

8676i Soares, José Francisco

Introdução à estatística médica / José Francisco Soares,  
Arminda Lucia Siqueira. - 1.ed. - Belo Horizonte: Departamento de Estatística - UFMG, 1999.  
vii, 300p.:i1

1. Estatística médica 2. Bioestatística I. Siqueira, Arminda Lucia. II. Título.

CDD610.21  
CDU61:31

Ficha catalográfica preparada pelo Centro de Extensão da Escola de Biblioteconomia da UFMG

Registro do ISBN: solicitado

Apoio: Pró-Reitoria de Graduação - UFMG

Capa: Lais Freire dos Reis

# Sumário

<b>1</b>	<b>O Papel da Estatística na Medicina</b>	<b>1</b>
1.1	Introdução . . . . .	1
1.2	Alguns exemplos . . . . .	1
1.3	Bioestatística . . . . .	11
1.3.1	Planejamento . . . . .	12
1.3.2	Análise . . . . .	12
1.4	Considerações finais . . . . .	13
<b>2</b>	<b>Organização da Pesquisa Médica</b>	<b>14</b>
2.1	Introdução . . . . .	14
2.2	Estudo descritivo . . . . .	16
2.2.1	Estudo de casos . . . . .	16
2.2.2	Estudo de uma série de casos . . . . .	17
2.2.3	Estudo baseado em dados institucionais . . . . .	18
2.3	Estudo caso-controle . . . . .	18
2.4	Estudo de coorte . . . . .	20
2.4.1	Seleção das coortes . . . . .	21
2.5	Ensaios clínicos aleatorizados . . . . .	22
2.5.1	Aleatorização . . . . .	24
2.5.2	Como implementar a alocação aleatória . . . . .	25
2.6	Estudos comparativos . . . . .	26
2.6.1	Grupo controle . . . . .	26
2.6.2	Confusão de efeitos . . . . .	27
2.6.3	Controle de fatores de confusão . . . . .	28
2.6.4	Generalização dos resultados . . . . .	29
2.7	Exercícios . . . . .	29

<b>3 Descrição e Apresentação de Dados</b>	<b>33</b>		
3.1 Introdução . . . . .	33	4.5.1 Sensibilidade e especificidade . . . . .	99
3.2 Conceitos básicos . . . . .	33	4.5.2 Valor das predições . . . . .	102
3.2.1 Variável . . . . .	33	4.5.3 Decisões incorretas . . . . .	104
3.2.2 Tipos de variáveis . . . . .	34	4.6 Combinação de testes diagnósticos . . . . .	106
3.2.3 Dados brutos . . . . .	36	4.6.1 Formas de combinação de testes . . . . .	106
3.3 Organização e apresentação de dados . . . . .	39	4.6.2 Necessidade da combinação de testes . . . . .	111
3.3.1 Tabelas de freqüências . . . . .	39	4.7 Testes diagnósticos baseado em variáveis contínuas . . . . .	113
3.3.2 Gráficos . . . . .	46	4.7.1 Efeito do ponto de corte na qualidade de um teste diagnóstico . . . . .	113
3.4 Síntese numérica . . . . .	52	4.7.2 Escolha do ponto de corte . . . . .	116
3.4.1 Medidas de tendência central . . . . .	52	4.8 Escolha entre testes diagnósticos . . . . .	117
3.4.2 Medidas de variabilidade . . . . .	55	4.9 Considerações finais . . . . .	119
3.4.3 Coeficiente de variação . . . . .	58	4.10 Exercícios . . . . .	120
3.4.4 Escore padronizado . . . . .	59		
3.4.5 Percentis . . . . .	61	<b>5 Caracterização Estatística de Variáveis</b>	<b>131</b>
3.4.6 Outras medidas descritivas . . . . .	63	5.1 Introdução . . . . .	131
3.5 Outros tópicos . . . . .	64	5.2 Conceitos fundamentais . . . . .	132
3.5.1 Coleta de dados . . . . .	64	5.2.1 Variáveis aleatórias e suas distribuições de probabilidade . . . . .	132
3.5.2 Banco de dados . . . . .	65	5.2.2 População e amostra . . . . .	137
3.5.3 Arredondamento . . . . .	67	5.3 Modelo de Poisson . . . . .	137
3.5.4 Boxplot . . . . .	68	5.4 Modelo de Gauss . . . . .	141
3.5.5 Observações atípicas ( <i>outlier</i> ) . . . . .	70	5.4.1 Curva de Gauss . . . . .	141
3.6 Exemplo comentado: Refluxo vesicureteral primário em crianças . . . . .	74	5.4.2 Distribuição gaussiana padrão . . . . .	142
3.6.1 Descrição da amostra no início do estudo . . . . .	75	5.5 Verificação da adequação do modelo . . . . .	148
3.6.2 Acompanhamento . . . . .	75	5.5.1 Modelo de Poisson . . . . .	148
3.7 Exercícios . . . . .	78	5.5.2 Modelo de Gauss . . . . .	150
<b>4 Probabilidade e Avaliação de Testes Diagnósticos</b>	<b>89</b>	5.6 Faixa de referência . . . . .	153
4.1 Introdução . . . . .	89	5.6.1 Obtenção da faixa de referência . . . . .	155
4.2 Probabilidade: conceitos fundamentais . . . . .	90	5.7 Exemplos comentados . . . . .	159
4.2.1 Espaço amostral e eventos . . . . .	90	5.7.1 Cartas para controle de processos em hospitais . . . . .	159
4.2.2 Cálculo de probabilidades . . . . .	91	5.7.2 Rastreamento de doadores de sangue . . . . .	163
4.2.3 Tipos especiais de eventos . . . . .	93	5.8 Exercícios . . . . .	166
4.2.4 Propriedades da probabilidade . . . . .	94		
4.3 Probabilidade condicional . . . . .	95	<b>6 Comparando Dois Grupos</b>	<b>171</b>
4.4 Eventos independentes . . . . .	97	6.1 Introdução . . . . .	171
4.5 Qualidade de testes diagnósticos . . . . .	99	6.2 Conceitos fundamentais . . . . .	176

6.2.3	Erros do tipo I e II, nível de significância e poder do teste . . . . .	179	7.2.3	Intervalo de confiança . . . . .	237
6.2.4	Probabilidade de significância (valor-p) . . . . .	180	7.2.4	Distribuição da média amostral . . . . .	238
6.3	Resposta dicotômica: amostras independentes . . . . .	181	7.3	Medida do efeito: resposta contínua . . . . .	241
6.3.1	Teste qui-quadrado ( $\chi^2$ ) . . . . .	181	7.3.1	Amostras pareadas . . . . .	241
6.3.2	Teste qui-quadrado com correção de continuidade	185	7.3.2	Amostras independentes . . . . .	243
6.3.3	Teste exato de Fisher . . . . .	186	7.4	Medida do efeito: resposta dicotômica . . . . .	244
6.3.4	Teste Z para comparação de proporções . . . . .	187	7.4.1	Risco relativo . . . . .	246
6.4	Resposta dicotônica: amostras pareadas . . . . .	189	7.4.2	Razão das chances . . . . .	247
6.4.1	Teste de McNemar . . . . .	190	7.5	Exemplo comentado: Desnutrição hospitalar . . . . .	251
6.5	Resposta contínua: amostras independentes . . . . .	192	7.6	Considerações finais . . . . .	253
6.5.1	Teste t . . . . .	192	7.6.1	Significância estatística e significância clínica .	253
6.5.2	Teste Z para comparação de médias . . . . .	199	7.6.2	Equivalência entre teste de hipóteses e intervalo de confiança . . . . .	254
6.6	Resposta contínua: amostras pareadas . . . . .	201	7.6.3	Diferentes tempos de seguimento . . . . .	254
6.7	Testes não-paramétricos . . . . .	204	7.6.4	O termo <i>odds-ratio</i> . . . . .	255
6.7.1	Teste de Mann-Whitney . . . . .	205	7.7	Exercícios . . . . .	255
6.7.2	Teste de Wilcoxon . . . . .	206			
6.8	Exemplo comentado: Imunogenicidade de uma vacina	208	<b>A Tabelas</b>		<b>260</b>
6.9	Considerações práticas sobre testes de hipóteses . . . . .	211	<b>B Respostas dos exercícios</b>		<b>266</b>
6.9.1	Valor-p . . . . .	211	<b>C Referências bibliográficas</b>		<b>278</b>
6.9.2	Hipóteses bilaterais versus unilaterais . . . . .	211			
6.9.3	Conclusões a partir de testes de hipóteses . . . . .	214			
6.9.4	Significância estatística e significância clínica .	214			
6.9.5	Fatores de confusão . . . . .	215			
6.9.6	Apresentação dos resultados . . . . .	215			
6.9.7	Pressupostos dos testes . . . . .	215			
6.10	Considerações finais . . . . .	217			
6.10.1	Denominações de testes . . . . .	217			
6.10.2	Amostras pareadas e amostras independentes .	218			
6.10.3	Sobre pareamentos . . . . .	218			
6.10.4	Amostragem . . . . .	219			
6.10.5	Extensões . . . . .	219			
6.11	Exercícios . . . . .	221			
<b>7</b>	<b>Medida do Efeito de uma Intervenção ou Exposição</b>	<b>234</b>			
7.1	Introdução . . . . .	234			
7.2	Conceitos fundamentais . . . . .	235			
7.2.1	Parâmetro de interesse . . . . .	235			
7.2.2	Estimadores . . . . .	236			

# Prefácio

Uma consulta aos periódicos mais relevantes de qualquer área médica mostra que a Estatística é hoje a linguagem predominante para a comunicação de resultados recentes que devem ser incorporados à prática médica. A Estatística é também instrumento básico para a realização das pesquisas que geram estes resultados. Além disso, a capacidade de atualização é característica essencial do profissional bem formado.

Por todos estes motivos, uma disciplina de Estatística é oferecida nos cursos de Medicina, que infelizmente, durante longos anos foi tratada nos cursos médicos como um conjunto de regras de apresentação de dados. Nada se falava sobre formas de organização de estudo clínico, nem das análises de dados típicos da pesquisa médica. Isto tornava a disciplina desinteressante e por outro lado não habilitava os alunos no essencial: ler com facilidade e de forma crítica e proveitosa a literatura médica recente. Desde 1988, um grupo de professores do Departamento de Estatística da UFMG vem procurando adequar a maneira de ensinar a esta necessidade. Este livro é o fruto de todo este processo.

O primeiro capítulo ilustra a inserção da Estatística na literatura médica através de exemplos simples e relevantes. O Capítulo 2 é sobre o planejamento do estudo clínico e é uma introdução de um tema de grande importância. O conhecimento médico avança através de estudos bem planejados e não através de análise de dados coletados de acordo com conveniências administrativas. O Capítulo 3 trata da síntese numérica e gráfica dos dados tendo como referência a necessidade de criação e uso de padrões de referência em estudos clínicos. O Capítulo 4 apresenta índices que medem a qualidade de testes diagnósticos: sensibilidade e especificidade. Discutimos a questão de incertos e erros no diagnóstico clínico, mostrando que estes conceitos são também expressões de conceitos estatísticos clássicos. O conteúdo do Capítulo 5 pode parecer excessivamente teórico, mas não há como tratar problemas reais sem tais informações. São introduzidas duas importantes distribuições: Poisson e gaussiana. A principal aplicação apresentada no capítulo é a construção de faixas de referências de uso rotineiro na prática médica. Finalmente, nos capítulos 6 e 7 são apresentadas formas estatísticas para comparação de tratamentos ou

genericamente de grupos de pacientes. Vários testes estatísticos que aparecem na literatura estão ali descritos.

Fizemos um esforço de trazer exemplos de nosso trabalho de interação com pesquisadores e clínicos que temos atendido ao longo do tempo no Departamento de Estatística da UFMG. Foram experiências muito ricas e gratificantes que ficaram aqui registradas e que representam o papel da interface da Estatística. Além disso, incluímos diversos exemplos cuidadosamente selecionados da literatura, vários deles provenientes de pesquisa com dados brasileiros. O exemplo da Seção 6.8 refere-se a um excelente trabalho publicado no respeitado periódico *The Lancet* e foi apresentado em um seminário por uma de nossas melhores alunas da pós-graduação da Faculdade de Medicina da UFMG. Sua inclusão é uma homenagem à Laura Maria Braga Borges de Mattos que morreu vítima de acidente em abril de 1998.

Ao apresentar este material em português, com ilustrações dentro da prática médica, esperamos estar contribuindo para um ensino de qualidade da Estatística para a Medicina. O texto apresenta a parte elementar da Estatística Médica e, embora seguindo o mesmo conteúdo de outros textos de Estatística, seu enfoque é diferente. Começa com o problema médico e depois apresenta as técnicas estatísticas necessárias para sua solução. Ao final dos capítulos estão propostos muitos exercícios para possibilitar a fixação dos conceitos apresentados e em alguns casos desenvolver o raciocínio necessário em uma análise de dados. Pode ser utilizado em disciplinas de graduação e, se apropriadamente complementado, em pós-graduação. Pode ainda ser útil para profissionais que atuem ou tenham interesse na área da Saúde.

Um trabalho como este não se concretiza sem a ajuda de muitas pessoas. Agradecemos às variás gerações de alunos que usaram o texto na forma de apostila e a todos que com suas sugestões e incentivo ajudaram a produzir um texto mais relevante. Embora fosse nosso desejo não citar nomes, dada sua contribuição especial, queremos agradecer à acadêmica de Medicina Fabiana Barreto Utsch de Matos, que aprendeu Estatística, editoração, padronização e com enorme competência e dedicação nos possibilitou terminar este livro.

*José Francisco Soares e Arminda Lucia Siqueira  
Belo Horizonte, março de 1999*

# Capítulo 1

## O Papel da Estatística na Medicina

### 1.1 Introdução

O objetivo deste capítulo é ilustrar o papel que conceitos e técnicas estatísticas têm na formação do médico, na sua prática profissional e de modo especial no avanço do conhecimento na área médica.

Este objetivo será alcançado principalmente através da apresentação e discussão dos resultados de estudos clínicos relevantes. Através destes exemplos, argumenta-se que conceitos estatísticos têm importante papel a desempenhar, tanto no estudo da saúde de populações, como no tratamento de pacientes individuais.

Outros assuntos correlatos, como o possível uso de computadores no diagnóstico clínico e o impacto do processo de quantificação da Medicina moderna, serão sinteticamente apresentados.

### 1.2 Alguns exemplos

Como em qualquer outra área da ciência, os avanços no conhecimento médico são primeiramente relatados em artigos publicados em revistas científicas. Cada artigo, antes de ser publicado, é revisto por um grupo de especialistas da área que, através de avaliação criteriosa, procura garantir a qualidade dos trabalhos publicados e, consequentemente, o prestígio da revista e da sociedade científica associada.

Nesta seção e no resto do texto, usaremos como exemplos artigos de periódicos médicos respeitados internacionalmente: *New England Journal of Medicine*, *Journal of The American Medical Association*, *The Lancet*, *British Medical Journal*, etc. Estas são também fontes da mídia para notícias sobre avanços na Medicina, tão presentes no noticiário diário.

### Exemplo 1.1: Aspirina e prevenção de doença coronariana

As propriedades analgésicas, antitérmicas e antiinflamatórias da aspirina foram verificadas há muitos anos. Só muito recentemente, entretanto, foi reconhecido o seu potencial na redução do risco de doenças cardíovasculares. Além de um possível mecanismo bioquímico que justificasse o efeito protetor da aspirina, evidências preliminares fornecidas por alguns estudos clínicos e a crença no seu efeito, difundida na sociedade americana, criaram a necessidade e a oportunidade de se verificar, cientificamente, a existência ou não deste efeito.

Com esta finalidade, foi organizado em 1982 um estudo conhecido como *Physicians' Health Study*. Neste estudo, 22.071 médicos americanos voluntários com idade entre 40 e 84 anos foram divididos aleatoriamente em dois grupos. O primeiro grupo, com 11.037 médicos, tomou a cada dois dias aspirina (Bufferin, Bristol-Myers Products, 325 mg), enquanto o segundo, com 11.034 médicos, tomou placebo: um comprimido semelhante ao da aspirina, mas sem o seu princípio ativo.

Depois de cinco anos o comitê que, independentemente dos pesquisadores responsáveis pela rotina do projeto, acompanhava e monitorava os dados obtidos, decidiu terminar o estudo. Os resultados preliminares mostravam que os efeitos protetores da aspirina eram evidentes. Essa informação deveria, pois, ser comunicada aos clínicos e à população em geral.

A Tabela 1.1, reproduzida do artigo original (relato final), inclui apenas os casos confirmados de infarto do miocárdio e acidente vascular cerebral (AVC) e mostra os dados que fundamentalmente contribuíram a decisão de parar o estudo e divulgar os resultados.

A proporção de infartos no grupo que tomava aspirina 139/11.037, enquanto no outro grupo era de 239/11.034. A questão crucial do

Tabela 1.1: Resultados relatados pelo *Steering Committee of Physicians' Health Study Group* em 1989

Resposta	Aspirina	Placebo	Risco	Intervalo	Valor-p
Infarto	139	239	0,53	0,42-0,67	< 0,00001
AVC	119	98	1,15	0,84-1,58	0,41

estudo, respondida pelo uso de métodos estatísticos, foi: será que a diferença observada nas proporções é devida ao uso de aspirina ou aparece simplesmente devido ao acaso? Em outras palavras, o mesmo efeito seria observado em outro grupo de pessoas se o estudo fosse repetido? Testes estatísticos mostram que a probabilidade do efeito observado ser proveniente apenas do acaso é menor que 1 em 10.000. Ou seja, há uma forte evidência do efeito protetor da aspirina.

Quantifica-se o efeito da aspirina calculando-se a razão do risco de infarto nos dois grupos. Neste estudo obteve-se 0,53, ou seja, o risco de quem toma aspirina é quase metade do de quem não toma.

Um estudo desta natureza é cercado de todos os cuidados: no seu planejamento, na sua execução e na análise dos dados. A Estatística fornece, além dos métodos para tomada de decisões na presença de incerteza, as formas de planejamentos do estudo. Através da alocação aleatória dos participantes aos grupos, obtiveram-se dois grupos semelhantes em todas as suas características medidas ou não. Entretanto, a decisão final de se usar este resultado na prática clínica considera outros fatores além daqueles obtidos através da análise estatística dos dados, como, por exemplo, a facilidade de uso da droga e seu preço.

### Exemplo 1.2: Qualidade de testes para detecção do HIV

A Síndrome da Imunodeficiência Adquirida (AIDS), logo que apareceu em 1981, constituiu-se de imediato em um sério problema de saúde pública para o Brasil e outros países do mundo. O sistema de saúde não estava preparado para a nova e difícil tarefa de diagnosticar e atender os doentes. Além disto, a possibilidade de contágio através de transfusão de sangue trouxe uma preocupação extra.

Em início de 1985, tornaram-se disponíveis comercialmente os primeiros testes ELISA para determinar a presença do vírus da imunodeficiencia humana (HIV). A partir daí, garantiu-se maior segurança para pessoas dependentes de hemoderivados, como os hemofílicos.

O sucesso do teste criou, entretanto, a falsa impressão de que ele era perfeito, ou seja, podia-se acreditar completamente no seu resultado. Logo percebeu-se a existência de grande número de resultados positivos incorretos, tornando necessárias alternativas diagnósticas, como melhorias técnicas, desenvolvimento de novas metodologias e combinação de testes.

O Ministério da Saúde brasileiro, procurando aumentar a confiabilidade para emissão de laudos dos testes e sistematizar os procedimentos utilizados, publicou no Diário Oficial de 19 de janeiro de 1998 o regulamento técnico para diagnóstico sorológico do HIV em indivíduos acima de dois anos a ser adotado por todos os serviços de saúde, públicos e privados.

Segundo essa sistematização, cada amostra de sangue deverá ser submetida a dois testes de triagem (por exemplo ELISA) distintos, com抗ígenos ou metodologia diferentes. Se negativos, a amostra é considerada negativa para HIV. Se discordantes, os testes devem ser repetidos em duplicata. No caso de positividade em ambos os testes, é necessária a realização de um teste confirmatório (imuno-fluorescência indireta ou Western Blot) que, uma vez positivo, obriga a coleta de uma segunda amostra para repetição dos testes de triagem e confirmação dos resultados da primeira amostra.

Estes testes, ao resolverem o problema de detecção, criaram vários outros, associados à sua utilização. Faz sentido seu uso em larga escala como defendido por alguns grupos de opinião? Como comunicar um resultado positivo do teste a uma pessoa sem outras evidências clínicas da doença? É ético avisar o companheiro de uma pessoa com resultado positivo? Para enfrentar estas questões ajuda bastante conhecer o grau de confiabilidade destes testes.

### Exemplo 1.3: Cirurgia conservadora em câncer de mama

A partir da década de 20, começaram a surgir evidências de que era possível, no caso de câncer de mama, optar por uma abordagem cirúrgica menos radical que a mutiladora mastectomia preconizada por Halsted no fim do século passado. Verificou-se que se podia conservar a mama da paciente, melhorando desta forma a qualidade de sua vida, sem piorar o prognóstico da doença.

Com o passar do tempo, evidências cada vez mais convincentes foram aparecendo. Hoje, a abordagem cirúrgica conservadora, seguida de tratamento radioterápico com ou sem quimioterapia, tornou-se o tratamento padrão em várias situações da patologia.

Nos Estados Unidos, entretanto, em 1976, a equipe de Bernard Fisher, ao analisar as práticas que conservavam o seio da paciente, era bastante crítica:

*“Apesar de esforços antigos e recentes, questões importantes continuam sem resposta, o que impede a aceitação universal do uso de uma abordagem mais conservadora no câncer de mama. Mesmo os poucos estudos clínicos aleatorizados, que avaliaram as diferentes opções cirúrgicas que mantêm o seio da paciente, não conseguiram estabelecer além da dúvida razoável, o valor desta opção terapêutica”*

Estavam bem estabelecidas a necessidade e a oportunidade de organização de um ensaio clínico sobre o assunto. Por um lado, havia muitas evidências históricas de eficácia, por outro, dúvidas sobre o valor real da opção.

Sob a coordenação daquele famoso cirurgião, de abril de 1976 até janeiro de 1984, 2.163 mulheres com câncer de mama em estágios I e II foram aleatoriamente alocadas a uma das três opções: mastectomia total (MT), tumorectomia e tumorectomia seguida de radioterapia (Tc+RTx). Todos as pacientes fizeram estadiamento axilar e aquelas nas quais se constataram linfonodos positivos receberam, adicionadamente, quimioterapia. A completa descrição dos detalhes de seleção de pacientes e de seu tratamento pode ser lida no artigo original, publicado no New England Journal of Medicine em março de 1985.

A Figura 1.1 mostra que os dois tratamentos não se equivalem no que se refere ao tempo até a recidiva ou tempo livre de doença. A inclusão de radioterapia nas pacientes que se submeteram a tumorectomia aumenta o seu tempo de sobrevida.

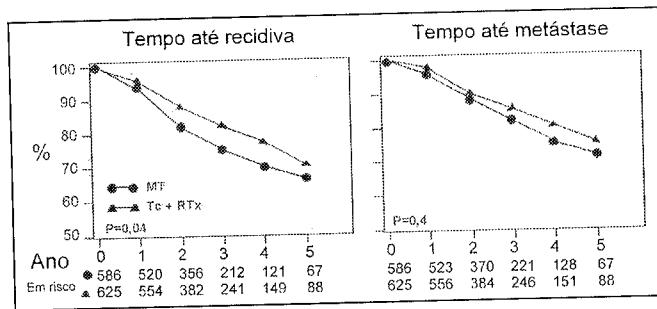


Figura 1.1: Tempo até recidiva e tempo até metástase para um grupo de pacientes que receberam dois tipos de tratamento: mastectomia total (MT) e tumorectomia seguida de radioterapia (Tc+ RTx)

Neste exemplo, a Estatística fornece uma forma eficiente de quantificação da resposta de interesse (recidiva e metástase) e de métodos de comparação das curvas referentes a essas variáveis para diferentes tratamentos. Como no caso da aspirina (Exemplo 1.1), a forma de organização deste estudo é o acompanhamento prospectivo das pacientes e alocação aleatória aos tratamentos, procedimentos básicos para a produção de uma pesquisa com grande poder de resolução da questão em pauta.

#### Exemplo 1.4: Consumo de carne e câncer do cólon

A Figura 1.2 mostra a incidência (por 100.000 mulheres) de câncer no cólon e o consumo médio de carne na população de vários países (Armstrong & Doll, 1975). Uma associação é claramente sugerida: menor consumo de carne, maior incidência de câncer no cólon.

A Estatística desempenha aqui o seu papel mais tradicional, o de instrumento de coleta de dados para o funcionamento do Estado. A produção total de carne e a quantidade exportada são números rotineiramente coletados. Por outro lado, todo país tem uma lista de doenças que são de registro obrigatório. Estes dados são coletados

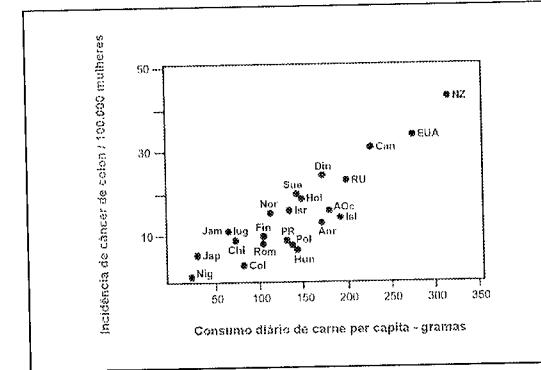


Figura 1.2: Incidência de câncer de cólon (1975) em função do consumo de carne nos seguintes países: Alemanha Ocidental (AOc), Alemanha Oriental (AOr), Canadá (Can), China (Chi), Colômbia (Col), Dinamarca (Din), Estados Unidos da América (EUA), Finlândia (Fin), Holanda (Hol), Hungria (Hun), Islândia (Isl), Israel (Isr), Iugoslávia (Iug), Jamaica (Jam), Japão (Jap), Nigéria (Nig), Nova Zelândia (NZ), Noruega (Nor), Polônia (Pol), Porto Rico (PR), Reino Unido (RN), Romênia (Rom), Suécia (Sue)

dos porque trazem informação útil para o planejamento econômico e da saúde da população. Podem, entretanto, como o gráfico mostra, produzir indicações de associações que merecem estudos mais aprofundados.

A partir desses dados, nada se pode concluir definitivamente, em termos de associação do fator consumo de carne com a doença câncer do cólon. Os valores usados são médias dos indivíduos e, portanto, não se sabe se quem consome a carne é a mesma pessoa que desenvolve o câncer. Além disto, outras variáveis não consideradas como, por exemplo, a quantidade de fibras na dieta, podem ser as responsáveis pelo câncer.

#### Exemplo 1.5: Fumo e câncer de pulmão

Um grande número de estudos clínicos é dedicado à pesquisa da etiologia das doenças, isto é, de sua origem. Etapa importante neste trabalho é a descoberta de fatores, usualmente chamados de fatores

de risco, associados com o aparecimento da doença.

Uma dificuldade básica nesse tipo de estudo é que não se pode, por motivos principalmente éticos mas também operacionais, alocar um fator aos indivíduos. É necessário, pois, usar dados obtidos de suas histórias. Examinam-se seus hábitos pessoais e as opções clínicas sugeridas por seus médicos. Os estudos que seguem essas linhas gerais são chamados de *estudos observacionais*.

Apresentamos abaixo, sinteticamente, os resultados dos estudos conduzidos por R. Doll e A. B. Hill, bioestatísticos britânicos, sobre a associação entre fumo e câncer do pulmão. Estas pesquisas iniciaram-se devido ao grande aumento da mortalidade causada por este câncer, observado na Inglaterra desde o fim dos anos 20.

No primeiro estudo, publicado em 1950, compararam-se dois grupos de pacientes. O primeiro, constituído de todos os casos de câncer de pulmão de um conjunto de hospitais londrinos. Assistentes socionômicas, especialmente treinadas para o estudo, entrevistaram todos os pacientes internados com diagnóstico de câncer de pulmão, levantando sua história clínica e seus hábitos tabagistas. O segundo grupo, chamado de grupo controle, foi constituído por pacientes dos mesmos hospitais, sem diagnóstico de câncer, na mesma faixa etária, do mesmo sexo e da mesma região de residência que os casos anteriores.

A Tabela 1.2 mostra os resultados obtidos. A associação é clara e forte.

Tabela 1.2: Número de fumantes e não fumantes entre pacientes do sexo masculino com diagnóstico de câncer pulmonar e controles

Grupo	Fumantes	Não fumantes
Câncer pulmonar	647	2
Controle	622	27

A forma de organização deste estudo é chamada de *Estudo Caso-Controle*. À época, esta metodologia era muito criticada, pois, com facilidade, podia produzir associações espúrias. Por isto, a evidência, embora clara e forte, não foi convincente o suficiente.

Assim, os mesmos autores iniciaram um segundo estudo, cuja característica básica é o fato de que o acompanhamento dos pacientes

foi prospectivo. Este tipo de planejamento é chamado de *Estudo de Coorte*.

Em outubro de 1951, os pesquisadores enviaram um questionário simples a todos os médicos da Inglaterra, aproximadamente 60.000 indivíduos. O questionário perguntava se o respondente já havia fumado ou não. Em caso afirmativo, pedia informações sobre o que e quanto. Mais de dois terços dos médicos responderam com detalhe suficiente para que seus dados pudesse ser incluídos no estudo.

As respostas possibilitaram aos pesquisadores classificar cada respondente como fumante ou não fumante. A definição de não fumante usada foi: um não fumante é uma pessoa que fumou até no máximo um cigarro diário, em média, por período inferior a um ano.

Através de complexo sistema de acompanhamento, observaram-se nos primeiros 10 anos, 136 mortes associadas ao câncer pulmonar entre os médicos incluídos no estudo. Destas, apenas 3 eram de não fumantes. Para equalizar os períodos de acompanhamento nos vários grupos, trabalhou-se com a taxa de incidência por 1.000 pessoas-ano de exposição. Os valores desta taxa estão na Tabela 1.3. O risco de morte por câncer pulmonar das pessoas que fumam mais de 25 cigarros diários é quase 32 vezes maior do que o mesmo risco para quem não fuma.

Tabela 1.3: Taxa de mortalidade por 1.000 pessoas-ano devida a câncer pulmonar (número de mortes entre parênteses) para não fumantes e fumantes

Não fumantes	Cigarros diários		
	1-14	15-24	25+
0,07(3)	0,57(22)	1,39(54)	2,27(57)

Estes dois planejamentos, estudo caso-controle e estudo de coorte, são as formas usuais de organização da pesquisa etiológica. Muitos outros estudos como os descritos acima foram feitos em populações variadas para se verificar o possível papel do fumo. Em todos, a evidência foi clara. Hoje, o papel de agente causador do câncer de pulmão é amplamente reconhecido e justifica a militância cada vez mais organizada contra o fumo.

### Exemplo 1.6: Distribuição do ácido úrico em homens

Talvez a tarefa mais rotineira no exercício da Medicina seja a comparação de um valor de uma medida de importância clínica, tais como temperatura, pressão, ritmo cardíaco e parâmetros hematológicos, com um padrão. Como há entre diferentes indivíduos uma grande variação nessas medidas, mesmo quando pertencentes a um grupo homogêneo, o padrão não é um número, mas uma faixa.

Os dados para a construção desta faixa são obtidos através da medição da característica de interesse em um grupo representativo de pessoas supostamente saudáveis. Estes valores são sintetizados em um gráfico, a partir do qual constrói-se a faixa de referência.

Este esquema foi seguido por pesquisadores britânicos na determinação de uma faixa de referência para o ácido úrico em homens. O resultado foi publicado na conceituada revista *The Lancet* (Finn et al., 1966). O valor do ácido úrico foi medido em 267 doadores de sangue, gozando de boa saúde, obtendo-se os dados apresentados na Figura 1.3.

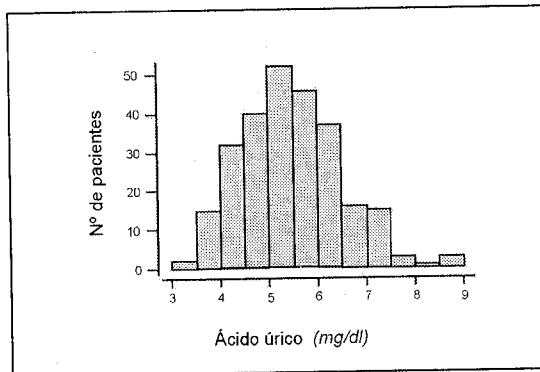


Figura 1.3: Distribuição de ácido úrico (mg/dl) em 267 homens

A faixa de referência é escolhida fixando-se uma pequena porcentagem de pacientes cujos resultados não serão incluídos e determinando-se o intervalo que engloba todos os outros indivíduos. Por exemplo, para os dados do ácido úrico excluindo-se 2,5% de cada extremo, obtém-se a faixa [3,7 - 7,5] mg/dl.

Nesta situação, é o conceito de distribuição estatística que ajuda no estudo do fato médico.

### 1.3 Bioestatística

Métodos estatísticos são essenciais no estudo de situações em que as variáveis de interesse estão sujeitas, inherentemente, a flutuações aleatórias. Este é o caso da Medicina. Mesmo tomando-se um grupo de pacientes homogêneos, observa-se grande variabilidade, por exemplo, no tempo de sobrevida após um tratamento adequado. Dosagens de características hematológicas flutuam não só entre indivíduos, como também no mesmo indivíduo em ocasiões diferentes. Na realidade, há variações entre diferentes pacientes para qualquer variável de interesse clínico. Portanto, para se estudar problemas clínicos, é necessária uma metodologia capaz de tratar a variabilidade de forma adequada.

Convencionou-se chamar de Bioestatística o conjunto de métodos estatísticos usados no tratamento da variabilidade nas ciências médicas e biológicas. A Bioestatística fornece métodos para se tomar decisões ótimas na presença de incerteza, estabelecendo faixas de confiança para a eficácia dos tratamentos e verificando a influência de fatores de risco no aparecimento de doenças.

Há uma aparente contradição entre a Estatística e a Medicina. Por um lado, todo ato médico é feito no sentido de se fornecer o melhor a um paciente específico. Isto se contrapõe à necessidade de consideração de um conjunto de indivíduos para a construção de um fato estatístico.

Deve-se notar, entretanto, que ao tratar um paciente, o médico se vale da experiência de eventos anteriores, vivenciada pessoalmente ou transmitida por outros através de livros e artigos. Assim, a Estatística pode ser vista como ferramenta de organização e validação do conhecimento médico.

Resumidamente, o objeto de estudo da Bioestatística é o planejamento e a análise de estudos médicos e biológicos. Ambas as etapas visam reduzir a possibilidade de vícios no estudo.

### 1.3.1 Planejamento

A primeira classificação útil das formas de organização de estudos clínicos é dada pela dicotomia: estudos descritivos versus estudos comparativos.

Os primeiros descrevem uma dada situação, sem preocupação de comparação. É o caso do exemplo sobre o consumo de carne e câncer no cólon (Exemplo 1.4) e das enquetes destinadas à criação de padrões de referência. Qualquer conclusão não meramente descritiva desses estudos está sujeita a críticas.

No segundo tipo, existe internamente um grupo de comparação. É útil fazer a distinção entre estudos observacionais, aqueles feitos com dados produzidos pela história clínica de indivíduos, e estudos experimentais, em que o pesquisador interfere na alocação do tratamento. Tanto o estudo sobre a aspirina (Exemplo 1.1) como o estudo sobre cirurgia conservadora de câncer de mama (Exemplo 1.3) são experimentais e, portanto, capazes de produzir evidências potencialmente muito convincentes. Já o exemplo sobre o câncer no pulmão (Exemplo 1.5) pertence à categoria de estudos observacionais.

### 1.3.2 Análise

O número de técnicas usadas em Bioestatística é muito grande. Vamos tratar neste texto apenas de formas elementares de análise de dados. Por exemplo, no estudo da aspirina, a resposta considerada de maior interesse era a ocorrência ou não de infarto, uma variável dicotómica. Como vimos na reprodução da tabela do artigo original, a técnica utilizada para a análise dos dados foi o *teste qui-quadrado* e o cálculo da *razão das chances*. O teste nos permite decidir se o tratamento é eficaz e a razão das chances medir o seu efeito.

Se a resposta de interesse não é dicotómica, embora o objetivo da análise seja o mesmo, por exemplo, testar a existência de efeito e quantificar a sua eficácia, são usados outros testes, entre eles o *teste t*. Há ainda testes para comparar curvas de sobrevidas, como no Exemplo 1.3 sobre cirurgia conservadora no câncer de mama.

Um planejamento e análise bem feitos são cruciais. Críticas sobre a inadequação de um estudo impedem que seu resultado modifique a prática clínica. Neste caso, todo esforço despendido no desenvolvimento do estudo estará perdido.

## 1.4 Considerações finais

Neste capítulo introduzimos, sinteticamente, algumas idéias importantes associadas à Bioestatística. Seu tratamento mais completo não é objetivo deste texto.

A Bioestatística pode ser vista como ferramenta de avaliação crítica da evidência clínica ou de acumulação criteriosa de conhecimentos médicos. A atualização dos conhecimentos se dá através da interação com especialistas no assunto, mas principalmente por leitura constante da literatura recente na área. A literatura médica usa hoje um jargão estatístico que será explicado ao longo dos capítulos deste texto. Acompanhar esta literatura é fundamental para todos os que querem ser bons profissionais. Na Medicina, entretanto, a necessidade é mais premente, já que o conhecimento muda com grande rapidez. Assim, a longo prazo, a capacidade de ler e entender é fundamental.

A tecnologia e a informática estão mudando a forma de exercício de muitas profissões. O impacto na Medicina em termos de aparelhagens para diagnóstico é impressionante. Um bom exemplo é a rapidez com que se incorporaram o raio-X, o ultra-som, a tomografia computadorizada e a ressonância magnética. Existe, entretanto, um outro movimento, ainda pequeno e muito controverso, que é o uso do computador como auxiliar no processo de diagnóstico clínico. A experiência de um grupo de especialistas é agregada em um programa, usado durante o processo de diagnóstico. É cedo para se saber onde essas experiências vão chegar. Sua existência deve entretanto ser conhecida.

Outro movimento que começa a ter importância na forma de exercício do médico é a preocupação com a qualidade do serviço. No exterior, os hospitais são cada vez mais administrados, incorporando-se na rotina esforços para a melhoria dos processos usados. O objetivo é não só proporcionar melhores serviços, mas também diminuir os custos. Este movimento de promoção da qualidade no oferecimento do serviço médico envolve muitas atividades de quantificação, para os quais são úteis as noções de Estatística aqui apresentadas.

## Capítulo 2

# Organização da Pesquisa Médica

### 2.1 Introdução

Jornais, noticiários de televisão e revistas têm usualmente uma seção sobre saúde. Neste espaço são relatados os avanços da Medicina tais como novos tratamentos, associações recentemente descobertas entre dieta ou outros hábitos pessoais e alguma doença ou forma de prevenção, etc.

