \quad Y_i = Y_i; \quad a = a.$$

Geométrica:

$$X_i = \ln(X_i); \quad Y_i = \ln(Y_i); \quad a = \ln(a).$$

Médias:

$$\text{Linear: } \bar{X} = \Sigma X/n; \quad \bar{Y} = \Sigma Y/n;$$

$$\text{Exponencial: } \bar{X} = \Sigma X/n; \quad Gm(Y) = \text{antiln} [\Sigma \ln(Y)/n];$$

$$\text{Logarítmica: } Gm(X) = \text{antiln} [\Sigma \ln(X)/n]; \quad \bar{Y} = \Sigma Y/n;$$

$$\text{Geométrica: } Gm(X) = \text{antiln} [\Sigma \ln(X)/n]; \quad Gm(Y) = \text{antiln} [\Sigma \ln(Y)/n].$$

22.15.2 – REGRESSÃO LINEAR

$$\bar{X} = \frac{\Sigma X_i}{n}; \quad \bar{Y} = \frac{\Sigma Y_i}{n};$$

$$b_Y = \frac{\Sigma XY - (\Sigma X)(\Sigma Y)/n}{\Sigma X^2 - (\Sigma X)^2/n};$$

$$a = \bar{Y} - b_Y \bar{X}; \quad \hat{Y} = a + b_Y X;$$

$$\Sigma x^2 = \Sigma X^2 - (\Sigma X)^2/n; \quad \Sigma y^2 = \Sigma Y^2 - (\Sigma Y)^2/n;$$

$$\Sigma xy = \Sigma XY - (\Sigma X)(\Sigma Y)/n;$$

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}}; \quad r^2 = (r)^2;$$

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}; \quad gl = n-2.$$

22.15.3 – REGRESSÃO MÚLTIPLA

Modelo de Regressão Múltipla:

$$y = \alpha + \sum_{j=1}^k \beta_j x_j + \varepsilon, \text{ sendo } k \text{ o número de variáveis independentes.}$$

Os parâmetros $\alpha, \beta_1, \beta_2, \dots, \beta_k$ são estimadas pelos coeficientes a, b_1, b_2, \dots, b_k , pelo método dos mínimos quadrados. Matrizes:

$$X = \begin{bmatrix} 1 & X_{11} & X_{12} & \Lambda & X_{1m} \\ 1 & X_{21} & X_{22} & \Lambda & X_{2m} \\ \text{M} & \text{M} & \text{M} & & \text{M} \\ 1 & X_{n1} & X_{n2} & \Lambda & X_{nm} \end{bmatrix}; \quad Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \text{M} \\ Y_n \end{bmatrix}; \quad \mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ \text{M} \\ b_{p-1} \end{bmatrix};$$

$$X'X = \begin{bmatrix} n & \sum x_1 & \sum x_2 & \Lambda & \sum x_m \\ \sum x_1 & \sum x_1^2 & \sum x_1 x_2 & \Lambda & \sum x_1 x_m \\ \sum x_2 & \sum x_2 x_1 & \sum x_2^2 & \Lambda & \sum x_2 x_m \\ \text{M} & \text{M} & \text{M} & & \text{M} \\ \sum x_m & \sum x_m x_1 & \sum x_m x_2 & \Lambda & \sum x_m^2 \end{bmatrix};$$

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1} (\mathbf{X}'\mathbf{Y});$$

$p \times 1 \quad p \times p \quad p \times 1$

$$a = \bar{Y} - b_1 \bar{X}_1 - b_2 \bar{X}_2 - \dots - b_{p-1} \bar{X}_{p-1}; \quad \hat{y}_i = a + \sum_{j=1}^k b_j x_{ij};$$

Teste F :

$$H_0: b_1 = b_2 = b_3 = \dots = b_{p-1} = 0; \quad H_1: b_i \neq 0;$$

$$SQ_{total} = \sum_{i=1}^n (y_i - \bar{y})^2; \quad SQ_{erro} = \sum_{i=1}^n (y_i - \hat{y})^2;$$

$$SQ_{regressão} = SQ_{total} - SQ_{erro};$$

$$gl_{regressão} = k; \quad gl_{erro} = (n - k - 1);$$

$$QM_{regressão} = \frac{SQ_{regressão}}{gl_{regressão}}; \quad QM_{erro} = \frac{SQ_{erro}}{gl_{erro}};$$

$$F_{regressão} = \frac{QM_{regressão}}{QM_{erro}}; \quad F(gl) = k, n - k - 1.$$

22.15.4 – REGRESSÃO LOGÍSTICA

$$\text{logit}(p) = \ln[p/(1-p)];$$

$$\text{logit}(p) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$$

$$p = \frac{1}{1 + e^{-(\alpha + \beta_1 x_1 + \dots + \beta_k x_k)}};$$

$$\text{logit}(p_A) = \alpha + \beta_1 x_1 + \dots + \beta_{j-1} x_{j-1} + \beta_j(\mathbf{1}) + \beta_{j+1} x_{j+1} + \dots + \beta_k x_k$$

$$\text{logit}(p_B) = \alpha + \beta_1 x_1 + \dots + \beta_{j-1} x_{j-1} + \beta_j(\mathbf{0}) + \beta_{j+1} x_{j+1} + \dots + \beta_k x_k$$

$$\beta_j = \ln \left[\frac{p_A / (1 - p_A)}{p_B / (1 - p_B)} \right];$$

$$OR = e^{\beta_j}; \quad z = \frac{\beta_j}{SE(\beta_j)}$$

$$p\text{-valor} = 2 \times [1 - \Phi(z)], \text{ quando } z \geq 0;$$

$$p\text{-valor} = 2 \times \Phi(z), \quad \text{quando } z < 0.$$

22.15.5 – POLINOMIAL

$\hat{Y} = b_0 + b_1X + b_2X^2 + \dots + b_pX^p$, sendo p o número de termos da regressão.

Análise de variância:

Modelos de Regressão

i) Linear:

$$\hat{Y} = a + b_1X;$$

$$\text{Estatística } F = \frac{R_1^2}{(1 - R_1^2)/(N - 2)}$$

$$GL \text{ regressão} = 1$$

$$GL \text{ erro} = N - 2$$

ii) Quadrática:

$$\hat{Y} = a + b_1X + b_2X^2;$$

$$\text{Estatística } F = \frac{R_2^2 / 2}{(1 - R_2^2)/(N - 3)};$$

$$GL \text{ regressão} = 2$$

$$GL \text{ erro} = N - 3$$

iii) Cúbica:

$$\hat{Y} = a + b_1X + b_2X^2 + b_3X^3;$$

$$\text{Estatística } F = \frac{R_3^2 / 3}{(1 - R_3^2)/(N - 4)};$$

$$GL \text{ regressão} = 3$$

$$GL \text{ erro} = N - 4$$

iv) Quártica:

$$\hat{Y} = a + b_1X + b_2X^2 + b_3X^3 + b_4X^4;$$

$$\text{Estatística } F = \frac{R_4^2 / 4}{(1 - R_4^2) / (N - 5)};$$

$$\begin{aligned} GL \text{ regressão} &= 4 \\ GL \text{ erro} &= N - 5 \end{aligned}$$

22.16 – TAMANHO DAS AMOSTRAS

22.16.1 – TESTE t : UMA AMOSTRA

a) Teste unilateral:

$$n = \frac{\sigma^2 (z_{1-\alpha} + z_{1-\beta})^2}{(\mu_0 - \mu_1)^2};$$

b) Teste bilateral:

$$n = \frac{\sigma^2 (z_{1-\alpha/2} + z_{1-\beta})^2}{(\mu_0 - \mu_1)^2}.$$

22.16.2 – TESTE t : DUAS AMOSTRAS PAREADAS

a) Teste unilateral:

$$n = \frac{s_d^2 (z_{1-\alpha} + z_{1-\beta})^2}{(\bar{D})^2};$$

b) Teste bilateral:

$$n = \frac{s_d^2 (z_{1-\alpha/2} + z_{1-\beta})^2}{(\bar{D})^2}.$$

22.16.3 – TESTE t : DUAS AMOSTRAS INDEPENDENTES

a) Teste unilateral:

$$n = \frac{(\sigma_1^2 + \sigma_2^2) (z_{1-\alpha} + z_{1-\beta})^2}{(\mu_2 - \mu_1)^2}, \text{ para cada amostra};$$

b) Teste bilateral:

$$n = \frac{(\sigma_1^2 + \sigma_2^2) (z_{1-\alpha/2} + z_{1-\beta})^2}{(\mu_2 - \mu_1)^2}, \text{ para cada amostra.}$$

22.16.4 – TESTE DE McNEMAR

$$n = \frac{\left(z_{1-\alpha/2} + 2z_{1-\beta}\sqrt{p_A q_A}\right)^2}{4(p_A - 0.5)^2 p_D} = \text{número de pares};$$

$2n$ = número de indivíduos;

p_A = pares discordantes tipo A;

p_D = total de todos os pares discordantes.

22.16.5 – PROPORÇÃO BINOMIAL: UMA AMOSTRA

a) Teste unilateral:

$$n = \frac{p_o q_o \left(z_{1-\alpha} + z_{1-\beta} \sqrt{\frac{p_1 q_1}{p_o q_o}}\right)^2}{(p_1 - p_o)^2};$$

b) Teste bilateral:

$$n = \frac{p_o q_o \left(z_{1-\alpha/2} + z_{1-\beta} \sqrt{\frac{p_1 q_1}{p_o q_o}}\right)^2}{(p_1 - p_o)^2}.$$

22.16.6 – PROPORÇÃO BINOMIAL: DUAS AMOSTRAS

a) Teste unilateral:

$$n_1 = \left[\sqrt{\bar{p}\bar{q}} \left(1 + \frac{1}{k}\right) z_{1-\alpha} + \sqrt{p_1 q_1 + \frac{p_2 q_2}{k}} z_{1-\beta} \right]^2 / (p_2 - p_1)^2;$$

b) Teste bilateral:

$$n_1 = \left[\sqrt{\bar{p}\bar{q}} \left(1 + \frac{1}{k}\right) z_{1-\alpha/2} + \sqrt{p_1 q_1 + \frac{p_2 q_2}{k}} z_{1-\beta} \right]^2 / (p_2 - p_1)^2.$$

$n_2 = n_1 k$;

$k = 0.5, 1, 2$, etc. = vezes que a segunda amostra deve ser em relação à primeira.

22.16.7 – CORRELAÇÃO LINEAR

$$z_r = 0.5 \times \ln \left(\frac{1+r}{1-r} \right);$$

$$n = \left[\frac{z_{1-\alpha/2} + z_{1-\beta}}{z_r} \right]^2 + 4.$$

22.17 – TRANSFORMAÇÃO DE DADOS

22.17.1 – BOX-COX

$$Y'' = (Y^\lambda - 1)/\lambda \quad (\text{para } \lambda \neq 0);$$

$$Y'' = \ln Y \quad (\text{para } \lambda = 0);$$

$\lambda = 1 \rightarrow$ transformação linear;

$\lambda = 1/2 \rightarrow$ transformação em raiz quadrada;

$\lambda = 0 \rightarrow$ transformação logarítmica;

$\lambda = -1 \rightarrow$ transformação recíproca;

$\hat{\lambda}$ é um valor maximizado por procedimento iterativo pela equação:

$$L = -\frac{v}{2} \ln s_T^2 + (\lambda - 1) \frac{v}{n} \sum \ln Y;$$

v = graus de liberdade;

s_T^2 = variância do valor de Y transformado;

n = tamanho da amostra.

22.18 – DISTRIBUIÇÃO CIRCULAR

22.18.1 – ESTATÍSTICA DESCRITIVA

$$Y = \frac{\sum \text{sen } a_i}{n}; \quad X = \frac{\sum \text{cos } a_i}{n}$$

$$r = \sqrt{X^2 + Y^2}$$

$$\text{cos } \bar{a} = \frac{X}{r}; \quad \text{sen } \bar{a} = \frac{Y}{r}$$

Ângulo Médio: arcos ou arcsen de \bar{a}

$$s^2 = 2(1 - r)$$

$$s = \frac{180}{\pi} \sqrt{2(1-r)} \quad ; \quad R = nr$$

$$d = \arccos = \frac{\sqrt{\frac{2n(2R^2 - 0nX_{\alpha,1}^2)}{4n - X_{\alpha,1}^2}}}{R}$$

$$\text{Intervalo de Confiança (95\%)} = \bar{a} \pm d$$

22.18.2 – TESTES DE HIPÓTESES

Teste de *Rayleigh*:

$$R^2 = nr;$$

$$z = \frac{R^2}{n};$$

p-valor = Tabelado em função de *n*

$$P = \exp \left[\sqrt{1 + 4n + 4(n^2 - R^2)} - (1 + 2n) \right]$$

Teste V de *Rayleigh*:

$$V = R \cos(\bar{a} - \mu_o)$$

$$u = V \sqrt{\frac{2}{n}}$$

Teste de *Watson-Williams* para duas (2) Amostras Independentes:

$$F = K \frac{(N-2)(R_1 + R_2 - R)}{N - R_1 - R_2}, \text{ onde } R = n_1 + n_2$$

Teste de *Watson-Williams* para três (3) Amostras Independentes:

$$F = K \frac{(N - k) \left(\sum_{j=1}^k R_j - R \right)}{(k - 1) \left(N - \sum_{j=1}^k R_j \right)}$$

Teste para 2 Amostras Pareadas (paramétrico):

$$X_j = \cos a_{2j} - \cos a_{1j}; \quad Y_j = \sin a_{2j} - \sin a_{1j}$$

$$\sum xy = \sum X_j Y_j - \frac{\sum X_j \sum Y_j}{k}$$

$$F = \frac{k(k-2)}{2} \left[\frac{\bar{X}^2 \sum y^2 - 2 \bar{X} \bar{Y} \sum xy + \bar{Y}^2 \sum x^2}{\sum x^2 \sum y^2 - (\sum xy)^2} \right]$$

Teste para Correlação:

$$r_{aa} = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n \sin(a_i - a_j) \sin(b_i - b_j)}{\sqrt{\sum_{i=1}^{n-1} \sum_{j=i+1}^n \sin^2(a_i - a_j) \sum_{i=1}^{n-1} \sum_{j=i+1}^n \sin^2(b_i - b_j)}}$$

CAPÍTULO 23

ESCOLHA DO TESTE ESTATÍSTICO – SUGESTÕES

23.1 – INTRODUÇÃO

A escolha do teste adequado aos dados obtidos em uma investigação é um passo que deve ser decidido no planejamento de uma pesquisa. A experiência acumulada em trabalhos anteriores, a indicação do orientador, quando é o caso, e o conhecimento pleno da Estatística propiciam elementos capazes de se efetuar a escolha do teste mais ajustado ao estudo em questão.

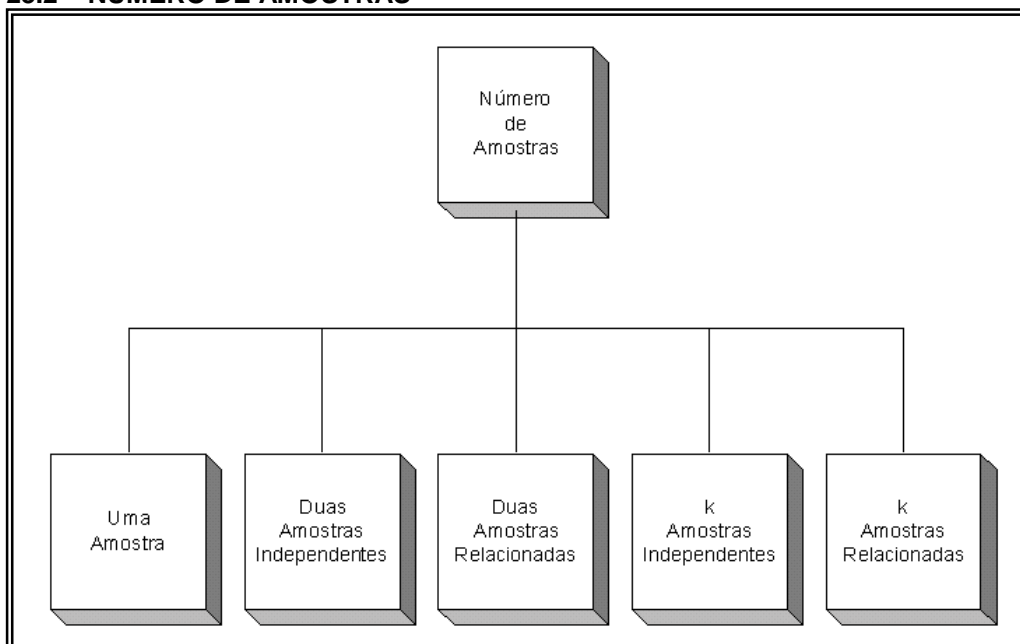
As sugestões oferecidas pelo *BioEstat* não esgotam todas as possibilidades na escolha de um teste estatístico, destinando-se, sobretudo, aos iniciantes em investigações científicas, mas deve ser precedida do planejamento da pesquisa, com formulação da hipótese a ser testada, da cuidadosa definição do universo de onde a(s) amostra(s) será(ão) coletada(s), do estabelecimento do nível de significância do teste (erro α), do levantamento ou do experimento a ser efetuado, ou seja, de todas as etapas que possibilitem, ao final, comprovar ou refutar a hipótese formulada, prevendo-se, nessa fase, o teste estatístico que atenda ao objetivo do trabalho proposto.

Os fluxogramas apresentados em seguimento são auto-explicativos, obedecendo os seguintes passos:

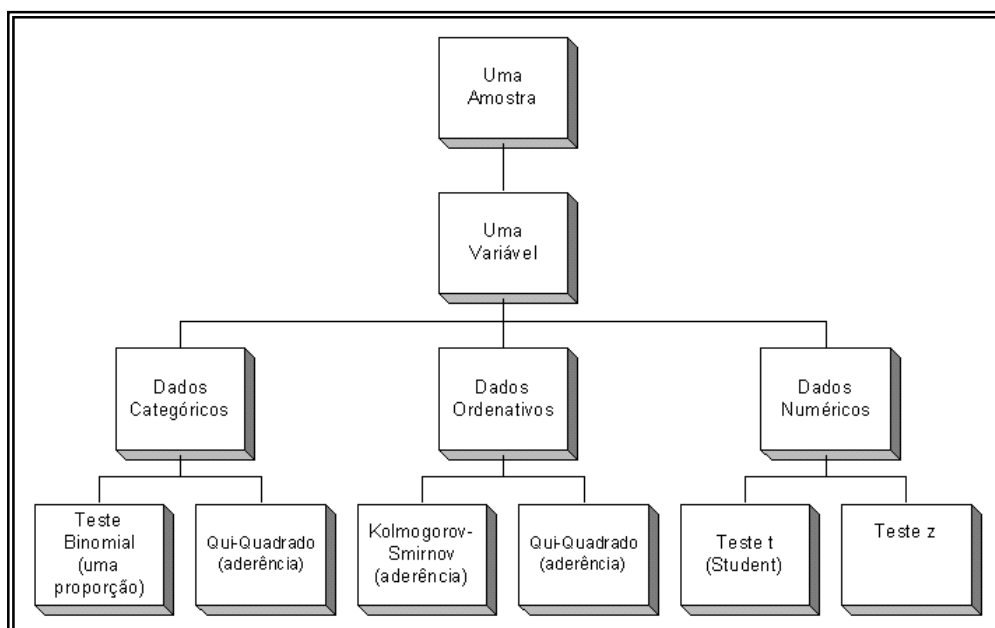
- a) Número de amostras:
 - Uma amostra;
 - Duas amostras independentes;
 - Duas amostras relacionadas;
 - k amostras independentes;
 - k amostras relacionadas.
- b) Número de variáveis:
 - Uma variável;
 - Duas variáveis;
 - k variáveis.
- c) Tipo de dados:
 - Dados categóricos;
 - Dados ordenativos;
 - Dados numéricos.

Os dados categóricos e ordenativos são aqueles mensurados em escala nominal e ordinal, respectivamente, e estão geralmente relacionados a testes de distribuição livre, abrangendo os chamados testes não-paramétricos ou estatísticas firmes, de largo emprego nas áreas de biologia e de medicina. Os dados numéricos, mensurados em escala intervalar ou de razões, dizem respeito a modelos de distribuição bem definidos, como os referentes à distribuição normal.