Um exemplo típico é a notícia da primeira página do Jornal do Brasil de 28/07/89, dia da abertura do 13º Congresso Interamericano de Cardiologia no Rio de Janeiro, transcrita a seguir:

#### **Incor diz que safena e droga têm efeito igual**

Uma Pesquisa feita pelo Incor, publicada em janeiro no *American Journal of Cardiology* e apresentada no 13º Congresso Interamericano de Cardiologia, concluiu que os pacientes com problemas cardíacos submetidos a cirurgias de ponte de safena apresentam uma evolução clínica semelhante a dos pacientes que são tratados com dietas e medicamentos. No entanto, o médico argentino René Favaloro, o primeiro a realizar uma operação de ponte de safena no mundo, classifica de “confusos” os resultados de pesquisas como a realizada pelo Incor.

Neste capítulo damos o primeiro passo para o desenvolvimento de uma avaliação crítica de notícias como esta. Primeiramente identificando os problemas típicos da pesquisa clínica: o estudo da associação entre a exposição a um fator e o eventual desenvolvimento de uma doença, a comparação de opções terapêuticas e o estudo de fatores de prognósticos para pacientes submetidos a um dado tratamento. Depois, apresentando as quatro formas básicas de pesquisa utilizadas na solução desses problemas: estudos descritivos, estudos caso-controle, estudos tipo coorte e ensaio clínico aleatorizado.

Nos estudos descritivos, apresentam-se de maneira organizada informações sobre pacientes atendidos em um serviço ou organizam-se sob uma ótica médica dados produzidos pelas instituições estatais de coleta de dados.

Os estudos caso-controle e de coorte são comparativos, ou seja, a conclusão final é obtida através de comparação entre dois grupos de pacientes. Nos estudos caso-controle, como o nome indica, a comparação se faz entre um grupo de doentes e um grupo de pessoas sem a doença estudada. No estudo de coorte compara-se um grupo exposto ao fator de risco em estudo com outro sem o fator. Esses estudos são observacionais, ou seja, o pesquisador não impõe um tratamento mas usa os dados produzidos pela história do paciente. Essas formas de organização são freqüentemente, por razões éticas ou práticas, a única estratégia de pesquisa viável.

A forma mais conclusiva de pesquisa clínica é o ensaio clínico em que, além da presença do grupo de controle, a comparabilidade dos grupos é garantida pela alocação aleatória dos pacientes aos tratamentos.

O estudo detalhado dos planejamentos de estudos clínicos em geral é objeto da epidemiologia. Textos da área, como Breslow & Day (1980, 1987) e Rothman & Greenland (1998), devem ser consultados para descrição completa de cada forma de pesquisa apresentada aqui.

A generalização dos resultados de um estudo clínico a grupos diferentes daqueles usados originalmente é tarefa que merece cuidados especiais. De forma geral podemos dizer que, antes de ser incorporado à clínica, os resultados de um trabalho de pesquisa são replicados em outras instituições, freqüentemente em outros países.

## 2.2 Estudo descritivo

Em um estudo descritivo o objetivo é a pura descrição de um fato médico. Sua principal característica é a ausência de um grupo de comparação. Um grande número de estudos se encaixam, grosso modo, nesta classificação. Sem preocupação de sermos exaustivos, citamos os seguintes tipos: estudo de casos, estudo de uma série de casos e estudos baseados em dados institucionais.

### 2.2.1 Estudo de casos

O estudo de casos consiste na cuidadosa e minuciosa descrição, por um ou mais médicos, do diagnóstico e evolução da doença de um pequeno número de pacientes. É o tipo mais básico de estudo clínico e tem grande tradição na Medicina. Muitos periódicos possuem uma seção especial para o relato de casos interessantes. Inúmeros grupos reúnem-se periodicamente para analisar casos específicos. Muitos tratamentos, hoje incorporados à prática médica, começaram com a observação cuidadosa de um ou alguns poucos casos.

#### Exemplo 2.1: Associação entre rubéola e catarata congênita

O oftalmologista N. Gregg, de Sydney, na Austrália, observou, no primeiro semestre de 1941, vários recém-nascidos com catarata congênita. As características pouco usuais dos casos o levaram a procurar uma explicação que envolvesse as mães. Verificou que todas haviam sido acometidas de rubéola em uma grande epidemia que havia atingido a Austrália no ano anterior, exatamente durante o primeiro trimestre de gestações que resultaram no nascimento das crianças com catarata congênita (Gregg, 1941). As observações clínicas do Dr. Gregg criaram as condições para que várias pesquisas posteriores concluíssem hoje, além de qualquer dúvida razoável, haver associação entre rubéola no primeiro trimestre da gravidez e defeitos congênitos. Essa constatação e o desenvolvimento de vacinas determinaram a introdução do procedimento de vacinação das mulheres em idade fértil.

### 2.2.2 Estudo de uma série de casos

Clínicos com inclinação científica usam os dados dos arquivos de suas instituições para produzir artigos que consistem na apresentação organizada dos resultados dos seus tratamentos preferidos, na descrição dos pacientes tratados e na divulgação de outras informações que julgam relevantes. Os artigos com estas características são chamados de Estudos de uma Série de Casos. São de boa qualidade, se baseados em um conjunto de casos em que há uniformidade de tratamento e quando todos os casos, satisfazendo a um critério objetivo, são incluídos.

Estudos de uma série de casos não fornecem informações confiáveis para comparação entre tratamentos, nem podem ser a base para opiniões sobre etiologia. Isto porque não são estudos comparativos. As opiniões sobre etiologia, prevenção e terapêutica, freqüentemente encontradas nas conclusões desses artigos, devem ser vistas como um primeiro elo de uma cadeia de evidências a serem obtidas, e não como opinião definitiva. Mesmo com tais limitações, esses estudos são fundamentais para o avanço do conhecimento médico em patologias muito raras.

#### Exemplo 2.2: Fatores prognósticos em linfomas não-Hodgkin

Os serviços de Oncologia Clínica da Santa Casa de Misericórdia de Belo Horizonte e do Centro de Quimioterapia Antiblástica e Imunoterapia do Hospital Belo Horizonte trataram, de 1977 a dezembro de 1984, 150 pacientes com linfomas agressivos usando o esquema quimioterápico CHOP-BLEO. Brandão et al. (1987) analisaram os dados correspondentes com vista à identificação de fatores de prognóstico para a obtenção de remissão completa e maior tempo livre de doença. Em relação à remissão completa, a análise conjunta dos fatores considerados identificou três fatores independentes: histologia, estádio e existência de sintomas. A partir destes fatores os pacientes foram classificados em 18 grupos; a remissão completa variava de 17% a 95,7%.

### 2.2.3 Estudo baseado em dados institucionais

Com este nome agrupamos os estudos que apresentam, com uma ótica clínica, dados coletados por instituições estatais.

#### Exemplo 2.3: Câncer no Brasil

O Ministério da Saúde publicou em 1991 o livro "Registro Nacional de Patologia Tumoral - Diagnóstico de Câncer - Brasil 1981/85". Um total de 442 laboratórios espalhados por todo o território nacional contribuíram com mais de meio milhão de diagnósticos, que foram codificados e compilados. Apesar de essa publicação não apresentar todos os diagnósticos durante um dado período de tempo, nem uma amostra probabilística, gera muita informação útil. De forma particular, destaca-se a distribuição segundo idade, sexo e órgão de desenvolvimento dos cânceres.

## 2.3 Estudo caso-controle

Estudo caso-controle é uma forma de pesquisa que visa verificar se indivíduos, selecionados porque têm uma doença - os casos - diferem significativamente, em relação à exposição a um dado fator de risco, de um grupo de indivíduos comparáveis, mas que não possuem a doença - os controles.

Este estudo começa com o levantamento da história clínica de todos os pacientes selecionados. Com as informações, procura-se saber se o fator de risco está presente mais frequentemente ou em nível mais elevado entre os casos que entre os controles. Se a evidência for suficiente, o pesquisador concluirá que existe uma associação entre o fator de risco e a doença.

Idealmente, os casos devem ser todos os que ocorreram durante um período de tempo em uma população definida. Os controles devem ser pessoas comparáveis aos casos, mas sem a doença, ou seja, pessoas que, se desenvolvessem a doença, seriam escolhidas como casos.

Os grupos de casos e de controles podem ser formados de forma cimpareslhada ou de forma independente. No primeiro esquema, para cada caso um ou mais controles semelhantes são escolhidos. Já na

formação de grupos de forma independente, os controles não são escolhidos de forma associada a um caso específico. Preocupa-se apenas em garantir que o grupo de casos seja, na sua totalidade, parecido com o grupo de controles.

O caso particular do emparelhamento em que há apenas um controle para cada caso é denominado pareamento. No pareamento, para cada caso, um controle com características básicas semelhantes ao do caso é escolhido.

Os estudos de caso-controle são uma forma de pesquisa simples e eficiente, por isso muito utilizada. Através dela, já se verificaram ou se confirmaram associações entre fatores de risco e vários tipos de câncer. Não há dificuldades éticas para sua implementação e, portanto, os dados usados são os de seres humanos. Isto livra o pesquisador de difíceis generalizações inerentes aos estudos desenvolvidos em animais. O tempo gasto e os custos associados são relativamente pequenos, já que, normalmente, são utilizados dados preexistentes. São particularmente adequados ao estudo de doenças raras, porque o pesquisador começa com um grupo de pessoas que comprovadamente têm a doença.

Sua grande limitação é a suscetibilidade aos vícios de informação e de seleção. Um estudo detalhado desses vícios, formas de evitá-los na etapa de coleta de dados ou durante a análise dos dados, regras claras para a escolha dos casos e controles extrapolam o objetivo deste texto, mas são fundamentais para quem quiser utilizar esta forma de pesquisa. Existem vários livros sobre o assunto, entre eles, Breslow & Day (1980) e Rothman & Greenland (1998).

#### Exemplo 2.4: Fatores de risco em câncer de mama

Gomes (1992), em tese de doutorado na Faculdade de Medicina da UFMG, realizou estudo caso-controle com o objetivo de avaliar a influência de fatores de risco no câncer de mama, utilizando dados de nosso meio. Como casos foram escolhidas pacientes portadoras de câncer de mama registradas no Hospital das Clínicas da UFMG, que satisfizeram os seguintes critérios de inclusão: idade à época do diagnóstico entre 25 e 75 anos, diagnóstico feito entre 1978-1987 e confirmado por exame anatomo-patológico, tumor originário do tecido epitelial e ter sido submetida a algum tipo de cirurgia na mama.

Como controles foram escolhidas pacientes com idade igual à do caso (mais ou menos dois anos), data de admissão ao hospital igual à data de confirmação do diagnóstico do caso (mais ou menos seis meses) e exame clínico da mama sem indicação de patologias mamárias. De acordo com esses critérios, foram selecionados dois controles para cada caso. O primeiro foi selecionado no ambulatório de ginecologia e o segundo no registro geral do hospital. A análise estatística, entre várias outras conclusões, mostrou que a presença na paciente de história familiar de câncer de mama aumenta o risco desta patologia em 8,84 vezes.

## 2.4 Estudo de coorte

Estudo de coorte é uma forma de pesquisa que visa verificar se indivíduos, selecionados porque foram expostos ao fator de risco, desenvolvem a doença em questão, em maior ou menor proporção do que um grupo de indivíduos, comparáveis, mas não expostos ao fator de risco.

Em contraste com um estudo caso-controle, um estudo de coorte avança no tempo e coloca ênfase no fator de risco. Identificam-se um grupo exposto ao fator e o grupo controle, constituído de pessoas que não foram expostas a ele. Os dois grupos são acompanhados por um período de tempo e as taxas de incidência da doença calculadas. Se essas taxas são significativamente diferentes nos dois grupos, o pesquisador conclui que há associação entre a doença e o fator.

Os estudos de coorte possuem várias vantagens. O pesquisador tem a possibilidade de usar critérios uniformes, tanto na identificação da presença ou não do fator de risco ao início do estudo, quanto na verificação da ocorrência da doença nos vários exames de acompanhamento. A comparabilidade dos dois grupos pode ser verificada no início do estudo e identificadas as variáveis para as quais são necessários ajustamentos na análise dos dados.

Em um estudo prospectivo, o pesquisador tem muito mais liberdade sobre o que medir e como medir, já que não se restringirá ao uso de dados já coletados. Uma outra vantagem, que só poderá ser completamente apreciada ao se estudar a metodologia de análise de dados, é que os estudos de coorte nos permitem obter diretamente uma estimativa de magnitude do risco relativo. Isto significa que é

possível quantificar o risco de desenvolver a doença comparando-se o grupo de expostos ao fator de risco com o grupo de não expostos.

Estudos de coorte são grandes, longos e normalmente caros. Quanto mais rara a doença em questão, maior o número de pacientes que precisam ser examinados. Embora, do ponto de vista teórico, os estudos de coorte sejam melhores que os estudos caso-controle, estes últimos são mais comuns.

### Exemplo 2.5: Personalidade e desenvolvimento de doença coronariana

Um estudo de coorte com o objetivo de avaliar o possível efeito da personalidade no risco de desenvolvimento de doença coronariana foi conduzido entre 3.154 trabalhadores do sexo masculino com idade de 30 a 59 anos (Brand et al., 1976). Os indivíduos entraram no estudo entre 1960-61 e foram acompanhados por um período médio de 8 anos e meio. Através de entrevista no início do estudo, foram classificados em dois tipos de personalidade, A e B, sendo os primeiros mais agressivos, competitivos e ansiosos.

Os resultados da Tabela 2.1 indicam que nas duas faixas etárias consideradas os percentuais de indivíduos do tipo A que desenvolveram doença coronariana são aproximadamente o dobro dos encontrados no outro grupo.

Tabela 2.1: Percentual de indivíduos que desenvolveram doença coronariana segundo faixa de idade e tipo de personalidade

Faixa de idade	Personalidade	
	A	B
39-49	8,9	4,2
50-59	15,9	7,6

### 2.4.1 Seleção das coortes

Grupos de pessoas são selecionadas para o estudo de coorte por uma variedade de razões. Apresentamos a seguir dois exemplos.

O estudo sobre os efeitos da radiação nos sobreviventes da bomba atômica descrito por Miller (1969), ilustra uma situação em que o grupo foi escolhido para compor a coorte por ter sofrido uma exposição de intensidade pouco comum e o objetivo do estudo era verificar os efeitos desta exposição.

O clássico estudo de Doll & Hill (1964) sobre associação entre cancer no pulmão e fumo ilustra a escolha da coorte pelo fato de o grupo possuir características que facilitam a obtenção dos dados sobre a exposição ao fator e o seguimento dos pacientes. Foram acompanhados médicos da Inglaterra, um grupo fácil de contactar e no qual todas as mortes são rotineiramente bem documentadas.

Em muitos estudos de coorte os grupos de comparação são obtidos após o início do estudo, de acordo com o nível de exposição ao fator. Por exemplo, as coortes do Estudo de Framingham sobre doenças coronárias (Kannel et al., 1972), foram construídas dividindo-se o grupo acompanhado de acordo com hábitos de fumo, níveis de colesterol, etc. Nestes casos não há necessidade de um grupo externo de comparação.

Em outras situações, particularmente quando um grupo submetido a uma exposição pouco comum é estudado, é importante comparar o resultado observado com aquele esperado, caso os indivíduos não tivessem sido submetidos ao fator de risco. Usa-se a experiência da população em geral, ao tempo em que a coorte é formada, como padrão de comparação.

Finalmente, usa-se também como base de comparação outra coorte formada por pessoas não expostas, parecidas nas características demográficas com o grupo exposto. Por exemplo, para uma coorte de radiologistas, Seltser e Sartwell (1965) usaram como padrão de comparação dados de oftalmologistas e otorrinolaringologistas.

Detalhes adicionais sobre a seleção de coortes e outros aspectos relacionados a esse tipo de estudo podem ser encontrados em Breslow & Day (1987) e Rothman & Greenland (1998).

## 2.5 Ensaios clínicos aleatorizados

Ensaios clínicos aleatorizados é um experimento médico, realizado com o objetivo de ~~verificar~~, entre dois ou mais tratamentos, qual é o mais efetivo.

São usados quando é incerto o valor de uma nova terapia ou os méritos da terapia existente estão em disputa. É a metodologia apropriada para a comparação de tratamentos.

Após um critério de admissão ter sido definido, os pacientes são, à medida que entram no experimento, alocados de maneira aleatória no grupo controle, que recebe a terapêutica padrão, ou ao grupo tratamento, que recebe a terapêutica sendo testada. Todo esforço deve ser feito para oferecer os mesmos cuidados aos dois grupos. Terminado o experimento, técnicas estatísticas são usadas para se decidir se há ou não diferença na eficácia das terapias envolvidas.

Esta é uma forma experimental de pesquisa, isto é, o pesquisador interfere de maneira deliberada no curso natural dos acontecimentos, em contraposição aos estudos observacionais, em que o pesquisador se restringe à coleta de dados, sem alterar a dinâmica do processo em consideração. Por isto está sujeita a constrangimentos éticos, disciplinados pela convenção de Helsinque.

O conhecimento pelo paciente ou pelo médico do tratamento a ser administrado pode influir na evolução da doença, não através de fatores extrínsecos, mas diretamente pelo processo de auto-sugestão. É o efeito placebo, que traz vícios ao estudo clínico.

A única maneira de eliminar o efeito placebo nos grupos comparados é realizar, sempre que possível, experimentos cegos, nos quais o paciente ou o médico não conhece o tratamento, ou duplo-cego, no qual ambas as partes desconhecem o tratamento. Este ideal, entretanto, nem sempre é factível.

### Exemplo 2.6: Tamoxifeno e câncer de mama

Muitos avanços no tratamento do câncer de mama têm sido estabelecidos através de estudos realizados pelo *National Surgical Adjuvant Breast and Bowel Project* (NSABP). Em 1985, discutia-se a necessidade de se submeter à quimioterapia ou hormonioterapia, pacientes de bom prognóstico, logo após a cirurgia. Em particular, havia evidências de várias origens de que o tamoxifeno poderia melhorar ainda mais o prognóstico dessas pacientes. Naquela época, o tamoxifeno já era usado largamente no tratamento de pacientes com câncer de mama. Diante disto, o NSABP iniciou o protocolo B-14 para determinar a eficácia do tamoxifeno em pacientes considerados

de bom prognóstico, isto é, aquelas com tumores com receptor de estrógeno positivo ( $> 10 \text{ fmol}$ ), idade inferior a 70 anos, com câncer de mama operável e linfonodos axilares negativos ao exame histológico.

Foram criados dois grupos de pacientes através de aleatorização feita dentro de estratos definidos por idade e tamanho do tumor na análise clínica, tipo de cirurgia e concentração de receptor de estrógeno. Um grupo recebeu tamoxifeno (10 mg por dia, via oral, duas vezes ao dia) e o outro, placebo, indistinguível do tamoxifeno em aparência e gosto. Além disto, o ensaio foi organizado na forma duplo cego.

Após acompanhamento das pacientes por um período de até 4 anos, constatou-se uma diferença significativa em termos de tempo livre de doença em favor das pacientes que receberam o tamoxifeno. No grupo tratamento, 83% estavam livres da doença aos quatro anos após a cirurgia, enquanto que no grupo placebo esta porcentagem era de 73%. Este estudo, publicado por Fisher et al. (1989), contribuiu decisivamente para a adoção do tamoxifeno como quimioterapia adjuvante logo após a cirurgia em pacientes de bom prognóstico.

### 2.5.1 Aleatorização

A resposta de um indivíduo a uma terapêutica varia de acordo com suas características pessoais. Assim, para que as diferenças porventura detectadas sejam realmente devidas à diferença na eficácia das terapias, é fundamental que os grupos de tratamento e controle sejam comparáveis.

A única maneira completamente segura de se atingir a comparabilidade é a alocação aleatória dos pacientes participantes no experimento aos grupos de controle e de tratamento. Existem várias técnicas de aleatorização, como discutidas por exemplo em Soares & Wu (1983).

A grande vantagem do uso da aleatorização está no fato de que os dois grupos obtidos serão semelhantes em todos os aspectos, exceto, é claro, no que diz respeito ao tratamento recebido. Isto acontece porque a tendência é aleatorização fazer com que os grupos apresentem as mesmas distribuições para as características dos indivíduos, conhecidas ou não. Isto é especialmente verdade quando o tamanho dos grupos é grande.

Além disto, a aleatorização elimina os vícios de seleção. É claro que se os indivíduos são alocados aleatoriamente aos diferentes grupos, não há a possibilidade de que preferências pessoais do investigador, no que se refere a qual é o melhor tratamento para um determinado paciente, venham a alterar os resultados do estudo.

Quando o número de pacientes é pequeno, a aleatorização torna os grupos apenas aproximadamente comparáveis. Para melhorar a capacidade de detecção do efeito pelo experimento, dois procedimentos são usados nessa situação: as técnicas de estratificação e de ajuste.

Na estratificação dividimos os pacientes em subgrupos de acordo com variáveis que reconhecidamente afetam a resposta ao tratamento e fazemos alocação aleatória dentro de cada subgrupo.

As técnicas de ajuste, geralmente métodos estatísticos multivariados, fornecem meios de comparação entre os grupos, válidas mesmo quando estes diferem em uma ou mais características conhecidas. Não é demais ressaltar que apenas a aleatorização pode produzir comparabilidade em variáveis que, embora afetem a resposta, não são ainda reconhecidas como influentes pelo pesquisador.

### 2.5.2 Como implementar a alocação aleatória

A alocação aleatória pode ser implementada através de esquemas bastante simples tais como o lançamento de moedas ou dados. No entanto é mais comum o uso de números aleatórios, gerados por programas de computador ou lidos de tabelas especiais, como a apresentada na Tabela A1 do apêndice.

Para usar esta tabela basta escolher uma linha e uma coluna e seguir em qualquer direção de forma sistemática. O número de dígitos a ser usado depende da situação. Para se alocar 100 indivíduos a dois tratamentos (A e B) são necessários dois dígitos. Os 100 indivíduos são primeiramente associados aos números 00 a 99. Por exemplo, os números ímpares são alocados ao tratamento A e os pares ao tratamento B.

A alocação aleatória é amplamente utilizada em ensaios clínicos e em estudos experimentais com cobaias e trata-se de um procedimento com regras claras. O termo aleatório não deve ser entendido como alocação desordenada ou feita de forma ambígua ou confusa.

## 2.6 Estudos comparativos

Estudo comparativo é um termo genérico usado para designar os estudos caso-controle, os estudos de coorte e os ensaios clínicos. Os dois primeiros são principalmente estudos etiológicos: procuram descobrir associação entre fatores de risco ambientais ou características pessoais e a doença. Criam mais um elo na determinação da causalidade da doença em estudo.

Podem ser estudos observacionais se os pacientes não são alocados aleatoriamente aos fatores em estudo. Nesse caso, usam-se apenas os dados obtidos no exame da vida pregressa dos pacientes. Isto em contraposição aos ensaios clínicos, em que os pacientes são alocados aleatoriamente aos tratamentos.

Ao se iniciar um estudo comparativo é fundamental identificar a resposta de interesse. Podemos estar interessados apenas na ocorrência da cura da doença ou na remissão do tumor, caso em que a resposta é uma variável dicotômica. Alternativamente, a resposta de maior interesse pode ser o tempo de sobrevivência, ou o tempo da remissão até a recidiva. Nos últimos anos, índices de qualidade de vida começaram a ser considerados como respostas de interesse, ao lado de medidas clínicas de efeito de tratamentos.

### 2.6.1 Grupo controle

Idealmente, a comparação dois tratamentos, A e B, consiste em se aplicar o tratamento A em um grupo de indivíduos e medir o efeito do tratamento. A seguir, determina-se o efeito do tratamento B, no mesmo grupo. Qualquer diferença observada seria devida unicamente aos tratamentos. Isto permitiria estabelecer diferenças na eficácia entre eles.

Este experimento ideal, por razões práticas ou éticas, é impossível de ser realizado na maioria das vezes. Por exemplo, como aplicar o segundo tratamento se o paciente morreu? A solução é a criação de dois grupos de pacientes, tão parecidos quanto possível. Cada grupo recebe um tratamento e os resultados são comparados. Como usualmente os dois tratamentos comparados são o padrão e um tratamento alternativo, os grupos usados recebem os nomes de *grupo controle* e *grupo tratamento*. Estes nomes são mantidos em todas as circunstâncias, mesmo quando a comparação não é mais entre dois tratamentos,

como no caso de estudos caso-controle.

Hoje é aceita por todos a necessidade de um grupo de controle. Historicamente, entretanto, isto nem sempre foi assim. Muitos procedimentos médicos ineficientes, como o congelamento do estômago no tratamento de úlcera gástrica (Ruffin et al., 1969), foram usados por longo tempo simplesmente porque, ao se estabelecer sua eficácia, não se fez nenhuma comparação.

O estabelecimento de um padrão de comparação é usual na prática médica. O clínico compara o paciente sob seus cuidados, hoje, com outros que já tratou ou estudou. Baseado nessas experiências, decide o procedimento que usará. Assim, a exigência de grupo controle é mera explicitação de procedimento médico padrão.

A aceitação do resultado do estudo depende, fundamentalmente da comparabilidade dos dois grupos. Parte ponderável dos cuidados no planejamento dos estudos comparativos visa garantir esta comparabilidade.

### 2.6.2 Confusão de efeitos

Quando a contribuição de um fator, identificada em uma análise individual, pode ser total ou parcialmente atribuída a outro, dizemos que estamos diante de uma situação em que há confusão de efeitos (Datta, 1993).

Um fator de confusão pode ocorrer em qualquer tipo de estudo, incluindo os experimentais. Em muitos estudos epidemiológicos, a idade e o nível sócio econômico são potencialmente importantes fatores de confusão. Por exemplo, suponha que a idade é um fator de risco para uma doença e um estudo tem como objetivo avaliar se esta doença está associada a uma determinada exposição (potencialmente fator de risco). Se pessoas expostas são mais velhas que pessoas não expostas, como se poderia saber se o efeito observado para a exposição é realmente devido à exposição ou à idade? Então neste caso a idade é um fator de confusão.

Para que se possa dizer que uma variável é um fator de risco para uma patologia, é preciso produzir evidências de que não existem fatores de confusão, isto é, variáveis associadas ao fator e à doença que explicam a associação observada. Em geral, uma manifestação necessária, mas não suficiente, de um fator de confusão em um dado estudo é que ele esteja associado ao fator de risco e à doença, como

mostrado na Figura 2.1.

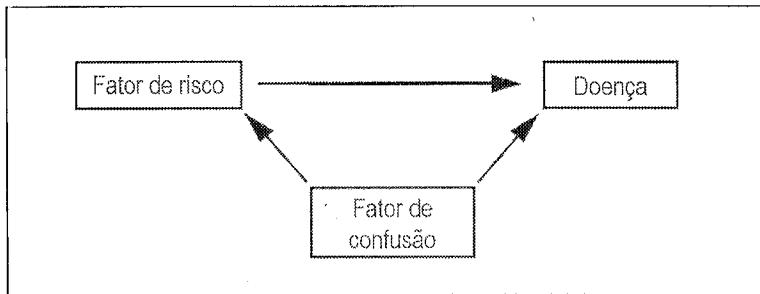


Figura 2.1: Diagrama ilustrando a relação entre fator de confusão, fator de risco e doença

Existem duas maneiras de abordar a influência dos fatores de confusão: controlá-los na etapa de planejamento do estudo ou na etapa de análise dos dados.

### 2.6.3 Controle de fatores de confusão

As principais estratégias para contornar o problema criado pelos fatores de confusão são descritas a seguir.

A primeira, na fase de planejamento do estudo, é usar a aleatorização e/ou a homogeneização.

A aleatorização faz com que os outros fatores que, além do tratamento, afetam a resposta tenham a mesma distribuição nos diversos grupos de comparação, deixando assim de constituir fatores de confusão.

A homogeneização consiste em selecionar indivíduos com características semelhantes em relação aos fatores que afetam a resposta, de tal forma que a única diferença relevante entre os diversos grupos fique sendo o tratamento. Esta estratégia, no entanto, dificulta o uso na prática dos resultados obtidos.

A segunda estratégia é usar, na etapa de análise dos dados, técnicas estatísticas para tratar a questão da não comparabilidade. A primeira é a construção de análises estratificadas, isto é, em subgrupos homogêneos de pacientes.

Considere um estudo caso-controle planejado para estimar o possível efeito do uso de pílula anticoncepcional e câncer de mama. Como o nível sócio-econômico (NSE) é um conhecido fator de risco para a doença, para avaliar a associação com o uso de pílulas anticoncepcionais, é recomendada uma análise estratificada por NSE.

Um outro exemplo é um estudo de coorte para estimar o efeito da exposição à serragem da madeira na ocorrência de doenças respiratórias crônicas em trabalhadores de fábricas de móveis, do sexo masculino e de meia idade. Como o fumo é conhecidamente um fator de risco para a doença, deve-se considerar o hábito de fumar como um fator de confusão e usar uma análise estratificada por este fator.

Outra forma consiste na utilização de técnicas multivariadas especialmente desenvolvidas para isolar a contribuição específica de cada fator. Elas exigem o uso de programas de computador específicos e só conseguem, entretanto, controlar o efeito de fatores conhecidos.

A identificação de fatores de confusão e de estratégias de controle é tema importante da epidemiologia. O leitor interessado pode ler mais sobre o assunto, por exemplo, em Kleinbaum et al. (1982) e Rothman & Greenland (1998).

### 2.6.4 Generalização dos resultados

Os pacientes cujos dados são usados em um estudo comparativo quase nunca constituem uma amostra representativa de uma população de interesse. Assim, a generalização de resultados de pesquisa clínica não se dá, usualmente, via inferência estatística, mas através da repetição do estudo em outros grupos de pacientes. Recentemente, a agregação de resultados de vários estudos recebeu o nome de *meta-análise* e é hoje importante forma de acumulação de conhecimento médico. É admirável a síntese feita sobre quimioterapia em câncer de mama, publicada em janeiro de 1992, e preparada pelo *Early Breast Cancer Trialists Collaborative Group*. Hedges & Olkin (1985) é uma referência sobre métodos estatísticos para meta-análise e Rothman & Greenland (1998) trata o assunto de forma mais conceitual.

## 2.7 Exercícios

1. Para cada uma das situações descritas a seguir, indique qual das seguintes formas de pesquisa foi utilizada: estudo descritivo, ex-

tudo caso-controle, estudo de coorte, ensaio clínico aleatorizado ou nenhuma das anteriores..

- (a) Para avaliar o efeito protetor da vacina BCG em crianças com menos de 15 anos, estudaram-se as crianças que receberam algum tratamento antituberculose durante o ano de 1981 na cidade de Buenos Aires, tanto internados em hospitais ou tratados na forma ambulatorial. Para cada caso, encontrou-se uma criança da mesma idade, sexo, condição sócio-econômica e que tinha alguma doença aguda, diferente de tuberculose, no mesmo período e que havia sido tratada no mesmo estabelecimento. Em ambos os grupos considerou-se como vacinados os que tinham a cicatriz correspondente à vacina BCG em uma ou ambas regiões deltidianas. Esta descrição foi publicada no Boletim OPAS (Calvete et al., 1986).
  - (b) Os primeiros 700 casos de câncer de fígado diagnosticados que ocorreram em vários estados americanos após o 1º de janeiro de 1990 foram seguidos até 31 de dezembro de 1994. Os níveis de consumo de álcool foram medidos para cada caso. Trezentos casos morreram durante esse período de cinco anos.
  - (c) Um estudo foi conduzido em um grupo de gêmeos monozigotos e dizigotos do mesmo sexo, sendo que um gêmeo do par tinha câncer no cólon e o outro não. Todos foram casos incidentes observados em 6 grandes centros médicos de universidades do sudoeste dos Estados Unidos durante um período de 5 anos. Informações sobre o teor de fibras da dieta foram coletadas para cada um dos indivíduos afim de estimar seu efeito na ocorrência de câncer no cólon.
2. Considere a seguinte hipótese: “O DIU predispõe para o câncer de cólon uterino”. Resumidamente descreva alternativas de planejamento para conduzir um estudo que discute este problema de pesquisa.
3. Baseado na observação de vários casos, um clínico suspeita que o fato de a mãe ter menos de 20 anos de idade está associado

com o nascimento de bebês com baixo peso ao nascer, isto é, menos de 2.500 g.

- (a) Como você encaminharia um estudo de coorte e um estudo caso-controle para avaliar a conjectura do pesquisador? Identifique claramente os grupos de comparação.
- (b) Cite pelo menos 3 fatores que você deveria controlar para tornar os grupos comparáveis.
- (c) Por que, neste caso, não se pode usar a forma de ensaio clínico?

4. Considere os seguintes problemas de interesse médico:

- (a) A pílula anticoncepcional é um fator de risco para tromboembolia?
- (b) Usinas nucleares podem causar defeitos congênitos nos filhos de mães que moram em áreas próximas a usinas?
- (c) A poluição do ar pode causar câncer de pulmão?
- (d) A anestesia com o halotano pode causar problemas de fígado?
- (e) O fumo é um fator de risco para problemas coronarianos?
- (f) Para algumas formas de doença coronariana o uso de medicamentos é tão eficiente quanto a cirurgia?

Explique que tipo de estudo você usaria para estudar estas questões, aqui no Brasil. Descreva os grupos que seriam usados e possíveis fontes de vícios.

5. Que tipo de estudo você usaria para verificar a associação entre um fator ambiental e um câncer raro. Justifique sua resposta.
6. Freqüentemente ouve-se a afirmativa que a única opção realmente correta para se investigar uma hipótese clínica é o uso prospectivo de uma coorte de pacientes. Explique porque você concorda ou não com esta afirmação.

7. Em um Congresso de Oncologia, dois trabalhos foram apresentados comparando-se tratamentos para o mesmo tipo de câncer. Os autores, em ambos os casos, apresentaram os resultados obtidos em suas clínicas. Um participante, analisando as duas comunicações, resolveu adotar o tratamento com a melhor taxa de resposta. Esta é uma decisão correta? Justifique sua resposta.
8. Critique as seguintes afirmativas:
  - (a) Em um ensaio clínico garante-se a comparabilidade dos grupos através de uma escolha aleatória dos dois grupos.
  - (b) Em um ensaio clínico deve-se usar uma amostra aleatória de pacientes.
  - (c) Este estudo é válido, pois os pacientes foram escolhidos de forma aleatória.
9. Como em um estudo caso-controle obtém-se a comparabilidade dos dois grupos de pacientes a serem contrastados?
10. Como pode ser controlado o efeito de um fator de confusão em um estudo caso-controle?
11. Diga o que é um grupo controle e qual é a sua importância em pesquisas clínicas. Exemplifique sua resposta usando um ensaio clínico.
12. Proponha um esquema de aleatorização nas seguintes situações:
  - (a) Em um ensaio clínico, dois tratamentos (A e B) serão testados através da comparação de dois grupos de igual tamanho, cada um com 30 pacientes.
  - (b) Em um ensaio clínico envolvendo 100 pacientes, decidiu-se que 40% deles receberiam o tratamento A e os restantes o tratamento B.
  - (c) Repita o item anterior para 50 e 20 pacientes mantendo o mesmo percentual em cada tratamento.

## Capítulo 3

# Descrição e Apresentação de Dados

### 3.1 Introdução

Os dados obtidos dos pacientes participantes de estudos médicos devem ser analisados e interpretados com o auxílio de métodos estatísticos. A primeira etapa desta tarefa é a organização e síntese dos dados. Para isto, foram desenvolvidos métodos que recebem o nome genérico de estatística descritiva, ou ainda, análise descritiva, tema deste capítulo.

A análise descritiva consiste basicamente na organização e descrição dos dados, na identificação de valores que traduzem o elemento típico e na quantificação da variabilidade presente nos dados. Os elementos básicos para essa análise são: tabelas, gráficos e medidas ou sínteses numéricas.

O uso de técnicas descritivas deve sempre preceder análises mais avançadas. Além de propiciar a familiarização com os dados, possibilita a detecção de estruturas interessantes e, eventualmente, a presença de valores atípicos nos dados.

### 3.2 Conceitos básicos

#### 3.2.1 Variável

Neste texto a palavra variável deve ser entendida como a quantidade

cação da característica de interesse do estudo. Corresponde ao termo parâmetro utilizado em muitas áreas da Medicina.

Um questionário de um inquérito epidemiológico que contém as seguintes perguntas: Qual é a sua idade? Qual o número de pessoas da sua família? Qual é a renda total de sua família? Qual é o seu estado civil? Você tem emprego fixo? produz informações nas variáveis: idade, tamanho da família, renda familiar, estado civil e emprego.

Em geral, a mera identificação da característica de interesse leva, de forma natural, a uma quantificação adequada. Não há problema na definição de variáveis quando o interesse recai sobre parâmetros hematológicos, pressão sanguínea ou dados antropométricos. Em outros casos, entretanto, antes que a quantificação da característica possa ser feita é preciso padronizar os procedimentos de avaliação. Por exemplo, não é imediatamente claro como quantificar intensidade de dor, falta de ar (dispnéia) e estado geral do paciente. Escalas específicas foram desenvolvidas para isto.

Em outras situações, embora a variável medida seja claramente definida, a variável de real interesse é obtida através da manipulação de outras variáveis. Para avaliar a obesidade de um indivíduo é comum a utilização do índice de massa corporal (IMC) definido como a razão entre o peso ( $kg$ ) e a altura ao quadrado ( $m^2$ ).

Chamamos *variável resposta* aquela a ser explicada no estudo. Além dela, há sempre outras variáveis que devem ser consideradas na análise, denominadas *variáveis de controle* ou *covariáveis*.

### 3.2.2 Tipos de variáveis

Facilita o tratamento estatístico de variáveis classificá-las em categóricas e quantitativas.

#### Variáveis categóricas

Podem ser do tipo *nominal* ou *ordinal* (caso exista uma ordem entre as classes). Sexo e estadiamento de uma doença são exemplos de variável nominal e ordinal, respectivamente.

Variáveis categóricas podem ter duas ou mais categorias (por exemplo sexo e tipo sanguíneo, respectivamente). O número de categorias pode depender do interesse do estudo. Por exemplo, para a

variável hábito de fumar, pode-se classificar um indivíduo como fumante ou não fumante ou ainda usar uma classificação mais detalhada: não fumante, ex-fumante, fumante ocasional, fumante inveterado, etc. A classificação de pessoas através de níveis pressóricos, pode ser hipertenso ou normotensão ou, entre outras, normal, hipertenso leve, hipertenso moderado e hipertenso grave.

Freqüentemente os dados são reduzidos a duas categorias para simplificar a análise e a sua apresentação, mas isto pode em alguns casos resultar em considerável perda de informação.

#### Variáveis quantitativas

Variáveis *quantitativas* podem ser *discretas* ou *contínuas*.

Em uma variável *discreta*, os valores diferem entre si por quantidades fixas. Tamanho da família é uma variável discreta, já que indivíduos diferentes têm tamanhos de família que diferem por exemplo por 0, 1, 2, 3 ou 4 indivíduos. Nenhum valor intermediário é possível. Variáveis discretas são geralmente o resultado de contagens, como por exemplo, número de bactérias em um volume de urina, número de consultas médicas durante um período e número de batimentos cardíacos por minuto.