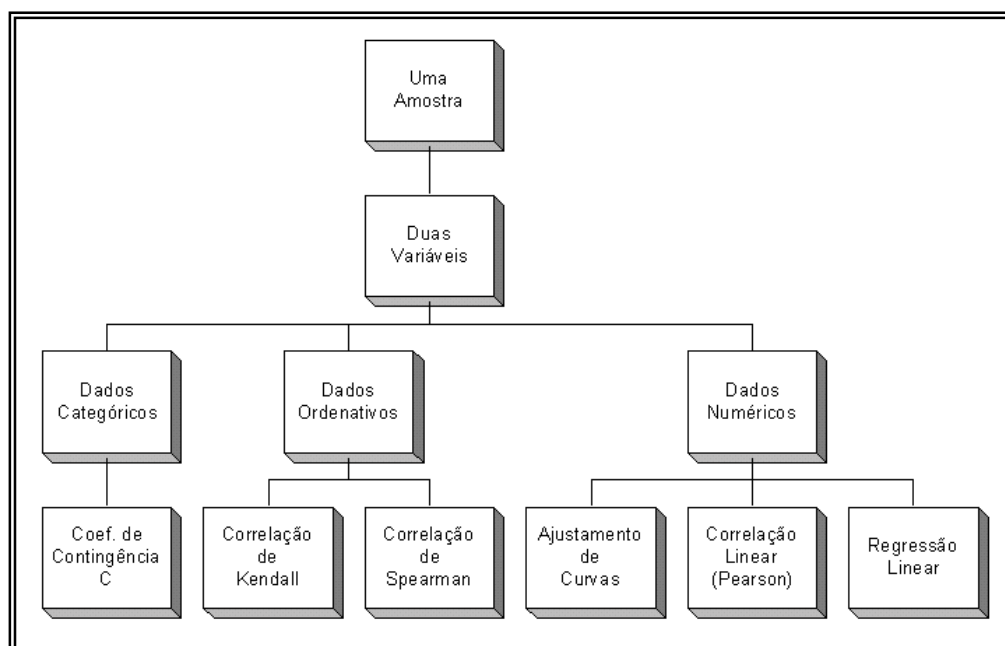
23.2 – NÚMERO DE AMOSTRAS



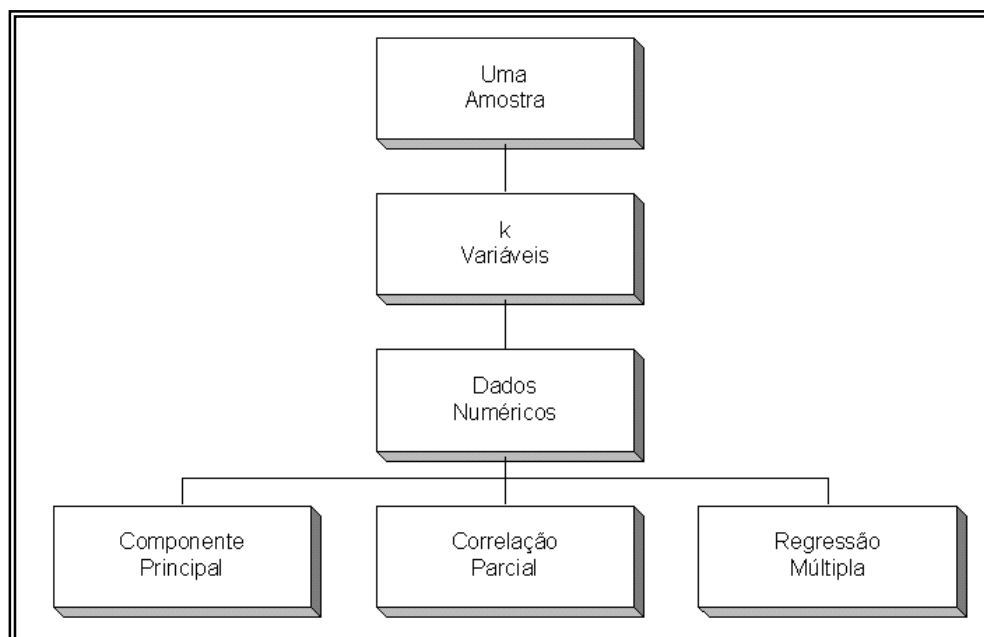
23.3a – UMA AMOSTRA: UMA VARIÁVEL

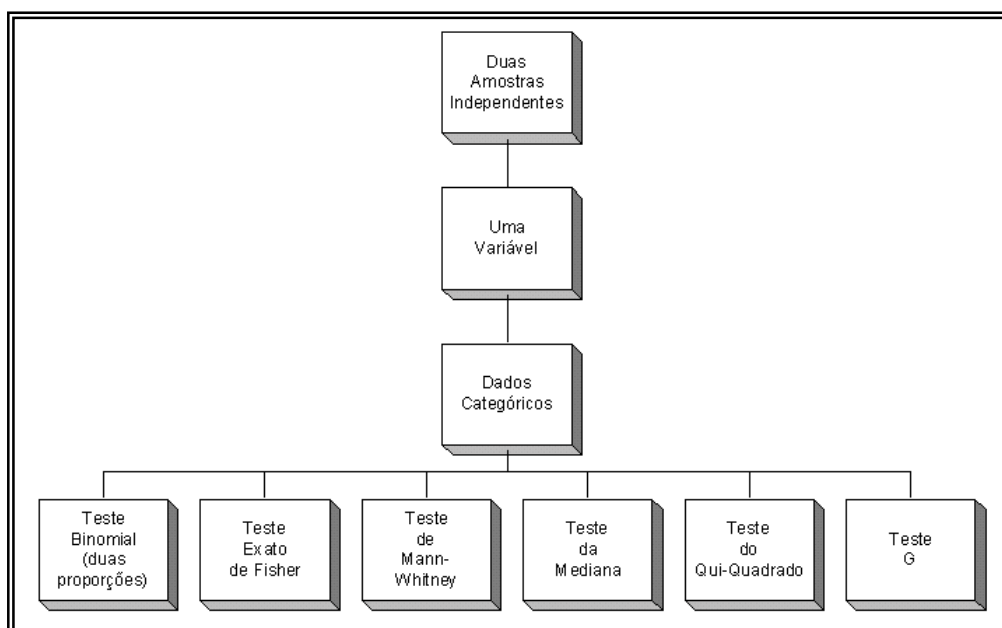
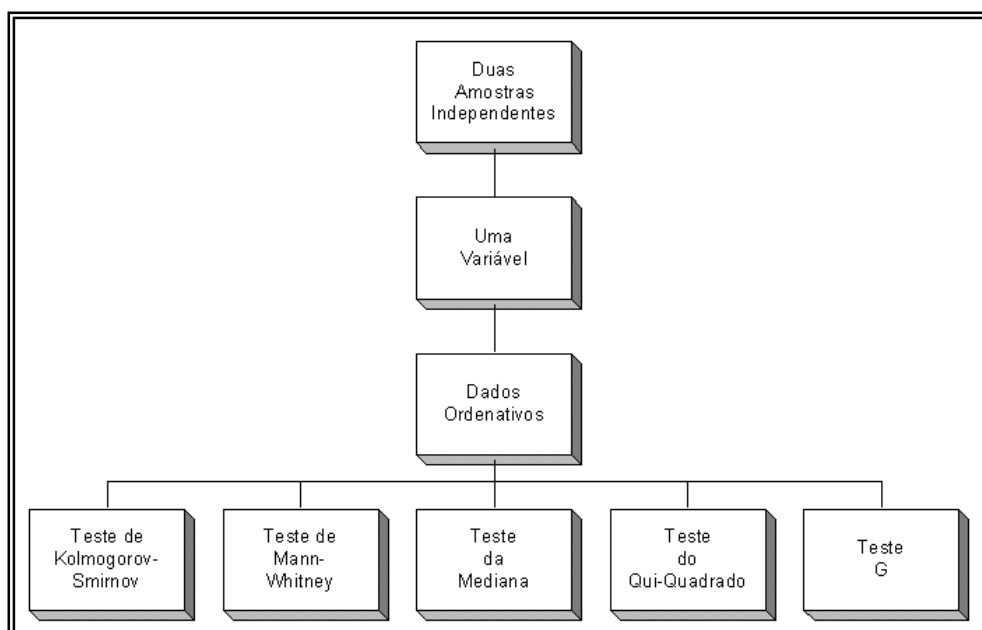


23.3b – UMA AMOSTRA: DUAS VARIÁVEIS

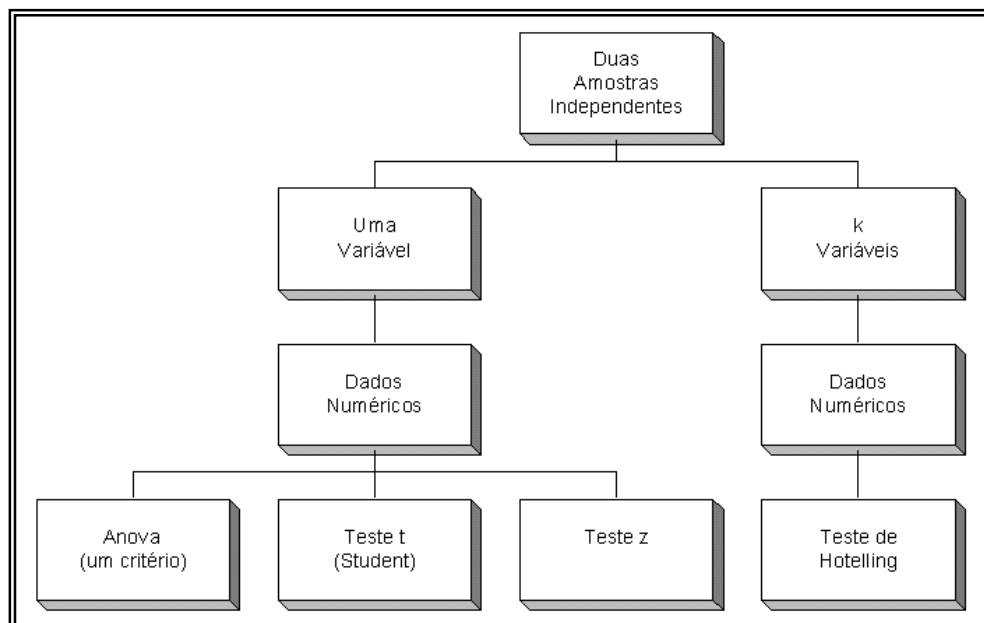


23.3c – UMA AMOSTRA: k VARIÁVEIS

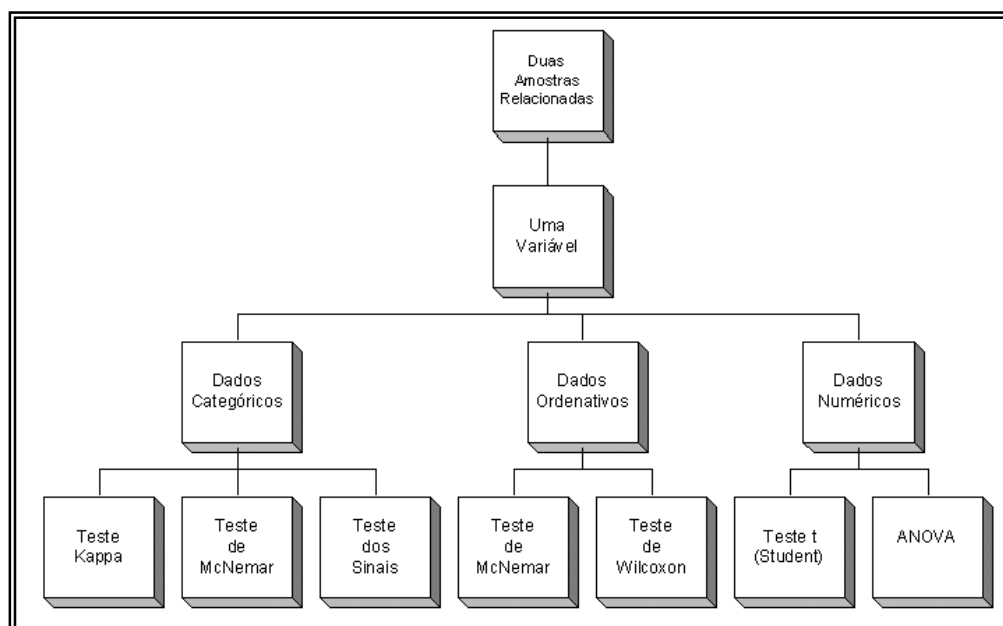


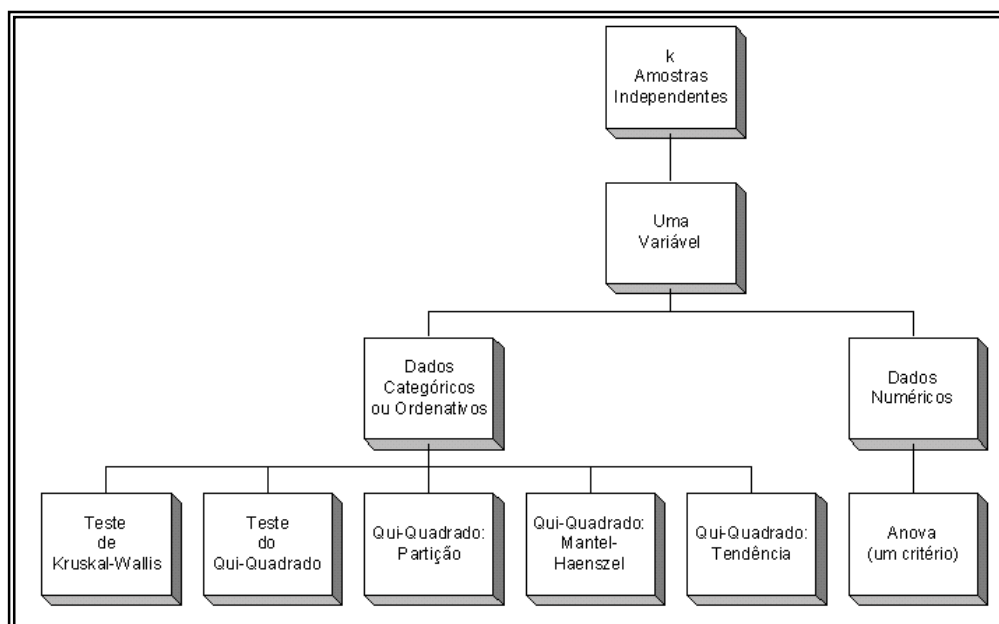
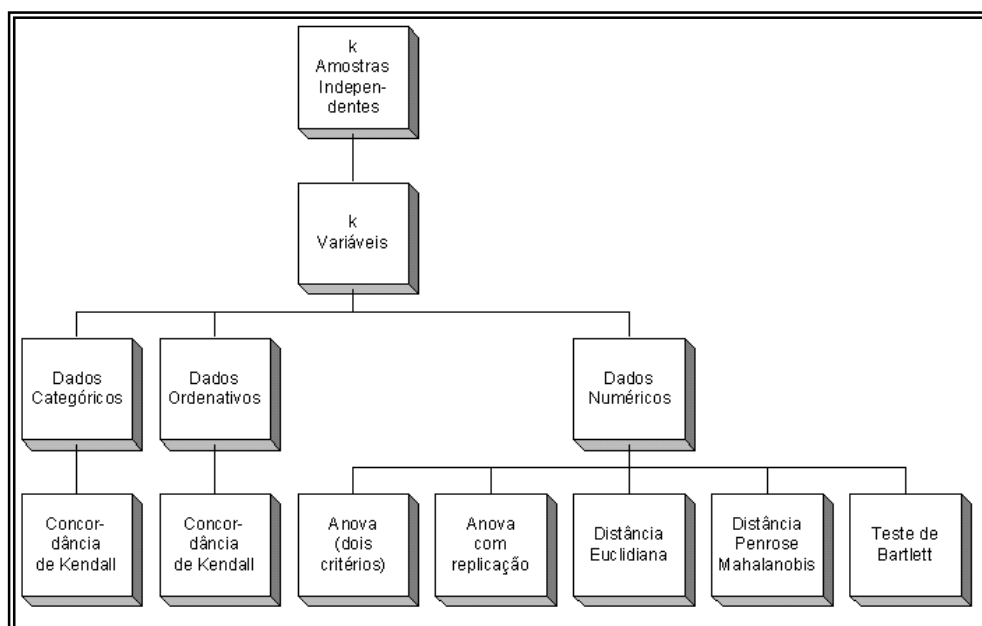
23.4a – DUAS AMOSTRAS INDEPENDENTES: UMA VARIÁVEL (dados categóricos)**23.4b – DUAS AMOSTRAS INDEPENDENTES: UMA VARIÁVEL (dados ordenativos)**