Variáveis *contínuas*, tais como pressão sanguínea, peso e altura, são usualmente medidas através de um aparelho (aparelho de pressão, balança, etc.). A diferença entre medidas de indivíduos distintos pode ser arbitrariamente pequena.

Na prática, as variáveis contínuas são anotadas até a precisão da medida usada. Por exemplo, quando dizemos que o tempo de sobrevivência de um indivíduo foi de 2 anos, estamos provavelmente simplificando a unidade de medida tomando-a em ano, mas na verdade poderia ser 2 anos, 1 mês e 10 dias. Quando dizemos que o tempo de reação de um indivíduo foi de 1 minuto, a leitura em um cronômetro com maior precisão poderia ter sido 1 minuto, 27 segundos e 10 milésimos de segundo.

Algumas variáveis podem ser classificadas em mais de um tipo dependendo da unidade usada. Por exemplo, idade é uma variável contínua (tempo é intrinsecamente uma medida contínua), mas pode ser classificada como discreta. Renda medida em salários mínimos também pode ser classificada como uma variável discreta, mas pode ser ordinal se categorizada como baixa, média e alta.

A diferença entre uma variável discreta e uma variável categórica ordinal é que a ordenação tem um significado diferente, como ilustra o exemplo a seguir. Suponha duas variáveis: estadiamento de câncer de mama (I, II, III e IV) e número de crianças (0, 1, 2, 3, etc.). Não se pode dizer que o estágio IV é duas vezes pior do que o II, nem que a diferença entre o I e II é equivalente àquela entre III e IV. Por outro lado, 4 crianças corresponde exatamente ao dobro de 2 crianças (obviamente não necessariamente o dobro de trabalho para os pais) e diferenças no número de crianças têm sempre o mesmo significado para todos os valores, isto é, uma família com 4 crianças tem exatamente uma criança a mais que uma família com 3, que por sua vez tem uma a mais que uma família com 2 crianças, e assim por diante.

Finalmente, todos estes tipos de variáveis são encontrados comumente em dados biológicos e da área médica. A importância dessa classificação justifica-se porque cada tipo de variável exige um tratamento estatístico específico. Portanto, é importante ter em mente que o tipo de variável resposta é um ponto de partida para se determinar os métodos de análise mais apropriados ou mesmo válidos.

### 3.2.3 Dados brutos

Dados brutos são aqueles obtidos diretamente da pesquisa, isto é, que ainda não sofreram qualquer processo de síntese ou análise. Em geral são apresentados em tabelas e freqüentemente omitidos na maioria das publicações, quase sempre por questão de espaço.

O conjunto de dados constitui uma *amostra*. O tamanho da amostra é geralmente denotado por  $n$ .

#### Exemplo 3.1: Teor de gordura fecal em crianças

Embora os gastroenterologistas infantis reconhecessem a utilidade diagnóstica do teor de gordura fecal, até 1984 não existia um padrão de referência desta medida para crianças brasileiras. Para preencher esta lacuna, o Prof. Francisco Penna, titular de Pediatria da UFMG, examinou 43 crianças saudáveis que produziram os valores da Tabela 3.1 expressos em  $g/24\text{ horas}$  (Penna, 1984; Penna et al. 1987).

Tabela 3.1: Teor de gordura ( $g/24\text{ horas}$ ) em 43 crianças

3,7	1,6	2,5	3,0	3,9	1,9	3,8	1,5	1,1
1,8	1,4	2,7	2,1	3,3	3,2	2,3	2,3	2,4
0,8	3,1	1,8	1,0	2,0	2,0	2,9	3,2	1,9
1,6	2,9	2,0	1,0	2,7	3,0	1,3	1,5	4,6
2,4	2,1	1,3	2,7	2,1	2,8	1,9		

Apesar de todos estes valores terem sido obtidos em crianças saudáveis, nota-se uma grande variação nos resultados. Esta variabilidade exige que o padrão de referência procurado seja expresso por uma faixa e não por um único número.

#### Exemplo 3.2: Nível de colesterol

Framingham é uma pequena cidade americana localizada perto de Boston, Massachusetts. Em 1948 foi selecionada como local adequado para desenvolvimento de um estudo prospectivo cujo objetivo era verificar como os hábitos de vida das pessoas influenciam o risco de desenvolvimento de doenças cardíacas. Muitos resultados, hoje completamente integrados à prática cardiológica, como a necessidade de controle do nível do colesterol, foram primeiramente obtidos neste estudo.

Os valores da Tabela 3.2 são as medidas em  $mg/dl$  do colesterol, referentes ao segundo exame realizado em 1952 em 80 pacientes. Os pacientes que não compareceram estão representados por 9.

Observando estes dados, na maneira como estão apresentados, pouco apreendemos sobre o nível do colesterol à época do exame. É impossível saber o valor em torno do qual as medidas estão agrupadas, a forma da distribuição e a extensão da variabilidade.

#### Exemplo 3.3: Doenças cardiovasculares em Honolulu

Em 1969 foi realizado um estudo na população de Honolulu (descrito em Kuzma, 1998). Para 7.683 indivíduos, foram pesquisadas as seguintes variáveis: nível educacional, peso ( $Kg$ ), altura ( $cm$ ), idade

Tabela 3.2: Taxa de colesterol ( $mg/dl$ ) em 80 indivíduos

278	182	247	227	277	194	196	276	244	192
118	219	255	201	9	209	219	228	209	209
171	213	233	226	209	200	200	363	209	200
179	167	192	277	317	146	217	292	217	255
212	233	250	243	150	209	184	199	250	479
175	194	221	233	9	184	217	150	167	265
242	180	255	170	209	161	196	165	234	179
248	184	291	185	242	276	243	229	242	250

(anos), glicemia ( $mg/dl$ ), colesterol sérico ( $mg/dl$ ) e pressão sistólica ( $mmHg$ ). Cada indivíduo foi classificado quanto ao hábito de fumar (fumante ou não) e quanto à atividade em casa (pouca, moderada e muita). Os dados referentes a uma amostra de 100 indivíduos estão disponíveis no site do Departamento de Estatística da UFMG (<http://www.est.ufmg.br/~estmed>).

#### Exemplo 3.4: Estudo de prevalência da toxoplasmose

Em sua tese de Mestrado em Pediatria da Faculdade de Medicina da UFMG, Andrade (1994) estudou a prevalência da toxoplasmose entre crianças de 0 a 12 anos de idade matriculadas no ambulatório geral de pediatria do Hospital das Clínicas da UFMG.

O banco de dados das 278 crianças estudadas, também disponível no site citado no Exemplo 3.3, compreende 12 variáveis: idade, sexo, motivo do exame, resultado do hemograma, contato com gatos domésticos ou vadios no domicílio, contato com areia e/ou terra na região peridomiciliar, ingestão de carne crua ou mal cozida, RIFI-IgG (título e classificação: positiva ou negativa), RHA (título e classificação: positiva, negativa ou não realizada) e resultado do exame de fezes.

Nos exemplos 3.3 e 3.4 é ainda mais difícil tirar alguma conclusão a partir dos dados brutos. Torna-se necessária uma descrição dos dados, e nestes casos o uso do computador é imprescindível.

Nos quatro exemplos pode-se perceber a grande variabilidade en-

tre os indivíduos considerados. Os métodos estatísticos são fundamentais para o estudo de situações em que a variabilidade é inherente. A Estatística Descritiva ajuda na percepção, avaliação e quantificação da variabilidade.

### 3.3 Organização e apresentação de dados

Nesta seção apresentaremos formas de organização e apresentação de dados através de tabelas e de gráficos. Construídos a partir de determinado conjunto de dados, proporcionam informações semelhantes ou que se complementam, constituindo formas apropriadas de se organizar dados.

Existem normas para a construção de tabelas e gráficos, mas neste texto daremos ênfase à interpretação. Ressaltamos, entretanto, ser essencial a apresentação de forma clara e compreensível, inclusive dos títulos.

A seguir ilustramos através de exemplos a construção de tabelas de freqüências e gráficos para uma variável ou para o cruzamento de variáveis, considerando os vários tipos apresentados na seção anterior.

#### 3.3.1 Tabelas de freqüências

Uma maneira de sintetizar os dados é através de *distribuição de freqüências*, que consiste na construção de uma tabela a partir dos dados brutos em que se leva em conta a freqüência com que cada observação ocorre.

A interpretação dos resultados obtidos em tabelas de freqüências pode ser auxiliada pela análise de gráficos construídos a partir delas.

#### Exemplo 3.5: Tentativas de suicídio

Em um estudo retrospectivo, Fernandes et al. (1995) fizeram levantamento de tentativas de suicídio por intoxicação aguda registradas no Centro de Assistência Toxicológica do Hospital de Base do São Paulo. No período de 01/92 a 02/93 conseguiram caracterizar 302 casos, que correspondem a 27% do total de atendimentos no período.

Das tentativas de suicídio, há predomínio do sexo feminino (67%). Isso difere dos êxitos letais, em que dois dos três óbitos são masculinos.

A Tabela 3.3 mostra a distribuição da profissão de pacientes potencialmente suicidas. A maior incidência foi observada entre profissionais mal remunerados e com sobrecarga de trabalho, sem perspectiva de ascensão social. O alto percentual de menores e estudantes ( $6,6\% + 4,6\% = 11,2\%$ ) confirma o fato relatado na literatura de que o índice de tentativas de suicídio entre adolescentes é preocupante.

Tabela 3.3: Distribuição de profissões entre pacientes potencialmente suicidas

Profissão	Freqüência	Proporção
Serviços gerais *	75	0,248
Doméstica **	55	0,182
Do lar	53	0,175
Indeterminada	29	0,096
Emprego especializado***	23	0,076
Menor	20	0,066
Desempregado	15	0,050
Estudante	14	0,046
Lavrador	12	0,040
Autônomo	04	0,013
Aposentado	02	0,007
Total	302	1

\* garçom, encanador, pedreiro, frentista, operário, padeiro, açogueiro, borracheiro, etc.

\*\* copeira, faxineira, costureira e bordadeira

\*\*\* enfermeira, modelo, protético, escrivão, professor, digitador e vendedor

A Tabela 3.4 mostra a distribuição de tentativas de suicídio segundo faixa etária. A faixa etária mais prevalente é de 20 a 29 anos tanto no geral (37,42%) como para cada sexo separadamente: 36,12% e 38,06% para o sexo masculino e feminino, respectivamente.

Observe que as idades foram agrupadas em 7 classes, quase todas, exceto a última, de mesma amplitude (10 anos). O símbolo  $\vdash$  indica a

Tabela 3.4: Distribuição de idade dos pacientes potencialmente suicidas

Idade (anos)	Freqüência absoluta	Freqüência relativa
10 $\vdash$ 20	57	18,54
20 $\vdash$ 30	113	37,42
30 $\vdash$ 40	57	18,87
40 $\vdash$ 50	32	10,62
50 $\vdash$ 60	19	6,29
60 $\vdash$ 70	7	2,29
$\geq 70$	2	0,67
Indeterminada	13	4,3
Total	302	100

inclusão do extremo inferior da classe e a exclusão do limite superior. Por exemplo, na classe 10  $\vdash$  20 são incluídas crianças a partir de 10 anos de idade e pessoas com até 19 anos. Uma indicação alternativa poderia ser 10 a 19.

#### Etapas para a construção de tabelas de freqüências para dados agrupados

1. Encontrar o menor e o maior valores (denominados mínimo e máximo) do conjunto de dados.
2. Escolher um número de subintervalos ou classes, preferencialmente de igual tamanho (amplitude), que englobem todos os dados sem haver superposição dos intervalos. Os extremos dos intervalos são conhecidos como limites de classes.
3. Contar o número de elementos que pertencem a cada classe; este número é denominado freqüência absoluta, usualmente denotado por n.
4. Determinar a freqüência relativa de cada classe, dividindo a freqüência da classe pelo número total de observações.

Não existem normas fixas para a construção de tabelas de distribuição de freqüências, entretanto, algumas regras práticas e empíricas podem ser citadas. O número de classes deve variar de 5 a 15. Além disso, a experiência tem mostrado que o número de classes fixado como  $\sqrt{n}$  ou  $1 + \log_2 n$ , para tamanho da amostra  $n$ , são escolhas razoáveis. Feita a escolha do número de classes, o tamanho de cada classe é escolhido como o quociente entre a amplitude (diferença entre o maior e o menor números) do conjunto e o número de classes escolhido. Se for conveniente, este valor pode ser modificado de forma a facilitar a construção e interpretação da tabela. O limite inferior da primeira classe deve ser um pouco menor que a menor observação e o limite superior da última classe deve ser um pouco maior que a maior observação.

### Exemplo 3.6: Câncer de mama

De 1.938 pacientes atendidas no Departamento de Mastologia do Hospital Amaral Carvalho de Jaú/SP, no período de 1987 a 1990, 149 (7,7%) tiveram diagnóstico de câncer de mama confirmados pelo exame anátomo-patológico.

O tipo histológico mais comum foi o carcinoma infiltrativo de ductos mamários sem outras especificações. Ocorreu em 128 casos (85,9%), sendo que em 38 (29,7%) as pacientes tinham 60 anos ou mais.

Caldeira & Budin (1995) preocuparam-se em avaliar o número de casos avançados (estádios III e IV) nas faixas etárias. A idade foi dicotomizada como menos de 60 anos e 60 anos ou mais; 45,6% das pacientes pertenciam à faixa de idades mais avançadas.

A distribuição do estadiamento para cada uma das duas faixas de idade consideradas está na Tabela 3.5. A maioria das mulheres com menos de 60 anos teve uma classificação I e II do estadiamento, enquanto o percentual de mulheres nos estádios III e IV da outra faixa de idade foi apenas ligeiramente superior (51,5%) do que o do grupo com estádios mais iniciais (48,5%).

Tabela 3.5: Distribuição de idade das pacientes com câncer de mama segundo estadiamento

Idade (anos)	Estádios I e II		Estádios III e IV		Total	
	n	%	n	%	n	%
< 60	53	65,4	28	34,6	81	100
≥ 60	33	48,5	35	51,5	68	100
Total	86	57,7	63	42,3	149	100

### Exemplo 3.7: Levantamento epidemiológico em asilos

Em trabalho realizado na década de 80, Cunha et al. (1985) avaliaram 308 idosos (261 mulheres e 47 homens com idade variando entre 60 e 100 anos) residentes em dez asilos da região da Grande Belo Horizonte.

Para análise do grau de distúrbio cognitivo (demência) aplicou-se um teste mental de doze perguntas e, de acordo com a soma das perguntas respondidas corretamente, os residentes foram classificados em: sem demência (9 a 12 pontos), demência leve (7 a 8 pontos) e demência grave (0 a 6 pontos).

As funções do comportamento foram testadas através da escala de avaliação geriátrica de Crichton, composta por onze itens: mobilidade, orientação, comunicação, cooperação, agitação, vestuário, alimentação, continência urinária, sono, humor objetivo (idéias delirantes e/ou alucinações) e humor subjetivo. Em cada item, as notas variaram de 1 (normal) a 5 (muito alterado) e a partir da soma dos 11 itens foi adotada a seguinte classificação do grau de deterioração: sem deterioração (11 pontos), deterioração leve (12 a 20 pontos), deterioração moderada (21 a 30 pontos) e deterioração grave ( $\geq 31$  pontos).

As tabelas 3.6 e 3.7 apresentam a distribuição de variáveis da função de comportamento de maior interesse relacionadas com a presença ou ausência de demência. Note que a coluna “totalmente alterado” é parte da coluna “com alguma alteração” e foi separada para destacar os casos de total incapacidade. Os percentuais referentes a alguma incapacidade das quatro primeiras funções do comportamen-

to listadas são significativamente menores para idosos sem demência. Observe também que não faz sentido calcular o percentual da última linha (Total) pois não fica claro o total do denominador, já que um residente pode apresentar mais de uma disfunção de comportamento.

Tabela 3.6: Distribuição dos 81 idosos sem demência para algumas funções de comportamento

Funções de comportamento	Grau de incapacidade		
	Sem alteração	Com alguma alteração	Totalmente alterado
Mobilidade	61 (75,3%)	20 (24,7%)	7 (8,6%)
Agitação	60 (74,1%)	21 (25,9%)	1 (1,2%)
Continência urinária	75 (92,6%)	6 (7,4%)	2 (2,5%)
Sono	56 (69,1%)	25 (30,9%)	0 (0,0%)
Humor objetivo	46 (56,8%)	35 (43,2%)	6 (7,4%)
Total	298	107	16

Tabela 3.7: Distribuição dos 227 idosos com demência para algumas funções de comportamento

Funções de comportamento	Grau de incapacidade		
	Sem alteração	Com alguma alteração	Totalmente alterado
Mobilidade	117 (51,5%)	110 (48,5%)	40 (17,6%)
Agitação	121 (53,3%)	106 (46,7%)	6 (2,6%)
Continência urinária	145 (63,9%)	82 (36,1%)	59 (26,0%)
Sono	131 (57,7%)	96 (42,3%)	7 (3,1%)
Humor objetivo	23 (10,1%)	204 (89,9%)	55 (24,2%)
Total	537	598	167

A maioria dos idosos sem demência não apresenta alteração no grau de incapacidade. No grupo com demência, a maioria (89,9%) apresenta alguma alteração de humor objetivo.

### Exemplo 3.8: Melanoma

Melanoma é uma neoplasia de grande potencial maligno responsável por grande parte das mortes por câncer de pele. O reconhecimento clínico do tumor em estágio inicial é difícil e de fundamental importância, já que o único procedimento terapêutico realmente eficaz é a excisão cirúrgica precoce.

Na análise de 35.695 biópsias realizadas entre 1982 e 1992 no Hospital Universitário da UFES em Vitória-ES, Lucas et al. (1994) encontraram 55 casos, sendo 30 primitivos de pele, 4 primitivos do globo ocular e 21 metastáticos com sede primária não confirmada.

A Tabela 3.8 mostra a distribuição da localização anatômica dos melanomas primitivos de pele, sendo a cabeça/pescoço mais comum.

Tabela 3.8: Distribuição de melanomas por localização anatômica

Localização anatômica	Nº de casos	Percentual
Cabeça/pescoço	10	33,3
Tronco	7	23,3
Membros superiores	6	20,0
Membros inferiores	2	6,7
Acral	5	16,7
Total	30	100

A Tabela 3.9 compara os casos confirmados de melanoma e o diagnóstico clínico. O grande percentual de diagnósticos incorretos (23,3%) pode prejudicar a adequação do tratamento.

Tabela 3.9: Distribuição de melanomas segundo diagnóstico clínico

Diagnóstico clínico	Nº de casos	Percentual
Correto	13	43,3
Duvidoso	10	33,3
Incorreto	7	23,3
Total	30	100

### 3.3.2 Gráficos

Existem inúmeras formas gráficas interessantes (ver por exemplo Tufte, 1983). Limitaremos a mostrar tipos de gráficos mais simples, mas possivelmente os mais úteis para a identificação da forma de um conjunto de dados e de sua descrição: diagrama de barras, histograma, polígono de freqüências, ogiva e gráfico de linhas.

#### Representação gráfica para variáveis categóricas

As variáveis categóricas podem ser representadas por um diagrama de barras, em que o tamanho de cada barra é proporcional ao número de indivíduos na categoria. Um exemplo ilustra melhor o processo de construção e a utilidade desta descrição gráfica.

#### Exemplo 3.9: Fraturas de face

Em uma amostra de 130 casos de fraturas de face registrados no Pronto Socorro do Hospital das Clínicas da Faculdade de Medicina da USP, Almeida et al. (1995) encontraram grande maioria de vítimas do sexo masculino (78,4%) e na faixa etária entre 20 e 40 anos. A Figura 3.1 apresenta o diagrama de barras com os agentes etiológicos.

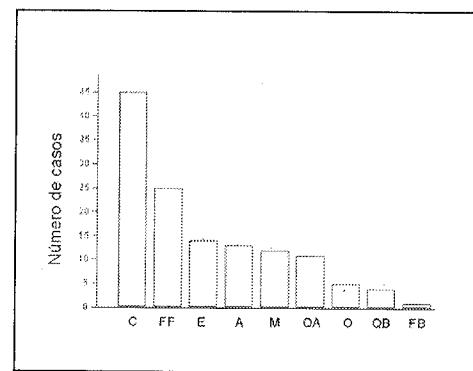


Figura 3.1: Diagrama de barras dos seguintes agentes etiológicos: carro (C), ferimentos de arma de fogo (FF), espancamento (E), atropelamento (A), moto (M), queda de altura (QA), queda de bicicleta (QB), ferimentos de arma branca (FB) e outros (O)

#### Representação gráfica para variáveis quantitativas

Resultados referentes a variáveis contínuas freqüentemente são organizados em tabelas de distribuições de freqüências através do agrupamento dos dados em classes. Três tipos de gráficos geralmente são utilizados neste caso: histograma, polígono de freqüências e ogiva.

O *histograma* é um gráfico de barras justapostas em que no eixo horizontal está a variável de interesse, dividida em classes geralmente de mesmo tamanho. No eixo vertical, constrói-se uma barra para cada classe com altura igual à freqüência absoluta ou relativa correspondente. A barra é centrada no ponto médio da classe.

Para construir um histograma, temos, portanto, que ordenar os dados e escolher uma escala apropriada para sua descrição. Estas tarefas, embora simples, são trabalhosas se o conjunto de dados é grande. Por isto é usual construir-se a tabela de freqüências como etapa preliminar na construção do histograma.

#### Exemplo 3.10: Nível de colesterol (continuação)

Para os dados do Exemplo 3.2,  $n = 78$ , o maior valor é 479 e o menor, 118. Escolhas razoáveis para o número e o tamanho de classe são 8 e 50, aproximação de  $(479 - 118)/8 = 45,125$ . Escolhendo o início da primeira classe 100, obtemos a Tabela 3.10.

Tabela 3.10: Distribuição do nível de colesterol

Nível de colesterol	Freqüência absoluta		Freqüência relativa	
	simples	acumulada	simples	acumulada
100 $\vdash$ 150	2	2	0,03	0,03
150 $\vdash$ 200	24	26	0,31	0,34
200 $\vdash$ 250	35	61	0,45	0,79
250 $\vdash$ 300	14	75	0,18	0,97
300 $\vdash$ 350	1	76	0,01	0,98
350 $\vdash$ 400	1	77	0,01	0,99
400 $\vdash$ 450	0	77	0,00	0,99
450 $\vdash$ 500	1	78	0,01	1,00
Total	78	-	1,00	-

Com uma simples inspeção do histograma da Figura 3.2 vemos que a grande maioria dos indivíduos tem colesterol em torno de 225. Temos ainda uma boa descrição de como os diferentes níveis se distribuem em torno deste valor.

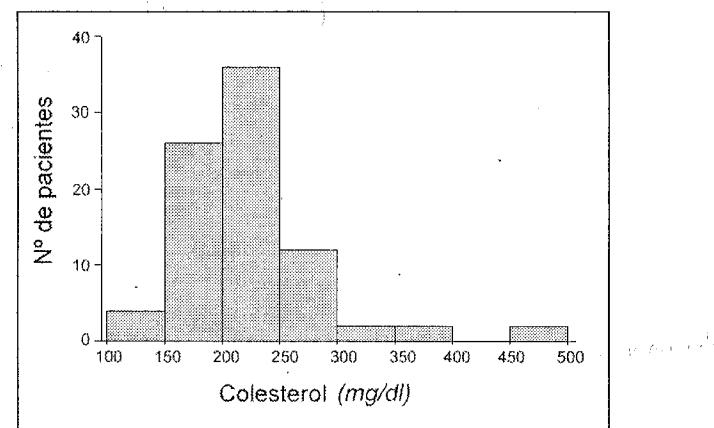


Figura 3.2: Histograma para o nível de colesterol

A partir do histograma pode-se construir o *polígono de freqüências*, que consiste em unir através de segmentos de reta as ordenadas correspondentes aos pontos médios de cada classe.

A *ogiva* é um gráfico de freqüências acumuladas (usualmente relativas). Para construí-la, coloca-se no eixo horizontal os intervalos de classe nos quais a variável em estudo foi dividida. Para cada limite de intervalo é assinalado no eixo vertical sua porcentagem acumulada. Em seguida, os pontos marcados são ligados por segmentos de reta. Embora a palavra ogiva não sugira, trata-se de uma poligonal ascendente.

O histograma e o polígono de freqüências servem para visualizar a forma da distribuição da variável estudada. Através da ogiva pode-se estimar *percentis* da distribuição, isto é, o valor que é precedido por certa porcentagem de interesse pré-estabelecida. Por exemplo, estimar o valor da variável abaixo do qual se tem 50% dos indivíduos.

### Exemplo 3.11: Dosagem de ácido úrico

Uma das características da *gota*, distúrbio hereditário que ocorre especialmente em homens, é a hiperuricemia (excesso de ácido úrico no sangue). Embora nem todos os indivíduos com hiperuricemia desenvolvam manifestações clínicas de gota, há muito tempo tem-se reconhecido que o componente hereditário da doença está ligado a uma tendência para hiperuricemia. A forma da distribuição de ácido úrico é um aspecto importante para o estudo dessa relação, especialmente quando se compara diferentes tipos de indivíduos, por exemplo, indivíduos doentes e sadios.

Tabela 3.11: Distribuição de freqüência da dosagem de ácido úrico

Ácido úrico (mg/dl)	Freqüência absoluta	Porcentagem	
		simples	acumulada
3,0 ← 3,5	2	0,7	0,7
3,5 ← 4,0	15	5,6	6,3
4,0 ← 4,5	33	12,4	18,7
4,5 ← 5,0	40	15,0	33,7
5,0 ← 5,5	54	20,2	53,9
5,5 ← 6,0	47	17,6	71,5
6,0 ← 6,5	38	14,2	85,7
6,5 ← 7,0	16	6,0	91,7
7,0 ← 7,5	15	5,6	97,3
7,5 ← 8,0	3	1,1	98,4
8,0 ← 8,5	1	0,4	98,8
8,5 ← 9,0	3	1,1	100,0
Total	267	100,0	—

A Tabela 3.11 apresenta a distribuição de freqüências das dosagens de ácido úrico sérico em 267 homens sadios, obtidas por Finn et al. (1966). O histograma, o polígono de freqüências e a ogiva são apresentados na Figura 3.3.

Estes gráficos mostram que a distribuição da dosagem de ácido úrico em indivíduos normais do sexo masculino é razoavelmente

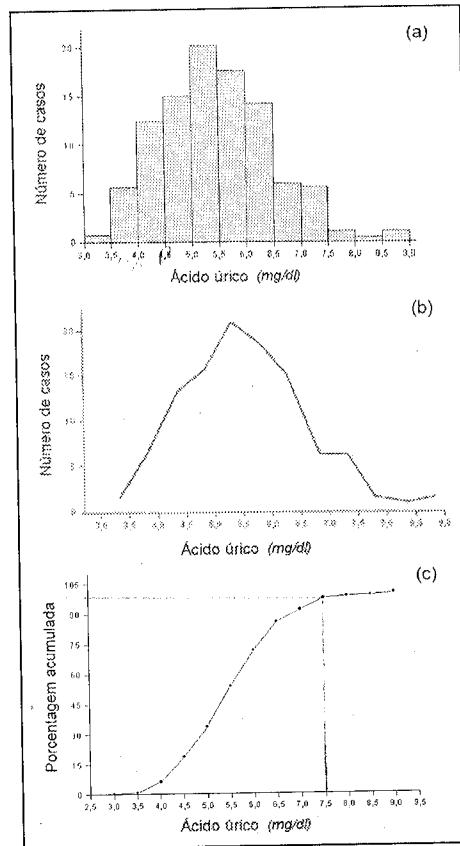


Figura 3.3: Distribuição da dosagem de ácido úrico: (a) histograma, (b) polígono de freqüências e (c) ogiva

simétrica. A variação desta dosagem está entre 3,0 e 9,0 mg/dl com uma maior concentração em torno de 5,0 a 5,5 mg/dl (20,2%). Além disso, 96,6% dos homens desta amostra apresentaram valores de ácido úrico sérico de 3,5 a 7,5 mg/dl.

Pode-se estimar que 50% dos indivíduos normais do sexo masculino apresentam uma dosagem de ácido úrico abaixo de 5,4 mg/dl e 95% abaixo de 7,25 mg/dl. Essas estimativas, denominadas percentis, foram obtidas pela projeção das porcentagens desejadas no eixo horizontal da ogiva. Este assunto será tratado novamente na Seção 3.4.5.

## Representação gráfica de dados temporais

Dados coletados ao longo do tempo são muito comuns em pesquisas médicas e em registros de Saúde Pública. Tais dados são denominados temporais ou longitudinais, ou ainda de medidas repetidas.

É comum o uso do diagrama de barras, discutido anteriormente, especialmente quando o período está agrupado (por exemplo, menos de 1 ano, 1 a 5 anos, 5 a 10 anos, etc.). Entretanto, o *gráfico de linhas* é uma representação mais apropriada para este tipo de dados. Consiste em colocar no eixo horizontal do gráfico a escala temporal (ano, mês, dia, etc.) e no eixo vertical a variável a ser estudada (freqüência, taxa ou medida tomada). É usual unir os pontos através de segmentos de reta, daí o nome recebido.

Através desses gráficos é possível constatar algum tipo de tendência e identificar alguns eventos inusitados, como por exemplo, o surto de uma determinada doença.

### Exemplo 3.12: AIDS no Brasil no período de 1980 a 1992

O número de casos de AIDS notificados ao Ministério da Saúde é periodicamente publicado no Boletim Epidemiológico de AIDS. Na Figura 3.4 vemos o número anual de casos notificados no Brasil no período de 1983 a 1995 segundo o sexo.

A incidência no sexo masculino é bem maior que no feminino. Essa tendência é mantida, mas há maior proximidade das curvas.

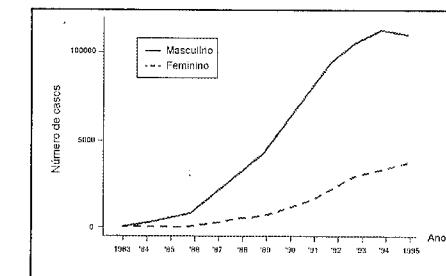


Figura 3.4: Número de casos notificados de AIDS no período de 1983 a 1995 segundo o sexo

### Exemplo 3.13: Hormônio luteinizante

O hormônio luteinizante (LH) é produzido na adenohipófise e exerce efeito trófico sobre testículo e ovário. Uma de suas principais ações na mulher consiste no estímulo à síntese de progesterona, ação influenciada pela variação da concentração dessa substância ao longo do ciclo menstrual.

Em estudo relatado em Diggle (1990), pesquisadores avaliaram os níveis de LH em uma mulher por um período de oito horas através de dosagens sucessivas com intervalos de dez minutos. Os valores das 48 medidas estão dispostos na Figura 3.5.

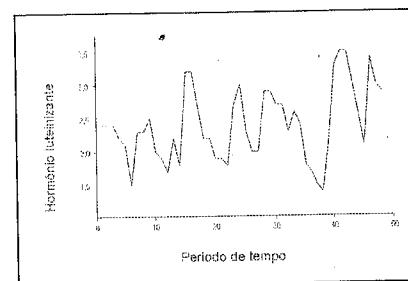


Figura 3.5: Dosagens de hormônio luteinizante obtidas em uma mulher num período de 8 horas com intervalos de 10 minutos

## 3.4 Síntese numérica

A análise descritiva de dados, além da construção de tabelas e gráficos, consiste também no cálculo de medidas que ajudam na produção de uma visão global dos dados. Tais medidas recebem o nome genérico de *estatísticas*.

### 3.4.1 Medidas de tendência central

Medidas de tendência central vão sintetizar em um único número o conjunto de dados. Procura-se definir um valor que represente bem

a distribuição da variável de interesse. No caso de uma distribuição simétrica, tal elemento é a *média aritmética*. Se a distribuição for assimétrica, a medida a ser usada é a *mediana*. Em distribuições simétricas, a média e a mediana na amostra devem ser aproximadamente iguais.

### Média

Existem vários tipos de média (aritmética, ponderada, geométrica, harmônica, etc.), mas estudaremos apenas a média aritmética. Como simplificação, a média aritmética será chamada apenas média daqui por diante.

A média aritmética de  $n$  observações  $(x_1, x_2, \dots, x_n)$  é denotada por  $\bar{x}$  (lê-se x barra) e é dada por

$$\bar{x} = \frac{x_1 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n} \quad (3.1)$$

A média pode ser interpretada como o “centro de gravidade”, isto é, o ponto de equilíbrio das discrepâncias positivas e negativas.

### Exemplo 3.14: Peso de recém-nascidos

Para exemplificar os cálculos, vamos utilizar os seguintes pesos em  $kg$  de 10 recém-nascidos:

3,2 3,2 2,8 2,1 2,9 3,1 3,2 3,0 3,5 4,0

A média dos pesos dos recém-nascidos é

$$\bar{x} = \frac{3,2 + 3,2 + 2,8 + 2,1 + \dots + 4,0}{10} = 3,1$$

O peso médio é de  $3,1\ kg$  ou  $3100\ g$ . Obviamente alguns recém-nascidos têm peso abaixo e outros acima deste valor, mas a média é um valor típico.

## Mediana

Uma medida de centro adequada para distribuições assimétricas é a mediana, que será denotada por  $\tilde{x}$  (lê-se x til). Por definição a mediana é o valor que divide a distribuição ao meio. Em outras palavras, 50% das observações ficam acima da mediana e 50% abaixo.

Para calcularmos a mediana é necessário primeiramente ordenar a amostra para que possamos localizar a *posição da mediana* e assim encontrar seu valor. Para amostras ordenadas, a posição da mediana é dada pelo elemento de ordem  $\frac{n+1}{2}$  se  $n$  for ímpar e pela média dos elementos de ordens  $\frac{n}{2}$  e  $\frac{n+2}{2}$  se  $n$  for par.

Como ilustração, considere as seguintes amostras ordenadas:

$$1 \ 2 \boxed{5} \ 6 \ 7 \quad 1 \ 2 \ \boxed{5 \ 6} \ 7 \ 7$$

Os elementos marcados localizam a mediana. Na primeira amostra, a mediana é o próprio elemento 5, isto é,  $\tilde{x}_1 = 5$ . Na segunda, a mediana é dada pela média aritmética dos elementos grifados, isto é

$$\tilde{x}_2 = \frac{5 + 6}{2} = 5,5$$

Voltando aos pesos dos recém-nascidos e dispondo-os em ordem crescente, obtemos:

$$2,1 \ 2,8 \ 2,9 \ 3,0 \ 3,1 \ 3,2 \ 3,2 \ 3,2 \ 3,5 \ 4,0$$

Como o número de observações é par,  $n = 10$ , a mediana é

$$\tilde{x} = \frac{3,1 + 3,2}{2} = 3,15$$

No cálculo da média, todos os valores da amostra são levados em conta, ao passo que no caso da mediana isto não acontece. Por essa razão, valores muito grandes ou muito pequenos, comparados com demais valores da amostra, causam grandes variações na média, o que em geral não ocorre com a mediana. Por isso, dizemos que a mediana é uma medida robusta, isto é, resistente a valores atípicos.

### 3.4.2 Medidas de variabilidade

Quase nenhuma única medida de tendência central é suficiente para descrever, de modo satisfatório, um conjunto de dados. Não basta saber o valor em torno do qual os dados se concentram. É preciso conhecer também o grau de agregação, ou seja, definir e usar medidas da dispersão dos dados.

Podem ser definidas várias medidas de variabilidade. Apresentaremos a seguir as duas mais utilizadas em análise de dados: variância e desvio-padrão.

#### Variância

A variância é uma medida da variabilidade dos dados em torno da média. É natural procurar uma medida de dispersão que dependa dos desvios de cada observação em relação à média ( $x_i - \bar{x}$ ), e seria razoável considerar a soma de todos estes desvios.

A idéia seria que, quanto maiores fossem os desvios, maior seria a variabilidade presente nos dados. Entretanto, pela definição de média,  $\sum_{i=1}^n (x_i - \bar{x}) = 0$  para qualquer conjunto de dados.

Uma alternativa seria, portanto, elevar os desvios de cada observação em relação à média ao quadrado, isto é, a quantidade a ser calculada seria  $\sum_{i=1}^n (x_i - \bar{x})^2$ . Desta forma somamos somente valores positivos.

Torna-se necessário considerar o número de observações, pois o acréscimo de observações aumenta o valor deste somatório. Uma idéia seria considerar a média, isto é,  $\sum_{i=1}^n (x_i - \bar{x})^2 / n$ . Em palavras, a variância é a média dos desvios ao quadrado das observações em relação à média.

Neste texto usaremos uma outra definição de variância que consiste em dividir por  $n - 1$  ao invés de  $n$ . A justificativa é meramente técnica. O importante é ser consistente, usando sempre uma mesma definição dentro de um mesmo trabalho.

Recomendamos a divisão por  $n - 1$  por ser o padrão das calculadoras e programas de computador. Se as amostras são grandes, os valores obtidos dividindo-se por  $n$  ou  $n - 1$  são praticamente iguais.

A notação para a variância é  $s^2$  e é calculada como

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \quad (3.2)$$

que pode ser reescrita como

$$s^2 = \frac{\sum_{i=1}^n x_i^2 - n\bar{x}^2}{n-1} \quad (3.3)$$

ou ainda

$$s^2 = \frac{\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2/n}{n-1} \quad (3.4)$$

Como as fórmulas (3.2), (3.3) e (3.4) são algebricamente equivalentes, a variância pode ser calculada por qualquer uma delas. Entretanto, em geral, a fórmula (3.4) induz a menos erros de aproximação e a expressão (3.2) deve ser usada se os números envolvidos são grandes.

### Desvio-padrão

Note que a unidade de medida de  $s^2$  é a unidade de medida das observações elevada ao quadrado. Por exemplo, se as observações foram medidas em  $cm$ , a unidade de  $s^2$  é  $cm^2$ . Logo, para obter uma medida de variabilidade com a mesma unidade das observações, extraímos a raiz quadrada. Esta medida é denominada desvio-padrão e é dada por

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad (3.5)$$

Portanto, para obtermos o desvio-padrão, calculamos primeiramente a variância e então extraímos a raiz quadrada.

### Exemplo 3.15: Peso de recém-nascidos (continuação)

Para o cálculo da variância dos pesos de recém-nascidos do Exemplo 3.14, organizamos os dados na Tabela 3.12. Indicamos na Tabela 3.13 o cálculo da variância através das três fórmulas.