23.4c – DUAS AMOSTRAS INDEPENDENTES: UMA VARIÁVEL e k VARIÁVEIS (dados numéricos)

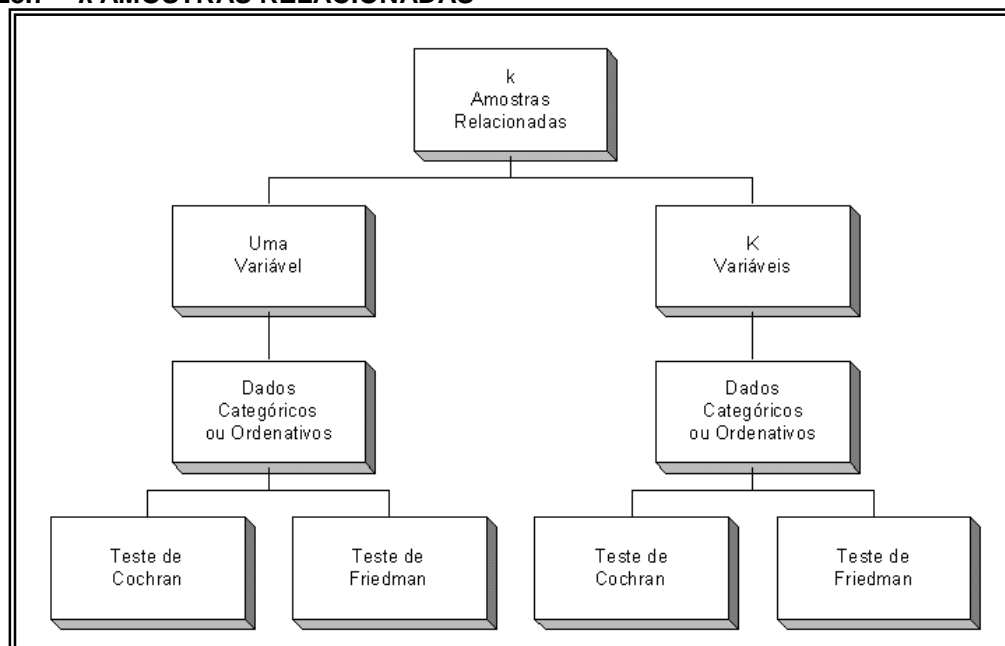


23.5 – DUAS AMOSTRAS RELACIONADAS: UMA VARIÁVEL



23.6a – k AMOSTRAS INDEPENDENTES: UMA VARIÁVEL (dados categóricos, ordenativos e numéricos)**23.6b – k AMOSTRAS INDEPENDENTES: k VARIÁVEIS**

23.7 – k AMOSTRAS RELACIONADAS



GLOSSÁRIO

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

ARR: Aumento do Risco Relativo ($1 - RR$).

Ajustamento de curvas: teste de regressão destinado a verificar o modelo de dependência entre duas variáveis: linear, exponencial, logarítmica ou geométrica.

Amostra: parte da população.

Amostra aleatória, casual, randômica: amostra escolhida sem a interferência do pesquisador, obtida por sorteio, pelos números aleatórios ou por procedimento computadorizado (pseudo-randômico), constituindo, do ponto de vista matemático, amostra probabilística.

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 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 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.

Amostras não-probabilísticas: amostras escolhidas por critérios pessoais, seja por conveniência, por julgamento ou por quotas.

Amostras independentes: dois ou mais subconjuntos randômicos da população retirados de modo independente e de tamanhos iguais ou desiguais.

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

Amostragem: procedimento de estudo de uma parte do universo. Sinônimo: Sondagem.

Amplitude total: diferença entre o maior e o menor escores de uma amostra ou universo.

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, através de dendograma.

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, etc. As instituições securitárias utilizam esse modelo de análise. O *BioEstat* apresenta os seguintes programas de análise de sobrevivência: Atuarial, e *Cox* (Risco Proporcional), *Cox-Mantel*, *Gehan* (Wilcoxon generalizado), *Kaplan-Meier* e *Log-Rank test*.

Análise de resíduos: auxiliar na interpretação das caselas em tabelas de contingência $l \times c$.

Análise discriminante: separação de dois ou mais grupos de indivíduos em função das medidas de suas variáveis por eles apresentadas.

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

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 desiguais, no caso de k tratamentos (um critério), e devem ser iguais quando for o caso de k tratamentos e r blocos (dois critérios). Este teste é também conhecido como teste F , em homenagem a R. A. Fisher, e destina-se a comparar diferenças entre médias através das variâncias, cujos escores amostrais devem ser mensurados a nível intervalar ou de razões.

Bioestatística ou Biometria: estatística aplicada às ciências biológicas e da saúde.

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 alfa de decisão. A comparação é efetuada após a análise da variância e somente se o valor de F for significativo.

Bootstrap: procedimento que consiste em retirar de uma pequena amostra numerosas amostras, com reposição, cada uma com probabilidade $1/n$ de ser obtida, sendo também conhecido como técnica de reamostragem ou de simulação.

Caixa de texto: espaço destinado à entrada de uma determinada informação, requerida para o teste.

Clicar: ato de pressionar um dos botões do *mouse* (em geral o da esquerda).

Coefficiente de assimetria: medida de forma de distribuição dos escores de variáveis aleatórias contínuas, podendo ser simétrica, assimétrica positiva (curva cuja cauda está desviada para a direita) e assimétrica negativa (curva cuja cauda está desviada para a esquerda).

Coefficiente de curtose: medida de forma de distribuição de variáveis aleatórias contínuas, podendo ser mesocúrtica (em forma de sino), leptocúrtica (afilada) ou platicúrtica (achatada).

Coefficiente de determinação (R^2): é o quadrado do coeficiente de correlação (r), representando a quantidade da variação de uma variável dependente explicada pela variável preditiva.

Coefficiente 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.

Coefficiente de variação: razão entre o desvio padrão e a média aritmética, vezes 100 (%).

Coefficiente Phi (r_ϕ): coeficiente de correlação entre duas variáveis qualitativas e dicotômicas, dispostas em tabela de contingência 2×2 .

Concordância de Kendall (W): teste de associação entre k variáveis dispostas em vários conjuntos de postos e mensuradas a nível ordinal.

Contingência C: teste de associação entre conjuntos de variáveis mensuradas a nível nominal, dispostas em tabelas de contingência l (linhas) \times c (colunas), isto é, com quaisquer números de categorias.

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.

Correlação Canônica: análise multivariada que permite investigar o relacionamento entre duas séries de variáveis: um grupo de variáveis X e outro do grupo Y .

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).

Correlação de Kendall: teste de associação entre duas variáveis mensuradas a nível ordinal (postos), calculando-se a correlação pelo coeficiente de *Kendall* τ .

Correlação de Spearman: teste de associação entre duas variáveis mensuradas a nível ordinal (postos), calculando-se a correlação pelo coeficiente de *Spearman* r_s .

Correlação intraclasse: a Correlação intraclasse 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. É, portanto, medida de replicabilidade amostral.

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 Parcial: teste de associação de duas variáveis quantitativas – X e Y – as quais são testadas juntamente com uma outra variável – Z –, a fim de se verificar se a correlação existente entre as primeiras é alterada pela presença da terceira variável introduzida.

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.

Curva ROC (Receiver Operating Characteristic Curve): é um método estatístico e gráfico para determinar o melhor ponto de corte (*cutoff point*) de um teste diagnóstico.

Dados: escores obtidos de observações ou de experimentos, podendo ser de fontes primárias ou secundárias e de amostras ou de populações.

D'Agostino: teste empregado para testar a normalidade dos dados amostrais.

D'Agostino-Pearson: teste destinado a verificar a normalidade dos dados amostrais.

Desvio padrão: raiz quadrada da variância.

Diagrama pontual (dotplot): diagrama que descreve graficamente duas ou mais amostras (até sete), permitindo a visualização pontual dos escores - em sentido vertical - e as respectivas comparações.

Diagrama IC: diagrama indicada 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 (95% ou 99%) da variável de cada categoria.

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 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 exponencial: modelo de distribuição de probabilidades de variáveis aleatórias contínuas, referente ao *intervalo de tempo* decorrido entre eventos raros e discretos.

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 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.

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 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 tipo I (alfa): rejeição da hipótese de nulidade quando ela é verdadeira.

Erro tipo II (beta): aceitação da hipótese de nulidade quando ela é falsa.

Escala de razões: caracteriza-se por apresentar unidade constante e comum de mensuração, atribuindo-se um número real a cada score, 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 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 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 ordinal ou por postos: escala na qual as modalidades de uma variável são ordenadas em graus ou 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*.

Escore padronizado: transformação dos escores brutos em escores z , onde a média é igual a zero e o desvio padrão igual à unidade (1), simbolizados por $N(0,1)$.

Especificidade: percentagem de indivíduos sem o evento (sem a afecção investigada, por exemplo), cujo teste – tuberculínico, por exemplo – é negativo (–).

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.

Estatísticas (no plural): valores numéricos das amostras: média, mediana, variância, coeficiente de assimetria, etc, constituindo nas amostras probabilísticas estimativas não enviesadas dos parâmetros, conforme é demonstrado pelo Teorema do Limite Central.

Estimação de parâmetros: parte da inferência estatística, cujo procedimento indutivo consiste em generalizar os valores numéricos amostrais para o universo investigado.

Estudo longitudinal: é aquele no qual se coletam informações sobre os indivíduos selecionados ao longo de um intervalo de tempo especificado.

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 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 transversal: dados coletados de um grupo de indivíduos em um momento definido, avaliando-se sobretudo a prevalência de uma determinada afecção.

Experimento: estudo analítico no qual há intervenção do pesquisador nos indivíduos submetidos à pesquisa, com controle efetivo dos fatores causais e dos respectivos efeitos.

Fenótipo: característica de um indivíduo resultante do produto dos genes e expressada de diversas maneiras. As pessoas do sistema sanguíneo ABO, por exemplo, são classificadas em fenótipos dos grupos A, B, AB ou O.

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.

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.

Gráfico de área: compara as áreas abrangidas pelos escores de duas ou mais amostras.

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 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 a) simples; b) justapostas e c) superpostas.

Gráfico de dispersão: utilizado para duas variáveis – X e Y – retiradas da mesma unidade, como ocorre nas análises de correlação e regressão.

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 da cada amostra.

Gráfico polar: representa períodos cíclicos – meses, semanas, horas, graus, etc –, com representação circular, indicado nas séries temporais.

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.

Graus de liberdade (gl): são parâmetros indexadores estatísticos correspondentes ao número de observações independentes, como se observam nas distribuições *t* de Student, *F* da ANOVA, Qui-Quadrado e *r* da correlação linear de Pearson.

Grid geral: é uma espécie de planilha eletrônica do BioEstat onde os dados, da maior parte dos testes estatísticos, devem ser introduzidos. Sinônimo: Matriz de Dados.

Grid específico: similar ao *grid* geral, utilizado apenas para determinados testes estatísticos.

Hipótese de nulidade (H_0): é a hipótese que se testa considerando-se não haver diferenças entre os grupos específicos objetos do estudo, representada, por exemplo: $\mu_1 = \mu_2$.

Hipótese alternativa (H_1): 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: $\mu_1 \neq \mu_2$.

Heterogeneidade das amostras: é um modelo probabilístico que indica se as amostras investigadas não são oriundas da mesma população, sendo utilizado nos testes G , do Qui-Quadrado, da correlação linear e da regressão linear.