Tabela 3.12: Cálculos necessários para obtenção do desvio-padrão

$i$	$x_i$	$x_i^2$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	3,2	10,24	0,1	0,01
2	3,2	10,24	0,1	0,01
3	2,8	7,84	-0,3	0,09
4	2,1	4,41	-1,0	1,00
5	2,9	8,41	-0,2	0,04
6	3,1	9,61	0,0	0,00
7	3,2	10,24	0,1	0,01
8	3,0	9,00	-0,1	0,01
9	3,5	12,25	0,4	0,16
10	4,0	16,00	0,9	0,81
Total	31,0	98,24	0,0	2,14

Tabela 3.13: Cálculo da variância e do desvio-padrão

Fórmula	Variância	Desvio-padrão
(3.2)	$s^2 = \frac{2,14}{9} = 0,24$	$s = \sqrt{0,24} = 0,49$
(3.3)	$s^2 = \frac{98,24 - 10(3,1)^2}{9} = 0,24$	$s = \sqrt{0,24} = 0,49$
(3.4)	$s^2 = \frac{98,24 - (31)^2/10}{9} = 0,24$	$s = \sqrt{0,24} = 0,49$

Portanto, a variância e o desvio-padrão deste conjunto de dados são  $0,24 \text{ kg}^2$  e  $0,49 \text{ kg}$ , respectivamente.

### 3.4.3 Coeficiente de variação

Uma pergunta que pode surgir é se um desvio-padrão é grande ou pequeno; questão revelante, por exemplo, na avaliação da precisão de métodos. Um desvio-padrão pode ser considerado grande ou pequeno dependendo da ordem de grandeza da variável. Por exemplo, um desvio-padrão de 10 pode ser insignificante se a observação típica for 10.000, mas será um valor bastante significativo para um conjunto de dados cuja observação típica é 100.

Portanto, por vezes, é conveniente exprimir a dispersão em termos relativos, ou seja, expressar a variabilidade dos dados tirando a influência da ordem de grandeza da variável.

Pode-se obter um índice relativo de dispersão comparando-se o desvio-padrão ( $s$ ) com a média ( $\bar{x}$ ). A medida utilizada é denominada *coeficiente de variação* e é definida por

$$CV = \frac{s}{\bar{x}} \quad (3.6)$$

O coeficiente de variação (CV) é adimensional, isto é, um número puro e é usualmente expresso em porcentagem. É zero quando não houver variabilidade entre os dados, ou seja, quando  $s = 0$ , o que ocorre quando todos os valores da amostra são iguais.

Sua utilidade é fornecer uma medida para a homogeneidade do conjunto de dados. Quanto menor o coeficiente de variação, mais homogêneo é o conjunto de dados.

Um valor de CV menor ou igual a 0,25 geralmente indicará que o conjunto de dados é razoavelmente homogêneo. Entretanto, esse padrão varia de acordo com a aplicação. Uma possível classificação é a seguinte: baixo (inferior a 0,10); médio (de 0,10 a 0,20); alto (de 0,20 a 0,30) e muito alto (superior a 0,30).

Se por um lado pode ser difícil classificar um coeficiente de variação como baixo, médio, alto ou muito alto, esta medida pode ser bastante útil na comparação de duas variáveis ou dois grupos que a princípio não são comparáveis (por exemplo, com ordens de grandeza das variáveis diferentes).

### Exemplo 3.16: Comparaçāo do colesterol em dois grupos

Em um grupo de jovens médicos residentes obteve-se, ao medir o colesterol, a média de 205 mg/dl e um desvio-padrão de 22 mg/dl. Para um grupo de médicos especialistas, entretanto, a média obtida foi de 244 mg/dl e desvio-padrão de 45 mg/dl. O grupo de médicos mais idosos apresenta não só uma média mais alta como também maior variabilidade em torno da média. O coeficiente de variação capta esta diferença. Neste caso, o coeficiente de variação é 10,7% para os residentes e 18,4% para os especialistas.

### 3.4.4 Escore padronizado

Na seção anterior vimos como relacionar a média e o desvio-padrão para caracterizar a homogeneidade de um grupo. Esta seção também trata de uma forma de relacionar estas duas estatísticas, mas para cada indivíduo.

A idéia é que, na comparação dos resultados de dois indivíduos, é importante a padronização em relação ao grupo. Por exemplo, um aluno que tenha obtido nota 7 numa prova em que a média da classe foi 5 apresentou melhor rendimento do que numa prova em que tirou 8 mas a média da classe foi 9. Além da comparação da nota individual com a média da classe também é importante avaliar se em cada caso a variabilidade das notas foi grande ou não. Assim, o desempenho deste aluno que obteve nota 7 seria bom se o desvio-padrão da classe fosse 2 e apenas razoável se o desvio-padrão da classe fosse 4.

Vamos tomar como referência os mesmos 10 bebês do Exemplo 3.14 e suponhamos que nasça um bebê de 4,1 kg. Este peso está praticamente dois desvios-padrão acima da média porque

$$\frac{x_i - \bar{x}}{s} = \frac{4,1 - 3,1}{0,49} = 2,04$$

Sejam  $x_1, \dots, x_n$  os dados observados em uma amostra de tamanho  $n$ ,  $\bar{x}$  e  $s$  a média e o desvio-padrão, respectivamente. Definimos

$$z_i = \frac{x_i - \bar{x}}{s}, \quad i = 1, \dots, n \quad (3.7)$$

O valor  $z_i$  é denominado *escore padronizado* da observação  $x_i$  e é o desvio da  $i$ -ésima observação em relação à média dividido pelo desvio-padrão. Portanto, esta medida fornece o número de desvios-padrão que a observação dista da média da amostra. Esta interpretação justifica o uso do termo desvio-padrão como a unidade segundo a qual os desvios em relação à média devem ser medidos.

Surge então o interesse em saber quando um escore padronizado deve ser considerado grande. Para responder a esta pergunta necessitamos de conhecimentos que ultrapassam o nível pretendido para este texto. Através de um resultado de Estatística Matemática conhecido como Desigualdade de Chebyshev, pode-se obter a seguinte regra:

*Para a maioria dos conjuntos de dados, 80% dos dados estão no intervalo centrado na média com amplitude de 4 desvios-padrão. Poucas observações estão além de dois desvios-padrão e raramente há uma observação além de 3 desvios-padrão.*

Diante disto, o bebê de 4,1 kg tem um peso muito diferente dos outros dez. Não se esperava, a priori, um valor tão grande como o que ocorreu. Devem-se procurar razões substantivas para este fato.

### Exemplo 3.17: Dosagens laboratoriais

A Tabela 3.14 apresenta os resultados de exames laboratoriais solicitados a duas pacientes, mãe (A) e filha (B), com respectivamente 60 e 40 anos de idade. Também são apresentados os resultados padronizados pelo grupo de adultos do sexo feminino.

Tabela 3.14: Média ( $\bar{x}$ ) e desvio-padrão ( $s$ ) para adultos do sexo feminino e resultados de exames laboratoriais de mãe (A) e filha (B)

Exame	$\bar{x}$	$s$	Resultado original		Escore padronizado	
					A	B
Glicemia em jejum	85	12,5	90	79	0,40	-0,48
Ácido úrico	4,2	0,9	3,5	3,1	-0,78	-1,22
Triglicérides	105	30	97	66	-0,27	-1,30
Colesterol total	200	25	251	185	2,04	-0,60

A paciente A apresentou um resultado de colesterol bastante alto (dois desvios-padrão acima da média). A paciente B não apresentou nenhum resultado preocupante.

### 3.4.5 Percentis

Dados que produzem histogramas simétricos são adequadamente descritos e sintetizados pela média e pelo desvio-padrão. Neste caso, várias perguntas sobre o comportamento dos dados podem ser respondidas usando-se somente estes dois números.

Isto não ocorre quando os dados são assimétricos. Vimos como a mediana identifica mais adequadamente o “centro” de um conjunto de dados com distribuição assimétrica. Além disso, para entender bem uma distribuição, precisamos conhecer valores acima ou abaixo dos quais se encontra uma determinada porcentagem dos dados: os percentis.

A mediana é o percentil de ordem 50. Usando uma formulação mais precisa que a da Seção 3.4.1, definimos mediana como o valor que deixa *pelo menos* 50% das observações acima de si e *pelo menos* 50% abaixo. Esta definição, embora possa parecer desnecessariamente complicada, é a forma correta de se tratar qualquer tipo de dados.

De forma geral, o percentil de ordem  $x$ , representado por  $P_x$ , é o valor que é precedido (maior ou igual) por  $(xn)/100$  dos valores e seguido (menor ou igual) por  $[(100 - x)n]/100$ .

Os percentis de ordem 25, 50 e 75 são chamados, respectivamente primeiro, segundo e terceiro quartis porque dividem a distribuição em  $1/4$ ,  $2/4 = 1/2$  e  $3/4$ . São representados por  $Q_1$ ,  $Q_2$  e  $Q_3$  e, evidentemente,  $Q_2$  é outra notação para a mediana.

O cálculo dos percentis é feito aplicando-se a definição ou podem ser obtidos de forma aproximada a partir de projeções da ogiva.

### Exemplo 3.18: Nível de albumina no sangue

O nível de albumina no sangue, um indicador do estado nutricional, foi medido em um grupo de 60 pacientes, obtendo-se os resultados ( $g/dl$ ) apresentados em forma ordenada na Tabela 3.15.

Tabela 3.15: Nível de albumina no sangue (g/dl)

4,44	4,47	4,48	4,51	4,54	4,54	4,61	4,64	4,66	4,68
4,68	4,69	4,71	4,73	4,76	4,76	4,76	4,81	4,86	4,86
4,87	4,88	4,90	4,90	4,95	4,95	4,96	4,97	4,98	4,98
4,99	5,00	5,01	5,01	5,01	5,02	5,04	5,05	5,08	5,09
5,09	5,10	5,11	5,11	5,16	5,17	5,18	5,18	5,19	5,24
5,24	5,26	5,27	5,27	5,29	5,32	5,35	5,46	5,50	5,85

$P_{25}$

O primeiro quartil, conforme a definição anterior, deixa pelo menos 25% dos dados abaixo e pelo menos 75% dos dados acima dele. Como 25% de 60 é igual a 15,  $Q_1$  tem pelo menos 15 valores abaixo de si e pelo menos 45 acima. Contando-se 15 valores do menor para o maior, teremos o paciente com valor de albumina igual a 4,76 g/dl. Começando do maior para o menor e contando 45 valores, encontramos também 4,76 g/dl. Assim, a dosagem 4,76 g/dl é o valor do primeiro quartil. Podemos dizer que pelo menos 25% dos pacientes da amostra apresentam nível de albumina inferior ou igual a 4,76 g/dl.

O percentil de ordem 80 ( $P_{80}$ ) deixa pelo menos 80%, ou seja 48 valores abaixo de si e pelo menos 12 acima. Contando 48 valores a partir do menor, chegamos ao valor de 5,18 g/dl e contando 12 valores do maior para o menor obtemos 5,19 g/dl. Portanto,  $P_{80}$  é qualquer valor entre 5,18 e 5,19. Por convenção, usa-se a média destes dois valores, ou seja, o valor 5,185 g/dl.

O cálculo dos percentis para dados agrupados é feito através do gráfico das frequências relativas acumuladas, a ogiva.

Para obter o percentil traçamos uma reta paralela ao eixo horizontal, a partir do percentil que se quer calcular. No encontro desta reta com o gráfico traçamos uma reta perpendicular ao eixo horizontal, encontrando assim, no eixo horizontal, o valor procurado. A Figura 3.6 exemplifica a regra de cálculo.

A noção de percentil é especialmente usada em Pediatria, em que a altura e o peso são utilizados como índices do estado geral de crescimento. Por isto existem tabelas e gráficos apresentando, para cada

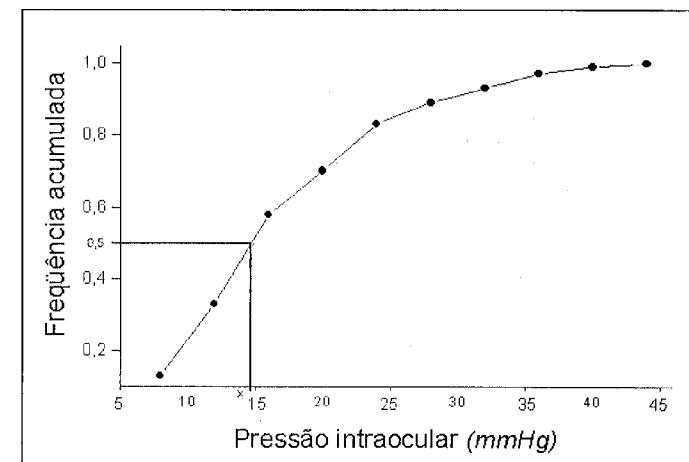


Figura 3.6: Cálculo de percentil através da ogiva

idade, o valor médio para o peso e altura, assim como a sua variação normal. Dada a importância destas informações, foram desenvolvidos padrões em vários países do mundo. Hoje, no entanto, há uma clara tendência para a adoção de um único padrão mundial. Este está disponível em um antigo programa de computador, distribuído pela Organização Mundial da Saúde.

#### Exemplo 3.19: Taxa de ácido úrico (continuação)

Através da Figura 3.3, vemos que o percentil 95 da taxa de ácido úrico é igual a 7,25 mg/dl, isto é,  $P_{95} = 7,25 \text{ mg/dl}$ . Isto significa que cerca de 5% dos indivíduos da amostra têm taxas acima de 7,25 mg/dl.

#### 3.4.6 Outras medidas descritivas

Existem muitas outras medidas descritivas. São particularmente importantes os coeficientes que descrevem as condições da saúde pública de um país. Existem, por exemplo, inúmeros indicadores epidemiológicos de morbidade e mortalidade.

É socialmente muito importante saber, para todos os nascidos vivos de uma região, sua condição (se vivo ou morto) ao fim do primeiro ano de vida. Estes dados permitem a construção do coeficiente de mortalidade infantil (*CMI*) da região dado por

$$CMI = \frac{\text{nº de óbitos de menores de 1 ano, área A, ano } t}{\text{nº de nascidos vivos, área A, ano } t} \times 1000$$

A medida que vão melhorando as condições de vida e de saúde da população, em uma área, vai diminuindo a mortalidade infantil. A morte de menores de um ano é diretamente influenciada por condições de saneamento, nutrição, educação, habitação, assistência pré-natal e no parto. Enfim, pode-se dizer que o *CMI* está ligado diretamente às condições sócio-econômicas da população.

O estudo de índices como o *CMI* é objeto da disciplina de Epidemiologia ou Saúde Pública.

## 3.5 Outros tópicos

### 3.5.1 Coleta de dados

Existem várias maneiras de se coletar dados, sendo que o instrumento de medida pode depender do tipo de estudo ou tipo de variável analisada. Por exemplo, em um experimento envolvendo cobaias, é comum a medida direta das variáveis de interesse. Em um inquérito, o questionário é o instrumento de medida mais utilizado. Em pesquisas clínicas, o instrumento de medida pode ser um formulário de informações, o prontuário do paciente, ou a ficha que será preenchida na unidade.

Vários cuidados devem ser tomados na elaboração e utilização de um instrumento de pesquisa. A definição de quais dados devem ser coletados, como captar as informações, a clareza das questões do questionário são extremamente importantes. Para esse último item, são referências úteis Mucchielli (1979), Barros & Victora (1991) e Bennett & Ritchie (1975).

### 3.5.2 Banco de dados

Após a coleta dos dados, o próximo passo é a construção do banco de dados. Como, em geral, a consolidação dos dados é feita através de uso do computador, são necessários os conceitos básicos de entrada de dados em programas computacionais e de uso de planilhas eletrônicas.

O grau de dificuldade para a construção de um banco de dados pode variar, dependendo do número de variáveis, tipos de dados, tamanho da amostra, etc.

Nunca é demais lembrar a importância da realização da conferência de dados especialmente para amostras grandes e quando várias pessoas estão envolvidas na coleta, organização e preparação do banco de dados. Essa conferência deve ser baseada em informações existentes, como, por exemplo, verificar a existência de valores fora da amplitude esperada da variável e verificações lógicas. Observações suspeitas devem ser conferidas e os erros encontrados, corrigidos. Quando não for possível fazer a correção no dado identificado, recomenda-se codificá-lo como se fosse um dado omissão. Tempo gasto no início da análise para verificação dos dados é um tempo bem empregado; erros nos dados não identificados no curso da análise principal dos dados exigirão que toda a análise seja refeita.

### Codificação

Para variáveis categóricas devem-se criar códigos para cada categoria. Eles são arbitrários e a escolha pode depender dos objetivos da análise. A mudança de códigos essencialmente não muda os resultados, mas uma escolha adequada pode facilitar a interpretação.

Para variáveis dicotômicas, os códigos mais convenientes são 0 e 1. Por exemplo 0 pode representar ausência e 1 presença de determinada característica. Para variáveis com  $k$  categorias devem ser criadas  $k - 1$  variáveis indicadoras (0 e 1), pois elas definem completamente cada categoria. São assim denominadas porque apenas indicam a que categoria o dado pertence. Usualmente a categoria de referência corresponde à situação em que todas as variáveis indicadoras são iguais a 0 (ou a 1).

Para variáveis contínuas é recomendável entrar com os dados originais e não os codificados para classes de interesse, pois pode haver necessidade de mudança nas classes durante a análise. Por exemplo, para idade só se deve entrar com faixas etárias se houver interesse apenas em faixas pré-fixadas.

Para dados omissos, deve-se usar um código que facilmente identifique esse tipo de dado. A forma mais usual consiste em completar o número de dígitos com o valor 9 (9, 99, 999, etc.), naturalmente se esse não é um possível valor para a variável. Por exemplo, para pressão arterial usamos 999 e para sexo o código seria 9.

### Exemplo 3.20: Tentativas de suicídio (continuação)

A Tabela 3.16 exemplifica o banco de dados do Exemplo 3.5. Os códigos usados para sexo foram 0 para masculino e 1 para feminino e para a variável profissão foram os seguintes: 1 - serviços gerais, 2 - doméstica, 3 - do lar, 4 - indeterminada, 5 - emprego especializado, 6 menor, 7 - desempregado, 8 - estudante, 9 - lavrador, 10 - autônomo e 11 - aposentado. O indivíduo nº 1 é um homem de 25 anos de idade, trabalhador de serviços gerais. O indivíduo nº 2 é uma doméstica de 48 anos e o indivíduo nº 302 é uma estudante de 13 anos.

Tabela 3.16: Banco de dados do estudo de tentativa de suicídio

Indivíduo	Sexo	Profissão	Idade
1	0	1	25
2	1	2	48
:	:	:	:
302	1	8	13

### Exemplo 3.21: Câncer de mama (continuação)

A Tabela 3.17 mostra possíveis códigos para as variáveis idade e estadiamento conforme as categorias usadas no artigo descrito no Exemplo 3.6.

Tabela 3.17: Códigos para idade e estadiamento

Categorias	Códigos
Idade	
< 60	0
≥ 60	1
Estadiamento	
Estádios I e II	0
Estádios III e IV	1

Suponha agora que os estádios não sejam agrupados e que o estadiamento tenha 4 classificações (I, II, III e IV). Neste caso, podemos criar três variáveis indicadoras ( $I_1$ ,  $I_2$  e  $I_3$ ), como indicadas na Tabela 3.18.

Tabela 3.18: Códigos para estadiamento com 4 classificações

Estádio	$I_1$	$I_2$	$I_3$
I	0	0	0
II	1	0	0
III	0	1	0
IV	0	0	1

### 3.5.3 Arredondamento

Na regra básica de arredondamento, os dígitos excedentes devem ser descartados se o último deles é menor que 5 e, em caso contrário o último dígito retido é acrescido de 1. Por exemplo, se os dados forem anotados como 87,72; 90,58 e 98,04 e quisermos usar apenas uma casa decimal, então os arredondamentos serão respectivamente 87,7; 90,6 e 98,0.

Como cada vez mais os cálculos são feitos através de computadores e calculadoras que fazem os arredondamentos automaticamente, este assunto não é tão importante em termos de cálculos. Entretanto, para

uma boa apresentação dos resultados, as seguintes recomendações devem ser observadas:

1. Sempre que possível, usar o mesmo número de casas decimais em uma análise.
  2. Em geral duas casas decimais são suficientes, a não ser que o número seja menor que 1 e muito pequeno, tal que um número maior de casas decimais seja necessário para distinguir uma observação de outra (por exemplo 0,0001 e 0,0002).
  3. Não se deve arredondar um número mais de uma vez, pois pode resultar em um arredondamento incorreto. Por exemplo, o valor 87,348074 arredondado para um número com duas casas decimais é 87,35 e se este número for arredondado novamente para um número com uma casa decimal é 87,4. Entretanto, se o valor original for arredondado para um número com uma casa decimal, o número resultante é 87,3 (diferente do primeiro esquema de arredondamento).
  4. Arredondamentos não devem ser feitos até o final apresentação dos resultados, isto é, a precisão completa deve ser mantida durante toda a análise.

### 3.5.4 Boxplot

Um tipo de gráfico muito útil para a descrição de dados, visualização de sua variabilidade, comparação entre diferentes grupos é o gráfico em caixas, *boxplot*, em inglês. Foi introduzido pelo estatístico americano John Tukey em 1977.

Para a construção do *boxplot* obtém-se primeiro as seguintes estatísticas: 1º quartil ( $Q_1$ ), mediana ( $Q_2$ ), 3º quartil ( $Q_3$ ) e a distância interquartílica ( $DQ$ ), definida como  $DQ = Q_3 - Q_1$ . O *boxplot* é obtido seguindo-se os seguintes passos:

1. Numa reta são marcados o 1º quartil ( $Q_1$ ), a mediana ( $Q_2$ ) e o 3º quartil ( $Q_3$ ).
  2. Acima dessa reta constrói-se um retângulo com limites iguais às posições do primeiro e terceiro quartis, cortado por um segmento de reta na posição relativa à mediana.

3. A partir dos limites do retângulo, traçam-se linhas até:

  - (a) encontrar um extremo (valor máximo ou mínimo) ou
  - (b) um valor correspondente a 1,5 DQ, se o extremo correspondente estiver a mais de 1,5 DQ do quartil respectivo.

Os pontos que estão a mais de 1,5 DQ do quartil correspondente até 3,0 DQ são chamados *pontos externos* e os que estão a mais de 3,0 DQ, *pontos soltos*. Existem símbolos especiais para representar no *boxplot* os pontos externos e os soltos. Serão utilizados os símbolos \* e 0 para indicar os pontos externos e soltos respectivamente. Um esquema de *boxplot* é apresentado na Figura 3.7.

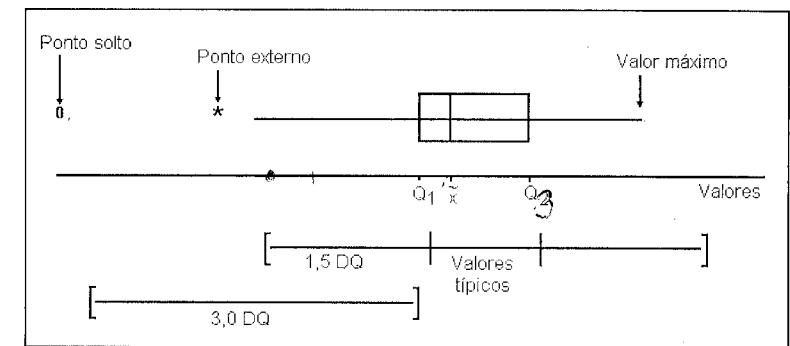


Figura 3.7: Esquema para construção do *boxplot*

O *boxplot* também fornece informações importantes sobre o comportamento do conjunto de dados, como simetria e variabilidade. Se a amplitude for muito maior que a distância interquartílica e a mediana estiver mais próxima do 1º quartil do que do 3º quartil, há fortes indicações de assimetria positiva e de grande dispersão das observações.

### **Exemplo 3.22: Doença de Chagas**

Mota (1992) faz um estudo comparativo de crianças filhas de mães chagásicas com filhas de mães não chagásicas (grupo controle).

Neste estudo foram consideradas dosagens obtidas em exames laboratoriais, medidas de variáveis vitais, antropométricas e de desenvolvimento motor (firmar a cabeça, sentar e andar).

Para o grupo de mães chagásicas, os dados já ordenados da dosagem de bilirrubina ( $mg/dl$ ) são os apresentados na Tabela 3.19

Tabela 3.19: Dosagem de bilirrubina ( $mg/dl$ ) de 46 crianças filhas de mães chagásicas

1,3	1,5	1,9	2,0	2,1	2,8	2,8	2,9	2,9	3,1
3,1	3,2	3,2	3,2	3,2	3,2	3,3	3,7	3,7	3,8
3,9	4,0	4,3	4,3	4,4	4,9	6,0	6,1	6,2	6,3
6,3	6,6	6,7	6,8	6,8	7,8	8,2	8,3	8,8	9,5
9,7	10,1	10,7	11,2	11,3	12,5				

As estatísticas descritivas são  $\bar{x} = 5,404$ ,  $s = 2,967$ ,  $x_{\min} = 1,300$ ,  $x_{\max} = 12,500$ ,  $Q_1 = 3,175$ ,  $Q_2 = 4,300$ ,  $Q_3 = 7,050$ . O boxplot correspondente é apresentado na Figura 3.8.

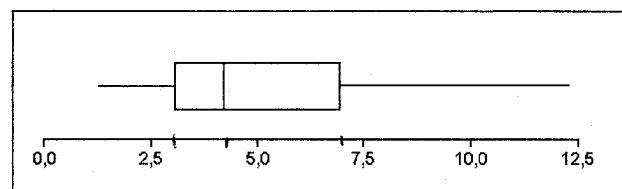


Figura 3.8: Boxplot da dosagem de bilirrubina ( $mg/dl$ )

### 3.5.5 Observações atípicas (*outlier*)

É muito comum aparecerem entre os dados coletados, observações atípicas (*outliers*), isto é, valores muito grandes ou muito pequenos em relação aos demais. Um conjunto de dados pode apresentar apenas um ou vários *outliers*.

Observações atípicas alteram enormemente as médias e variabilidade dos grupos a que pertencem e podem até mesmo distorcer as

conclusões obtidas através de uma análise estatística padrão. Portanto, é de fundamental importância detectar e dar um tratamento adequado a elas. É sempre boa prática fazer-se uma inspeção dos dados no início da análise estatística. Técnicas descritivas de dados têm um papel importante nesta fase.

### Causas do aparecimento de *outliers*

Dentre as possíveis causas do aparecimento de *outliers*, podem-se citar as seguintes:

- Leitura, anotação ou transcrição incorreta dos dados.
- Erro na execução do experimento ou na tomada da medida.
- Mudanças não controláveis nas condições experimentais ou dos pacientes.
- Característica inerente à variável estudada (por exemplo, grande instabilidade do que está sendo medido).

Por exemplo, em um experimento oftalmológico sobre os efeitos do trauma devido à fotocoagulação com laser na íris de coelhos, o pesquisador mediu, entre outras variáveis, a concentração de proteínas totais no humor aquoso (líquido contido no espaço entre a córnea e o cristalino). Observou que uma medida era muito maior do que as outras. Procurando a causa, verificou que havia caído um mosquito na amostra analisada aumentando consideravelmente seu teor protéico.

### Como detectar *outliers*

As questões básicas são quais observações devem ser consideradas como *outliers* e como detectá-los. Existem procedimentos para responder a essas perguntas.

Os *outliers* podem ser detectados simplesmente por uma verificação lógica dos dados, através de gráficos específicos ou ainda através de testes apropriados. Uma forma gráfica usual é o boxplot apresentado na Seção 3.5.4.

Na verificação lógica dos dados, pode-se testar se as observações estão dentro de faixas de valores esperados, confirmam-se classificações dos dados, entre outros procedimentos. Em consequência é possível eliminar inconsistências e erros encontrados.

### Exemplo 3.23: Doença de Chagas (continuação)

A distribuição de idade em que as crianças filhas de mães chagásicas firmaram a cabeça é apresentada na Tabela 3.20.

Tabela 3.20: Distribuição da idade ao firmar a cabeça de 51 crianças filhas de mães chagásicas

Idade ao firmar a cabeça (meses)	Frequência absoluta	Porcentagem	
		simples	acumulada
2	10	19,60	19,60
3	28	54,90	74,50
4	10	19,60	94,10
5	1	1,96	96,06
6	1	1,96	98,02
17	1	1,96	100,00
Total	51	100,00	-

O valor 17 meses está muito fora do padrão (3 meses segundo a literatura), o que é confirmado no *boxplot* apresentado na Figura 3.9.

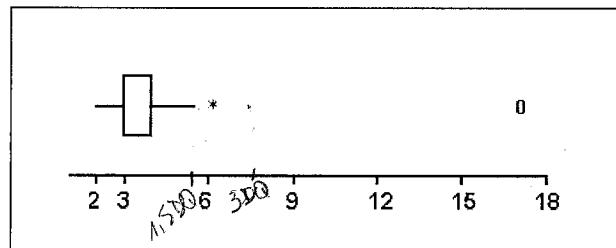


Figura 3.9: Boxplot da idade ao firmar a cabeça (meses)

Foram identificados dois *outliers*: um ponto externo (o valor 6) e um ponto solto (o valor 17). Baseado na literatura médica, o primeiro valor detectado não foi tido como aberrante e assim optou-se por não retirá-lo da análise. A criança que firmou a cabeça aos 17 meses era portadora de paralisia cerebral.

### Medidas a serem tomadas

Quando um *outlier* é detectado, duas medidas podem ser tomadas: abandoná-lo ou conservá-lo. Existem justificativas para cada uma dessas medidas e o tipo de análise pode variar, dependendo se o *outlier* foi ou não eliminado.

Um *outlier* deve ser eliminado da análise quando houver uma justificativa convincente para isto, por exemplo quando a observação é incorreta ou houve erro na execução do experimento ou na medida tomada. Após a eliminação do *outlier* pode-se fazer a análise estatística usando-se apenas as observações restantes, ou uma análise mais sofisticada, assunto que foge ao nível deste texto.

Por outro lado, se nenhuma explicação pode ser dada à observação atípica, o *outlier* pode refletir uma característica do que está sendo estudado. Neste caso, tal observação deve ser incluída na análise e um tratamento especial deve ser dado aos dados. Por exemplo, pode-se usar uma ponderação da influência das observações ou alternativamente uma transformação ( $\sqrt{x}$ ,  $\log x$ , etc.) da variável estudada. Para mais detalhes consultar Snedecor & Cochran (1989) e Altman (1991).

### Exemplo 3.24: Contagem de linfócitos T CD<sub>4</sub>

A Tabela 3.21 mostra os resultados de Shapiro et al. (1986) referentes a contagens de linfócitos T CD<sub>4</sub> de 20 pacientes em remissão de doença de Hodgkin ( $\bar{x} = 823,2$ ;  $s = 566,4$ ) e de 20 pacientes em remissão de malignidades disseminadas não Hodgkin ( $\bar{x} = 522,1$ ;  $s = 293,0$ ).

O valor 2415 do paciente nº 20 do grupo com doença de Hodgkin parece estar um pouco alto. Se for considerado como um *outlier* e

descartado do conjunto de dados, a média e desvio-padrão das 19 observações caem respectivamente para 739 e 436.

Tabela 3.21: Número de células T4 /  $mm^3$  em amostras de sangue

Nº	Hodgkin	Não Hodgkin
1	171	116
2	257	151
3	288	192
4	295	208
5	396	315
6	397	375
7	431	375
8	435	377
9	554	410
10	568	426
11	795	440
12	902	503
13	958	675
14	1004	688
15	1104	700
16	1212	736
17	1283	752
18	1378	771
19	1621	979
20	2415	1252

### 3.6 Exemplo comentado: Refluxo vesicureteral primário em crianças

No período compreendido entre 1971 e 1985, a Unidade de Nefrologia Pediátrica do Hospital das Clínicas da UFMG desenvolveu um estudo com crianças portadoras de infecção do trato urinário (ITU) e refluxo vesicureteral (RVU) descrito em Diniz et al. (1996).

Todas foram estudadas a partir de surto de ITU, suspeitado pela sintomatologia, sendo hipertensão o fato mais importante. Cultura quantitativa com número de colônias igual ou superior a  $10^5$  por ml de urina, colhida em condições adequadas, confirmou o diagnóstico.

Estudos anatômicos do aparelho urinário realizaram-se a partir de urografia excretora (UE) e uretrocistografia miccional (UCM). Os refluxos foram classificados a partir da UCM nos graus I, II, III, IV e V conforme Comitê Internacional de Refluxo. Avaliaram-se ainda medidas renais (diâmetros longitudinal e transversal e espessura do córtex) e presença de cicatrizes renais.

#### 3.6.1 Descrição da amostra no início do estudo

Um total de 48 crianças foram acompanhadas sob protocolo por período médio de 4 anos. Todas as crianças eram portadoras de RVU sem outra anormalidade do trato urinário, sendo que em 26 pacientes era unilateral e 22 bilaterais, totalizando 70 unidades renais. Considera-se unidade renal o conjunto envolvido no fenômeno de refluxo vesicureteral representado pelo rim, sistema excretório intrarenal, pelve e ureter até sua junção com bexiga.

Houve predominância do sexo feminino (81%), com a idade no início do estudo variando entre um mês e oito anos e nove meses.

#### 3.6.2 Acompanhamento

Seguindo o diagnóstico de ITU, as crianças foram submetidas à terapêutica antimicrobiana por 10 dias. Após melhora sintomática, a droga foi suspensa por 5 dias e nova urulcultura realizada para confirmação de esterilização da urina. Pacientes com urina comprovadamente “livre de bactéria” receberam quimioprofilaxia por tempo prolongado com sulfametoxazol associado ao trimetoprim ou nitrofurantoína.

Exames radiológicos dos 48 pacientes foram repetidos com intervalos de um a três anos durante o período de observação, tendo cada paciente pelo menos dois estudos radiológicos no período de estudo.

A Tabela 3.22 mostra a distribuição dos graus do refluxo no início do estudo e a distribuição de casos de cura e de quadros estáveis.

O RVU foi tratado clínica ou cirurgicamente, salientando-se que os pacientes submetidos à correção cirúrgica do refluxo possuíam, em

Tabela 3.22: Distribuição dos graus do refluxo no início do estudo, de casos de cura e de quadros estáveis segundo o grau de refluxo

Grau de RVU	Inicial		Cura		Estável	
	n	%	n	%	n	%
I	10	14,3	3	6,8	5	33,3
II	25	35,7	18	40,9	5	33,3
III	8	11,4	5	11,4	1	6,7
IV	15	21,4	7	15,9	3	20,0
V	12	17,01	11	25,0	1	6,7
Total	70	100	44	100	15	100

grande maioria, graus mais graves de refluxo além de apresentarem com maior freqüência lesões de cicatrizes renais antes mesmo da intervenção terapêutica.

Apesar do tratamento (conservador ou cirúrgico) estabelecido, houve 8 casos de piora: duas unidades passando de I para II, duas unidades de III para IV e quatro unidades de IV para V. Em três unidades houve melhora: duas de II para I e uma de IV para I.

A partir de grau do RVU observado a cada verificação radiológica e do gráfico de linhas correspondente, criou-se um índice de evolução do refluxo (*IERVU*). Foi definido como o quociente entre a área observada (*AO*) e a área sob a curva que seria obtida caso o grau inicial fosse mantido (*AI*), isto é,  $IERVU = AO/AI$ . Esse índice reflete três situações distintas, ilustradas na Figura 3.10.

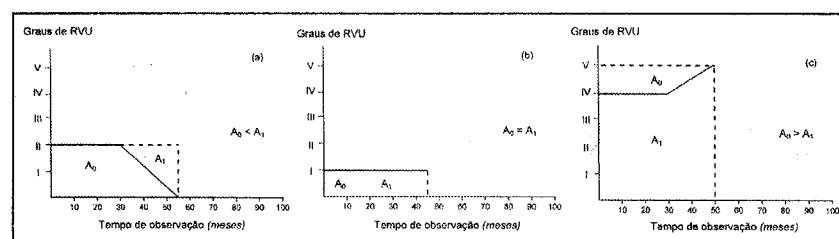


Figura 3.10: Gráficos exemplificando a evolução do grau de refluxo: (a) melhora, (b) estável e (c) piora

Há indicação de melhora do RVU se  $IERVU < 1$  ( $AO < AI$ ), de RVU estável se  $IERVU = 1$  ( $AO = AI$ ) e de piora do RVU se  $IERVU > 1$  ( $AO > AI$ ).

A relação temporal entre o início dos sintomas e os diagnósticos de ITU e RVU é mostrada na Tabela 3.23.

Tabela 3.23: Distribuição segundo faixa etária no início da sintomatologia, à época do diagnóstico de ITU e de RVU

Grupo etário (anos)	Início da sintomatologia		Diagnóstico de ITU		Diagnóstico de RVU	
	n	%	n	%	n	%
0 - 1	32	67	30	63	13	27
1 - 5	14	29	16	33	28	58
$\geq 5$	2	4	2	4	7	15
Total	48	100	48	100	48	100

Não houve retardos consideráveis para o diagnóstico do ITU, não ocorrendo o mesmo com o diagnóstico do RVU. Esses fatos são claramente identificados ao se observar que 63% das crianças estudadas tiveram diagnóstico de ITU com idade inferior a 1 ano e que, no mesmo período, somente 27% tiveram o diagnóstico de RVU estabelecido. É interessante destacar que o diagnóstico de ITU em crianças muito novas não é tão freqüente, devido à má interpretação de hipertensão nessa faixa etária.

A distribuição de cicatrizes renais iniciais, sua evolução ao longo do acompanhamento e o surgimento de novas cicatrizes são apresentadas na Tabela 3.24. Das 21 unidades renais com cicatrizes no início do estudo, houve progressão em 16. Em cinco casos houve piora concomitante do grau de refluxo (quatro de IV para V e um de I para II). Nas sete unidades renais em que apareceram novas cicatrizes, o refluxo já havia sido curado, três com tratamento conservador (grau II) e quatro cirúrgicos (graus III e IV).

A Tabela 3.25 mostra a distribuição etária dos pacientes segundo a ocorrência de cicatrizes no início do estudo, com notável predominio no primeiro ano de vida.

Tabela 3.24: Distribuição de cicatrizes renais segundo grau do refluxo

Cicatrizes renais	Grau de RVU					Total
	I	II	III	IV	V	
Iniciais	2	2	1	10	6	21
Estáveis	1	0	0	4	0	5
Progressão	0	3	1	2	10	16
Novas	0	3	0	1	3	7
Final	1	6	1	7	13	28

Tabela 3.25: Distribuição etária dos pacientes segundo a ocorrência de cicatrizes no início do estudo

Grupo etário (anos)	Rins com cicatrizes	
	n	%
0 † 1	12	57
1 † 5	8	38
≥ 5	1	5
Total	21	100

### 3.7 Exercícios

- Classifique as variáveis dos Exemplos 3.3 e 3.4. (use os tipos descritos na Seção 3.2.2).
- Esboce um banco de dados para a pesquisa descrita na Seção 3.6, incluindo a codificação das variáveis envolvidas.
- Use a sua calculadora para obter o desvio-padrão dos 10 pesos de recém-nascidos, usados como exemplo na Seção 3.4.2. Está sendo usada a fórmula que contém  $n$  ou  $n - 1$  no denominador?
- Diga com suas próprias palavras o que é o desvio-padrão de um conjunto de números.
- Uma conseqüência da regra da Seção 3.4.4 é que a amplitude dos dados (a diferença entre o maior e menor valores) vale aproxi-

madamente 6s. Usando essa regra, obtenha intuitivamente uma estimativa para o desvio-padrão das idades de estudantes que cursam no primeiro ano de Medicina.

- O que acontece com a média e com o desvio-padrão:
  - Se um mesmo número é somado a todos os elementos de um conjunto de dados?
  - Se cada elemento de um conjunto de dados for multiplicado por um valor constante?
- Comente as seguintes afirmativas:
  - Sempre a metade dos dados está abaixo da média.
  - A média é o valor típico de um conjunto de dados.
  - Enquanto tivermos alunos com rendimento abaixo da média, não poderemos descansar.
- Apresentamos à seguir os resultados da segunda prova da turma C do 1º semestre de 1990 da disciplina de Estatística para Medicina.
 

6	12	12	14	15	15	15	15	16
17	18	18	19	19	19	20	21	21
22	22	22	23	23	23	23	23	23
24	25	25	25	27	27	28	28	32

  - Calcule a média e a mediana.
  - Calcule o primeiro e terceiro quartis. Explique o significado destes números.
  - Considere também os resultados das turmas A e B. Compare as notas das três turmas quanto a sua homogeneidade.

Turma	Média	Desvio-padrão
A	22,5	4,5
B	24,0	5,4

9. Utilizando os dados do Exemplo 3.2 construa uma tabela de freqüência com os dados agrupados para um número e uma amplitude de classes convenientes.

10. Consideremos 12 observações (ordenadas) do tempo de internação (dias) de pacientes acidentados no trabalho, em um certo hospital: 1, 4, 7, 9, 10, 13, 15, 17, 17, 18, 19, 21. Obtenha os quartis e interprete estes valores.

11. O tempo (em meses) entre a remissão de uma doença e a recidiva de 48 pacientes de uma determinada clínica médica foi registrado. Os dados ordenados são apresentados a seguir, separadamente para os sexos masculino (M) e feminino (F):

M	2	2	3	4	4	4	4	7	7	7	8	9
	9	10	12	15	15	15	16	18	18	22	22	24
F	2	2	3	3	4	4	5	5	6	6	7	7
	7	7	8	8	8	8	10	10	11	11	12	18

- (a) Calcule a média, o desvio-padrão, a mediana e o coeficiente de variação para cada sexo. Interprete.  
 (b) Repita os cálculos pedidos em (a) para todos os 48 pacientes. Compare com os resultados de (a).

12. Um laboratório resolveu divulgar, além do valor da dosagem, a ordem do percentil para pessoas sadias a ela associada. Interprete os seguintes resultados de uma pessoa que fez exames neste laboratório:

Albumina 5,3 g/dl percentil 95  
 Colesterol 180 mg/dl percentil 5

13. A tabela a seguir apresenta os dados de um teste de psicoanalogia (um teste de inteligência em que um indivíduo tem que resolver uma série de analogias). A amostra contém 158 indivíduos que receberam notas de acordo com o rendimento no teste.

Escore	Freqüência absoluta	
	simples	acumulada
33 ⊢ 36	1	1
36 ⊢ 39	3	4
39 ⊢ 42	1	5
42 ⊢ 45	4	9
45 ⊢ 48	3	12
48 ⊢ 51	7	19
51 ⊢ 54	15	34
54 ⊢ 57	23	57
57 ⊢ 60	16	73
60 ⊢ 63	24	97
63 ⊢ 66	21	118
66 ⊢ 69	21	139
69 ⊢ 72	15	154
72 ⊢ 75	4	158

- (a) Construa o histograma e a ogiva.  
 (b) Localize as classes que contêm o 1º quartil ( $Q_1$ ), a mediana e o percentil de ordem 90 ( $P_{90}$ ).  
 (c) Acima de que nota encontram-se 80% dos indivíduos? A que percentil corresponde este valor?

14. Os dados da tabela a seguir referem-se à taxa de colesterol ( $mg/dl$ ) de um grupo A de 50 estudantes com idade entre 10 e 19 anos de determinada escola pública encontra-se na tabela:

Taxa de colesterol	Freqüência absoluta		Freqüência relativa	
	simples	acumulada	simples	acumulada
112 ⊢ 132	1			
132 ⊢ 152	3			
	6			
	9			
	14			
	9			
	5			
	2			
	1			
Total	50			

- (a) Complete a tabela.  
 (b) Construa o histograma e a ogiva.  
 (c) Calcule graficamente  $P_{90}$ . Qual é o seu significado?  
 (d) Outro grupo de indivíduos (B), com faixa etária entre 50 e 60 anos, apresentou uma taxa de colesterol média igual a  $254 \text{ mg/dl}$  com um desvio-padrão de  $48 \text{ mg/dl}$ . Qual dos dois grupos é mais homogêneo? Por que?  
 (e) O que podemos suspeitar, comparando as médias dos dois grupos?
15. Em uma pesquisa sobre a concentração de minerais no leite materno, foram coletados no período de 1984 a 1985 dados de 55 mães do Hospital Maternidade Odete Valadares em Belo Horizonte. As mães foram divididas em dois grupos, segundo o período de lactação: colostro e leite maduro. Os minerais considerados foram, cobre, magnésio e zinco. Os dados a seguir, extraídos de Nunes et al. (1988), referem-se ao cálcio e zinco:

Cálcio ( $\mu\text{g/ml}$ de leite) - Grupo colostro									
113	181	254	311	334	145	221	256	312	344
163	225	275	313	372	163	231	296	323	375
167	241	303	325	375	437				
Cálcio ( $\mu\text{g/ml}$ de leite) - Grupo maduro									
159	175	181	188	200	206	213	214	217	231
238	238	242	244	256	259	260	263	264	275
277	279	281	293	302	303	314	344	394	
Zinco ( $\mu\text{g/ml}$ de leite) - Grupo colostro									
1,07	1,20	1,30	3,13	3,20	3,70	4,40	4,57	5,20	5,82
6,07	6,13	6,50	6,82	6,82	6,90	7,23	7,42	7,43	7,43
7,45	8,25	8,40	8,77	9,54					
Zinco ( $\mu\text{g/ml}$ de leite) - Grupo maduro									
0,52	0,60	0,78	0,86	1,08	1,23	1,28	1,31	1,40	1,55
1,55	1,60	1,92	2,15	2,27	2,40	2,41	2,52	2,57	2,60
2,63	2,92	3,02	3,05	3,57	3,88	5,50	7,38		

Para as concentrações de cálcio e zinco em cada grupo:

- (a) Calcule medidas de tendência central e de variabilidade, o coeficiente de variação e os quartis. Apresente essas estatísticas em uma tabela contendo inclusive o título. Comente os resultados.
- (b) Verifique, construindo boxplots, se há outliers e, em caso afirmativo, classifique-os.
- (c) Quais gráficos seriam indicados para esses dados?
16. Uma das formas que um pediatra usa para avaliar o grau de controle do diabetes é a dosagem da glicohemoglobina, que é medida em % da hemoglobina total. Valores acima de 12 indicam que o paciente não está sob controle. Em um grupo de 15 crianças obtiveram-se os seguintes resultados para 13 delas:

10,5	11,2	11,7	12,1	12,3	12,3	12,3
12,4	12,4	12,5	13,1	13,1	13,1	13,5

Por motivos associados à limitação do aparelho de análise, sabe-se apenas que os valores observados de dois pacientes são maiores que 14.

Para mostrar que este grupo tem um alto valor de glicohemoglobina, ele igualou a 15,5 os dois valores maiores que 14 e tirou a média dos 15 valores. Comente o procedimento usado, proponha e obtenha uma forma alternativa de síntese dos dados. ~~Médiana~~ — MED — ? ?

17. O gráfico a seguir mostra as composições, ordenadas por ano, de Robert Schumann, compositor alemão que viveu entre 1810 e 1856 (Jaminson, 1995). As obras estão representadas por seus números de opus.



142	145
127	141
077	133
057	137
059	138
051	106
049	102
041	101
035	098
043	096
042	094
040	092
039	091
038	095
030	093 136
035	092 123
034	091 121
033	095 119
036	095 148 117
035	092 130 113
034	079 120 112 143
032	079 125 111 134
031	079 097 110 133
030 120	075 094 109 148 132
032 020 064	072 084 115 074 060 107 147 131
024 021 022 027 054	050 080 081 073 050 105 146 126
004 022 017 018 023 026 052 047	058 065 071 070 068 104 139 123
003 010 011 014 012 016 020 026 038 041 050	056 061 063 068 069 087 103 126 116
007 001 005 002 005 090 009 013 006 016 019 024 027 041 046	056 059 052 068 067 093 100 122 114
1329 139 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 1656	1657 1658 1659 1660 1661 1662 1663 1664 1665 1666 1667 1668 1669 1670 1671 1672 1673 1674 1675 1676 1677 1678 1679 1680 1681 1682 1683 1684 1685 1686

- (a) Aproximadamente a qual percentil corresponde uma das mais conhecidas peças para piano do compositor: Cenas Infantis (opus nº 15)?
- (b) Em que ano Schumann completou 75% de seu trabalho? Qual o percentil correspondente e qual a composição que marca essa realização?
- (c) Em 1833, Schumann tentou suicidar-se pela primeira vez. Se ele tivesse conseguido realizar seu intento, qual seria o percentual de perda de suas obras para a humanidade?
- (d) Os anos 1840 e 1849 foram extremamente produtivos. Qual o percentual de sua obra foi composto nesse período?
18. No final de 1988, o Prof. Eder Quintão da Escola Paulista de Medicina, submeteu dez amostras do mesmo sangue a dez laboratórios da cidade de São Paulo. Segundo declaração do pesquisador (Folha de São Paulo de 3/08/89) os laboratórios escolhidos foram “os dez maiores e mais conhecidos” da cidade. Foram obtidos os seguintes resultados para o nível de colesterol e triglicérides ( $mg/dl$ ).

Colesterol	284	295	330	228	365
	335	361	333	294	318
Triglicérides	-	166	158	202	162
	135	82	150	86	121

Nota: - significa dado ausente

(a) Comente e quantifique a variação encontrada nestes resultados.

(b) O jornal apresentou a notícia com a seguinte manchete: “Médico constata erro em exames para colesterol”. O jornalista que escreveu esta manchete usou uma expressão adequada para apresentar os resultados?

19. A tabela a seguir mostra a distribuição do peso ao nascer dos 7.225 nascidos vivos no ano de 1992 na região urbana do distrito sede de Juiz de Fora (Souza & Costa, 1993). Não foram incluídas na tabela 99 crianças com peso ao nascer ignorado.

Peso ao nascer (g)	Freqüência absoluta		Percentual	
	simples	acumulada	simples	acumulado
500 à 1000	33			
1000 à 1500	60			
1500 à 2000	147			
2000 à 2500	614			
2500 à 3000	2077			
3000 à 3500	2875			
3500 à 4000	1113			
4000 à 4500	184			
4500 à 5000	20			
5000 à 5500	1			
5500 à 6000	2			
Total				

- (a) Complete a tabela.
- (b) Em que classe está a mediana da distribuição?
- (c) Acima de que peso encontram-se 90% dos nascidos vivos? Qual o percentil correspondente?
- (d) Construa o histograma e comente.
20. Obtenha os escores padronizados para os dados do Exemplo 3.14. Comente.
21. A seguir apresentamos tabelas obtidas pelo cruzamento de variáveis do estudo descrito no Exemplo 3.7.

Faixa de idade	Sexo		
	Masculino	Feminino	Total
60 + 75	30 (21,1%)	119 (79,9%)	
75 + 100	17 (10,7%)	142 (89,3%)	
Total			

Grau de deterioração	Demência			Total
	Sem	Leve	Grave	
Nulo	16 (94,1%)	1 (5,9%)	0 (0,0%)	
Leve	51 (41,1%)	27 (21,8%)	46 (37,1%)	
Moderado	10 (11,5%)	16 (18,4%)	61 (70,1%)	
Grave	4 (5,0%)	7 (8,8%)	69 (86,3%)	
Total				

Funções de comportamento	Alteração		
	Nenhuma	Alguma	Completa
Mobilidade	178 (57,8%)	130 (42,2%)	47 (15,3%)
Orientação	95 (30,8%)	213 (69,2%)	34 (11,0%)
Comunicação	84 (27,3%)	224 (72,7%)	17 (5,5%)
Cooperação	89 (28,9%)	219 (71,1%)	95 (30,8%)
Agitação	181 (58,8%)	127 (41,2%)	7 (2,3%)
Vestuário	203 (65,9%)	105 (34,1%)	40 (13,0%)
Alimentação	242 (78,6%)	66 (21,4%)	21 (6,8%)
Continência urinária	220 (71,4%)	88 (28,6%)	61 (19,8%)
Sono	187 (60,7%)	121 (39,3%)	7 (2,3%)
Humor objetivo	69 (22,4%)	239 (77,6%)	61 (19,8%)
Humor subjetivo	70 (22,7%)	238 (77,3%)	36 (11,7%)

- (a) Complete os totais das tabelas.
- (b) Comente os resultados apresentados nas tabelas.
- (c) Qual a prevalência da demência? humor objetivo
- (d) Qual o percentual de idosos com algum grau de deterioração nas funções de comportamento?

(e) Segundo os autores, 54,2% dos residentes que tinham algum grau de deterioração nas funções do comportamento necessitavam de ajuda importante nas tarefas diárias. Identifique esses casos.

(f) Baseado nessas tabelas e/ou na tabela apresentada no Exemplo 3.7, complete se possível o seguinte parágrafo: "A análise das variáveis em separado mostrou que .....% dos residentes apresentavam distúrbios da mobilidade; .....% eram agitados; .....% necessitavam de ajuda para se alimentar; .....% eram incontinentes de urina; .....% apresentavam distúrbios do sono e .....% alucinavam e/ou tinham idéias delirantes".

22. Para o exemplo sobre refluxo vesicureteral primário em crianças, apresentado na Seção 3.6:

- (a) Quantas meninas havia no grupo de pacientes desse estudo?
- (b) Justifique a importância da utilização do índice de evolução do refluxo (IERVU).
- (c) Em quantas unidades renais houve constatação de cicatrizes no início do estudo?
- (d) Para as tabelas a seguir, compare os grupos com ou sem cicatrizes em relação ao padrão de infecção urinária e ao acometimento unilateral ou bilateral dos rins.

Infecção urinária	Sem cicatrizes		Com cicatrizes	
	n	%	n	%
Sem recorrência	8	32,0	8	34,8
Recorrência ocasional	6	24,0	8	34,8
Recorrente	8	32,0	3	13,0
Persistente	3	12,0	4	17,0
Total	25	100	23	100

Ocorrência de cicatrizes	RVU			
	Unilateral		Bilateral	
	n	%	n	%
Sim	18	69	5*	23
Não	8	31	17	77
Total	26	100	22	100

\* cicatrizes em ambos os rins

23. Faça uma análise descritiva completa dos dados apresentados na Tabela 3.2 da Seção 3.2.3 (recomendável o uso de computador).
24. Construa o *boxplot* para os dados do Exemplo 3.24. Comente.
25. Com uso de um programa computacional, faça uma análise descritiva completa dos dados dos Exemplos 3.3 e 3.4.
26. Apresente exemplos de aplicação do que foi estudado nesse capítulo retirados de revistas, jornais e publicações científicas. Comente a adequacidade do uso da estatística nesses exemplos e se for o caso sugira modificações da análise e/ou da apresentação.

## Capítulo 4

# Probabilidade e Avaliação de Testes Diagnósticos

### 4.1 Introdução

Uma das experiências mais rotineiras da prática médica é a solicitação de um teste diagnóstico. Os objetivos são vários, incluindo a triagem de pacientes, o diagnóstico de doenças e o acompanhamento ou prognóstico da evolução de um paciente. Para chegar ao diagnóstico, o médico considera várias possibilidades, com níveis de certeza que variam de acordo com as informações disponíveis.

Um objetivo deste capítulo é, usando a linguagem da Probabilidade, mostrar como se mede o nível de certeza da ocorrência de um evento, por exemplo, a presença de uma doença após a observação de um teste positivo. Consideraremos o teste *positivo* quando indicar a presença da doença e *negativo* quando indicar a ausência.

Não existe teste perfeito, aquele que com certeza absoluta determina a presença ou ausência da doença. Assim, o objetivo principal deste capítulo é estudar os índices nos quais o conceito de qualidade de um teste diagnóstico é usualmente desmembrado.

Freqüentemente, um único teste não é suficiente, e portanto deve-se combinar dois ou mais testes. O ideal seria que, para cada patologia, fossem determinados os testes a serem incluídos no processo diagnóstico e a melhor forma de combiná-los. Neste capítulo também são apresentadas as formas mais comuns de combinação de testes e como medir a qualidade do teste conjunto.

Para atingir estes objetivos introduzimos de forma intuitiva conceitos elementares da Teoria das Probabilidades. Apresentamos formalmente os conceitos de sensibilidade, especificidade, valores de predição positivo e negativo.

## 4.2 Probabilidade: conceitos fundamentais

Conceitos probabilísticos são necessários para se estudar fenômenos aleatórios, isto é, situações em que os resultados possíveis são conhecidos, mas não se pode saber a priori qual deles ocorrerá.

Os primeiros resultados da Teoria das Probabilidades foram obtidos quando matemáticos do século XVII tiveram que responder perguntas de amigos sobre jogos de azar. Usaremos esta mesma motivação para ilustrar os conceitos básicos. Nossa objetivo, entretanto, é analisar situações muito mais complexas, colocadas pela medicina clínica.

Nesta seção apresentamos alguns resultados, essenciais para o cálculo de probabilidades na maioria das situações práticas.

### 4.2.1 Espaço amostral e eventos

O conjunto de todos os resultados possíveis de um experimento aleatório é chamado de *espaço amostral* e será denotado por  $E$ . Um *evento* é um subconjunto do espaço amostral. Quando constituído de apenas um elemento, chama-se *evento simples*.

#### Exemplo 4.1: Lançamento de um dado

Consideremos o experimento “lançamento de um dado”. Trata-se de um experimento aleatório. O espaço amostral é  $\{1, 2, 3, 4, 5, 6\}$ . São eventos simples:  $\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}$ . Outros eventos:  $A = \{1, 3\}$ ,  $B = \{2, 4, 6\} = \{\text{números pares}\}$ ,  $C = \{3, 5, 6\}$ .

#### Exemplo 4.2: Lançamento de dois dados

A situação é menos simples quando se lançam dois dados. O espaço amostral tem agora  $6 \times 6 = 36$  elementos, já que cada face pode ocorrer associada com todas as outras. Muitos podem ser os

eventos de interesse, por exemplo,  $F$ : “a soma dos dois valores é 10”;  $G$ : “os dois valores são iguais”;  $H$ : “uma das faces é maior que 3”. É fácil ver que  $F$  tem três elementos, isto é,  $F = \{(4, 6), (5, 5), (6, 4)\}$ ,  $G$  e  $H$  têm seis e vinte e sete elementos, respectivamente.

#### Exemplo 4.3: Dados genéricos de um teste clínico

Na etapa de pesquisa para a determinação do grau de confiabilidade de um teste diagnóstico, o pesquisador utiliza-o primeiramente em dois grupos muito específicos de pessoas: um de portadores da doença perfeitamente caracterizada e outro de pacientes sem a doença em questão. O diagnóstico, nesta etapa, deve ser feito por um meio diferente do teste em estudo, o chamado *padrão ouro*. Os resultados desta etapa da pesquisa podem ser resumidos na forma da Tabela 4.1.

Tabela 4.1: Esquema padrão de síntese dos dados para verificação da qualidade de um teste clínico

Doença	Teste		Total
	Positivo	Negativo	
Presente	$a$	$b$	$a + b$
Ausente	$c$	$d$	$c + d$
Total	$a + c$	$b + d$	$n$

Para definir os índices que descrevem o grau de confiabilidade de um teste, precisamos trabalhar com os seguintes eventos:

$T_+$  corresponde a teste positivo

$T_-$  corresponde a teste negativo

$D_+$  corresponde a indivíduos portador da doença

$D_-$  corresponde a indivíduo não portador da doença.

### 4.2.2 Cálculo de probabilidades

Nos exemplos 4.1 e 4.2, os eventos simples são equiprováveis. Suas probabilidades são, portanto, calculadas contando-se o número de

resultados favoráveis ao evento de interesse e dividindo-o pelo número total de possibilidades. Nestas circunstâncias, o cálculo da probabilidade de um evento requer simplesmente a contagem do número de resultados favoráveis e o número total de possibilidades.

#### Exemplo 4.4: Lançamento de dados

No Exemplo 4.1, a probabilidade de ocorrência de cada face é  $1/6$  e consequentemente:

$$Pr(A) = \frac{2}{6} \quad Pr(B) = \frac{3}{6} \quad Pr(C) = \frac{3}{6}$$

No Exemplo 4.2 os eventos simples são também equiprováveis. Portanto, as probabilidades dos eventos  $F$ ,  $G$ , e  $H$  são:

$$Pr(F) = \frac{3}{36} \quad Pr(G) = \frac{6}{36} \quad Pr(H) = \frac{27}{36}$$

#### Freqüência relativa

Na maioria das situações práticas, os elementos do espaço amostral não são igualmente prováveis e, deste modo, a probabilidade dos eventos deve ser calculada utilizando-se da noção de *freqüência relativa*, isto é, a proporção do resultado de interesse em sucessivas observações sob mesmas condições do fenômeno aleatório.

Por exemplo, a probabilidade de infecção nos leitos de uma ala cirúrgica de um hospital só pode ser calculada através da freqüência relativa de infecções, já que diferentes pacientes têm riscos diferentes.

Na realidade, a freqüência relativa apenas converge para o valor da probabilidade do evento. Esta idéia só pode ser colocada sob bases sólidas através do estudo da teoria axiomática da Probabilidade. Isto não será feito neste texto. Registrmos simplesmente que a freqüência relativa é diferente do valor da probabilidade, mas não faremos, de agora em diante, diferença entre os dois valores. Estaremos admitindo que o número de repetições é suficientemente grande para que a diferença seja desprezível.

#### 4.2.3 Tipos especiais de eventos

A ocorrência simultânea dos eventos  $A$  e  $B$  é chamada de evento interseção e representada por  $A \cap B$ . Contém todos os pontos do espaço amostral comuns a  $A$  e  $B$ , sendo ilustrada pela área sombreada na Figura 4.1.

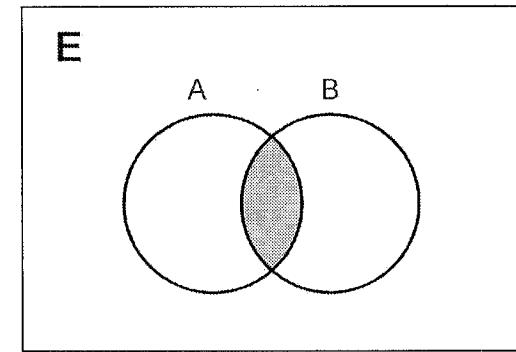


Figura 4.1: Diagrama ilustrando a interseção de dois eventos

O *evento união* de  $A$  e  $B$  equivale à ocorrência de  $A$  ou de  $B$ , ou de ambos. Contém os elementos do espaço amostral que estão em pelo menos um dos dois eventos. Denota-se por  $A \cup B$ , representada pela área sombreada da Figura 4.2.

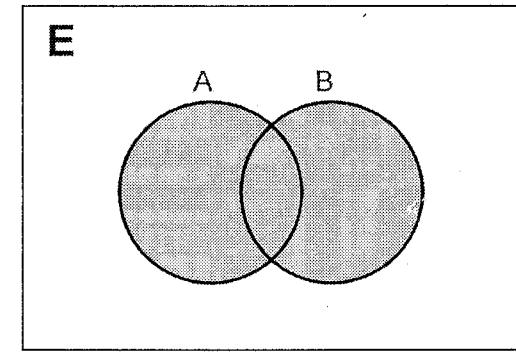


Figura 4.2: Diagrama ilustrando a união de dois eventos

O evento complementar de  $A$  contém todos os elementos do espaço amostral que não pertencem a  $A$ , ou seja, trata-se da negação do evento  $A$ . É usualmente indicado por  $\bar{A}$ .

Dois eventos  $A$  e  $B$  são *mutuamente excludentes* se a ocorrência de um deles impossibilita a ocorrência do outro. Em outras palavras, os dois eventos não têm nenhum elemento do espaço amostral em comum, isto é,  $A \cap B = \emptyset$ .

#### 4.2.4 Propriedades da probabilidade

1.  $0 \leq Pr(A) \leq 1$ , para qualquer evento  $A$
2.  $Pr(E) = 1$ , onde  $E$  é o espaço amostral
3. Se  $A$  e  $B$  são eventos mutuamente excludentes, então

$$Pr(A \cup B) = Pr(A) + Pr(B)$$

A partir das regras anteriores, pode-se mostrar que valem as seguintes relações:

1. Se  $A$  e  $B$  são dois eventos quaisquer, então:

$$Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$$

É fácil justificar o termo subtrativo considerando a Figura 4.2. Se tomassemos apenas  $Pr(A \cup B) = Pr(A) + Pr(B)$ , estariamos considerando duas vezes a probabilidade de interseção.

2. Para o evento complementar vale a seguinte relação

$$Pr(\bar{A}) = 1 - Pr(A).$$

Como a dificuldade de cálculo da probabilidade dos eventos  $A$  e  $\bar{A}$  é usualmente muito diferente, esta propriedade tem muitas aplicações na prática.

### 4.3 Probabilidade condicional

Em muitas situações o objetivo é calcular a probabilidade de um evento restrito a determinada condição. É a chamada *probabilidade condicional*. Por exemplo, na avaliação de um teste diagnóstico, interessa conhecer a probabilidade de o teste ser positivo, dado que o paciente é ou não doente. Portanto, podemos calcular probabilidades *condicionais*, representadas por:

$$Pr(T_+|D_+) \text{ e } Pr(T_+|D_-)$$

Para introduzir o conceito de probabilidade condicional, voltamos entretanto à motivação simples fornecida pelo lançamento de dois dados.

Suponhamos que se lancem dois dados não viciados. Existem 36 resultados possíveis, todos com probabilidade  $1/36$ . Se dispusermos, entretanto, da informação de que o resultado do primeiro dado é 3, qual será a probabilidade de a soma dos dois dados ser 8?

Para calcular esta probabilidade, raciocinamos da seguinte maneira: como o resultado do primeiro dado é 3, existem somente seis resultados possíveis em nosso experimento:

$$(3,1) \quad (3,2) \quad (3,3) \quad (3,4) \quad (3,5) \quad (3,6)$$

Note que agora o espaço amostral ficou reduzido a apenas 6 pontos. Intuitivamente, como cada um desses resultados era equiprovável antes de se conhecer o resultado do primeiro dado, o mesmo deve continuar acontecendo, após considerar-se aquela informação. Como apenas o par  $(3,5)$  satisfaz a condição desejada, a probabilidade procurada é  $1/6$ . Naturalmente a probabilidade condicional, dado que o primeiro dado produziu 3, dos outros 30 resultados é então zero, já que sua ocorrência é impossível.

Vemos que a probabilidade de um evento  $A$  modifica-se quando se dispõe de uma informação sobre a ocorrência de um outro evento associado. A probabilidade do evento  $A$ , quando se sabe que o evento  $B$  ocorreu, é chamada probabilidade condicional de  $A$  dado  $B$ ; denota-se por  $Pr(A|B)$  e se  $Pr(B) > 0$  é calculada por

$$Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)} \quad (4.1)$$

Ajudaria a entender a expressão (4.1) pensar que, quando escrevemos  $Pr(A|B)$ , o espaço amostral ficou restrito ao evento  $B$ . Reescrevendo (4.1), obtemos a expressão da regra da multiplicação.

$$Pr(A \cap B) = Pr(B)Pr(A|B) \quad (4.2)$$

#### Exemplo 4.5: Distribuição do peso e pressão arterial

Um grupo de pessoas foi classificado quanto a peso e pressão arterial de acordo com as proporções mostradas na Tabela 4.2.

Tabela 4.2: Distribuição de um conjunto de pacientes segundo peso e pressão arterial

Pressão arterial	Peso			Total
	Excesso	Normal	Deficiente	
Elevada	0,10	0,08	0,02	0,20
Normal	0,15	0,45	0,20	0,80
Total	0,25	0,53	0,22	1,00

Qual a probabilidade de uma pessoa escolhida ao acaso naquele grupo ter pressão elevada? Considerando que a pessoa escolhida tem excesso de peso, qual a probabilidade de ter também pressão elevada?

Chamando de  $A$  o evento *ter pressão elevada*, e lembrando que a pessoa é escolhida ao acaso em um grupo em que 20% tem pressão elevada, temos que  $Pr(A) = 0,20$ .

Chamemos  $B$  o evento *ter excesso de peso*. Diante da informação de que a pessoa tem excesso de peso, passam a não nos interessar as colunas correspondentes às pessoas de peso normal e deficiente. Na primeira coluna, vemos que, no grupo dos que têm excesso de peso, a proporção dos que têm pressão elevada é  $0,10/0,25 = 0,4$ . O

que estamos fazendo é precisamente estabelecer  $Pr(A|B)$  a partir de  $Pr(A \cap B) = 0,10$  e  $Pr(B) = 0,25$ . Assim,  $Pr(A|B) = 0,10/0,25 = 0,4$  é a probabilidade condicional de  $A$  dado  $B$ . Em palavras, entre as pessoas com excesso de peso, 40% têm pressão elevada.

#### 4.4 Eventos independentes

Dois eventos  $A$  e  $B$  são *independentes* se o fato de um deles ter ocorrido não altera a probabilidade de ocorrência do outro. Esta definição expressa em fórmula é

$$Pr(A|B) = Pr(A) \quad (4.3)$$

ou seja, a probabilidade condicional de  $A$  dado  $B$  é igual à probabilidade não condicional de  $A$ . Em outras palavras, a ocorrência do evento  $B$  não fornece nenhuma informação sobre a ocorrência do evento  $A$ .

#### Exemplo 4.6: Extrações de cartas de baralho

Para a extração sucessiva sem reposição de duas cartas de um baralho comum, vamos calcular a probabilidade de ambas serem “ás”.

Antes da primeira extração, há 4 ases em 52 cartas. Portanto, a probabilidade do evento  $A$ : “ás na primeira extração” é

$$Pr(A) = \frac{4}{52} = \frac{1}{13}$$

Já na segunda extração, o espaço amostral se reduziu porque a primeira carta extraída não foi repostada. Restam 51 cartas, e entre elas apenas 3 ases. Chamando  $B$  o evento “ás na segunda extração”

$$Pr(B|A) = \frac{3}{51}$$

e então

$$Pr(A \cap B) = Pr(A)Pr(B|A) = \frac{4}{52} \times \frac{3}{51} = \frac{1}{221}$$

Os eventos não são independentes. A probabilidade do segundo evento depende do resultado do primeiro. Se considerarmos a mesma situação, porém com reposição da primeira carta, o espaço amostral se manterá inalterado após a primeira extração, como também a probabilidade de extração de um ás

$$Pr(B|A) = Pr(B) = \frac{1}{13}$$

ou seja os eventos são independentes e neste caso

$$Pr(A \cap B) = \frac{1}{52} \times \frac{1}{13} = \frac{1}{676}$$

#### Exemplo 4.7: Problema histórico de Probabilidade

Este é o primeiro problema que De Meré propôs ao matemático Blaise Pascal, que ao resolvê-lo iniciou o desenvolvimento da Teoria das Probabilidades. Qual é o resultado mais provável: obter a face 6 pelo menos uma vez em quatro jogadas de um dado, ou obter a soma 12 pelo menos uma vez em 24 jogadas de dois dados? Como as jogadas do dado são independentes, temos:

$$\begin{aligned} Pr\left(\begin{array}{l} \text{ao menos um 6} \\ \text{em 4 jogadas} \end{array}\right) &= 1 - Pr\left(\begin{array}{l} \text{nenhum 6} \\ \text{em 4 jogadas} \end{array}\right) \\ &= 1 - \left(\frac{5}{6}\right)^4 = 0,518 \end{aligned}$$

Na jogada de dois dados temos:

$$\begin{aligned} Pr\left(\begin{array}{l} \text{ao menos uma soma} \\ 12 \text{ em 24 jogadas} \end{array}\right) &= 1 - Pr\left(\begin{array}{l} \text{nenhuma soma} \\ 12 \text{ em 24 jogadas} \end{array}\right) \\ &= 1 - \left(\frac{35}{36}\right)^{24} = 0,491 \end{aligned}$$

Assim, a primeira situação apresenta maior probabilidade.

#### Exemplo 4.8: Sexo dos filhos

Um casal possui 2 filhos. Qual a probabilidade de ambos serem do sexo masculino?

Os eventos  $M$ : “nascer uma criança do sexo masculino” e  $F$ : “nascer uma criança do sexo feminino” são equiprováveis. Logo a probabilidade de nascer um filho do sexo masculino é  $1/2$ . A ocorrência do evento  $A$ : “o primeiro filho é do sexo masculino” não influencia a ocorrência do evento  $B$ : “o segundo filho é do sexo masculino”, e então:

$$P(A \cap B) = P(A)P(B) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

#### 4.5 Qualidade de testes diagnósticos

O bom uso de um teste diagnóstico requer, além de considerações clínicas, o conhecimento de medidas que caracterizam a sua qualidade intrínseca: a sensibilidade, a especificidade e os parâmetros que refletem a sua capacidade de produzir decisões clínicas corretas: valor da predição positiva e o valor da predição negativa.

Usando os resultados da Teoria de Probabilidades apresentados nas seções anteriores, vamos nos ocupar desses quatro índices nos quais é usualmente decomposta a idéia de qualidade de um teste diagnóstico.

##### 4.5.1 Sensibilidade e especificidade

Na análise da qualidade de testes diagnósticos, interessa conhecer duas probabilidades condicionais que, pela sua importância, recebem nomes especiais: sensibilidade e especificidade.

A *sensibilidade*, denotada por  $s$ , é definida como

$$s = Pr[T_+ | D_+] \quad (4.4)$$

ou seja, a probabilidade de o teste ser positivo sabendo-se que o paciente que está sendo examinado é doente.

A *especificidade*, denotada por  $e$ , é definida como

$$e = \Pr[T_- | D_-] \quad (4.5)$$

ou seja, a probabilidade de o teste ser negativo sabendo-se que o paciente examinado não é portador da doença.

Os nomes são descritivos: sensibilidade mede a capacidade de reação do teste em um paciente doente enquanto que especificidade, a não reação do teste em pacientes não portadores da doença, isto é, o teste é específico para a doença em questão.

A análise da definição desses dois índices ( $s$  e  $e$ ) mostra que, subjacentemente a estes conceitos, estamos assumindo a existência de um *padrão ouro*, ou seja, um teste diagnóstico que sempre produz resultados corretos. Além disso, assumimos que os pacientes são classificados apenas como doentes e não doentes, não se admitindo estágios intermediários.

Usando a notação da Tabela 4.1 e a definição de probabilidade condicional, os índices  $s$  e  $e$  são estimados por:

$$s = \frac{a}{a + b} \quad (4.6)$$

$$e = \frac{d}{c + d} \quad (4.7)$$

#### Exemplo 4.9: Diagnóstico de doença coronariana

Wiener et al. (1979) compararam os resultados do teste ergométrico de tolerância a exercícios entre indivíduos com e sem doença coronariana. O teste foi considerado positivo quando se observou mais de 1 mm de depressão ou elevação do segmento ST, por no mínimo 0,08 s, em comparação com os resultados obtidos com o paciente em repouso. O diagnóstico definitivo foi feito através de angiografia. A Tabela 4.3 sintetiza os resultados encontrados.

Tabela 4.3: Resultados da aplicação do teste ergométrico de tolerância a exercícios em 1.465 pessoas

Doença coronariana	Teste ergométrico		Total
	Positivo ( $T_+$ )	Negativo ( $T_-$ )	
Presente ( $D_+$ )	815	208	1023
Ausente ( $D_-$ )	115	327	442
Total	930	535	1465

A sensibilidade e a especificidade são estimadas por:

$$s = \frac{815}{1023} = 0,797$$

$$e = \frac{327}{442} = 0,740$$

O teste ergométrico tem uma sensibilidade de 79,7%, ligeiramente superior que sua especificidade (74%).

#### Exemplo 4.10: Metástase de carcinoma hepático

Lind & Singer (1986) estudaram a qualidade da tomografia computadorizada para o diagnóstico de metástase de carcinoma de fígado, obtendo os resultados sintetizados na Tabela 4.4. Um total de 150 pacientes foram submetidos a dois exames: a tomografia computadorizada e a laparotomia. Este último é tomado como o padrão ouro, isto é, classifica o paciente sem erro.

Tabela 4.4: Resultados da tomografia computadorizada em 67 pacientes com metástase e 83 sem metástase do carcinoma hepático

Metástase de carcinoma hepático	Tomografia computadorizada		Total
	Positiva ( $T_+$ )	Negativa ( $T_-$ )	
Presente ( $D_+$ )	52	15	67
Ausente ( $D_-$ )	9	74	83
Total	61	89	150

A sensibilidade e a especificidade da tomografia computadorizada são estimados por:

$$s = \frac{52}{67} = 0,776$$

$$c = \frac{74}{83} = 0,892$$

Diferentemente do exemplo anterior, a especificidade (89,2%) é maior do que a sensibilidade (77,6%).

#### 4.5.2 Valor das predições

A sensibilidade e a especificidade, embora sendo índices ilustrativos e bons sintetizadores das qualidades gerais de um teste, têm uma limitação séria: não ajudam a decisão da equipe médica que, recebendo um paciente com resultado positivo do teste, precisa avaliar se o paciente está ou não doente. Não se pode depender apenas da sensibilidade e da especificidade, pois estes índices são provenientes de uma situação em que há certeza total sobre o diagnóstico, o que não acontece no consultório médico. Daí a necessidade destes dois outros índices que refletem melhor a realidade prática. Neste momento, interessa mais conhecer os seguintes índices denominados valor da predição positiva (*VPP*) e valor da predição negativa (*VPN*), definidos respectivamente por:

$$VPP = Pr[D_+|T_+] \quad (4.8)$$

$$VPN = Pr[D_-|T_-] \quad (4.9)$$

Em palavras, *VPP* é a probabilidade do paciente estar realmente doente quando o resultado do teste é positivo e *VPN*, a probabilidade do paciente não estar doente quando o resultado do teste é negativo. Estes valores são probabilidades condicionais, tal que o evento condicionante é o resultado do teste, aquele que na prática acontece primeiro.

A maneira mais fácil de se calcular o *VPN* e o *VPP* é através da Tabela 4.5, sugerida por Vecchio (1966). Seja  $p = Pr(D_+)$  a

Tabela 4.5: Probabilidades necessárias para o cálculo dos índices *VPP* e *VPN*

População	Proporção	Proporção com resultado	
		Positivo	Negativo
Docente	$p$	$ps$	$p(1-s)$
Sadia	$1-p$	$(1-p)(1-c)$	$(1-p)c$
Total	$p + (1-p)$	$ps + (1-p)(1-c)$	$p(1-s) + (1-p)c$

prevalência da doença na população de interesse, isto é, a proporção de pessoas doentes, ou a probabilidade de doença pré-teste.