Heterocedasticidade: vide Heterogeneidade das variâncias.

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 ANOVA, no t teste de *Student* para duas amostras independentes e na regressão linear. Sinônimo: *heterocedasticidade*.

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.

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.

Homocedasticidade: vide Homogeneidade das variâncias.

Homogeneidade das variâncias: é um modelo probabilístico que revela se as variâncias dos subconjuntos testados são iguais, oriundas, portanto, da mesma população, sendo empregado no F teste da ANOVA, no t teste de *Student* para duas amostras independentes e na regressão linear. Sinônimo: *homocedasticidade*.

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*.

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.

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.

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.

Intervalos 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.

Iterações (Runns Test): destina-se a verificar se a amostra em questão foi obtida de maneira aleatória. A técnica deste teste é baseada na *ordem* ou *seqüência* com que os escores originais forem coletados.

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.

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.

Lilliefors: teste de aderência empregado para testar a *normalidade* amostral.

Matriz de Dados: é uma espécie de planilha eletrônica do *BioEstat* onde os dados, da maior parte dos testes estatísticos, devem ser introduzidos. Sinônimo: *Grid Geral*.

Média aritmética: medida de tendência central de largo emprego em estatística, representada pela razão do somatório dos valores pelo número de escores.

Média geométrica: medida de tendência central representada pela raiz n do produto dos escores e indicada, sobretudo, quando os dados estão dispostos em progressão geométrica.

Média harmônica: medida de tendência central para grandezas inversamente proporcionais como, por exemplo, tempo e velocidade.

Mediana: medida de tendência central que separa os escores em dois grupos: 50% inferiores e 50% superiores à mediana.

Menu: lista de opções do *BioEstat* que é ativada com um clique do *mouse*. É encontrado na parte superior da tela.

Meta-Análise: conjunto de procedimentos destinados a examinar, de modo simultâneo, vários estudos, diversas pesquisas, sobre um mesmo tópico. Largamente utilizado na esfera das ciências médicas.

Moda: é o valor mais freqüente de um conjunto de dados.

Modelo: forma simbólica de um princípio físico expressado por uma equação ou por uma fórmula.

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.

Nível de significância: valor probabilístico do limite de rejeição da hipótese de nulidade, do erro alfa (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 (5%) ou 1 em 100 (1%), respectivamente.

NNH: Número Necessário para causar (*Harm*) 1 (um) evento desfavorável num determinado período de tempo (t).

NNT: Número Necessário para Tratar e prevenir 1 (um) evento desfavorável num determinado período de tempo (t).

Observação censurada: observação na análise de sobrevivência que não cumpriu o seguimento estabelecido, desconhecendo-se o motivo (abandono, etc).

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.

p -valor: é a probabilidade obtida em um teste de hipótese – inferência estatística – 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.

Parâmetros: valores numéricos das populações representando constantes de cada variável do universo investigado.

Pirâmide populacional: diagrama de populações sob a forma piramidal, podendo representar grupos etários, proporções sexuais, locais de residências, estados civis e outras variáveis.

Poder do teste: capacidade de um teste estatístico de rejeitar a hipótese de nulidade quando de fato ela é falsa.

Ponto de corte: determinação do melhor Ponto de Corte (*Cutoff point*) para um teste diagnóstico, ou seja, aquele que mais se aproxime do chamado *Padrão Ouro*.

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 o universo que altera seu tamanho durante o período da pesquisa, em decorrência de nascimentos, óbitos e fenômenos migratórios.

População fechada: é o universo que não altera seu tamanho durante o período de investigação.

Precisão amostral: é a proximidade entre os valores das *estatísticas* obtidas de várias amostras do mesmo tamanho e da mesma 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*.

Probabilidade: é uma escala de mensuração usada para descrever a chance de ocorrência de um valor específico (evento) de uma variável aleatória.

Regra de Bayes: modelo de distribuição de probabilidade condicional onde se calcula a probabilidade *a posteriori* do evento (A) dada a ocorrência do (B) – $(A|B)$ –, em função do conhecimento *a priori* da probabilidade de ocorrência do evento B , desde que o evento A tenha ocorrido – $Pr(B|A)$. Para dois eventos o *BioEstat* apresenta o chamado crivo ou *screening test*, enquanto que para três ou mais, o programa dispõe do modelo da regra de Bayes.

Regressão linear simples: teste estatístico que determina o modelo estimador dos valores de Y a partir dos escores de X .

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 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, mas pode assumir valores contínuos. Este teste tem larga aplicação nas ciências médicas.

Regressão logística múltipla: semelhante à regressão logística simples, mas o número de variáveis independentes, em geral binárias, é múltiplo – dois (2) ou mais. A variável dependente é sempre binária.

Resumo amostral: testes paramétricos (t e z) para uma ou duas amostras, quando se dispõe apenas de informações sobre tamanho, médias e variâncias amostrais.

Risco relativo: teste estatístico simbolizado por RR , representando a *razão* entre o *coeficiente* de incidência de determinado evento em pessoas expostas (a uma doença, por exemplo) e o *coeficiente* de incidência do mesmo evento em indivíduos não expostos.

RAR: Redução Absoluta do Risco representado por $p_1 - p_2$.

RRR: Redução do Risco Relativo representado por $1 - RR$.

Sensibilidade de um teste: percentagem de indivíduos com o evento (determinada afecção, por exemplo) cujo teste – tuberculínico, por exemplo – é positivo (+).

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 geográfica: é uma série estatística na qual o local (fator geográfico) é variável, mantendo-se fixos o tempo e o fenômeno observado.

Shapiro-Wilk: teste empregado para verificar a *normalidade* dos dados amostrais.

Sondagem: procedimento de estudo de uma parte da população voltado, sobretudo, para pesquisas de opinião e de mercado. Sinônimo: amostragem.

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: a) Método Progressivo (*Forward Selection*) e b) Modelo Regressivo (*Backward Selection*).

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.

Tabelas de contingência: tabelas nas quais se dispõem as frequências observadas de duas ou mais amostras, cada uma com duas ou mais categorias, em tabelas de *l* linhas e *c* colunas (2 x 2, 3 x 2, 2 x 3, 3 x 3, etc).

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 testes *G* e do Qui-Quadrado, para uma amostra, são exemplos de testes de aderência.

Teste bilateral: teste estatístico no qual a hipótese alternativa não especifica a direção da diferença a ser detectada, com esta representação: $\mu_1 \neq \mu_0$, podendo μ_1 ser maior ou menor que μ_0 .

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 estatístico baseado na distribuição binomial ou na aproximação à curva normal, onde se compara a proporção de uma amostra com a do parâmetro.

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 de Cochran: teste estatístico de modelo livre de distribuição de probabilidades, onde os dados estão contidos em *n* linhas e *k* colunas. Todos os escores – nominais ou ordinais dicotomizados – são 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.

Teste de Friedman: teste estatístico de modelo livre de distribuição de probabilidades abrangendo três (3) ou mais amostras relacionadas, para comparação das respectivas médias, cujos escores são mensurados em escala nominal ou ordinal (postos). As amostras devem ser do mesmo tamanho.

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 dados devem ser mensurados, no mínimo, a nível ordinal.

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 2×2 , 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 $(- +)$.

Teste de Parker: teste de estimação do tamanho de população através 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 Schumacher e Eschmeyer: teste de estimação do tamanho da população através de amostragem múltipla em universos fechados.

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 como *Wilcoxon signed-rank test*.

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 $(l - 1) \times (c - 1)$ graus de liberdade.

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 $(+ \text{ ou } -)$, independentemente, portanto, da magnitude das diferenças.

Teste Exato de Fisher: teste estatístico, para duas amostras independentes cujas proporções estão dispostas em tabela de contingência 2×2 , devendo ser escolhido quando os valores observados são pequenos, inclusive com proporções cujos escores são iguais a zero (0).

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 diferentes categorias são independentes ou estão associadas. Os graus de liberdade neste teste são calculados como segue: $(l - 1) \times (c - 1)$.

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* dos resultados, no sentido de haver ou não concordância nas proporções observadas nos diferentes períodos. Os dados são dispostos em tabela de contingência 2×2 .

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.

Teste paramétrico: é aquele com pressuposto de um modelo de distribuição – normal, binomial, etc –, cujos parâmetros assumem um desses modelos, aplicando-se-lhe o Teorema do Limite Central.

Teste unilateral: teste estatístico no qual a hipótese alternativa especifica a direção da diferença a ser detectada, assim representado: $\mu_1 < \mu_0$ ou $\mu_1 > \mu_0$.

Testes *t* (Student): testes paramétricos 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.

Testes *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.

Transformação de dados: mudança de escala dos dados de uma variável, podendo ser efetuada pelos logaritmos, pela raiz quadrada, pela ordenação dos escores (*rank*), pelo quadrado dos valores, etc.

Unidade: indivíduo, elemento, item, membro ou unidade estatística onde se estuda uma ou mais variáveis. A unidade estatística pode ser simples: pessoa, objeto, animal, planta, protozoário, etc, ou coletiva: cardume, blocos residenciais, turmas escolares, colméias, colônias bacterianas e outras.

Universo: o mesmo que população.

Valor preditivo de um teste negativo: probabilidade de que um indivíduo com teste negativo (–) seja uma pessoa sem o evento objeto da investigação (determinada doença, por exemplo).

Valor preditivo de um teste positivo: probabilidade de que um indivíduo com teste positivo (+) apresente o evento objeto da investigação (determinada doença, por exemplo).

Valor extremo: escore 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. Na literatura inglesa denomina-se *outlier*.

Variância: soma dos quadrados dos desvios em relação à média, dividida pelo número de observações da amostra menos um ($n - 1$).

Variáveis independentes: são variáveis consideradas preditivas de uma outra variável denominada dependente.

Variável: determinada característica dos indivíduos ou elementos objetos da investigação, como são exemplos o sexo dos animais, o estado civil, a etnia, o peso de recém-nascidos, etc.

Variável dependente: é a variável decorrente de uma ou mais variáveis, estas denominadas independentes ou preditivas.

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 – superposição, inclusão ou exclusão de unidades –, da escolha não-aleatória das amostras, de erros de observação, de defeitos instrumentais e outros.

REFERÊNCIAS BIBLIOGRÁFICAS

- Abuabara, M.A.P. & Petrere Jr., M. (1997). *Estimativas da Abundância de Populações Animais*. EDUEM – Editora da Universidade Estadual de Maringá, Paraná.
- Allen, R.G.D. (1979). *Estatística para Economistas*. Zahar Editores, Rio de Janeiro.
- Ayres, M. & Ayres Jr., M. (1987). *Aplicações Estatísticas em BASIC*. McGraw-Hill do Brasil, São Paulo.
- Ayres, M., Ayres Jr., M., Ayres, D.L., Santos, A.S. (1998). *BioEstat*. Versão 1.0, Sociedade Civil Mamirauá, MCT – CNPq, Belém, Pará, Brasil.
- Beiguelman, B. (1979). *Genética Médica*, vol. 3. Edart-São Paulo Livraria Editora, Ed. da Universidade de São Paulo.
- Beiguelman, B. (1994). *Curso Prático de Bioestatística*. 3. ed., Revista Brasileira de Genética.
- Beiguelman, B. (1994). *Dinâmica dos Genes nas Famílias e nas Populações*. Sociedade Brasileira de Genética, Ribeirão Preto, São Paulo.
- Berkeley, E.C. (1965). *Sorte ou Azar*. Editora Universidade de Brasília, Brasília.
- Berquó, E.S., Souza, J.M.P. & Gotlieb, S.L.D. (1980). *Bioestatística*. Editora Pedagógica e Universitária, São Paulo.
- Blank, S., Seiter, C. & Bruce, P. (2001). *Resampling Stats in Excel*. Resampling Stats, Inc., Arlington, Virginia.
- Brown, F.L., Amos, J.R. & Mink, O.G. (1975). *Statistical Concepts*. 2nd. ed., Harper & Row, Publishers, New York.
- Burleson, D.R. (1980). *Elementary Statistics*. Winthrop Publishers, Inc., Cambridge, Massachusetts.
- Bussad, W.O. (1988). *Análise de Variância e Regressão*. 2. ed., Editora Atual, São Paulo.
- Butcher, H.J. (1966). *Sampling in Educational Research*. Manchester University Press, Manchester.
- Caldeira, A.M.S. & Ferreira, M.L.B. (1974). *Estatística – Instrução Programada*, vols. 1 e 2. Conquista, Rio de Janeiro.
- Callegari-Jacques, S. (2003). *Bioestatística: princípios e aplicações*. Artmed Editora S.A., Porto Alegre, Brasil.
- Calot, G. (1979). *Cours de Statistique Descriptive*. Deuxième édition, Dunod, Paris.
- Camel, F. (1991). *Estadística Médica y Planificación de la Salud, I y II*. Consejo de Publicaciones de la Universidad de Los Andes, Mérida, Venezuela.
- Campbell, R.C. (1994). *Statistics for Biologists*. 3rd. ed., Cambridge University Press, Cambridge.
- Canavos, G.C. (1988). *Probabilidad y Estadística*. McGraw-Hill, México.
- Castello, L. (2004). A method to Count Pirarucu Arapaima gigas: Fishers, Assessment, and Management. *North American Journal of Fisheries Management*, 24:379-389.
- Castle, W.M. (1977). *Statistics in Small Doses*. 2nd. ed., Churchill Livingstone, London.
- Caswell, H. (2001). *Matrix Population Models*. 2nd. ed., Sinauer Associates, Inc. Publishers, Sunderland, Massachusetts.
- Centeno, A.J. (1982). *Curso de Estatística Aplicada à Biologia*. Editora da Universidade Federal de Goiás, Goiânia.
- Chacón, E. (1972). *Curso de Estatística*. Livro Ibero-Americano, Rio de Janeiro.
- Chatfield, C. & Collins, A.J. (1986). *Introduction to Multivariate Analysis*. Chapman and Hall, London.