O valor de predição positiva é obtido dividindo-se a freqüência dos “verdadeiros” positivos, aqueles oriundos de pacientes doentes, pelo total de positivos. Obtém-se a seguinte expressão

$$VPP = \frac{ps}{ps + (1-p)(1-c)} \quad (4.10)$$

De forma análoga, considerando-se os “verdadeiros” negativos obtemos o valor da predição negativa

$$VPN = \frac{(1-p)c}{p(1-s) + (1-p)c} \quad (4.11)$$

Ambas as expressões dependem do conhecimento de  $p$ , uma estimativa da prevalência da doença na população de interesse. Estas são probabilidades de resultados corretos de diagnóstico.

#### Exemplo 4.11: Metástase de carcinoma hepático (continuação)

Para uma população cuja prevalência de metástase de carcinoma de fígado é de 2%, os valores de predição da tomografia computadorizada são:

$$VPP = \frac{0,02 \times 0,78}{0,02 \times 0,78 + (1-0,02)(1-0,89)} = 0,13$$

$$VPN = \frac{(1 - 0,02) \times 0,89}{(1 - 0,02)0,89 + 0,02(1 - 0,78)} = 0,99$$

Portanto, o valor de predição positiva é baixo enquanto que o valor de predição negativa é bastante alto. Se o resultado da tomografia computadorizada é negativo, a chance de não haver metástase é de 99%. Se, antes de qualquer informação, o paciente tinha uma chance de 2% de apresentar a doença, após o resultado do teste negativo esta chance é de apenas 1%.

#### 4.5.3 Decisões incorretas

As probabilidades

$$PFP = Pr[D_-|T_+] = 1 - Pr[D_+|T_+] = 1 - VPP \quad (4.12)$$

$$PFN = Pr[D_+|T_-] = 1 - Pr[D_-|T_-] = 1 - VPN \quad (4.13)$$

referem-se, respectivamente, ao *falso-positivo* e *falso-negativo*, isto é, decisões incorretas baséadas no teste diagnóstico.

Infelizmente, não há na literatura padronização relativa a nomes dos índices de um teste diagnóstico. Por exemplo, as probabilidades de *falso-positivo* e *falso-negativo* são muito freqüentemente usadas para as quantidades  $1 - s$  e  $1 - e$ , quando deveriam ser reservados para  $1 - VPP$  e  $1 - VPN$ . Por isto, na medida do possível estes termos são evitados neste texto.

Outra dificuldade é que muitos autores admitem, implicitamente, que a prevalência que ocorre na tabela é a mesma na população de interesse e assim usam a tabela  $2 \times 2$  para calcular os valores de predição. Nada justifica esta hipótese e este procedimento, se usado sistematicamente, leva a erros sérios.

#### Exemplo 4.12: Teste ELISA para detecção do HIV

Entre as várias tecnologias para detectar a presença do HIV, a primeira a se difundir no Brasil foi o teste de ELISA (*Enzyme-linked immunosorbent assay*). Em 1985 esta técnica foi simultaneamente comercializada por vários laboratórios americanos: ABBOTT, DUPONT, ELECTRO-NUCLEONICS, LITTON, TRAVENOL. O Laboratório ABBOTT reportou, em seus testes preliminares, sensibilidade de 95% e especificidade de 99,8%. Os valores para os outros laboratórios são parecidos, segundo Marwick (1985).

A Tabela 4.6 apresenta os valores dos índices  $VPP$  e  $VPN$  para a implementação ABBOTT do teste ELISA e vários valores da prevalência.

Tabela 4.6: Valores da predição ( $VPP$  e  $VPN$ ) e proporção de falsos resultados ( $PFP$  e  $PFN$ ) do teste ELISA para detecção do HIV, versão ABBOTT

Prevalência	$VPP$ (%)	$VPN$ (%)	$PFP$ (%)	$PFN$ (%)
1/100.000	0,47	100,00	99,53	0,00
1/10.000	4,54	100,00	95,46	0,00
1/1.000	32,21	99,99	67,79	0,01
1/500	48,77	99,99	51,23	0,01
1/200	70,47	99,99	29,53	0,01
1/100	82,75	99,99	17,25	0,01
1/50	90,65	99,89	9,35	0,11

Considerando-se a população total de um país, a AIDS é uma doença de prevalência pequena. Os resultados da Tabela 4.6 mostram que em um programa de uso do teste em larga escala, grande parte dos pacientes com resultado positivo consiste na realidade de falsos-positivo; em outras palavras, o valor da predição positiva é muito pequeno. Por outro lado, pouquíssimos não doentes deixarão de ser detectados e, portanto, o valor de predição negativa é alto. Isto sugere um cuidado básico: um resultado positivo deve ser reconfirmado através de teste baseado em tecnologia diferente do ELISA.

## 4.6 Combinação de testes diagnósticos

Muitas vezes, para o diagnóstico de certa doença, dispomos apenas de testes com  $VPP$  ou  $VPN$  baixo ou, se existe um bom teste, este é muito caro ou oferece grande risco e/ou desconforto ao paciente. Nestas circunstâncias, uma opção freqüentemente usada é o uso de uma combinação de testes mais simples. A associação de testes eleva a qualidade do diagnóstico, diminuindo o número de resultados incorretos.

Quando dois ou mais testes são usados para se chegar a um diagnóstico é preciso saber como são obtidos os índices de qualidade do teste múltiplo, aquele composto pela agregação de dois ou mais testes individuais. Restringiremos ao caso de apenas dois testes e as idéias apresentadas a seguir podem ser estendidas para o caso de mais testes. Alguns detalhes podem ser encontrados em Hirsch & Riegelman (1996).

### 4.6.1 Formas de combinação de testes

As maneiras mais simples de se formar um teste múltiplo a partir dos resultados de dois testes são os esquemas em paralelo e em série. No caso do teste em paralelo, se um dos dois testes é positivo o teste conjunto também o é. O teste em série é considerado positivo apenas se os dois testes individuais são positivos. A associação de testes em série é bastante empregada, tanto em triagens como no diagnóstico individual, sendo de grande utilidade quando a questão do custo é relevante.

Chamando os testes originais de  $A$  e  $B$ , o teste em paralelo de  $T_p$  e o em série de  $T_s$ , e usando a linguagem de eventos da Seção 4.2.3 temos que:

$$T_{p+} = A_+ \cup B_+ \quad (4.14)$$

$$T_{s+} = A_+ \cap B_+ \quad (4.15)$$

As sensibilidades e especificidades de  $T_p$  e  $T_s$  são calculadas com o auxílio das regras de cálculo de probabilidades de eventos explicadas anteriormente.

### Combinação em paralelo

Nesse caso, o resultado do teste será considerado positivo, se pelo menos um dos testes apresentar resultado positivo. É de maior utilidade em casos de urgência, quando se necessita de uma abordagem rápida, ou por outro tipo de conveniência, como para pacientes provenientes de lugares distantes. A Tabela 4.7 apresenta de forma explícita o procedimento proposto.

Tabela 4.7: Resultado do teste em paralelo dependendo da classificação dos testes individuais A e B

Teste A	Teste B	Teste em paralelo
—	—	—
—	+	+
+	—	+
+	+	+

Em analogia à expressão (4.4) e aplicando a propriedade da probabilidade de união de eventos, a sensibilidade do teste em paralelo é calculada como:

$$\begin{aligned} Pr[T_{p+}|D_+] &= Pr[A_+ \cup B_+|D_+] \\ &= Pr[A_+|D_+] + Pr[B_+|D_+] - Pr[A_+ \cap B_+|D_+] \end{aligned}$$

ou seja, a sensibilidade do teste em paralelo ( $s_P$ ) é dada por

$$s_P = s_A + s_B - s_A \times s_B \quad (4.16)$$

Pode-se calcular a especificidade de um teste em paralelo usando-se a definição (4.5). Admitindo-se que os resultados dos dois testes são independentes, pode-se escrever

$$\begin{aligned} Pr[T_{p-}|D_-] &= Pr[A_- \cap B_-|D_-] \\ &= Pr[A_-|D_-] \times Pr[B_-|D_-] \end{aligned}$$

Assim, a especificidade do teste em paralelo ( $e_P$ ) é dada por

$$e_P = e_A \times e_B \quad (4.17)$$

Portanto, facilmente calculamos a sensibilidade e a especificidade de um teste em paralelo a partir das sensibilidades e especificidades dos testes  $A$  e  $B$ . Além disso, os parâmetros  $VPP$  e  $VPN$  são calculados através de (4.10) e (4.11) para testes isolados, utilizando-se agora a sensibilidade e especificidade da combinação em paralelo, obtidas através de (4.16) e (4.17), e a prevalência da população de interesse.

### Combinação em série

Nesse caso, os testes são aplicados consecutivamente, sendo o segundo teste aplicado apenas se o primeiro apresentar resultado positivo. O teste só será considerado positivo, se o resultado dos dois testes for positivo. Esse procedimento é indicado em situações em que não há necessidade de rápido atendimento e quando o paciente pode ser acompanhado ao longo do tempo, e se a consideração de custo é importante, seja pela questão financeira, pelo risco ou desconforto induzidos pelo exame. A Tabela 4.8 apresenta de forma explícita o procedimento proposto.

Tabela 4.8: Resultado do teste em série dependendo da classificação dos testes individuais A e B

Teste A	Teste B	Teste em série
-	desnecessário	-
+	-	-
+	+	+

Como só serão aplicados dois testes se o primeiro for positivo, o custo desse tipo de combinação é menor. Uma discussão sobre a ordem mais indicada para a aplicação dos testes a serem combinados em série pode ser encontrada em Soares & Parenti (1995).

Se os dois testes  $A$  e  $B$  são independentes, a sensibilidade ( $s_S$ ) e a especificidade ( $e_S$ ) para o teste combinado em série são obtidas

utilizando-se regras de cálculo para a probabilidade do evento interseção. Por exemplo, para o cálculo da sensibilidade temos que

$$Pr[T_{s+}|D_+] = Pr[A_+ \cap B_+|D_+] = Pr[A_+|D_+] \times Pr[B_+|D_+]$$

Portanto,

$$s_S = s_A \times s_B \quad (4.18)$$

A partir de raciocínio análogo, obtemos a expressão para a especificidade:

$$e_S = e_A + e_B - e_A \times e_B \quad (4.19)$$

Os valores de  $VPP$  e  $VPN$  são também obtidos usando-se as expressões (4.10) e (4.11) substituindo  $s$  e  $c$  calculados através de (4.18) e (4.19).

Para os cálculos da sensibilidade e especificidade da associação em série e em paralelo, a independência dos dois testes é crucial. Entretanto, não se pode garantir que isto ocorra sempre. Quando os testes não forem independentes, não há uma forma analítica simples para se obter tais índices para um teste composto.

### Exemplo 4.13: Diagnóstico de câncer pancreático

Imagine um paciente idoso com dores persistentes nas costas e no abdômen e perda de peso. Na ausência de uma explicação para estes sintomas, a possibilidade de câncer do pâncreas é freqüentemente levantada. É comum para se verificar esta possibilidade diagnóstica, que ambos os testes de ultra-som e tomografia computadorizada do pâncreas sejam solicitados. A Tabela 4.9 apresenta dados hipotéticos sobre os índices  $s$  e  $e$  dos testes, quando utilizados separadamente e em conjunto (Griner et al., 1981).

Note que os esquemas C e D correspondem respectivamente a testes em paralelo e em série. Admitindo que os resultados dos

Tabela 4.9: Sensibilidade e especificidade dos testes de ultra-som e tomografia computadorizada no diagnóstico do câncer de pâncreas, individualmente e em conjunto

Teste	Sensibilidade (%)	Especificidade (%)
A: Ultrasom	80	60
B: Tomografia	90	90
C: A ou B positivo	98	54
D: A e B positivo	72	96

dois testes sejam independentes e usando as expressões (4.16), (4.17), (4.18) e (4.19) temos as seguintes sensibilidades e especificidades:

$$s_C = 0,8 + 0,9 - 0,8(0,9) = 0,98$$

$$s_D = 0,8(0,9) = 0,72$$

$$e_C = 0,6(0,9) = 0,54$$

$$e_D = 0,6 + 0,9 - 0,6(0,9) = 0,96$$

Quando um ou outro teste é positivo, a sensibilidade combinada é maior que o mais sensível dos testes, mas a especificidade é menor. Ao contrário, quando o critério para a positividade do teste é que tanto o ultra-som como a tomografia sejam positivos, a especificidade combinada é maior que o mais específico dos dois, mas a sensibilidade é menor.

#### Exemplo 4.14: Sensibilidade e especificidade de testes em paralelo e em série

Consideremos dois testes A e B com sensibilidade e especificidade apresentados na Tabela 4.10 e suponhamos uma prevalência de 1%. Usando as expressões (4.16) a (4.19) e a seguir as expressões (4.10) e (4.11), preenchemos as outras entradas da Tabela 4.10.

Tabela 4.10: Sensibilidade, especificidade e valores de predição de testes individuais A e B e dos testes em série e em paralelo considerando-se uma prevalência de 1%

Teste	<i>s</i>	<i>e</i>	<i>VPP</i>	<i>VPN</i>
A	0,9500	0,9000	0,0876	0,9994
B	0,8000	0,9500	0,1391	0,9979
Paralelo	0,9900	0,8550	0,0645	0,9999
Série	0,7600	0,9950	0,6056	0,9976

A combinação em paralelo apresenta alta sensibilidade (0,99) mas sua especificidade é menor que a dos testes isolados. Já a combinação em série apresenta alta especificidade (0,995) enquanto que a sensibilidade é relativamente baixa comparada com testes isolados.

Como esperado, os valores de predição negativa são altos, tanto para os testes isolados como para as duas formas combinadas dos testes. Entretanto, os valores de predição positiva não são altos, sendo que a combinação em série proporcionou o melhor resultado (0,6056).

É importante lembrar que, na maioria das vezes, os testes usados na combinação são dependentes entre si. Portanto, os valores acima tendem a superestimar o verdadeiro valor dos índices dos testes combinados. Nesse caso, não é possível determinar os valores dos parâmetros de qualidade conhecendo-se apenas os valores de cada teste em separado. Torna-se necessário um trabalho de pesquisa realizado em um grupo de pacientes nos quais são utilizados o teste padrão ouro e teste combinado (DiMagno et al., 1977).

#### 4.6.2 Necessidade da combinação de testes

Há pelo menos duas situações em que a necessidade de combinação de testes surge naturalmente: triagem e diagnóstico individual.

##### Triagem

Triagem é um tipo de procedimento que visa classificar pessoas assintomáticas quanto à probabilidade de terem ou não a doença. É

aplicado em grande número de pessoas de uma população. Dessa forma, a triagem não faz por si só o diagnóstico, mas aponta as pessoas que tem maior probabilidade de estarem doentes. Para essas, é indicado um teste diagnóstico para comprovar ou não a presença da doença.

A triagem é indicada em caso de doença séria, se o tratamento na fase assintomática é mais benéfico do que na fase sintomática e em casos de alta prevalência. Na linguagem desta seção, o teste de triagem é considerado o primeiro dos testes de uma combinação em série, já que o segundo teste (no qual o diagnóstico definitivo será baseado) só será aplicado se o resultado da triagem for positivo.

### Diagnóstico individual

No diagnóstico individual, aparecem os dois tipos de combinação: em série e em paralelo. A combinação em paralelo é usada, como já foi citado, em casos de urgência ou para pacientes residentes em lugares distantes.

Como exemplo da necessidade de se usar testes em paralelo, podemos pensar em um paciente que chega ao hospital com suspeita de infarto. Como se trata de um caso de urgência, o clínico poderá indicar vários exames, tais como, eletrocardiograma, dosagem de LDH e dosagem de CPK-MB, que serão realizados imediatamente e ao mesmo tempo. Ao final dos testes, o clínico analisa os resultados e aponta o diagnóstico. Muitas vezes, é estabelecida uma norma para orientar a decisão do clínico. Por exemplo, se forem realizados 4 testes, o diagnóstico seria de infarto se 3 ou 4 resultados fossem positivos (Galen & Gambino, 1975).

A combinação em série é usada em consultórios e clínicas hospitalares e em casos de testes caros e arriscados. Nesse último caso, o teste em série se torna de grande validade. O paciente realiza primeiramente um teste simples e barato e, se apresentar resultado positivo, é indicado o teste de maior custo ou que oferece riscos ao paciente. Com esse procedimento, menor número de pessoas saudáveis deverá se submeter a esses testes.

## 4.7 Testes diagnósticos baseado em variáveis contínuas

Os índices usados neste capítulo para definir qualidade de um teste clínico requerem que o resultado do teste seja, por simplicidade, classificado como positivo ou negativo e que a classificação do diagnóstico também seja dicotómica (por exemplo doente ou sadio).

Para dados de variáveis contínuas, tais como dosagens, surge a questão de como dicotomizar de forma que os dados sejam colocados no formato da tabela 2 x 2 considerada anteriormente. É preciso então estabelecer o limite entre o que será considerado sadio ou doente, isto é, um valor a partir do qual o teste é considerado positivo ou negativo. Este limite é chamado de *valor de referência* ou *ponto de corte*. Assim, cada teste clínico ou critério de classificação é caracterizado por um valor de referência e variando-se este ponto de corte, a sensibilidade e a especificidade também variam. Para cada situação é preciso escolher o valor que fornece a combinação de sensibilidade e especificidade mais adequada.

### 4.7.1 Efeito do ponto de corte na qualidade de um teste diagnóstico

A cada valor crítico está associada uma sensibilidade e uma especificidade. Freqüentemente pode-se alterar a sensibilidade e a especificidade de um teste trocando-se o ponto de corte, como ilustra o exemplo a seguir, adaptado de Ingelfinger et al. (1987).

#### Exemplo 4.15: Qualidade do teste de intoxicação digitálica

Os digitálicos (digoxina e digitoxina) podem ser de grande benefício para pacientes com falha congestiva do coração porque produzem redução de batimento cardíaco e aumento da força de contração do miocárdio. A terapia com essas drogas pode, entretanto, resultar em uma complicação séria, a intoxicação, que ocorre em 5% a 30% de pacientes hospitalizados que fazem uso de digitálicos.

O diagnóstico da intoxicação pode ser difícil porque muitos dos sintomas de intoxicação, como por exemplo, contração ventricular

prematura, náusea e fadiga, são também sintomas de falha congestiva cardíaca. A concentração sérica de digitálicos fornece um diagnóstico útil para a toxicidade.

Um nível de digoxina de  $1,7 \text{ ng/ml}$  e um nível de digitoxina de  $25 \text{ ng/ml}$  são essencialmente equivalentes e são considerados como concentrações altas. O resultado do teste de toxicidade é considerado positivo para concentrações acima desses níveis.

Beller et al. (1971) examinaram a relação entre concentrações séricas de digitálicos e sua toxicidade. Das 931 admissões consecutivas do hospital *Boston City*, foram investigados todos os 135 pacientes que estavam fazendo uso de digoxina ou digitoxina. Na admissão, a concentração dessas drogas foi anotada e eletrocardiogramas forneceram informações do acompanhamento de cada paciente.

Através de critérios clínicos bastante apurados, os pacientes foram classificados como intoxicados ou não, como mostra a Tabela 4.11.

Tabela 4.11: Resultados do teste de toxicidade de digitálicos aplicado a 135 pacientes que faziam uso de digoxina ou digitoxina

Classificação de toxicidade baseada em critérios clínicos	Teste de toxicidade		
	positivo	negativo	Total
Intoxicado	25	18	43
Não intoxicado	14	78	92
Total	39	96	135

A sensibilidade e a especificidade do teste que usa o nível sérico de digoxina ou digitoxina para classificar o indivíduo como intoxicado ou não são:

$$s = \frac{25}{43} = 0,58$$

$$c = \frac{78}{92} = 0,85$$

Estes cálculos mostram que esse teste tem uma boa especificidade e uma sensibilidade apenas razoável.

Um estudo análogo ao anterior porém com outro grupo de pacientes foi conduzido. A Tabela 4.12 representa a distribuição de

21 pacientes intoxicados e 62 não intoxicados, sendo que ambos os grupos de pacientes faziam uso de digoxina.

Tabela 4.12: Distribuição da concentração sérica de digoxina ( $\text{ng/ml}$ ) para pacientes intoxicados ou não pelo uso da digoxina

Digoxina ( $\text{ng/ml}$ )	Pacientes	
	Intoxicados	Não intoxicados
0 $\vdash$ 0,5	2	6
0,5 $\vdash$ 1,0	1	18
1,0 $\vdash$ 1,5	4	30
1,5 $\vdash$ 2,0	4	8
2,0 $\vdash$ 2,5	5	0
2,5 $\vdash$ 3,0	1	0
3,0 $\vdash$ 3,5	1	0
3,5 $\vdash$ 4,0	1	0
4,0 $\vdash$ 4,5	1	0
4,5 $\vdash$ 5,0	1	0
Total	21	62

Suponha agora que o critério de classificação quanto à toxicidade fosse trocado, isto é, o ponto de corte do teste para o nível de digoxina fosse alterado. Considere os pontos de corte para classificação em relação aos níveis de digoxina para diagnóstico de intoxicação:  $1 \text{ ng/ml}$  (critério 1) e  $1,5 \text{ ng/ml}$  (critério 2).

Os resultados para estes dois critérios de classificação nos 83 pacientes em uso de digoxina estão organizados nas tabelas 4.13, 4.14.

Tabela 4.13: Teste de toxicidade digitálica usando o critério 1

Classificação de toxicidade baseada em critérios clínicos	Teste de toxicidade		
	positivo	negativo	Total
Intoxicado	18	3	21
Não intoxicado	38	24	62
Total	56	27	83

Tabela 4.14: Teste de toxicidade da digitálica usando o critério 2

Classificação de toxicidade baseada em critérios clínicos	Teste de toxicidade		
	positivo	negativo	Total
Intoxicado	14	7	21
Não intoxicado	8	54	62
Total	22	61	83

A Tabela 4.15 mostra a sensibilidade e especificidade de cada critério. Note que o critério 2 produziu perda de sensibilidade e ganho de especificidade.

Tabela 4.15: Sensibilidade e especificidade para os critérios 1 e 2

Critério	Sensibilidade ( $s$ )	Especificidade ( $e$ )
1	$\frac{18}{21} = 0,86$	$\frac{24}{62} = 0,39$
2	$\frac{14}{21} = 0,67$	$\frac{54}{62} = 0,87$

#### 4.7.2 Escolha do ponto de corte

Uma possibilidade para a escolha do ponto de corte seria calcular a sensibilidade e a especificidade para vários valores de referência e então adotar aquele que produz uma combinação mais desejável para essas duas medidas da qualidade de um teste. Por exemplo, Galen & Gambino (1975) sugeriram uma medida de eficiência definida pela soma da sensibilidade e especificidade.

Um procedimento alternativo é a análise das curvas de operação característica (em inglês ROC, a abreviação para *receiver operator characteristic*), nome recebido porque o método originou-se em estudos de detecção de sinais por operadores de radar. Esta é uma representação gráfica da sensibilidade (eixo vertical) e o complemento da especificidade (eixo horizontal) para diversos pontos de cortes. Detalhes sobre a construção e interpretação da curva como também a ilustração com exemplos reais podem ser encontrados por exemplo

em Beck & Shultz (1986), Fletcher et al. (1989) e no Capítulo 1 de Hirsch & Riegelman (1996).

## 4.8 Escolha entre testes diagnósticos

Idealmente, os testes utilizados devem ter alta sensibilidade e especificidade. Entretanto, na prática nem sempre existem testes disponíveis com características ótimas. É relativamente comum a situação em que há mais de um teste e então surge a necessidade da comparação entre eles para uma escolha mais adequada.

Suponha que dois testes diagnósticos estão disponíveis: um com alta sensibilidade mas relativamente baixa especificidade e o outro com alta especificidade e relativamente baixa sensibilidade. Qual seria a melhor escolha?

Embora a avaliação de um teste não seja trivial, já que vários fatores devem ser considerados, apresentaremos alguns argumentos baseados nas medidas de qualidade definidas neste capítulo, que podem ajudar na escolha de um teste diagnóstico.

O primeiro fato importante é que, como visto na Seção 4.5.2, os valores de predição ( $VPP$  e  $VPN$ ) dependem conjuntamente da sensibilidade ( $s$ ), da especificidade ( $e$ ) e da prevalência ( $p$ ). Soares & Parenti (1995) apresentam um estudo detalhado através de gráficos que ilustram claramente o efeito da prevalência nos valores de predição. O exemplo a seguir ilustra numericamente este fato.

#### Exemplo 4.16: Diagnóstico da gonorréia

O teste Gonosticon Dri-Drot, desenvolvido para o diagnóstico de gonorréia, tem sensibilidade de 0,80 e especificidade de 0,95. Os valores da predição positiva e negativa são apresentados na Tabela 4.16 para quatro populações com diferentes prevalências.

Embora o teste tenha parâmetros de qualidade razoáveis, o  $VPP$  é baixíssimo para populações de baixa prevalência, tornando o exame sem utilidade nessas condições.

Para simplificar o raciocínio sobre a escolha de testes diagnósticos, vamos inicialmente considerar duas situações extremas, obviamente sem interesse prático, e depois vamos estender para situações

Tabela 4.16: Valores de predição ( $VPP$  e  $VPN$ ) do teste Gonosticon Dri Drot para quatro prevalências ( $p$ )

População	$p$	$VPP$	$VPN$
A	0,500	0,940	0,830
B	0,100	0,640	0,980
C	0,020	0,250	0,990
D	0,001	0,020	1,000

mais gerais. Suponhamos que  $p = 0$ , isto é, ninguém tem a doença, ou que  $p = 1$ , isto é, toda a população tem a doença em questão. Intuitivamente ou através das expressões (4.10), (4.11), (4.12) e (4.13), facilmente chegamos nos valores apresentados na Tabela 4.17, para quaisquer valores de sensibilidade e especificidade.

Tabela 4.17: Valores de predição ( $VPP$  e  $VPN$ ) e proporção de resultados falsos ( $PFP$  e  $PFN$ ) para casos extremos de prevalência

$p$	$VPP$	$PFP$	$VPN$	$PFN$
0	0	1	1	0
1	1	0	0	1

A extensão para o caso mais geral de  $p$  pequeno (doença rara) e  $p$  grande (doença comum) pode ser feita pela análise das expressões (4.10) e (4.11) variando-se as quantidades envolvidas (sensibilidade, especificidade e prevalência), mas nos restringimos a apresentarmos uma análise numérica. A Tabela 4.18 mostra os valores de predição para prevalências de 1% e 90%.

Quando a prevalência é baixa, o valor de predição positiva ( $VPP$ ) é mais influenciado pela especificidade. O valor de predição negativa ( $VPN$ ) é pouco influenciado tanto pela sensibilidade quanto pela especificidade e é alto, como era de se esperar.

Para a prevalência alta, o valor de predição positivo ( $VPP$ ) é próximo de 1, independente dos valores da sensibilidade e da especificidade. Além disso, o valor de predição negativo ( $VPN$ ) é influen-

Tabela 4.18: Valores de predição ( $VPP$  e  $VPN$ ) para alguns valores de sensibilidade e especificidade para prevalências de 1% e 90%

$s$	$e$	Prevalência de 1%		Prevalência de 90%	
		$VPP$	$VPN$	$VPP$	$VPN$
0,99	0,99	0,5000	0,9999	0,9989	0,9167
0,99	0,90	0,0909	0,9999	0,9889	0,9091
0,99	0,80	0,0476	0,9999	0,9780	0,8989
0,90	0,99	0,4762	0,9990	0,9988	0,5238
0,90	0,90	0,0833	0,9989	0,9878	0,5000
0,90	0,80	0,0435	0,9987	0,9759	0,4706
0,80	0,99	0,4469	0,9980	0,9986	0,3548
0,80	0,90	0,0748	0,9978	0,9863	0,3333
0,80	0,80	0,0388	0,9975	0,9730	0,3077

ciado mais pela sensibilidade do que pela especificidade.

Embora este não seja um estudo exaustivo, existem evidências para as seguintes conclusões:

1. Um teste com alta especificidade deve ser usado quando a prevalência da doença é relativamente baixa (*doença rara*), mesmo que o teste tenha relativamente baixa sensibilidade.
2. Um teste com alta sensibilidade deve ser usado quando a prevalência da doença é alta (*doença comum*), mesmo que o teste tenha relativamente baixa especificidade.

É comum a idéia de que se a doença é rara, um teste com alta sensibilidade deve ser usado para achar os casos e para uma doença de alta prevalência, um teste com alta especificidade deve ser escolhido. Esse raciocínio não coincide com os argumentos apresentados.

No caso de associação de testes, outros procedimentos podem ser recomendados, como aqueles descritos em Soares & Parenti (1995).

## 4.9 Considerações finais

Na definição dos importantes índices  $VPP$  e  $VPN$ , o valor de  $p$ , a prevalência da doença na população de interesse, desempenha

papel crucial. Este valor pode variar muito dependendo do tipo de população considerada. O médico que trabalha em uma clínica de referência para doenças cardíacas encontra uma prevalência de angina muito maior do que aquele que atende a população em geral. Assim, o mesmo teste terá utilidades diferentes para estes dois profissionais.

Outra interpretação possível para  $p$  é tomá-lo como a probabilidade de ocorrência da doença, assumida pelo médico, antes do teste ser feito. Recebe neste caso o nome de *probabilidade a priori* ou *pré-teste*. Esta probabilidade não leva em conta apenas a freqüência da doença na comunidade a que o paciente pertence, mas também a sua história familiar, idade, sintomas, fatores de risco, etc.

Vários trabalhos têm sido desenvolvidos com objetivo de se determinar probabilidades pré-teste para vários tipos de pacientes, em relação a uma dada doença. Esse tipo de estudo auxilia o clínico na análise do estado do paciente e também é de grande utilidade, para o cálculo correto do valor da predição positiva. Diamond & Forrester (1979) realizaram ampla pesquisa bibliográfica e, para cada faixa etária, sexo e tipo de sintoma, estabeleceram a probabilidade pré-teste de um paciente ter doença coronariana. Por exemplo, a uma mulher com 35 anos de idade e sem queixa de angina associa-se uma probabilidade de ter a doença de 0,008 antes do teste. Já para um homem de 65 anos de idade e com angina típica associa-se, antes de realizar o teste, uma probabilidade de 0,906. Daí podemos notar como dois pacientes de uma mesma população podem apresentar probabilidades pré-teste bem diferentes que resultarão em valores de predição distintos.

Usando esta interpretação é possível quantificar quanto muda a certeza do diagnóstico depois de realizado o teste e, portanto, verificar a sua utilidade relativa. O leitor interessado deve consultar Diamond & Forrester (1983).

## 4.10 Exercícios

- É bem conhecido que daltonismo é hereditário. Devido ao fato do gene responsável ser ligado ao sexo, o daltonismo ocorre mais freqüentemente nos homens do que nas mulheres. Numa grande população humana, a distribuição de daltonismo da cor

vermelha-verde segundo sexo foi:

Daltonismo	Masculino	Feminino	Total
Presente	4,23%	0,65%	4,88%
Ausente	48,48%	46,64%	95,12%
Total	52,71%	47,29%	100,00%

- Uma pessoa é escolhida, ao acaso, dessa população. Calcule a probabilidade de ela ser:
  - Daltônica;
  - Do sexo feminino;
  - Daltônica sabendo-se que é do sexo feminino;
  - Daltônica sabendo-se que é do sexo masculino.
- Três pessoas são selecionadas ao acaso, com reposição, dessa população. Qual a probabilidade de:
  - Apenas uma ser daltônica;
  - Todas serem daltônicas;
  - Duas ou mais serem daltônicas;
  - No máximo uma ser daltônica.
- Os eventos ser daltônico e ser do sexo masculino são independentes? O que isto significa na prática?
- Se 1000 indivíduos são selecionados ao acaso, qual o número esperado de daltônicos?
- Se  $A$  e  $B$  são eventos mutuamente excludentes, qual é o valor de  $\Pr(A|B)$  e de  $\Pr(B|A)$ ?
- Durante o mês de julho de 1985, a imprensa, através de editoriais, tratou freqüentemente do assunto AIDS. Um dos pontos em questão era o teste recém-descoberto para a detecção do HIV.
  - A versão do teste do laboratório Abbot produziu 37 resultados positivos em 17.420 amostras de sangue de pessoas sadias e 123 positivos em 129 pacientes comprovadamente com AIDS. Calcule a sensibilidade e a especificidade do teste. Comente.

- (b) Se a prevalência da AIDS é de 15/100.000, qual o valor da predição positiva do teste?
4. A faixa usual de normalidade da pressão sanguínea diastólica para adultos jovens é de 75 a 85. Explique os riscos para o diagnóstico médico de se modificar as faixas para 75 a 120 ou 50 a 80. Use na sua resposta os conceitos de especificidade e sensibilidade e considere como doentes apenas os indivíduos com pressão alta.
5. (a) Os testes clínicos usuais têm sensibilidade e especificidade acima de 70%. Qual a influência nos valores das predições destes testes ao se passar de uma situação de baixa prevalência para uma de alta?  
 (b) De maneira geral o que se deve fazer para não se incorrer no erro falso-positivo muito frequentemente?  $e \uparrow 1$
6. Em que condições o valor de predição positiva ( $VPP$ ) é igual a  $a/(a + c)$  para os dados da Tabela 4.1?
7. Para o Exemplo 4.9, sobre o teste ergométrico, calcule  $VPP$ ,  $VPN$ ,  $PFP$  e  $PFN$  para as seguintes prevalências 0,05, 0,10 e 0,15. Comente os resultados.
8. Foram examinadas radiografias do tórax de 1.820 indivíduos, dos quais 30 estavam com tuberculose e 1.790 não apresentavam a doença, por diagnóstico feito de forma independente da leitura dos raios X e com uma margem de erro desprezível. Os resultados do experimento estão resumidos na tabela abaixo.

Tuberculose	Leitura dos raios X		
	Positivo	Negativo	Total
Presente	22	8	30
Ausente	51	1.739	1.790
Total	73	1.747	1.820

Pelo modo como estes 1.820 pacientes foram selecionados, a prevalência da doença, tanto neste grupo quanto na população de interesse é de  $p = 30/1.820 = 0,0165$ .

- (a) Calcule a sensibilidade e a especificidade do raio X como teste diagnóstico para tuberculose.  
 (b) Calcule as proporções de falso-positivo e falso-negativo.  
 (c) Se um médico tem 75% de certeza sobre o diagnóstico de tuberculose em um dado paciente, quanto aumenta esta probabilidade depois de se observar um teste de leitura de raios X positivo.  $P(D) = 0,75$ ,  $P(D|+) =$

9. Pesquisadores que tratam de doenças hepáticas em uma clínica especializada sugeriram um novo teste para detectar câncer no fígado. Os resultados do experimento, para uma amostra de 2.225 pacientes atendidos nessa clínica, foram:

Câncer hepático	Teste		
	Positivo	Negativo	Total
Presente	90	17	107
Ausente	39	2.079	2.118
Total	129	2.096	2.225

Pelo modo como os dados foram obtidos, a prevalência de câncer hepático, nesta clínica, pode ser calculada usando os dados da tabela.

- (a) Calcule a sensibilidade e a especificidade do teste.  
 (b) Calcule a probabilidade de um paciente, atendido nessa clínica e que não tem câncer no fígado, tenha um resultado positivo no teste.  
 (c) Calcule o valor da predição positiva e o valor da predição negativa.  
 (d) Os pesquisadores que desenvolveram o teste sugeriram seu uso como um meio simples para os clínicos em geral decidirem se devem ou não encaminhar o paciente para uma clínica especializada. Critique esta recomendação.  
 $P(D|+)$ ?  $PF_P$ ?  $e^{\text{bmx}}$ ?  $PF_N$  e  $e^{\text{bmx}}$ ?  
 10. Os valores da sensibilidade e especificidade de um teste para a detecção da diabetes são  $s = 0,95$  e  $e = 0,95$ .

(a) Um paciente assintomático submete-se ao teste. Se o resultado for positivo, qual a probabilidade de ele ter realmente a doença?  $P(D|+)$ ? ~~Por que está p?~~

(b) Um outro paciente que apresenta alguns dos sintomas da doença (como sede anormal e perda de peso) e possui história de diabetes na família também se submeteu ao teste. Se o resultado for positivo, o que você espera que aconteça com o valor do  $VPP$ ? Por quê?

(11.) Foi feita avaliação para câncer de próstata em 300 homens hospitalizados devido a sintomas de obstrução urinária. Um dos testes realizado foi o exame digital do reto. O resultado foi classificado como “positivo” segundo o critério padrão. Os resultados do exame digital e da biópsia são apresentados a seguir:

Resultado da biópsia	Resultado do exame digital		Total
	Positivo	Negativo	
Presente	48	21	69
Ausente	25	206	231
Total	73	227	300

- (a) Admita que o resultado da biópsia é completamente preciso na determinação de presença do câncer de próstata. Qual é a sensibilidade e a especificidade do exame digital na detecção de câncer de próstata?
- (b) Nesta população de homens, qual é o valor de predição positiva e o valor de predição negativa do exame digital?
- (c) Um clínico geral está considerando a possibilidade de realizar o exame digital do reto em todos os homens com mais de 50 anos que procuram seu consultório, mesmo não apresentando sintomas de obstrução urinária. Da literatura, ele sabe que a prevalência de câncer de próstata nesta faixa etária é de 0,005. Usando os valores para a sensibilidade e especificidade obtidos anteriormente, calcule a proporção de homens com resultado positivo no exame digital que realmente tem a doença.

$P(+|D)$

12. Comente as seguintes afirmações:

(a) O  $VPP$  deste teste é inaceitavelmente baixo. Jamais usaria um teste que em uma dada população produz  $VPP$  de 0,01%. ~~Não, VPP depende de p. Vide teste EUA~~

(b) A especificidade deste teste é alta, logo a maioria dos testes negativos correspondem a pessoas não doentes.  $P(-|D) \uparrow \rightarrow e \uparrow$

(c) O  $VPN$  abaixa quando a prevalência aumenta.  $TP = P(-|D) \downarrow$

13. No dia 28 de janeiro de 1995, O Jornal do Brasil publicou a seguinte notícia:

#### Teste detecta tumor de próstata em fase inicial

BOSTON, EUA - Um teste sanguíneo que detecta precocemente câncer de próstata a partir da presença de um antígeno no sangue é bastante preciso, apontando a presença da doença em 80% dos homens afetados pelo tumor, segundo estudo realizado por pesquisadores do *Brigham and Women's Hospital de Boston*. Quando o teste foi aplicado em amostras de sangue de 1464 homens coletadas em 1982, época em que o exame não estava disponível, verificou-se que ele poderia ter revelado 80% dos cânceres que só foram detectados cinco anos após a realização dos exames de sangue disponíveis na ocasião.

Critique a forma de divulgação deste novo teste.

14. Comente a seguinte afirmativa: A qualidade de um teste clínico é medida pelos índices  $s$  e  $e$ , enquanto que a qualidade do diagnóstico pelos índices  $VPP$  e  $VPN$ .

15. Suponha que a prevalência de uma dada doença é 1/100 e que um teste de diagnóstico para esta doença está sendo proposto. Se em uma verificação preliminar deste teste 99% das pessoas testadas que estavam realmente doentes e 1% das pessoas testadas que estavam realmente sãs produziram resultados positivos, quais são os percentuais de falso-positivo e falso-negativo? Na sua opinião este teste é confiável?

- (16) A detecção precoce do câncer cervical uterino é crucial para o tratamento e cura da paciente. O papanicolau é um dos testes utilizados no diagnóstico. Na tabela a seguir temos os resultados deste teste para 600 mulheres. As mulheres foram classificadas como portadoras ou não da doença através de biópsia cervical.

Paciente	Teste		
	Positivo	Negativo	Total
Com câncer	94	6	100
Sem câncer	250	250	500
Total	344	256	600

- (a) Calcule a prevalência de câncer na amostra.
- (b) Calcule a sensibilidade e a especificidade do papanicolau. Interprete os valores obtidos.
- (c) Calcule o VPP e o VPN usando a prevalência obtida no item (a).
- (d) Considere a seguinte afirmação:  
“O papanicolau é útil para excluir a presença da enfermidade, embora um resultado positivo não possa ser confiável para diagnosticar a presença da doença”. Você concorda com a afirmação? Justifique sua resposta.
17. Um clínico considera uma pessoa diabética se a concentração de glicose no sangue é maior que  $L_1$ . Um outro clínico toma como valor de referência  $L_2$  ( $L_2 > L_1$ ). Compare as sensibilidades e a especificidades dos testes utilizados pelos dois médicos.
18. A creatinina fosfocinase (CFC) é um marcador para o diagnóstico de infarto agudo do miocárdio. Pacientes com infarto agudo do miocárdio apresentam valores elevados de CFC. A tabela a seguir apresenta os valores de CFC para 360 pacientes de um hospital do coração, sendo 230 com infarto agudo do miocárdio.

Valores de CFC	Infarto agudo		Total
	Sim	Não	
$CFC < 80$	15	114	129
$80 \leq CFC < 280$	118	15	133
$CFC \geq 280$	97	1	98
Total	230	130	360

- (a) Considerando como resultado positivo valores de CFC menores ou iguais a 80 (teste 1), calcule a sensibilidade e a especificidade.
- (b) Considerando como resultado positivo valores de CFC menores ou iguais a 280 (teste 2), calcule a sensibilidade e a especificidade.
- (c) Considerando que a prevalência de infarto em hospitais do coração seja aproximadamente igual à deste estudo, calcule VPP e VPN para cada teste. Comente os resultados.
19. A hemorragia peri-intraventricular (HPIV) é uma das causas mais freqüentes de agressão ao sistema nervoso central no período neonatal e a segunda causa mais freqüente de morte em prematuros, precedida apenas pela doença da membrana hialina. Atualmente, o método de escolha para diagnóstico consiste em técnicas não invasivas (ultra-sonografia e tomografia computadorizada).
- Tavares (1995) estudou a ocorrência desta enfermidade em 120 dos 129 recém-nascidos com peso menor que 2000 g, nascidos no Hospital das Clínicas da UFMG, no período de 18/01/94 a 17/05/95. O critério único para exclusão foi morte neonatal precoce, anterior à possibilidade de realização do exame ultra sonográfico.
- (a) Para avaliar a eficácia do exame clínico neurológico, 118 crianças (com exceção de uma) foram examinadas por um pediatra clínico com formação em neonatologia, sem o conhecimento prévio dos exames ultrasonográficos, que confirmam ou não o diagnóstico fornecido pelo médico. Os

dados são apresentados na tabela a seguir.

HPIV	Exame neurológico		
	Alterado	Normal	Total
Presente	25	13	38
Ausente	31	50	81
Total	56	63	119

Calcule a sensibilidade e especificidade do exame clínico-neurológico e a prevalência dessa condição na população em questão.

- (b) O baixo peso ao nascer é usado freqüentemente para rastrear HPIV, sendo 1500 g o limite sugerido para tal fim. Considerando essa informação e os dados mostrados a seguir do peso ao nascer (g) de 39 crianças com HPIV, qual seria a sensibilidade resultante do rastreamento? Abaixo de que peso encontram-se 95% das crianças com HPIV? Quais as consequências de se alterar o valor de referência para o novo valor encontrado?

1530	1520	1350	1260	1160	1120	1500	1020
1000	1020	1690	1140	1160	1290	1410	1850
1280	1410	1750	800	1020	830	1620	1430
1150	1230	1660	1430	1590	1280	1040	1810
1420	1760	1670	1520	1680	1650	1180	

20. O'Shea et al. (1996) testaram a exatidão do diagnóstico de lesões traumáticas no joelho a partir da história do paciente, exame físico e radiográfico (comparados com artroscopia posterior) como forma de verificar até que ponto o uso da técnica de imagem ressonância magnética (exame de alto custo), que está se tornando o estudo complementar preferido na avaliação de lesões no joelho, é necessário. O estudo incluiu 156 pacientes com 156 lesões no joelho (72 agudas e 84 crônicas) que foram atendidos durante um ano no *Martin Army Community Hospital*, Georgia, EUA. Por se tratar de uma base do Exército, a maioria dos pacientes era composta de homens jovens, ativos e de uma maneira geral saudáveis. Cada paciente recebeu pelo menos um diagnóstico primário para qual o procedimento de

artroscopia foi indicado. Vinte e cinco pacientes tiveram um ou mais diagnósticos secundários, em que não se acreditava ser a causa da maioria dos sintomas ou pensava-se ser consequência do diagnóstico primário.

As tabelas a seguir representam os resultados obtidos para ruptura do menisco medial e condromalácia. Para cada tipo de lesão, calcule a sensibilidade e especificidade. Comente.

Ruptura do menisco medial	Teste		
	Positivo	Negativo	Total
Presente	60	8	68
Ausente	20	68	88
Total	80	76	156

Condromalácia	Teste		
	Positivo	Negativo	Total
Presente	11	18	29
Ausente	0	130	130
Total	11	148	149

21. Dor persistente no tendão de Aquiles (tendão calcanear) durante mais de seis meses pode ser causada por ruptura parcial do tendão. Essa lesão é geralmente refratária a tratamento conservador, sendo freqüente a necessidade de intervenção cirúrgica para alívio da dor. Para avaliar o valor diagnóstico da ultra-sonografia nesses casos, 37 pacientes tiveram seus achados do exame pré-operatório comparados aos da cirurgia (Kälebo et al., 1992). Os 30 resultados positivos encontrados na ultra-sonografia foram confirmados, mas dos 7 resultados negativos, 2 eram na verdade positivos.
- (a) Calcule a sensibilidade e especificidade da ultra-sonografia para diagnóstico de ruptura parcial de tendão de Aquiles.
- (b) Por que se pode afirmar que a prevalência calculada a partir desses estudo é diferente daquela encontrada na população em geral?

- (c) Uma bailarina chega ao consultório de seu ortopedista com história clínica, sinais e sintomas que o fazem ter 70% de certeza no diagnóstico de ruptura parcial do tendão de Aquiles. Com resultado negativo da ultra-sonografia qual a probabilidade dela não ter a lesão? O que aconteceria se o resultado fosse positivo?
22. Suponha que a sensibilidade e a especificidade de dois testes (*A* e *B*) para o diagnóstico de uma determinada doença sejam:

Teste	Sensibilidade	Especificidade
<i>A</i>	0,90	0,70
<i>B</i>	0,70	0,90

- (a) Assumindo que os testes sejam realizados de forma independente, qual a ordem que produz as melhores características gerais do teste conjunto em paralelo, isto é, o resultado é positivo se ambos são positivos?
- (b) Baseado na resposta de (a) e na informação de que os dois testes são igualmente invasivos e os custos para aplicá-los em um indivíduo são os mesmos, em que ordem eles deveriam ser aplicados para uma população cuja prevalência é  $p = 1.000/11.000$ ? (Sugestão: para as duas ordens, simule o que seria esperado para uma amostra de 11.000 indivíduos.)

## Capítulo 5

# Caracterização Estatística de Variáveis

### 5.1 Introdução

Neste capítulo vamos apresentar uma forma de caracterizar variáveis, que consiste em agrupar situações envolvendo variáveis do mesmo tipo em classes denominadas *modelos*. Descreveremos apenas os modelos de Poisson e de Gauss, respectivamente para variáveis discretas e contínuas. Eles têm grande importância porque podem descrever muitas situações práticas.

Não é possível conhecer a priori o valor de uma variável intrinsecamente sujeita à variabilidade, daí ela ser chamada *variável aleatória*. Portanto, para estudá-la, temos que identificar os valores que pode assumir e a correspondente freqüência de ocorrência, isto é, a sua *distribuição de probabilidade*, ou seja, o modelo que será usado para caracterizá-la.

Embora não exista um padrão geral para a distribuição de dados provenientes de variáveis biológicas, destaca-se uma forma de distribuição que, pelo menos de maneira aproximada, é encontrada muito freqüentemente. É a *distribuição gaussiana*, que recebe alternativamente o nome de distribuição normal, termo que será evitado neste texto pelas óbvias dificuldades que traz no estudo de variáveis com interpretação clínica. Esta distribuição será utilizada para a determinação de faixas de referência para medidas laboratoriais.

Este é um capítulo conceitual e nem sempre é possível apresentar

motivações simples e intuitivas. Apresentamos, entretanto, aplicações importantes, como a determinação de faixa de referência de medidas clínicas. Ressaltamos que as idéias aqui estudadas são absolutamente fundamentais para o entendimento das soluções de problemas médicos apresentados e discutidos a partir do próximo capítulo.

## 5.2 Conceitos fundamentais

### 5.2.1 Variáveis aleatórias e suas distribuições de probabilidade

Devido à variabilidade inerente aos seres humanos, a maioria das variáveis de interesse clínico é aleatória, ou seja, produz valores diferentes quando observadas em repetições feitas mesmo sob condições identicas. Daí a importância de se estudar o conceito de variável aleatória antes de se considerar outros problemas clínicos resolvidos com o auxílio de técnicas estatísticas.

Crianças de mesma idade, mesmo sexo, mesma classe sócio-económica, mesma raça, medidas pelo mesmo pediatra, têm pesos diferentes. Então, o peso de uma criança deve ser estudado através do conceito de variável aleatória. A palavra *aleatória* indica aqui apenas que o peso está sujeito à variabilidade e exige para seu estudo o conhecimento de seus valores possíveis e da respectiva freqüência.

Valores possíveis e freqüência são duas informações que identificam a *distribuição de probabilidade* da variável aleatória, conceito básico para todas as idéias apresentadas neste texto.

As variáveis aleatórias, usualmente representadas por letras maiúsculas, podem ser classificadas em discretas ou contínuas, se algumas tecnicidades forem desprezadas.

A variável discreta pode assumir um número finito de valores, cujas probabilidades de ocorrência são conhecidas. O número de consultas médicas anuais de um associado de um plano de saúde é uma variável aleatória discreta. Há uma probabilidade de um associado não fazer nenhuma consulta, de fazer uma, duas e assim por diante.

A variável contínua pode assumir todos valores de um intervalo e as probabilidades necessárias ao seu estudo são calculadas como a área abaixo da curva da distribuição, usualmente chamada de *função densidade de probabilidade*. A pressão arterial é um exemplo.

A partir da distribuição de probabilidade podemos calcular probabilidades da ocorrência de um determinado evento. Assim, frequentemente estaremos escrevendo:

$$Pr[a \leq X \leq b] \quad (5.1)$$

para indicar a probabilidade de a variável aleatória  $X$  tomar um valor no intervalo  $[a, b]$ .

Outros exemplos apresentados a seguir ajudarão a melhorar o entendimento do conceito de variável aleatória e de distribuição de probabilidade.

#### Exemplo 5.1: Número de meninos

Se  $X$  é o número de meninos em uma família com duas crianças, seus valores possíveis são 0, 1 e 2. A probabilidade das crianças serem do mesmo sexo é  $1/4$  e indicamos como  $Pr[X = 0] = 1/4$  para duas meninas e  $Pr[X = 2] = 1/4$  para dois meninos. A probabilidade de apenas um menino é  $2/4 = 1/2$  e indicamos por  $Pr[X = 1] = 2/4 = 1/2$ . A distribuição de probabilidade de  $X$  está apresentada na Tabela 5.1.

Tabela 5.1: Distribuição de probabilidade do número de meninos em uma família com duas crianças

Número de meninos ( $x$ )	$Pr(X = x)$
0	$1/4$
1	$2/4$
2	$1/4$
Total	1

#### Exemplo 5.2: Lançamento de um dado

Em um lançamento de um dado, seja  $Y$  o número da face voltada para cima. Os valores possíveis para  $Y$  são 1, 2, 3, 4, 5, 6 e sua distribuição de probabilidade é dada na Tabela 5.2.

Tabela 5.2: Distribuição de probabilidade dos resultados do lançamento de um dado

Resultado ( $y$ )	$\Pr[Y = y]$
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6
Total	1

### Exemplo 5.3: Número de consultas médicas

É cada vez maior o número de pessoas que utilizam planos saúde como forma de garantia de atendimento médico. Na elaboração destes planos, devemos considerar um grande número de variáveis aleatórias. Entre elas, por causa do alto volume de recursos que exige, destaca-se o número de consultas médicas por ano por associado ( $T$ ), cujos valores possíveis são 0, 1, 2, etc. O número máximo de consultas depende do plano. Neste caso, a distribuição de probabilidade não é conhecida mas podemos obter a distribuição empírica a partir dos dados.

### Exemplo 5.4: Pressão arterial

O termo pressão sanguínea refere-se à pressão do sangue nas artérias, daí a denominação pressão arterial usualmente empregada. Os níveis pressóricos alteram-se ao longo do ciclo cardíaco, atingindo seu ponto mais alto quando o músculo do coração está mais contraído, forçando sangue nas artérias, e seu menor valor quando o músculo relaxa, ao fim da batida. A contração do coração é chamada sístole e por isto a pressão medida naquela circunstância é chamada de pressão sistólica. Analogamente existe a diástole e a pressão diastólica.

A disponibilidade de métodos simples e fidedignos de medida associa à grande incidência e importância dos distúrbios relacionados

à pressão arterial fazem sua avaliação fundamental na prática médica.

A Figura 5.1 mostra a distribuição da pressão diastólica de 158.906 pessoas entre 30 e 69 anos de 14 comunidades nos Estados Unidos (*Circulation Research*, 1977). Para pessoas, mesmo pertencendo a um grupo homogêneo, são observados valores diferentes de pressão arterial, daí ser considerada uma variável aleatória. No entanto, mesmo com sua diversidade natural, percebe-se um comportamento com clara estrutura, que pode ser sintetizada na sua distribuição.

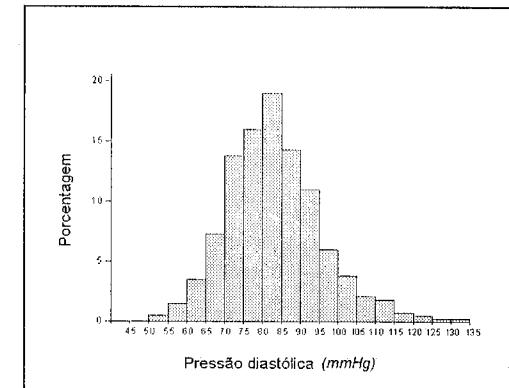


Figura 5.1: Distribuição da pressão diastólica (mmHg)

### Exemplo 5.5: Peso ao nascer

Como o peso ao nascer é uma variável com grande importância prognóstica, há todo interesse em estudá-la. Trata-se de uma variável aleatória contínua e para se chegar à sua distribuição é necessário assumir uma expressão matemática baseada em algum conhecimento prévio ou a partir de um trabalho empírico.

Na realização de um estudo é importante primeiramente ser mais específico e estabelecer critérios de inclusão, como crianças do sexo feminino nascidas de gravidez a termo, e critérios de exclusão, por exemplo, presença de diabetes na mãe.

Em uma maternidade, tomando os pesos de 100 recém-nascidos que satisfizeram os critérios estabelecidos e construindo o histograma de seu peso, obtemos o gráfico da Figura 5.2.

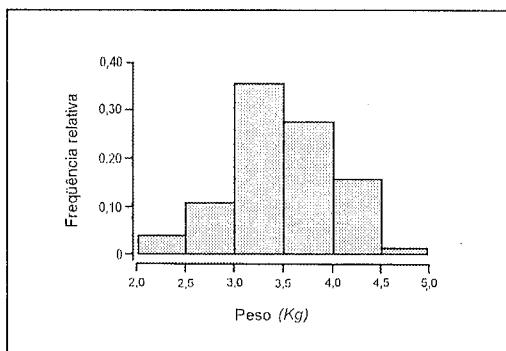


Figura 5.2: Histograma do peso ao nascer de 100 meninas

Embora o comportamento da variável possa ser sugestivo, 100 valores é pouco para se determinar a distribuição do peso ao nascer, pois o histograma não apresenta forma bem definida. Elevando-se o número de observações para 1.500, obtemos o histograma da Figura 5.3, cuja forma da distribuição é bem mais clara.

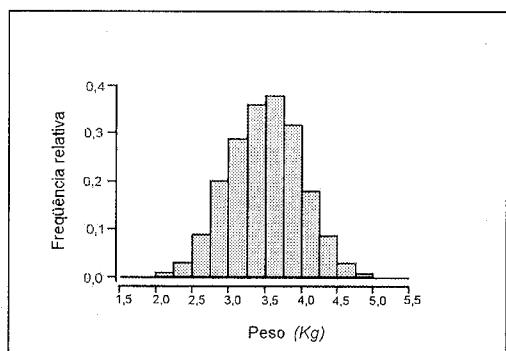


Figura 5.3: Histograma do peso ao nascer de 1500 meninas

Na realidade, por maior que seja o número de recém-nascidos considerados, só conhecemos de fato a distribuição desta variável aleatória quando identificamos a expressão matemática de sua distribuição, isto é, a sua função de densidade de probabilidade. Neste caso, guiados pela Figura 5.3, é razoável admitir que o peso ao nascer tem distribuição gaussiana, que será apresentada na Seção 5.4.

### 5.2.2 População e amostra

Uma variável aleatória e sua respectiva distribuição é muito frequentemente chamada de *população*, conceito estatístico diferente do significado usual desta palavra. Refere-se a todos os valores possíveis, com as respectivas freqüências, de uma medida específica, e não a um conjunto genérico de pessoas. Podemos, pois, nos referir à população da pressão diastólica de todos os alunos admitidos em uma universidade em determinado ano.

Neste contexto, *amostra* é um conjunto de observações extraídas da população. O processo de se obter uma amostra é chamado *amostragem* e visa garantir representatividade da população. Para isto são necessárias técnicas especializadas que geralmente incorporam especificidades das variáveis estudadas e do tipo de estudo.

Na prática, tipicamente dispomos de apenas uma amostra e a partir dela pretendemos fazer inferências para a população. Assim, baseados na amostra estimamos parâmetros populacionais de interesse (por exemplo média, variância e proporção que apresenta uma determinada característica). A teoria da estimação estatística foi desenvolvida para resolver este problema.

### 5.3 Modelo de Poisson

O modelo de Poisson é frequentemente usado para descrever variáveis aleatórias que se expressam através de contagens.

Suponha que a variável aleatória  $X$  represente o *número de ocorrências de um evento em um período de tempo*. Muitas variáveis desse tipo seguem o modelo desenvolvido pelo matemático francês Poisson, cuja distribuição de probabilidade é dada por

$$\Pr(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0, 1, \dots \quad (5.2)$$

onde  $e \approx 2,7183$ ,  $x! = 1 \times 2 \times \dots \times x$  e  $\lambda$  (lê-se lambda) é um valor positivo. O parâmetro  $\lambda$  é interpretado como a *t taxa média de ocorrências*, isto é, o número médio de ocorrências por unidade de tempo. Observe que neste modelo não há limite superior para as contagens que podem ser observadas.

A Figura 5.4 ilustra a distribuição de Poisson para alguns valores de  $\lambda$ .

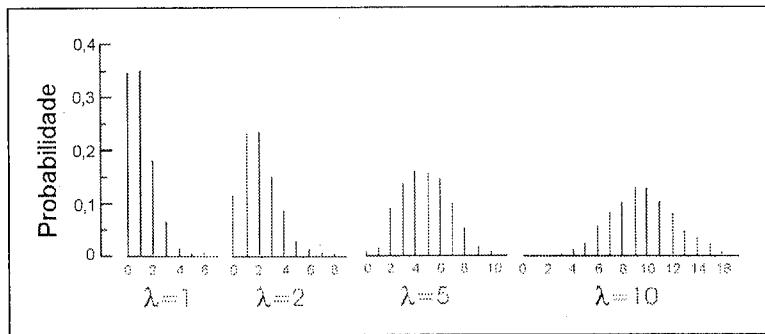


Figura 5.4: Distribuição de Poisson para valores selecionados de  $\lambda$

Note que a distribuição de Poisson pode ou não ser simétrica, dependendo do valor de  $\lambda$ . Tende a ser assimétrica para valores pequenos de  $\lambda$ . Além disso, pode-se perceber que a concentração das probabilidades ocorre em torno da média ( $\lambda$ ). Essas observações também podem ser notadas na Tabela A2 do apêndice.

Os pressupostos básicos para a utilização do modelo são:

- As condições permanecem estáveis no decorrer do tempo, isto é, a taxa média de ocorrências ( $\lambda$ ) é constante ao longo do tempo.
- Intervalos de tempo disjuntos são independentes, isto é, a informação sobre o número de ocorrências em um intervalo nada revela sobre o número de ocorrências em outro intervalo.

Além do tempo, pode-se considerar a *ocorrência do evento por certo volume ou por certa área*. Número de registros anuais de cânceres, número de óbitos diários em um grande hospital, número de bactérias por ml de urina e número de pacientes que chegam diariamente em centros de saúde são exemplos de situações que podem se ajustar à distribuição de Poisson.

### Exemplo 5.6: Chegadas de pacientes em um pronto socorro

Suponha que o número de pacientes que chegam a um pronto socorro de uma pequena cidade durante a madrugada tenha distribuição de Poisson com média 3 ( $\lambda = 3$ ). Para a obtenção da distribuição de probabilidade apresentada na Tabela 5.3, a probabilidade de chegadas de  $x$  pacientes é calculada como

$$\Pr(X = x) = \frac{e^{-3} 3^x}{x!}, \quad x = 0, 1, \dots$$

Tabela 5.3: Distribuição de Poisson com parâmetro  $\lambda = 3$

$x$	$\Pr(X = x)$	$x$	$\Pr(X = x)$
0	0,050	7	0,022
1	0,149	8	0,008
2	0,224	9	0,003
3	0,224	10	0,001
4	0,168	11	0,000
5	0,101	12	0,000
6	0,050	$\geq 13$	$\approx 0$

Usando o modelo de Poisson, pode-se dizer que:

- É muito pouco provável que nenhum paciente chegue em uma determinada madrugada, pois  $\Pr(X = 0) = 0,050$ .
- É muito provável que pelo menos um paciente chegue em determinada madrugada, pois  $\Pr(X \geq 1) = 1 - P(X = 0) = 1 - 0,050 = 0,950$ .
- É improvável que cheguem 13 ou mais pacientes durante a madrugada, já que  $\Pr(X \geq 13) \approx 0$ .

- A maior concentração da distribuição está em torno de  $x = 3$ , que é o número médio de chegadas de pacientes de madrugada.
- Se fossem consideradas duas madrugadas, o número médio de chegadas seria  $6 = 2 \times 3$ ; em um mês espera-se que cheguem  $90 = 30 \times 3$  pacientes.

#### Exemplo 5.7: Número de consultas médicas (continuação)

O número de consultas médicas anuais de um associado de um plano de saúde é, naturalmente, um número finito. Uma aproximação, que simplifica a especificação de sua distribuição, é supor que pode tomar qualquer valor do conjunto  $\{0, 1, 2, 3, 4, \dots\}$ .

Em um plano com 5.694 filiados, ao fim de um ano fizeram-se 13.098 consultas, segundo a distribuição mostrada na Tabela 5.4.

Tabela 5.4: Distribuição de consultas anuais de um plano de saúde

Nº de consultas	Freqüência	Nº de consultas	Freqüência
0	589	5	304
1	1274	6	126
2	1542	7	39
3	1144	8	10
4	663	9	3

A estimativa para o parâmetro  $\lambda$  é

$$\begin{aligned}\lambda &= \frac{\text{nº total de consultas}}{\text{nº total de associados}} = \\ &= \frac{1 \times 1274 + 2 \times 1542 + \dots + 8 \times 10 + 9 \times 3}{589 + 1274 + 1542 + \dots + 10 + 3} = \\ &= \frac{13.098}{5.694} = 2,3\end{aligned}$$

O valor 2,3 é a média de consultas anuais. Isto significa que o associado típico deste plano de saúde faz cerca de duas consultas anuais e espera-se aproximadamente 23 consultas em 10 anos.

## 5.4 Modelo de Gauss

### 5.4.1 Curva de Gauss

A curva que vamos estudar nesta seção deve-se a Gauss, eminentemente matemático alemão, e por isto ficou associada a seu nome. A curva da distribuição da variável  $X$  é caracterizada por dois parâmetros:  $\mu$  (lê-se mi) e  $\sigma$  (lê-se sigma). É definida matematicamente pela seguinte equação

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \quad (5.3)$$

A Figura 5.5 é a representação gráfica desta equação. A curva de Gauss é sempre positiva, embora pareça anular-se para valores fora do intervalo  $[\mu - 3\sigma, \mu + 3\sigma]$ , região em que assume valores muito pequenos. É simétrica em torno de  $\mu$  e engloba uma área total de 1, ou seja, 100%. As constantes  $\mu$  e  $\sigma$  são, respectivamente, a média e o desvio-padrão populacionais.

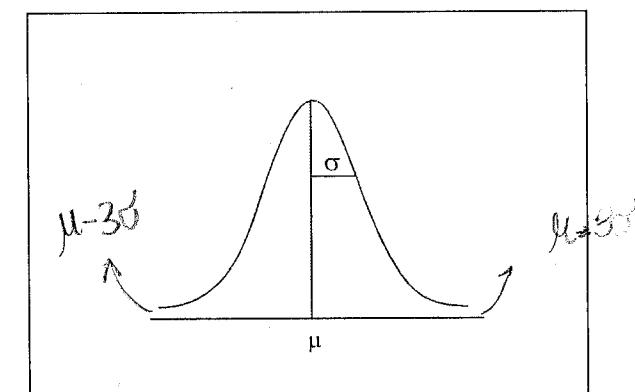


Figura 5.5: Curva de Gauss

Se uma variável aleatória  $X$  tem distribuição caracterizada pela curva de Gauss, representada na Figura 5.5, dizemos que segue a *distribuição gaussiana* e indicamos como

$$X \sim N(\mu, \sigma) \quad (5.4)$$

Inúmeras variáveis clínicas, pelo menos aproximadamente, seguem a distribuição gaussiana: valores da hemoglobina em pacientes sadios, pressão arterial sistólica, temperatura corporal, altura de crianças de uma mesma idade, etc.

#### 5.4.2 Distribuição gaussiana padrão

A distribuição gaussiana com média 0 e desvio-padrão 1, isto é,  $\mu = 0$  e  $\sigma = 1$ , é chamada de gaussiana padrão. Seu gráfico está na Figura 5.6.

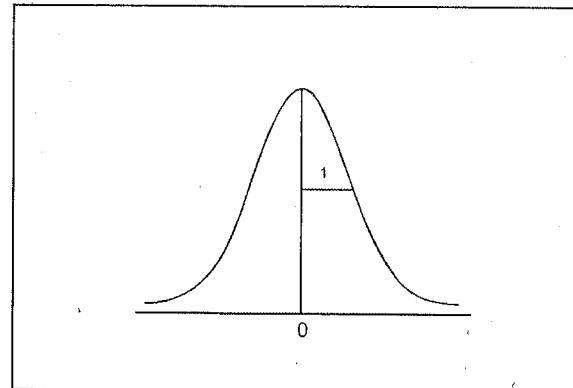


Figura 5.6: Curva de Gauss padrão

Uma variável que tem a curva de Gauss padrão como distribuição é usualmente identificada pela letra  $Z$  e é representada por

$$Z \sim N(0, 1) \quad (5.5)$$

Para o estudo e uso de técnicas estatísticas, é fundamental conhecer e saber usar a tabela da curva gaussiana padrão, reproduzida na Tabela A3 do Apêndice.

Esta tabela dá, para cada número  $x$  entre  $-3$  e  $3$ , a área abaixo da curva gaussiana padrão no intervalo  $(-\infty, x]$ . Ou seja, fornece o valor de  $Pr[Z \leq x]$ , correspondente à área sombreada da Figura 5.7.

Observe que números negativos têm valores tabelados menores que 0,5 enquanto números positivos, valores maiores que 0,5. Esta

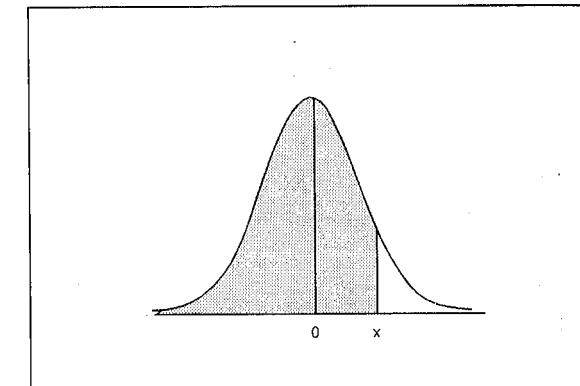


Figura 5.7: Área fornecida pela tabela da curva de Gauss padrão

propriedade decorre da simetria da curva gaussiana padrão em torno de zero.

Da Figura 5.8, vemos que, para calcularmos a probabilidade de  $X$  estar entre dois valores,  $a$  e  $b$ , precisamos fazer a seguinte conta

$$Pr[a \leq Z \leq b] = Pr[Z \leq b] - Pr[Z \leq a] \quad (5.6)$$

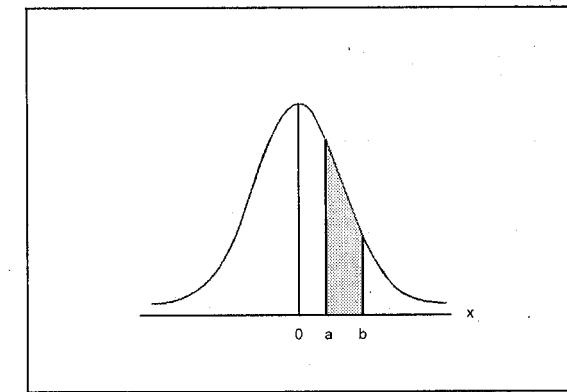


Figura 5.8: Probabilidade de ocorrência de valores entre  $a$  e  $b$

Portanto, para o cálculo da probabilidade desejada, basta consultar a tabela duas vezes: uma para obter  $Pr[Z \leq b]$  e outra para obter  $Pr[Z \leq a]$  e então fazer a diferença entre os dois valores lidos.

### Exemplo 5.8: Cálculo de probabilidades na gaussiana padrão

Neste exemplo ilustramos todos os tipos de uso da tabela da distribuição gaussiana padrão necessários em problemas práticos.

1.  $Pr[Z \leq 1]$  e  $Pr[Z \leq -1]$

Os valores dessas probabilidades são obtidos diretamente da tabela:  $Pr[Z \leq -1] = 0,1586$  e  $Pr[Z \leq 1] = 0,8413$ .

2.  $Pr[-1 \leq Z \leq 1]$

Através da expressão (5.6),  $Pr[-1 \leq Z \leq 1] = Pr[Z \leq 1] - Pr[Z \leq -1] = 0,8413 - 0,1586 = 0,6827$ .

Ou seja, o intervalo centrado na média com amplitude de dois desvios-padrão engloba 68,27% da distribuição.

3.  $Pr[1 \leq Z \leq 2]$

$Pr[1 \leq Z \leq 2] = Pr[Z \leq 2] - Pr[Z \leq 1] = 0,9773 - 0,8413 = 0,1360$ .

Ou seja, 13,6% dos valores da distribuição estão no intervalo  $[1; 2]$ .

4.  $Pr[Z \geq 2,33]$

Para o cálculo desta probabilidade podemos usar a simetria da curva, reconhecendo que  $Pr[Z \geq 2,33] = Pr[Z \leq -2,33] = 0,0099$  (valor lido diretamente da tabela). Alternativamente, podemos lembrar que a área total é 1 e portanto  $Pr[Z \geq 2,33] = 1 - Pr[Z \leq 2,33] = 1 - 0,9901 = 0,0099$ .

Ou seja, 2,33 é o percentil de ordem 0,99 da gaussiana padrão.

5.  $Pr[Z < 1,64]$  e  $Pr[Z < 1,96]$   $Pr[Z < 1,64] \cong 0,950$  e  $Pr[Z < 1,96] \cong 0,975$ .

Resumindo, 1,64; 1,96 e 2,33 são respectivamente os percentis de ordem 0,95; 0,975 e 0,99 da distribuição gaussiana.

Para calcular áreas sob a curva de Gauss com parâmetros  $\mu$  e  $\sigma$ , basta reduzir o problema ao uso da curva padrão através do seguinte resultado

$$X \sim N(\mu, \sigma) \implies \frac{X - \mu}{\sigma} \sim N(0, 1) \quad (5.7)$$

Por exemplo,

$$Pr[X \leq a] = Pr\left[\frac{X - \mu}{\sigma} \leq \frac{a - \mu}{\sigma}\right] = Pr\left[Z \leq \frac{a - \mu}{\sigma}\right] \quad (5.8)$$

Esta última probabilidade, como vimos, é obtida facilmente através de uma consulta à tabela da gaussiana padrão. A Figura 5.9 mostra graficamente o resultado enunciado em (5.7).

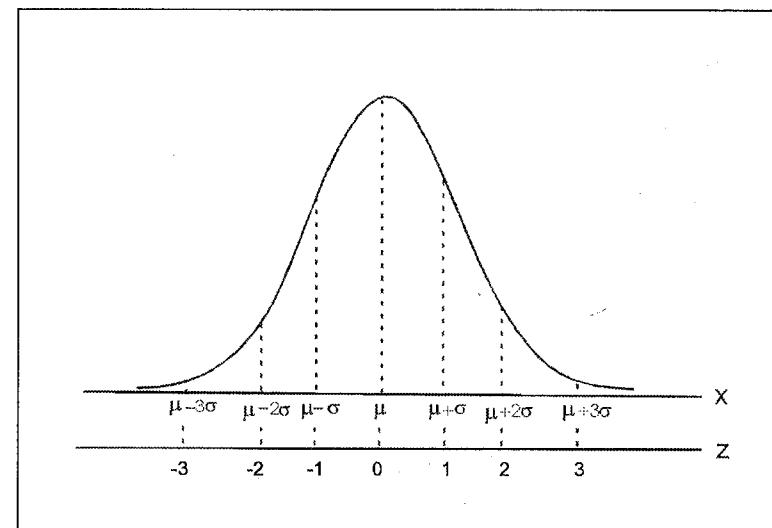


Figura 5.9: Correspondência de percentis de uma curva geral ( $x$ ) com os da curva padrão ( $z$ )

Freqüentemente, tem-se interesse nos intervalos cujos limites são média  $\pm 1$  desvio-padrão, média  $\pm 2$  desvios-padrão e média  $\pm 3$  desvios-padrão. Para o modelo gaussiano, as probabilidades da variável estudada estar nestes intervalos são mostradas a seguir:

$$Pr[\mu - \sigma \leq X \leq \mu + \sigma] = Pr[-1 \leq Z \leq 1] = 0,6826$$

$$Pr[\mu - 2\sigma \leq X \leq \mu + 2\sigma] = Pr[-2 \leq Z \leq 2] = 0,9544$$

$$Pr[\mu - 3\sigma \leq X \leq \mu + 3\sigma] = Pr[-3 \leq Z \leq 3] = 0,9973$$

Em palavras, em uma distribuição gaussiana, cerca de 68% da população apresenta resultado entre a média e o desvio-padrão; aproximadamente 95%, entre a média e 2 desvios-padrão e praticamente toda a população (99,7%), entre a média e 3 desvio-padrão.

### Exemplo 5.9: Pressão sistólica em jovens saudáveis

Suponha que a pressão arterial sistólica em pessoas jovens gozando de boa saúde tenha distribuição  $N(120, 10^2)$ .

- Qual é a probabilidade de se encontrar uma pessoa com pressão sistólica acima de 140 mmHg?

$$\begin{aligned} Pr[X \geq 140] &= Pr\left[\frac{X-120}{10} \geq \frac{140-120}{10}\right] = Pr[Z \geq 2] \\ &= 1 - Pr[Z \leq 2] = 1 - 0,9772 = 0,0228 \end{aligned}$$

Ou seja, 2,275% das pessoas jovens e saudáveis têm pressão sistólica acima de 140 mmHg.

- Quais são os limites de um intervalo simétrico em relação à média que engloba 95% dos valores das pressões sistólicas de pessoas jovens e saudáveis?

Neste caso queremos encontrar  $a$  e  $b$  tais que:  $Pr[a \leq X \leq b] = 0,95$ . Primeiro padronizamos essa probabilidade, isto é,

$$Pr[a \leq X \leq b] = Pr\left[\frac{a-120}{10} \leq \frac{X-120}{10} \leq \frac{b-120}{10}\right] = 0,95.$$

Ou seja,  $Pr[a' \leq Z \leq b'] = 0,95$ . Escolhendo uma solução simétrica temos  $-a' = b'$ . Como  $Pr[Z \leq b'] = 0,975$ , da tabela

da gaussiana padrão obtemos  $a' = -1,96$  e  $b' = 1,96$ . Conseqüentemente,  $\frac{a-120}{10} = -1,96$  e  $\frac{b-120}{10} = 1,96$ . Ou seja,  $a = 100,4$  e  $b = 139,6$ .

Para essa população de jovens saudáveis, o intervalo [100,4; 139,6] engloba 95% dos valores pressóricos, isto é, aproximadamente entre 100 mmHg e 140 mmHg.

### Exemplo 5.10: Investigação de infecção por *H. pylori*

Desde o isolamento em 1983 do *Helicobacter pylori* na mucosa gástrica humana, inúmeros estudos têm sido realizados objetivando determinar uma possível relação causal entre sua presença e algumas entidades gastroduodenais: úlceras, gastrite crônica, etc.

A presença do microorganismo tem sido diagnosticada através do exame de cultura (com uma margem de erro desprezível). Um outro método, mais simples e rápido é o teste respiratório que emprega uréia marcada com carbono-14 ( $C_{14}$ ). Por possuir uma urease, enzima capaz de degradar uréia a gás carbônico  $CO_2$ , a presença da bactéria pode ser evidenciada pela detecção de carbono marcado no ar expirado após administração, por via oral, da uréia marcada.