- Chernick, M.R. & Friis, R.H. (2003). *Introductory Biostatistics for the Health Science*. Wiley Interscience, New Jersey.
- Cochran, W.G. (1977). *Sampling Techniques*. 3rd. ed., John Wiley & Sons, Inc., New York.
- Cohen, J. (1977). *Statistical Power Analysis for the Behavioral Sciences*. 2nd. ed., Lawrence Erlbaum Associates, Publishers, New Jersey.
- Collett, D (1996). *Modelling Survival Data in Medical Research*. Chapman & Hall, London.
- Cooke, D., Craven, A.H. & Clarke, G.M. (1984). *Basic Statistical Computing*. Edward Arnold (Publishers), London.
- Costa, J.J.S. (1981). *Elementos de Estatística*. Editora Campus, Rio de Janeiro.
- Costa, S.F. (1988). *Introdução Ilustrada à Estatística*. Editora Harbra, São Paulo.
- Cox, D.R. (1972). Regression Models and Life-Tables. *J. Roy. Statist. Soc.* , **34**: 187-220.
- Crespo, A.A. (1997). *Estatística Fácil*. 15. ed., Editora Saraiva, São Paulo.
- Cunha, O.S. & Ramos, F.A.O. (1976). *Iniciação à Estatística: Volumes I e II*. Tipografia do Carvalhido, P.
- Cunha, S.E. (1978). *Estatística Descritiva na Psicologia e na Educação*. Forense-Universitária, Rio de Janeiro.
- Daniel, W.W. (1999). *Biostatistics: A Foundation for Analysis in the Health Sciences*. 7th. ed., John Wiley & Sons, New York.
- Dawson-Saunders, B. & Trapp, R.G. (1994). *Basic & Clinical Bioestatistics*. 2nd. ed., Appleton & Lange, Norwalk, Connecticut.
- DeGroot, M.H. (1988). *Probabilidade y Estadística*. 2. ed., Addison-Wesley Iberoamericana, Wilmington, Delaware.
- Di Dio, R.A.T. (1979). *Estatística, Instrução Programada*. Editora Pedagógica e Universitária, São Paulo.
- Dress, F. (2004). *Probabilités et Statistique de A à Z*. Dunod, Paris.
- Edwards, A.L. (1984). *An Introduction to Linear Regression and Correlation*. 2nd. ed., W. H. Freeman and Company, New York.
- Efron, B. (1979). Bootstrap Methods: Another Look at de Jackknife. *The Annals of Statistics* 7 (1): 1-26.
- Efron, B. & Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall/CRC, London.
- Elston, R.C. & Johnson, W.D. (1994). *Essentials of Biostatistics*. 2nd. ed., F. A. Davis Company, Philadelphia.
- Essex-Sorlie, D. (1995). *Medical Bioestatistics & Epidemiology*. Appleton & Lange, Norwalk, Connecticut.
- Fisher, L.D. & Van Belle, G. (1993). *Biostatistics*. John Wiley & Sons, Inc., New York.
- Fisher R.A. (1946a). The fitting of gene frequencies to data on Rhesus reactions. *Ann. Eug.* 13:150-155.
- Fisher R.A. (1946b). Note on the calculation of the frequencies of Rh allelomorphs. *Ann. Eug.* 13:223-224.
- Fisher, R.A. (1993). *Statistical Methods Experimental Design and Scientific Inference*. Oxford University Press, Oxford.
- Fisher, R.A. & Yates, F. (1954). *Tablas Estadísticas*. Aguilar Ediciones, Madrid.
- Fleiss, J.L. (1986). *The design and analysis of clinical experiments*. Wiley, New York.
- Fogiel, M. (1996). *The Statistics Problem Solver*. Research and Education Association, New Jersey.
- Freund, R.J. & Wilson, W.J. (1997). *Statistical Methods*. Revised Edition, Academic Press, New York.
- Galambos, J. (1984). *Introductory Probability Theory*. Marcel Dekker, Inc., New York.
- Galton, F. (1886). Rate of Regression in Hereditary Stature. *Journal of the Anthropological Institute*, Vol. 15, p. 248.

- Gattás, R.R. (1978). *Elementos de Probabilidade e Inferência*. Editora Atlas, São Paulo.
- Gemmell, I., Heller, R.F., McElduff P., Butler, G., Edwards, R., Roland, M. & Dumington, P. (2005). The population impact of stricter adherence to recommendations for pharmacological and lifestyle interventions over one year in patients with coronary heart disease. *Journal of Epidemiology and Community Health*; 59: 1041-1046.
- Gimenes, M. (2003). Interaction between visiting bees (Hymenoptera, Apoidea) and flowers of *Ludwigia elegans* (Camb.) hara (Onagraceae) during the year in two different areas in São Paulo, Brazil. *Braz. J. Biol., São Carlos*, v. 63, n. 4, 2003.
- Glantz, S.A. (1997). *Primer of Biostatistics*. 4th. ed., McGraw-Hill, New York.
- Gnedenko, B.V. & Khintchine, A.I. (1974). *Introdução à Teoria das Probabilidades e à Estatística*. Editorial Estúdios Cor, Lisboa.
- Gonick, L. & Smith, W. (1993). *The Cartoon Guide To Statistics*. HarperPerennial, HarperCollins Publishers, New York.
- Greenberg, R.S., Daniels, S.R., Flanders, W.D., Eley, J.W. & Boring, J.R. (1996). *Medical Epidemiology*. 2nd. ed., Appleton & Lange, Norwalk, Connecticut.
- Groeneveld, R.A. (1979). *An Introduction to Probability and Statistics Using Basic*. Marcel Dekker, Inc., New York.
- Groves, R.M. (1989). *Survey Errors and Survey Costs*. John Wiley & Sons, New York.
- Guedes, M.L.S. & Guedes, J.S. (1988). *Bioestatística para Profissionais de Saúde*. Ao Livro Técnico, Rio de Janeiro.
- Guimarães, R.C. & Cabral, J.A.S. (1997). *Estatística*. McGraw-Hill de Portugal, Porto.
- Gullon, A. (1971). *Introducción a la Estadística Aplicada*. Editorial Alhambra, Madrid.
- Hayslett Jr., H.T. (1968). *Statistics Made Simple*. Doubleday, New York.
- Hedges, L.V. & Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. Academic Press, New York.
- Heller, R.F., Dobson, A.J., Attia, J. & Page, J.H. (2002). Impact numbers: measures of risk factor impact on the whole population from case control and cohort studies. *J. Epidemiol Community Health* 2002; 56: 606-10.
- Heller, R.F., McElduff, P. & Edwards, R. (2002). Impact of upward social mobility on population mortality: analysis with routine data. *BMJ*; 325: 134.
- Heller, R.F., Edwards, R. & McElduff, P. (2003). Implementing guidelines in primary care: can population impact measures help? *Bmc Public Health*; 2: part 7.
- Heller, R.F., Buchan, L., Edwards R., Lyratzopoulos, G., McElduff P. and St Leger S. (2004). Communicating risks at the population level: application of population impact numbers. *British Medical Journal*: 327: 1162-5.
- Heller, R.F. (2005). *Evidence for Population Health*. Oxford University Press, Manchester, UK.
- Herzon, F.D. & Hooper, M. (1976). *Introduction to Statistics for the Social Sciences*. Thomas Y. Crowell Company, New York.
- Hine, J. & Wetherill, G.B. (1975). *A Programmed Text in Statistics: Book One, Summarizing Data*. Chapman and Hall, London.
- Hinkle, D.E., Wiersma, W. & Jurs, S.G. (1994). *Applied Statistics for the Behavioral Sciences*. 3rd. ed., Houghton Mifflin Company, Boston.
- Hoel, P.G. (1979). *Estatística Elementar*. Editora Atlas, São Paulo.
- Hosmer, D.W. & Lemeshow, S. (1989). *Applied Logistic Regression*. John Wiley & Sons, New York.
- Hosmer, D.W. & Lemeshow, S. (1998). *Applied Survival Analysis*, John Wiley & Sons, New York.
- Howell, D.C. (1993). *Fundamental Statistics for the Behavioral Sciences*. 2nd. ed., PWS-KENT Publishing Company, Boston.

- Huck, S.W. & Sandler, H.M. (1984). *Statistical Illusions: Solutions*. Harper & Row, Publishers, New York.
- Huff, D. (1993). *How to Lie with Statistics*. W. W. Norton & Company, Inc., New York.
- Ingelfinger, J.A., Mosteller, F., Thibodeau, L.A. & Ware, J.H. (1994). *Biostatistics in Clinical Medicine*. 3rd. ed., McGraw-Hill, New York.
- Jabardo, W.R.S. (1985). *Genética*. Editora McGraw-Hill do Brasil, São Paulo.
- Jammalamadaka, S. R. & Gupta, A. (2001). *Topics in Circular Statistic*. Word Scientific, London.
- Johnson, D. E. (1998). *Métodos Multivariados Aplicados al Análisis de Datos*. International Thomson Editores, México.
- Johnson, R.A. & Wichern, D.W. (1998). *Applied Multivariate Statistical Analysis*. 4th. ed., Prentice Hall, Inc., New Jersey.
- Kaar, A.F. (1986). *Point Processes and Their Statistical Inference*. Marcel Dekker, Inc., New York.
- Kachigan, S.K. (1991). *Multivariate Statistical Analysis – A Conceptual Introduction*. 2nd Ed.. Radius Press. New York.
- Kalbfleisch, J.G. (1985). *Probability and Statistical Inference*, vols. 1, 2. 2rd. ed., Springer-Verlag, New York.
- Kaplan, D.T. (1999). *Resampling Stats in MATLAB*. Resamplibg Stats. Inc., Arlington, Virginia.
- Katz, D.L. (2001). *Epidemiologia, Bioestatística e Medicina Preventiva*. Livraria e Editora Revinter Ltda., Rio de Janeiro.
- Kellner, S.R.O., Cavas, C.S.T. & Nick, E. (1979). *Exercícios de Estatística*. Editora Renes, Rio de Janeiro.
- Kirkwood, B.R. (1995). *Essentials of Medical Statistics*. Blackwell Science, Oxford.
- Lang, T.A. & Secic, M. (1997). *How to Report Statistics in Medicine*. American College of Physicians, Philadelphia.
- Lange, K. (1997). *Mathematical and Statistical Methods for Genetic Analysis*. Springer-Verlag, New York.
- Lapponi, J.C. (1997). *Estatística usando Excel 5 e 7*. Lapponi Treinamento e Editora, São Paulo.
- Laurenti, R., Mello Jorge, M.H.P., Lebrão, M.L. & Gotlied S.L.D. (1987). *Estatísticas de Saúde*. 2. ed., Editora Pedagógica e Universitária, São Paulo.
- Lee, E.T. (1992). *Statistical Methods for Survival Data Analysis*. 2nd. ed., John Wiley & Sons, Inc., New York.
- Levin, J. (1978). *Estatística Aplicada a Ciências Humanas*. Editora Harper & Row do Brasil, São Paulo.
- Levy, P.S. & Lemeshow, S. (1991). *Sampling of Populations*. John Wiley & Sons, Inc., New York.
- Lindley, D.V. & Scott, W.F. (1984). *New Cambridge Elementary Statistical Tables*. Cambridge University Press, Cambridge.
- Maletta, C.H.M. (2000). *Bioestatística – Saúde Pública*. 3ª edição, Editora Independente, Belo Horizonte, Minas Gerais.
- Manly, B.F.J. (2000). *Multivariate Statistical Methods- A Primer*, Third Ed., Chapman and Hall, London.
- Martinelli, J.M. (2005). *Estrutura Populacional dos Camarões Penaeidae no Estuário do Rio Caeté, litoral Norte do Brasil*, Tese de Doutorado, Belém, Pará, Brasil.
- Martins, G.A. & Donaire, D. (1979). *Princípios de Estatística*. Editora Atlas, São Paulo.
- Massad, E., Menezes, R.X., Silveira, P.S.P. & Ortega, N.R.S. (2004). *Métodos Quantitativos em Medicina*. Editora Manole, São Paulo, Brasil.
- Mattar, F.N. (1994). *Pesquisa de Marketing, 1 e 2*. Editora Atlas, São Paulo.

- Matthews, D.E. & Farewell, V.T. (1996). *Using and Understanding Medical Statistics*. 3rd. ed., Karger, Basel.
- Maxwell, A.E. (1977). *Multivariate Analysis in Behavioural Research*. Chapman and Hall, London.
- McCuen, R.H. (1993). *Microcomputer Applications in Statistical Hidrology*. Prentice Hall, New Jersey.
- Mendenhall, W. & Beaver, R.J. (1991). *Introduction to Probability and Statistics*. 8th. ed., PWS-KENT Publishing Company, Boston.
- Michelson, S. & Schofield, T. (1996). *The Biostatistics Cookbook*. Kluwer Academic Publishers, Boston.
- Miller, S. (1977). *Planejamento Experimental e Estatística*. Zahar Editores, Rio de Janeiro.
- Minium, E.W., King, B.M. & Bear, G. (1993). *Statistical Reasoning in Psychology and Education*. Third ed., John Wiley & Sons, Inc., New York.
- Morettin, P.A. & Bussab, W.O. (2004). *Estatística Básica*, 5a. Ed., Editora Saraiva, SãoPaulo.
- Mosteller, F. & Rourke, R.E.K. (1993). *Estatísticas Firms*. Edições Salamandra, Lisboa.
- Murteira, B.J.F. (1990). *Probabilidades e Estatística, vols. I e II*. 2. ed., McGraw-Hill de Portugal, Lisboa.
- Neave, H.R. (1985). *Elementary Statistics Tables*. George Allen & Unwin, London.
- Neter, J., Kutner, M.H., Nachtsheim, C.J. & Wasserman, W. (1996). *Applied Linear Regression Models*. 3rd. ed., IRWIN, Chicago.
- Neter, J., Kutner, M.H., Nachtsheim, C.J. & Wasserman, W. (1996). *Applied Linear Statistical Models*. 4th. ed., IRWIN, Chicago.
- Nick, E. & Kellner, S.R.O. (1971). *Fundamentos de Estatística para as Ciências do Comportamento*. Editora Renes, Rio de Janeiro.
- Oliveira, T.F.R. (1976). *Estatística na Escola*. Ao Livro Técnico, Rio de Janeiro.
- Ott, R.L. & Mendenhall, W. (1995). *Understanding Statistics*. 6th. ed., Duxbury Press, Belmont, California.
- Pagano, R.R. (1994). *Understanding Statistics in the Behavioral Sciences*. 4th. ed., West Publishing Company, New York.
- Peters, W.S. & Summers, G.W. (1978). *Análise Estatística e Processo Decisório*. 2. ed., Editora da Fundação Getúlio Vargas, Rio de Janeiro.
- Queiroz, H.L. (2000). *Natural history and conservation of pirarucu, Arapaima gigas, at the Amazonian Varzea: Red giants in muddy waters*. Thesis submitted for the degree of Doctor of Philosophy, University of St. Andrew.
- Rameau, C. (1971). *As Estatísticas: um Instrumento de Gestão, 1ª e 2ª*. Livraria Clássica Editora, Lisboa.
- Rees, D.G. (1990). *Essential Statistics*. 2nd. ed., Chapman and Hall, London.
- Reichmann, W.J. (1975). *Uso e Abuso das Estatísticas*. Editora Artenova, Rio de Janeiro.
- Reis, E. (1997). *Estatística Multivariada Aplicada*. Edições Sílabo, Ltda., Lisboa.
- Reis, M.D. (1978). *Elementos Básicos de Estatística*. 2. ed., Editora Estrutura, São Paulo.
- Reszczyński, O. (1969). *Sumatorias*. Centro Interamericano de Enseñanza de Estadística, Santiago.
- Rivera, D.P.S. (1989). *Estadística Modelos y Métodos, 1. Fundamentos*. Alianza Editorial, Madrid.
- Rocha, M.V. (1975). *Curso de Estatística*. 3. ed., Fundação Instituto Brasileiro de Geografia e Estatística (IBGE), Rio de Janeiro.
- Rosner, B. (2006). *Fundamentals of Biostatistics*. 6th. ed., Duxbury Press, Boston.
- Rothenberg, R.I. (1991). *Probability and Statistics*. Harcourt Brace Jovanovich, Publishers, New York.
- Rowntree, D. (1981). *Statistics Without Tears - A Primer for Non-mathematicians*. Penguin Books, New York.

- Sahai, H. & Martínez, W. (1996). *Tablas y Fórmulas Estadísticas para las Ciencias Biológicas, Sociales y Físicas*. Grupo Editorial Iberoamérica, de C.V., México.
- Salaru, N.N. (1989). Blood groups in large sample from city of São Paulo (Brazil): allele and haplotype frequencies for MNSs, Kell-Cellano, Rh and ABO systems. *Rev. Brasil. Genet.* 12, 2, 625-643.
- Shapiro, S.S. & Wilk, M.B. (1965). An Analysis of variance test for normality (complete samples). *Biometrika*, 52: 591- 411.
- Shapiro, S.S., Wilk, M.B. & Chen, H.J. (1968). A comparative study of various tests for normality. *J. Amer. Statist. Assoc.* 63: 1343 – 1372.
- Scheaffer, R.L., Mendenhall, W. & Ott, L. (1990). *Elementary Survey Sampling*. 4th. ed., PWS-KENT Publishing Company, Boston.
- Shiavi, R. (1991). *Introduction to Applied Statistical Signal Analysis*. Aksen Associates Incorporated Publishers, Boston.
- Shiffler, R. E. & Adams, A. J. (1999). *Succeeding Statistics*. Duxbury Press, New York.
- Siegel, S. & Castellan Jr., N.J. (1988). *Nonparametric Statistics for the Behavioral Sciences*. 2nd. ed., Mc Graw-Hill, Inc., New York.
- Sincich, T. (1993). *Statistics by Example*. 5th. ed., Prentice Hall, New Jersey.
- Snedecor, G.W. & Cochran, W.G. (1978). *Statistical Methods*. 6th. ed., The Iowa State University Press, Ames, Iowa.
- Soares, J.F. & Siqueira, A.L. (2002). *Introdução à Estatística Médica*. Departamento de Estatística – UFMG, Belo Horizonte, Brasil.
- Sokal, R.R. & Rohlf, F.J. (1969). *Biometry*. W. H. Freeman and Company, San Francisco.
- Sokal, R.R. & Rohlf, F.J. (1995). *Biometry*. 3rd. ed., W. H. Freeman and Company, New York.
- Sounis, E. (1985). *Bioestatística*. 3. ed., Livraia Atheneu, Rio de Janeiro.
- Spiegel, M.R. (1984). *Estatística*. 2. ed., McGraw-Hill do Brasil, São Paulo.
- Stansfield, W.D. (1985). *Genética*. 2. ed., McGraw-Hill do Brasil, São Paulo.
- Steel, R.G.D., Torrie, J.H. & Dickey, D.A. (1997). *Principles and Procedures of Statistics – A Biometrical Approach*. 3rd. ed., McGraw-Hill, New York.
- Stevens, W.L. (1938). Estimation of blood-groups gene frequencies. *Ann. Eug.* 8:362-217.
- Stevens, W.L. (1950). Statistical analysis of the A-B-O blood groups. *Hum. Biol.* 22:191-217.
- Sutton, A. J., Abrams, K. R., Jones, D. R., Sheldin, T. A. & Song. F. (2000). *Methods for Meta-Analysis in Medical Research*. John Wiley & Sons, Ltd., New York.
- Tanis, E.A. (1987). *Statistics II – Estimation and Tests of Hypotheses*. Harcourt Brace Jovanovich, Publishers, New York.
- Tippett, L.H.C. (1968). *Estatística*. 2. ed., Coleção SABBER, Publicações Europa-América, Lisboa.
- Toledo, G.L. & Ovalle, I.I. (1995). *Estatística Básica*. 2. ed., Editora Atlas, São Paulo.
- Triola, M.F. (1999). *Introdução à Estatística*. 7. ed., LTC – Livros Técnicos e Científicos Editora, Rio de Janeiro.
- Van Belle. G. (2002). *Statistical Rules of Thumb*. Wiley Interscience, New York.
- Vicente, P., Reis, E. & Ferrão, F. (1996). *Sondagens*. Edições Sílabo, Lisboa.
- Vieira, S. (1981). *Introdução à Bioestatística*. Editora Campus, Rio de Janeiro.
- Weaver, J. H. (2001). *Conquering Statistics – Numbers Without The Crunch*. Perseus Publishing, Cambridge, Massachusetts.
- Wetherill, G.B. (1967). *Elementary Statistical Methods*. Methuen & Co. Ltd., London.
- White, J., Yeats, A. & Skipworth, G. (1979). *Tables for Statisticians*. 3rd. ed., Stanley Thornes (Publishers) Ltd., Cheltenham.

- Wonnacott, R.J. & Wonnacott, T.H. (1985). *Introductory Statistics*. 4th. ed., John Wiley & Sons, New York.
- Zar, J.H. (1999). *Biostatistical Analysis*. 4th. ed., Prentice Hall, New Jersey.

ÍNDICE

A

Abrindo arquivos, 5
 Acurácia, 102, 107
 Aderência,
 G, teste, 119
 Kolmogorov-Smirnov, 120
 Lilliefors, 121
 Qui-Quadrado, 121, 122
 Ajustamento de curvas, 219
 Amostras, tamanho, 245
 Amostragem, tipos de, 9
 Aleatórias do grid, 13
 Aleatórias simples, 10
 Estratificadas, 11
 Por conglomerados, 11
 Sistemáticas, 10
 Análise multivariada, 17
 Componente principal, 23
 Conglomerados, 17
 Correlação canônica, 25
 Discriminante, 20
 Distância Euclidiana, 30
 Distância de Penrose e Mahalanobis, 31
 Teste de Hotelling, 33
 Teste de Bartlett, 33
 Teste de Mantel, 34
 Análise de resíduos, 212, 221
 Análise de sobrevivência, 39
 atuarial, 40
 Cox-Mantel, 41
 Cox: Modelo de Risco Proporcional, 42
 Gehan, 51
 Kaplan-Meier, 46
 Log-Rank test, 47
 Mantel-Haenszel, 52
 ANOVA, 59
 dois critérios, 64

duas amostras pareadas, 150
 fatorial a x b, 65
 fatorial a x b x c, 67
 Friedman, 68
 Kruskal-Wallis, 70
 Q de Cochran, 72
 um critério, 59
 Área, gráfico de, 260
 Aumento do Tamanho Amostral, 14

B

Bartlett, teste de, 33
 Bayes, regra de, 106
 Binomial, distribuição, 95
 Binomial, teste, 116, 131
 Bootstrap (Tira de bota), 75
 Box-plot, 262, 262

C

Casas decimais, expressão de, 1
 Caule e folha, diagrama, 263
 Circular, estatísticas, 235
 Cochran, teste Q de, 72
 Coeficientes de variação, teste de, 144
 Colunas, gráfico de, 265
 Configurando o *BioEstat*, 7
 Componente principal, 23
 Conglomerados, amostras por, 11
 Conglomerados, análise, 17
 Correlação, 85
 canônica, 25
 circular, 244
 concordância de Kendall, 91
 contingência C, 90
 de Kendall, 92
 intraclasse, 151

linear de Pearson, 85
linear $\mu < 0$ ou $\mu > 0$, 89
matriz de, 86
parcial, 88
Phi, 93
Spearman, 94
Curva, gráfico de, 275
Curva ROC, 103
Cox-Mantel, 41
Cox: Modelo de Risco Proporcional, 42
Crivo (*screening test*), 102

D

D'Agostino, teste, 203
D'Agostino-Pearson, teste, 204
Densidade e tamanho da população, 170
Diagrama de dispersão, gráfico, 267
Diagrama IC, 268
Diagrama Pontual, 271
Discriminante, análise, 20
Distância multivariada, 17
 Euclidiana, 30
 Penrose e Mahalanobis, 31
Distribuição de probabilidades, 95
 binomial, 95
 crivo (*Screening test*), 102
 curva ROC, 103
 exponencial, 98
 F, 107
 hipergeométrica, 96
 normal, 99
 Poisson, 97
 Qui-Quadrado, 108
 Regra de Bayes, 106
 Sensibilidade/Especificidade, 107
 t Student, 109
 Z, 109
Duas amostras independentes, 125
 coeficiente de variação, 144
 Odds Ratio (OR), 139
 Risco Relativo, (RR), 140
 taxa de Incidência, 143
 teste binomial, 131
 teste exato de Fisher, 132
 teste G, 133

teste de Kolmogorov-Smirnov, 134
teste de Mann-Whitney, 135
teste da Mediana, 137
teste de Poisson, 145
teste do Qui-Quadrado, 138
teste t de Student, 126, 128
teste Z, 129, 130
Duas amostras relacionadas, 147
 ANOVA, 150
 teste de McNemar, 152
 teste de Wilcoxon, 157
 teste dos Sinais, 154
 teste Kappa, 156
 teste t de Student, 147, 149

E

Entrando com os dados, 3
Escolha do teste estatístico, sugestões, 331
Escolhendo o teste, 6
Especificidade, 102, 107
Estatística circular, 235
 Correlação, 244
 Descritiva, 236
 Uma amostra,
 Rayleigh, 238
 Teste V, 239
 Duas amostras independentes, 240
 Duas amostras relacionadas, 243
 Três amostras independentes, 242
Estatística descritiva, 159
 Tendência central, dispersão e forma, 160
 Índice de Shannon-Wiener, 162
 Índice de Simpson, 163
Estimação de parâmetros, 165
 da diferença entre duas médias, 167
 da diferença entre duas proporções, 169
 da média, 165
 da proporção, 167
 densidade e tamanho da população, 170
 fenótipos e genótipos, 172
 Fourier (densidade), 173
 Hardy-Weinberg, 171
 Impacto populacional (PIN-ER-t), 174
 Matriz de Crescimento (Leslie), 175
 tamanho da população, 176, 177, 177, 179

Estratificadas, amostras, 11
Euclidiana, distância, 30
Executando o teste, 6
Exponencial, distribuição, 98

F

F, distribuição, 107
Fatorial, anova, 65, 69
Fenótipos e genótipos, 172
Fórmulas estatísticas, 289
 análise multivariada, 289
 análise de sobrevivência, 292
 análise de variância, 294
 correlação, 296
 distribuição circular, 328
 distribuição de probabilidades, 297
 duas amostras independentes, 301
 duas amostras relacionadas, 305
 estatística descritiva, 308
 estimação de parâmetros, 310
 meta-análise, 315
 normalidade, 318
 Qui-Quadrado, 319
 regressão, 321
 tamanho das amostras, 325
 transformação de dados, 327
 uma amostra, 300
Fisher, teste exato, 132
Fourier (densidade), 173
Frequência, tabelas de, 285
Friedman, teste de, 68

G

G, teste, 119, 133
Gehan, 51
Glossário, 339
Gráficos e tabela, 259
 área, 260
 avançados, 287
 box-plot (mediana), 261
 box-plot (média), 262
 caule e folha, 263
 coluna, 265
 curva, 275

diagrama IC, 268
diagrama pontual, 271
dispersão, 267
histograma, 272
manipulação, 283
máximo e mínimo, 276
pirâmide populacional, 277
polar, 281
resultados, 287
setor, 282
tabelas de frequência, 285

H

Hipergeométrica, distribuição, 96
Histograma, gráfico, 272
Hotelling, teste de, 33

I

Imprimindo, 6
Incidência, taxa de, 143
Independência, 125
 duas amostras, 126, 129
 Qui-Quadrado, 138
Índice, 357
Índice de Shannon-Wiener, 162
Índice de Simpson, 163
Instalando o *BioEstat*, 1
Iterações, (runs test), 117

K

Kaplan-Meier, 46
Kappa, teste, 156
Kendall, 91
 coeficiente de concordância, 91
 coeficiente de correlação, 92
Kolmogorov-Smirnov, teste de
 aderência, 120, 205
 independência, 134
 uma amostra, 120
Kruskal-Wallis, teste de, 70

L

Leslie (Matriz de crescimento), 175
 Lilliefors, teste de, 121, 205
 Likelihood ratio, 102, 107
 Linear, correlação, 85, 89
 Linear múltipla, regressão, 217
 Linear simples, regressão, 215
 Logística múltipla, regressão, 226
 Logística simples, regressão, 224
 Log-Rank test, 47

M

Mahalanobis, distância de, 31
 Manipulação, Gráficos, 283
 Mann-Whitney, teste de, 135
 Mantel, teste de, 34
 Mantel-Haenszel, 52, 187
 Máximo e mínimo, gráfico, 276
 McNemar, teste de, 152
 Média, estimação da, 165
 Mediana, teste da, 137
 Meta-Análise, 181
 DerSimonian-Laird, 182
 Dados Contínuos, 185
 Gráficos, 259
 Mantel-Haenszel, 177
 Qui-Quadrado, 191, 192
 p-valor, 190
 Regressão, 194
 Risco Relativo, 195
 Teste G, 197
 Várias Correlações, 198
 Várias Proporções, 200
 Kolmogorov-Smirnov, teste de, 120, 134
 Multivariada, análise, 17
 Múltipla, regressão linear, 217
 Múltipla, regressão logística, 226

N

NNT e NNH, 134, 140
 Normal, distribuição, 99
 Normalidade, 203

D'Agostino, teste de, 203
 D'Agostino-Pearson, teste de, 204
 Kolmogorov-Smirnov, teste de, 205
 Lilliefors, teste de, 205
 Shapiro-Wilk, teste de, 206
 Valores extremos, 207, 209

O

Obtendo ajuda, 7
 Odds-ratio, 139

P

P-valor (Meta-Análise), 190
 Parcial, correlação, 88
 Pearson, correlação de, 85
 Penrose, distância de, 31
 PIN-ER-t, 174
 Pirâmide Populacional, 53, 277
 Ponto de Corte, 105
 População, tamanho, 176
 amostragem simples direta, 176
 amostragem simples inversa, 177
 amostragem múltipla pop. aberta, 179
 amostragem múltipla pop. fechada, 177
 densidade e tamanho, 170
 Poisson, distribuição de, 97
 Poisson, teste, 145
 Polígono de frequência, 285
 Polinomial, regressão, 228
 Prevalência, 102, 107
 Probabilidades, distribuição, 95
 Proporção,
 estimação da, 167, 169
 uma amostra, 116
 duas amostras, 131
 várias proporções, 200

Q

Q, de Cochran, 72
 Qui-Quadrado, 211
 amostras independentes, 211
 análise de resíduos, 212

distribuição do, 108
homogeneidade, aderência, 211
homogeneidade, independência, 211
Mantel-Haensze, 52, 187
partição, 211
tendência, 213
uma amostra, 121, 122
várias proporções, 200

R

Rayleigh (circular, uma amostra), 238
Reamostragem: bootstrap, 15
Referências bibliográficas, 358
Regra de Bayes, 106
Regressão, 215
ajustamento de curvas, 219
análise de resíduos, 221
comparação de duas regressões, 233
homogeneidade, 194
linear múltipla, 217
linear simples, 215
logística múltipla, 226
logística simples, 224
polinomial, 228
seleção de regressores, 230
stepwise, 230
tamanho da amostra, 250
Requerimentos de sistema, 1
Risco relativo, 140, 195
Runs test (Iterações), 117

S

Salvando os dados, 4
Sensibilidade, 102, 107
Setor, gráfico de, 282
Sensibilidade/Especificidade, 107
Shapiro-Wilk, teste, 206
Sinais, teste dos, 154
Spearman, correlação de, 94
Sugestões, escolha do teste estatístico, 331

T

t de Student, distribuição, 109
Tabela de frequência, 285
Tamanho da população, 176
amostragem simples direta, 176
amostragem simples inversa, 177
amostragem múltipla pop. Aberta, 179
amostragem múltipla pop. fechada, 177
Tamanho das amostras, 245
análise de variância, 246
correlação linear, 246
McNemar, 247
proporção
duas amostras, 248
uma amostra, 249
Qui-Quadrado, 250
Regressão linear, 250
Teste *t*,
amostras pareadas, 253
duas amostras independentes, 252
uma amostra, 251
Taxa de incidência, 143
Teste binomial,
duas proporções, 131
uma proporção, 116
Teste da Mediana, 137
Teste de Bartlett, 33
Teste de Cox-Mantel, 41
Teste D'Agostino, 203
Teste D'Agostino-Pearson, 204
Teste de Gehan, 51
Teste de Friedman, 68
Teste de Hotelling, 33
Teste de Kappa, 156
Teste de Kaplan-Meier, 46
Teste de Kolmogorov-Smirnov,
duas amostras independentes, 134
uma amostra, 120, 205
Teste de Kruskal-Wallis, 70
Teste de Lilliefors, 121, 205
Teste de Mantel-Haenszel, 52, 187
Teste de Mann-Whitney, 135
Teste de McNemar, 152
Teste de Parker, 179
Teste de Poisson, 123, 145

Teste de Shapiro-Wilk,
Teste de Schumacher e Eschmeyer, 177
Teste de Wilcoxon, 157
Teste do Qui-Quadrado,
 amostras independentes, 138
 homogeneidade, 191
 partição, 211
 tendência, 213
 uma amostra, 121, 122
 várias proporções, 200
Teste dos sinais, 154
Teste exato de Fisher, 132
Teste *G*,
 amostras independentes, 133
 uma amostra, 119
 homogeneidade, 197
Teste Kappa, 156
Teste *Q* de Cochran, 72
Teste *t*,
 duas amostras independentes, 126, 128
 duas amostras relacionadas, 147, 149
 uma amostra, 111, 113
Teste *V* (circular, uma amostra), 239
Teste *Z*,
 duas amostras independentes, 129, 130
 uma amostra, 114, 115
Tira de bota (Bootstrap), 75
Transformação de dados, 255

U

Uma amostra, 111

binomial, teste, 116
Iterações, (runs test), 117
teste *G*, 119
teste de Kolmogorov-Smirnov, 120
teste de Lilliefors, 121
teste do Qui-Quadrado, 121, 122
teste de Poisson, 123
teste *t* de Student, 111, 113
teste *Z*, 114, 115

V

Valores extremos, 207, 209
Valores preditivos, 102, 107
Várias proporções, 200
Variação, coeficiente de, 144
Variância, análise de, 59

W

Wilcoxon, teste de, 157
Wilcoxon-Mann-Whitney, 135

Z

Z, distribuição, 109
Z, teste para duas amostras independentes, 129
Z, para uma amostra, 114