Suponha que a quantidade de  $C_{14}$ , liberada sob a forma de  $CO_2$  para pacientes não portadores da bactéria *H. pylori*, seja uma variável gaussiana com média 0,07 unidades de  $C_{14}$  e desvio-padrão igual a 0,03 unidades de  $C_{14}$ . A partir dessa informação podemos calcular:

- Probabilidade de uma pessoa não infectada liberar
  - Entre 0,04 e 0,10 unidade de  $C_{14}$ .
$$Pr[0,04 \leq Z \leq 0,10] = Pr\left[\frac{0,04-0,07}{0,03} \leq Z \leq \frac{0,10-0,07}{0,03}\right] = Pr[Z \leq 1] - Pr[Z \leq -1] = 0,8413 - 0,1587 = 0,6826.$$
  - Mais de 0,15 unidades de  $C_{14}$ .
$$Pr[X > 0,15] = Pr\left[Z > \frac{0,15-0,07}{0,03}\right] = Pr[Z > 2,67].$$

Há duas soluções:

- i.  $Pr[Z > 2,67] = Pr[Z < -2,67] = 0,0038$ .
- ii.  $1 - Pr[Z < 2,67] = 1 - 0,9962 = 0,0038$ .

(2) Quais valores, simétricos em relação à média, incluem 80% dos pacientes não infectados?

$$Pr[a < Z < b] = Pr\left[\frac{a-0,07}{0,03} \leq Z \leq \frac{b-0,07}{0,03}\right] = 0,80, \text{ ou seja,}$$

$$Pr[a' \leq Z \leq b'] = 0,80.$$

Como  $Pr[Z < a'] = Pr[Z > b'] = 0,10 \Rightarrow a' = -1,28$  e  $b' = 1,28$ . Assim,  $\frac{a-0,07}{0,03} = -1,28$  e  $\frac{b-0,07}{0,03} = 1,28 \Rightarrow a = 0,03$  e  $b = 0,11$ . O intervalo  $[0,03; 0,11]$  engloba 80% dos indivíduos não infectados.

## 5.5 Verificação da adequação do modelo

Após a escolha de um modelo para descrever uma situação é necessário verificar a sua adequação aos dados. A verificação pode ser baseada nas características do modelo, em gráficos especiais e em testes de aderência. Citaremos alguns desses procedimentos para os dois modelos tratados neste capítulo.

### 5.5.1 Modelo de Poisson

Uma forma para a verificação da adequação do modelo de Poisson consiste em comparar as proporções observadas com as probabilidades obtidas pelo modelo ou equivalentemente, as freqüências observadas com as esperadas segundo o modelo. Se as discrepâncias forem “pequenas” existem evidências de um bom ajuste do modelo aos dados.

#### Exemplo 5.11: Número de consultas médicas (continuação)

As proporções de ocorrência de cada número de consultas obtidas no Exemplo 5.7 são próximas das probabilidades obtidas pela expressão  $P(X = x) = (e^{-2,3} 2,3^x)/x!$ , como mostra a Tabela 5.5.

Tabela 5.5: Percentuais observados e os esperados segundo o modelo de Poisson

Nº de consultas	Percentual observado	Percentual esperado
0	10,34	10,03
1	22,37	23,06
2	27,08	26,52
3	20,09	20,33
4	11,64	11,69
5	5,34	5,38
6	2,21	2,06
7	0,68	0,68
8	0,18	0,19
9	0,05	0,05

A partir desses valores podemos calcular os valores previstos pelo modelo, conforme mostra a Tabela 5.6. Por exemplo, o número esperado de associados com duas consultas é obtido pelo seguinte cálculo:  $5.694 \times 0,2652 \cong 1.510$ . A comparação dos dois percentuais indica que o modelo de Poisson parece ser adequado.

Tabela 5.6: Freqüências de associados segundo número de consultas efetivamente feitas e previstas pelo modelo de Poisson

Nº de consultas	Freqüência		Nº de consultas	Freqüência	
	observada	prevista		observada	prevista
0	589	571	5	304	306
1	1274	1313	6	126	117
2	1542	1510	7	39	39
3	1144	588	8	10	11
4	663	666	9	3	3

### 5.5.2 Modelo de Gauss

A maneira mais simples para verificar a partir de uma amostra se a variável estudada pode ser adequadamente descrita por uma distribuição gaussiana consiste em examinar graficamente a distribuição dos dados, por exemplo, através do histograma. Se o gráfico apresenta razoável simetria e forma aproximada da curva gaussiana, tem-se indicação de adequacidade do modelo gaussiano aos dados.

Além disso, é importante verificar as proporções de observações em determinados intervalos que envolvem a média. Para uma distribuição gaussiana com média  $\mu$  e desvio-padrão  $\sigma$ , sabemos que os intervalos  $(\mu - \sigma; \mu + \sigma)$ ,  $(\mu - 2\sigma; \mu + 2\sigma)$  e  $(\mu - 3\sigma; \mu + 3\sigma)$  compreendem respectivamente 68,3%, 95,4% e 99,7% da distribuição. Em geral, tanto  $\mu$  quanto  $\sigma$  são desconhecidos, mas com uma amostra razoavelmente grande, digamos  $n \geq 30$ , é de se esperar que a média amostral  $\bar{x}$  esteja próxima de  $\mu$  e que o desvio-padrão amostral  $s$  esteja próximo de  $\sigma$ . Podemos então, com base nos dados amostrais, calcular  $\bar{x}$  e  $s$ , e contar o número de observações nos intervalos  $(\bar{x} - s; \bar{x} + s)$ ,  $(\bar{x} - 2s; \bar{x} + 2s)$  e  $(\bar{x} - 3s; \bar{x} + 3s)$ . Dividindo essas contagens por  $n$ , o tamanho da amostra, obtemos as freqüências relativas observadas que devem ser comparadas com as probabilidades teóricas (0,683; 0,954 e 0,997). Grandes discrepâncias entre as freqüências relativas observadas e as probabilidades acima indicam uma possível falta de adequação ao modelo gaussiano.

Em resumo, existe indicação de que os dados são adequados ao modelo gaussiano se:

- O intervalo  $\bar{x} \pm s$  contém cerca de 68% das observações
- O intervalo  $\bar{x} \pm 2s$  contém cerca de 95% das observações
- O intervalo  $\bar{x} \pm 3s$  contém quase todas as observações

Existem outros critérios mais precisos para se verificar se há uma boa aderência de um conjunto de dados ao modelo gaussiano. Uma boa referência é o Capítulo 1 do livro de Madansky (1988). Destaca-se um gráfico denominado *Q-Q plot*, disponível nos programas de

computador. É construído de tal forma que se há uma boa aderência dos dados ao modelo gaussiano, as observações estarão alinhadas em torno de uma reta destacada no gráfico.

#### Exemplo 5.12: Pressão sistólica de estudantes

Os alunos de uma turma de 60 estudantes do sexo masculino mediram a pressão sistólica (*mmHg*), uns dos outros, obtendo os resultados da Tabela 5.7.

Tabela 5.7: Pressão sistólica (*mmHg*) de 60 estudantes

142	142	134	110	98	130
136	120	118	130	116	140
118	122	128	128	114	138
104	116	110	100	128	128
124	140	108	146	116	114
152	118	140	128	116	110
138	132	118	120	122	120
108	112	94	130	130	118
120	128	108	120	124	110
124	132	132	130	102	118

O histograma desses dados, apresentado na Figura 5.10, embora ligeiramente assimétrico, tem aproximadamente a forma gaussiana. Portanto, é razoável pensar que a pressão sistólica possa ser descrita pela distribuição gaussiana.

Para esses dados tem-se:  $\bar{x} = 122,70$  e  $s = 12,21$ . Espera-se cerca de 68,3%, 95,4% e 99,7% da população respectivamente nas faixas  $(122,70 - 12,21; 122,70 + 12,21) = (110,49; 134,91)$ ,  $(122,70 - 2 \times 12,21; 122,70 + 2 \times 12,21) = (98,28; 147,12)$  e  $(122,70 - 3 \times 12,21; 122,70 + 3 \times 12,21) = (86,07; 159,33)$ , enquanto que os percentuais observados são 63,3%, 95,0% e 100%. Como não há grande discrepância entre estas porcentagens e as probabilidades teóricas, conclui-se que o modelo de Gauss para esta variável é razoável.

O *Q-Q plot* apresentado na Figura 5.11 indica uma excelente aderência dos dados ao modelo gaussiano.

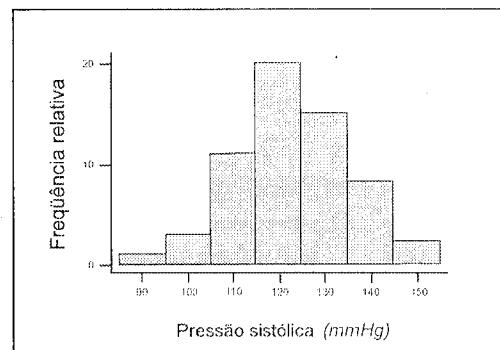


Figura 5.10: Histograma para a pressão sistólica ( $mmHg$ )

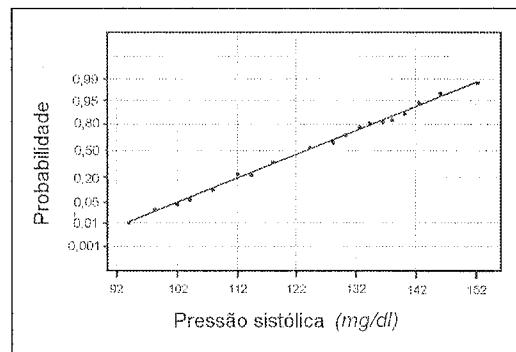


Figura 5.11:  $Q-Q$  plot para a pressão sistólica ( $mmHg$ )

#### Exemplo 5.13: Nível de colesterol do Estudo de Framingham

A média e o desvio-padrão para as 78 observações do Exemplo 3.2 do Capítulo 3 são  $\bar{x} = 221,27$  e  $s = 51,02$ . As faixas  $(221,27 - 3 \times 51,02; 221,27 + 51,02) = (170,25; 272,29)$ ,  $(221,27 - 2 \times 51,02; 221,27 + 2 \times 51,02) = (119,23; 323,31)$  e  $(221,27 - 3 \times 51,02; 221,27 + 3 \times 51,02) = (68,21; 374,33)$ , englobam 79,5%, 98,3% e 98,7% dos dados. O primeiro percentual difere significantemente do teórico (68,3%).

O histograma mostrado na Figura 3.1 é assimétrico, indicando

que neste caso o modelo gaussiano não é razoável. Isto é confirmado pelo gráfico  $Q-Q$  plot da Figura 5.12.

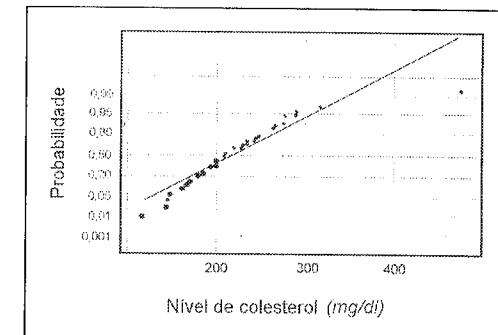


Figura 5.12:  $Q-Q$  plot para o nível de colesterol ( $mg/dl$ )

#### 5.6 Faixa de referência

A hemoglobina é um parâmetro hematológico de uso rotineiro na clínica médica. Tomando o histograma da Figura 5.13 como aproximação para a sua distribuição entre mulheres, vemos que valores menores que  $12\text{ g}/dl$  são pouco comuns. É, pois, razoável considerar um diagnóstico de anemia ao se observar uma paciente com valor de hemoglobina de  $11\text{ g}/dl$ .

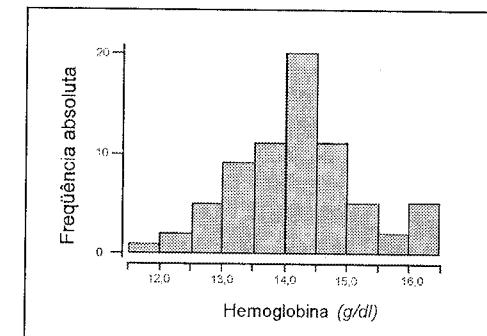


Figura 5.13: Histograma da hemoglobina ( $mg/dl$ )

Por causa da variabilidade entre indivíduos, a análise de um valor específico de hemoglobina em mulheres deve ser feita confrontando-o com a faixa 12-16 g/dl. É a chamada *faixa de referência*.

Toda medida laboratorial é analisada confrontando-se seu valor com uma faixa. Isto é tão comum que, na própria apresentação do resultado, muitos laboratórios já indicam os limites inferiores e superiores para o valor da medida que devem servir de base ao raciocínio clínico.

O procedimento de definição de faixa de referência baseia-se na hipótese de que as populações de pessoas saudáveis e as de pessoas doentes produzem, para aquela medida, valores que flutuam em torno de médias diferentes, gerando curvas com pequena interseção. A Figura 5.14 ilustra um possível modelo.

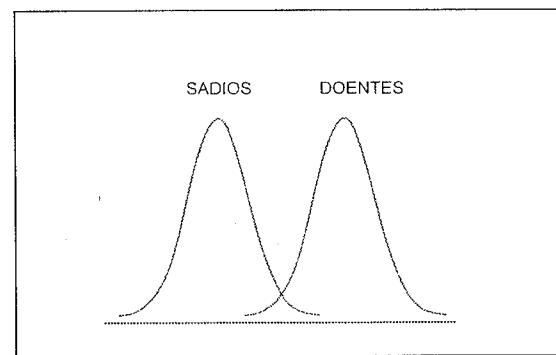


Figura 5.14: Modelo para construção de faixas de referência

Estas faixas devem ser utilizadas de maneira crítica. A associação de valores fora da faixa com a condição de doente não pode ser automática. O conceito de saúde é complexo e não pode ser reduzido nem a uma única dimensão, nem a uma sucessão de exames laboratoriais. É necessário unificar todas as informações disponíveis, tarefa do médico.

É fácil provar que a probabilidade de uma pessoa apresentar valores dentro da faixa de referência em todos os exames de uma grande bateria de testes é muito pequena. Ou seja, o uso ingênuo destas faixas pode levar a conclusões clinicamente sem sentido.

A construção de faixas de referência baseia-se no uso de informações de pessoas pertencentes a uma população saudável. Há, entretan-

to, inúmeras dificuldades para definição desta população e divergências entre os autores a esse respeito. É preciso considerar também outras características além daquela em estudo: idade, sexo, peso, nível de atividade, tabagismo, dieta, raça, local de residência e nível sócio-econômico. Esses elementos criariam uma forma de comparação dos valores com aqueles da população estudada. Todo médico que incorpora criticamente sua experiência anterior desenvolverá, ao longo da vida, critério de valorização de desvios dos valores de referência.

### 5.6.1 Obtenção da faixa de referência

Existem, essencialmente, duas maneiras de se obter uma faixa de referência. A primeira, baseada na curva de Gauss, só é apropriada para variáveis que têm distribuição definida por esta curva. Apesar disto, é a forma mais usada. A segunda, mais recomendada, é baseada nos percentis e vale para qualquer tipo de distribuição.

Em algumas circunstâncias, há outros fatores relevantes que tornam inadequada a simples utilização desses métodos na construção de faixas de referência. Este é o caso da pressão arterial, em que os limites para diagnóstico de hipertensão baseiam-se também no risco aumentado de complicações cardiovasculares e na intervenção médica para reduzi-lo. Nesses casos, torna-se necessário incorporar esses outros fatores na construção de faixas de referências através de modelos multivariados.

#### Método da curva de Gauss

Na Seção 5.4, vimos como obter valores, simétricos em torno da média  $\mu$ , tais que a área sob a curva de Gauss, delimitada por este intervalo, seja um valor pré-fixado. Em particular, sabemos que o intervalo  $(\mu - 2\sigma, \mu + 2\sigma)$  engloba 95,6% da área sob a curva gaussiana. Em outras palavras, uma variável com esta distribuição produzirá, em média, 95,6% de valores naquela faixa.

Como na prática lidamos sempre com uma amostra, a distribuição não é completamente conhecida e, portanto, utilizamos o intervalo

$$(\bar{x} - 2s; \bar{x} + 2s) \quad (5.9)$$

como faixa de referência. Na expressão acima, o número 2 está associado ao valor de cobertura 95,6%, valor próximo de 95%. Outros

multiplicadores produzem intervalos com diferentes porcentagens de cobertura. Por exemplo, 2,5 produz uma faixa de referência que engloba 98,75% dos valores da medida de interesse. Apresentamos a seguir as faixas de referências para as coberturas mais comuns.

Cobertura	Faixa de referência
90%	$(\bar{x} - 1,64s; \bar{x} + 1,64s)$
95%	$(\bar{x} - 1,96s; \bar{x} + 1,96s)$
99%	$(\bar{x} - 2,58s; \bar{x} + 2,58s)$

Em muitas situações só interessa o limite inferior ou o limite superior de referência. Nestes casos, faz-se a extensão do procedimento da construção de faixas de referência. Os limites são determinados facilmente utilizando-se a tabela da distribuição gaussiana padrão. Por exemplo, o limite superior de referência que engloba 95% dos valores da medida de interesse é  $\bar{x} + 1,64s$ .

#### Exemplo 5.14: Teor de gordura fecal

Utilizando os dados da Tabela 3.1 do Capítulo 3, referentes ao teor de gordura fecal de 43 crianças que não eram amamentadas pela mãe, calculamos a média e o desvio-padrão ( $\bar{x} = 2,303$ ,  $s = 0,872$ ) e obtemos os seguintes valores de referência com cobertura de aproximadamente 95%:

$$\text{Limite inferior de referência: } 2,303 - 2(0,872) = 0,559$$

$$\text{Limite superior de referência: } 2,303 + 2(0,872) = 4,047$$

#### Método dos percentis

Críticas podem ser feitas ao cálculo de faixas de referência pelo método da curva de Gauss. A principal delas é que os limites do intervalo  $(\bar{x} - 2s, \bar{x} + 2s)$  são muito dependentes da suposição de que a distribuição da medida de interesse é gaussiana e nem sempre isto é verdade. Usar este intervalo em população não gaussiana produz, às vezes, valores pouco razoáveis para os limites da faixa (como, por

exemplo, valor negativo para o limite inferior em uma situação em que as medidas tomam apenas valores positivos).

Como, contudo, nem sempre as pessoas que constroem padroes de referência estão atentas a esses problemas, a metodologia de Gauss é utilizada com grande freqüência, às vezes indevidamente.

Para superar as limitações do método baseado na curva de Gauss, basta usar o método dos percentis, adequado a qualquer tipo de distribuição. Por exemplo, para uma cobertura de 95% utilizamos como limites de referência os percentis de ordem 2,5% e de 97,5%, ou outros mais convenientes à situação.

No Capítulo 3 vimos como calcular os percentis. Entretanto, quando são poucos os dados, uma forma aproximada, mas bastante efetiva de se obter a faixa, é descrita a seguir. Ordenam-se os valores a serem usados e diante de cada um coloca-se o valor  $\frac{i-0,5}{n}$ , onde  $i$  é o número de ordem e  $n$  o número total de observações. Este valor é uma boa aproximação para a ordem do percentil, correspondente ao valor de ordem  $i$ . Procura-se então os percentis mais próximos aos percentuais fixados.

#### Exemplo 5.15: Pressão intraocular

Glaucoma é uma séria doença oftalmológica que pode levar à cegueira. A presença de pressão intraocular elevada é a principal indicação de que a doença está em desenvolvimento. Por isto é fundamental para a prática oftalmológica estabelecer faixas de referência para a medida da pressão intraocular.

Muitas populações já foram estudadas, o que permite estabelecer limites estatísticos de referência com bastante segurança. Quase todos estes estudos indicaram que a distribuição da pressão ocular entre indivíduos normais é assimétrica à direita, como ilustrado na Figura 5.15.

A forma de cálculo dos limites não pode ser, portanto, aquela baseada na curva de Gauss. Aqui só interessa o limite superior. A Tabela 5.8 apresenta percentis obtidos em algumas pesquisas.

Estes estudos justificam o uso do valor 21 mmHg como limite superior de normalidade para uma medida isolada da pressão ocular.

Para cada medida utilizada em sua prática, o médico criterioso deve conhecer estudos como estes que justificam os limites escolhidos.

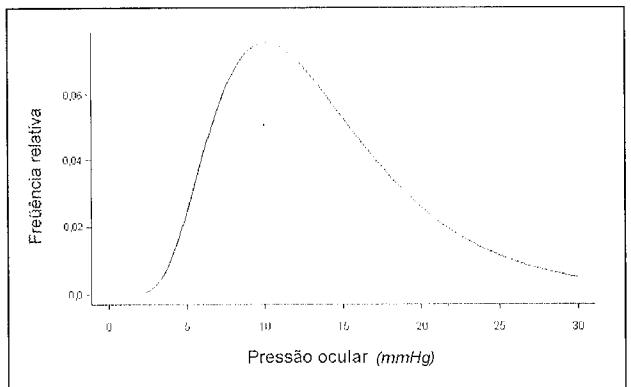


Figura 5.15: Distribuição da pressão ocular (mmHg)

Tabela 5.8: Informações sobre vários trabalhos que obtiveram percentis para a pressão intraocular

Autores, ano	Nº de pessoas	Tonômetro	Percentis		
			97,5	99	99,9
Bechrakis, 1969	1.235	Aplanação	21	24	30
Hollow & Graham, 1966	4.049	Aplanação	21	22	30
Armaly, 1965	4.652	Aplanação	20	21	28
Leydhecker et al., 1958	19.880	Schiotz	26	29	32

#### Exemplo 5.16: Teor de gordura fecal (continuação)

Os cálculos necessários para se obter a faixa de referência para o teor de gordura fecal através do método dos percentis estão na Tabela 5.9.

A partir desses resultados, uma faixa razoável seria  $(1,00 - 3,91)$  correspondente aos percentis de ordem 3,5% e 96,5%. Através do método de Gauss obtivemos  $(0,559 - 4,047)$ . Observe que métodos diferentes produzem valores diferentes para a faixa de referência.

Tabela 5.9: Cálculos necessários para o método dos percentis

Nº de ordem	Valor observado	Ordem do percentil
1	0,78	0,012
2	1,00	0,035
3	1,00	0,058
4	1,12	0,081
5	1,28	0,105
6	1,31	0,128
:	:	:
36	3,12	0,826
37	3,22	0,849
38	3,24	0,872
39	3,27	0,895
40	3,68	0,919
41	3,81	0,942
42	3,91	0,965
43	4,58	0,988

## 5.7 Exemplos comentados

### 5.7.1 Cartas para controle de processos em hospitais

Freqüentemente há a preocupação em monitorar problemas de Saúde Pública, como surtos de doenças, e em detectar desvios das normas aceitas em hospitais. Sellick Jr. (1993) apresenta um importante procedimento para o controle de processos em hospitais. Este controle é feito mantendo-se o resultado mensurável e importante do processo dentro de limites. Por exemplo, no processo de controle de infecção hospitalar, pode-se contar o número de infecções por um certo período. Analogamente para determinados acidentes de trabalho no hospital, óbitos, etc., pode-se considerar as freqüências de ocorrências, mas poderia ser definido um outro indicador ligado ao processo.

Os dados são as observações referentes ao processo tomadas no longo do tempo. Além de freqüências, pode-se também considerar medidas, como tempo de espera ou pressão sanguínea. Como a maioria

ria dos dados relacionados ao controle de processos em hospitais são atributos, por exemplo presença ou não de infecção, as freqüências são mais comuns.

O artigo de Sellick Jr. (1993) apresenta aplicações para variáveis que podem ser descritas através do modelo de Poisson e discute também outros casos de variáveis discretas. Não trata, entretanto, de algumas situações de interesse, como por exemplo de variáveis contínuas, embora apresente referências para esse caso.

O princípio básico do procedimento é que todos os processos têm uma variabilidade inerente e que podem ser descritos em termos estatísticos. A forma de avaliar se o processo está sob controle é através da construção de um gráfico conhecido como *carta de controle*. Consiste em construir um gráfico de linhas (discutido na Seção 3.3 do Capítulo 3) e adicionar linhas paralelas ao eixo do tempo na altura da média, da média  $\pm 2$  desvios-padrão e da média  $\pm 3$  desvios-padrão, conforme mostra a Figura 5.16.

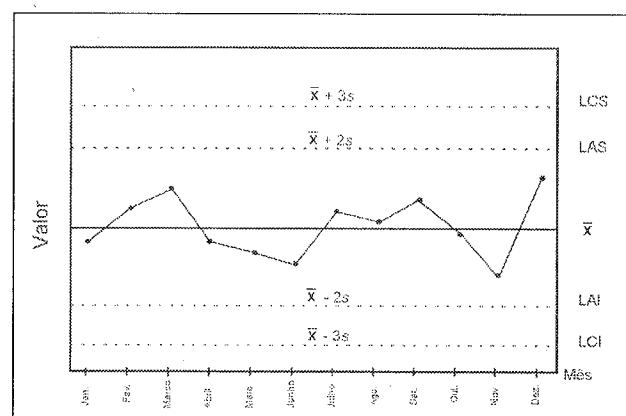


Figura 5.16: Carta de controle genérica

As linhas correspondentes a dois desvios-padrão são chamadas limites de atenção superior (LAS) e inferior (LAI), e as linhas correspondentes a 3 desvios-padrão, limites de controle superior (LCS) e inferior (LCI). Não há um critério claro sobre quantos períodos de tempo (mês, ano, etc.) devem ser considerados para o cálculo da média. Se o período de tempo é mês, geralmente são usados 12 a 24 meses e os períodos de epidemia devem ser excluídos.

De uma forma geral, considera-se que as seguintes situações merecem ser investigadas, pois podem indicar que o processo está fora de controle:

- Um ponto encontra-se acima da linha LCS ou abaixo da linha LCI.
- Dois de três pontos consecutivos estão entre as faixas LAS e LCS ou entre as faixas LAI e LCI.
- Quatro de cinco pontos consecutivos estão entre a linha média  $+1$  desvio-padrão e a linha LAS ou entre a linha média  $-1$  desvio-padrão e a linha LAI.
- Nove pontos consecutivos estão acima (abaixo) da média.
- Seis pontos consecutivos são crescentes ou decrescentes.
- Quinze pontos consecutivos estão dentro das linhas média e média  $+1$  desvio-padrão ou das linhas média e média  $-1$  desvio-padrão.

Estes testes refletem não só a distância de um dado valor em relação à média, mas também padrões não naturais para pontos dentro de linha de controle. Portanto, as cartas de controle podem ser usadas para detectar alguma causa de variação.

Essas variações são geralmente relacionadas com influências a lheias ao processo, tais como troca na população de pacientes, novo método de vigilância, novo pessoal de atendimento, ou podem resultar de características intrínsecas ao processo.

Os testes de controle podem também ser usados para monitorar mudanças em causas comuns de variação durante ou após a implementação de atividades de melhoramento de qualidade do processo.

#### Exemplo 5.19: Exame de fezes para *Clostridium difficile*

Colite associada a antibiótico relaciona-se com alterações na flora microbiana normal que podem levar a supercrescimento de *Clostridium difficile* e produção de toxina pelo mesmo. Os sintomas variam de diarréia leve à colite pseudomembranosa, condição potencialmente

fatal em que estão presentes febre, dor abdominal, diarréia, muco ou sangue nas fezes. O diagnóstico usualmente é feito pela demonstração de toxina nas fezes. A proctoscopia revela placas brancas e amarelas sobre a mucosa do cólon. Geralmente há remissão com a interrupção do uso do antibiótico, mas recidivas podem ocorrer. Algumas vezes é necessária terapia específica.

Em um hospital que está monitorando o problema de colite associada a antibióticos por *Clostridium difficile*, vamos assumir que o universo de pacientes é estável e que não houve troca de métodos de vigilância, nos procedimentos, nos exames e que é aproximadamente constante o número desses exames realizados por mês.

Na Tabela 5.10 encontra-se o número mensal de resultados positivos no exame de fezes para a toxina produzida pelo *Clostridium difficile*.

Tabela 5.10: Número mensal de resultados positivos de toxina produzida pelo *Clostridium difficile* em exames de fezes

Mês do exame	Nº de resultados positivos
Janeiro	2
Fevereiro	4
Março	1
Abril	3
Maio	1
Junho	4
Julho	0
Agosto	3
Setembro	4
Outubro	3
Novembro	6
Dezembro	3
Total	34

No mês de novembro houve o maior número de resultados positivos, aproximadamente duas vezes maior que a média mensal de 2,8. A pergunta é se esses seis resultados positivos devem ser motivo de

preocupação, ou seja, se é realmente uma contagem estatisticamente improvável. Para responder essa pergunta construímos a carta de controle com média 2,8 e desvio-padrão 1,6.

Como pode ser visto na Figura 5.17, nenhum ponto do gráfico atinge os critérios do teste estatístico de variação significante. Isto não significa que a média mensal de resultados positivos é apropriada ou aceitável, mas somente que, baseado na experiência anterior, o número está dentro da variação esperada.

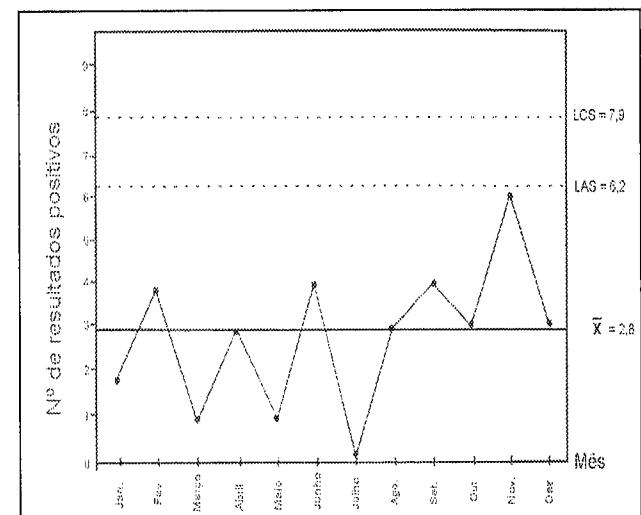


Figura 5.17: Carta de controle

### 5.7.2 Rastreamento de doadores de sangue

Esta seção tem como objetivo, além de ilustrar o uso da distribuição gaussiana, esclarecer alguns dos princípios básicos dos programas de rastreamento de massa.

A dosagem sérica da enzima alanina transaminase (ALT), anteriormente designada transaminase glutâmico pirúvica (TGP), é usada desde 1955 como indicadora de danos hepatocelulares, encontrando-se pronunciadamente aumentada nos casos de hepatites virais, agudas e crônicas, sendo mais específica para detectar doença hepática assintomática em pacientes não alcoólicos. Nos bancos de sangue, os doadores são rotineiramente avaliados para a elevação dessa enzima antes da transfusão.

Apresentamos a seguir uma adaptação de um exemplo descrito em Colton (1974), baseado no artigo de Prince & Gershon (1965) que

trata da transmissão da hepatite virótica na transfusão de sangue. Na época foi implantado um programa de rastreamento de sangue contaminado com vírus da hepatite em doadores de um banco de sangue, baseado na medida de ALT.

Resultados empíricos de várias pesquisas indicam que o logaritmo da determinação do ALT tem aproximadamente distribuição gaussiana (a distribuição do ALT é denominada log-normal) nas populações saudáveis e com danos hepatocelulares, respectivamente denominadas *saudável* e *doente*. Os parâmetros destas populações na escala logarítmica em  $UI/l$  estão mostrados no quadro abaixo:

População	Média	Desvio-padrão
Saudável	1,25	0,12
Doente	1,55	0,13

Para definir o rastreamento proposto, foi necessário estabelecer um ponto de corte de  $X = \log(\text{ALT})$ . Assim, o sangue de um doador em potencial com  $x$  acima deste ponto de corte seria rejeitado e abaixo seria aceito. Foi estabelecido que este ponto de corte seria tal que o procedimento de rastreamento aceitasse 95% dos doadores provenientes da população saudável. A determinação do ponto de corte ( $x$ ) usando este critério é facilmente obtida, lembrando-se que o percentil de ordem 95 da gaussiana padrão é 1,64. Assim,

$$1,64 = \frac{x - 1,25}{0,12} \Rightarrow x = 1,45$$

que corresponde à dosagem de ALT igual a 28,18.

Portanto, se para um doador,  $x$  fosse maior que 1,45, seu sangue seria rejeitado. Obviamente um doador nestas condições poderia ser ou não portador do vírus da hepatite. Pelo critério adotado, 5% dos sangues serão rejeitados indevidamente, isto é, a medida  $x$  seria maior que 1,45 mas o doador não é portador do vírus. Por outro lado, o fato de  $x$  ser menor que 1,45 não garante que o doador não seja portador do vírus. Surge então a pergunta: qual o percentual de sangues aceitos indevidamente? Para este cálculo, basta padronizar o ponto de corte na curva da população doente ( $z = (1,45 - 1,55)/0,13 = -0,77$ ) e verificar que a área abaixo deste percentil é 0,218. Portanto, espera-se que 21,8% dos portadores do vírus teriam seu sangue aceito. A Figura 5.18 mostra esquematicamente este programa de rastreamento

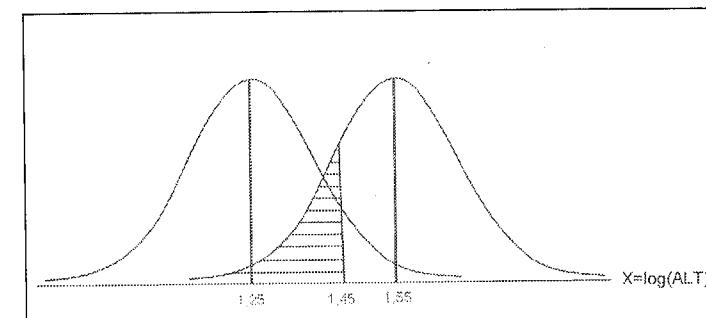


Figura 5.18: Distribuição da população saudável e doente

através da sobreposição das curvas correspondentes às populações saudável e doente.

O percentual de amostras aceitas indevidamente é bastante alto e pode ser diminuído mudando-se o ponto de corte para um valor menor. Idealmente deveria ser obtido um ponto de corte tal que o percentual de sangues contaminados fosse o menor possível. Entretanto, isto implicaria no aumento do percentual de sangues rejeitados indevidamente, o que poderia inviabilizar o programa de rastreamento. Uma solução seria então fazer uma análise variando-se o ponto de corte e então calcular os percentuais de sangues aceitos e rejeitados indevidamente. A partir destes cálculos seria escolhido um ponto de corte ótimo, definido a partir de algum critério (por exemplo, em termos de valores de predição).

No Capítulo 4 vimos como os valores de predição dependem da prevalência na população. Ilustraremos estes cálculos para uma prevalência de danos hepatocelulares de 12%. Podemos usar as fórmulas da Seção 4.6.2 ou alternativamente podemos simular uma população de determinado tamanho e construir uma tabela 2 x 2 a partir dos percentuais fixados. Por exemplo, em uma população de 1.000 indivíduos, seriam 120 doentes (correspondendo a 12% que é a prevalência) e 880 saudáveis. Considerando o ponto de corte de 1,45 e portanto os percentuais de sangue aceitos e rejeitados indevidamente respectivamente de 21,8% e 5%, espera-se que 44 (880 × 0,05) doadores saudáveis rejeitados e 26 (120 × 0,218) doadores doentes aceitos. Resumindo, temos a Tabela 5.11.

Tabela 5.11: Distribuição de sangues aceitos e rejeitados em uma população de 1.000 indivíduos

Doador proveniente de população	Sangue		
	Aceito	Rejeitado	Total
Sadia	836	44	880
Doente	26	94	120
Total	862	138	1.000

A efetividade do rastreamento pode então ser avaliada através dos valores preditivos ou equivalentemente de falsos-positivos e negativos. Entre os 862 doadores aceitos, apenas 26 (3%) são doentes. Por outro lado, dos 138 doadores rejeitados, 44 (32%) poderiam ter sido aceitos. Em resumo, usando o ponto de corte de 1,45 tem-se um alto percentual de falso-positivo (32%) e um baixo percentual de falso-negativo (3%). A efetividade deste programa é bastante boa, mas pode não ser viável na prática.

Se, em uma tentativa de minimizar a rejeição de sangue em doadores saudáveis, passarmos a aceitar 97,5% dos doadores originados da população sadia, o ponto de corte seria 1,49. Nesse caso, apenas 2,5% de sangue seriam rejeitados indevidamente, mas o percentual de doentes com sangue aceito passaria para 32,3%.

Para a prevalência de danos hepatocelulares anterior (12%), o percentual de sangue rejeitado de indivíduos sadios cairia para 20% e o percentual de doentes com sangue indevidamente aceito passaria para 4%.

## 5.8 Exercícios

1. A vida de certo aparelho cirúrgico pode ser descrita pela distribuição gaussiana com média de oito anos e desvio-padrão de 1,4 anos. O contrato de garantia diz que o fabricante substituirá os aparelhos que apresentarem defeito dentro do prazo de garantia. Se ele deseja substituir no máximo 5% dos aparelhos que fabrica, qual deve ser o prazo de garantia a ser estabelecido?
2. Em um exame de Estatística, a nota média foi 70 com desvio-

padrão de 4,5. Todos os alunos que obtiveram nota de 70 a 80 tiveram conceito B. Se as notas têm distribuição aproximadamente gaussiana e se 10 estudantes obtiveram conceito B, quantos se submeteram ao exame?

3. Suponha que o número de novos casos de câncer diagnosticados em uma determinada região siga a distribuição Poisson com parâmetro  $\lambda = 2,2$ . Interprete esse valor e apresente possíveis aplicações práticas do uso desse modelo neste caso.
4. Na curva gaussiana padrão obtenha o valor  $z_0$  tal que a área entre a média da curva e  $z_0$  seja 0,40.
5. Um teste de aptidão para o exercício de certa profissão exige uma seqüência de operações a serem executadas rapidamente uma após outra. Para passar no teste, o candidato deve completá-lo em 80 minutos no máximo. Admita que o tempo para completar o teste seja uma variável aleatória  $N(90, 20)$ .
  - (a) Qual a porcentagem dos candidatos tem chance de ser aprovados?
  - (b) Os melhores 5% receberão um certificado especial. Qual o tempo máximo para fazer jus a tal certificado?
6. A distribuição da altura de 500 estudantes do sexo masculino de uma escola é aproximadamente gaussiana, com média 1,70 m e desvio-padrão 2,5 cm.
  - (a) Quantos têm altura inferior a 1,75 m?
  - (b) Quantos têm altura entre 1,72 e 1,80 m?
7. Para os dados do Exemplo 5.12:
  - (a) Determine os valores de referência.
  - (b) Um livro de referência estabelece, para adultos jovens do sexo masculino entre 20 e 24 anos, os valores de 105 a 140 mmHg. Os resultados obtidos em (a) são compatíveis com estes limites?

8. Calcule a mediana e o percentil de ordem 75 para uma variável com distribuição gaussiana com média  $\mu$  e desvio-padrão  $\sigma$ .
9. Os seguintes valores representam os tempos gastos, em segundos, por pessoas com doença cardíaca para a realização de um teste simples de esforço.

289, 203, 359, 243, 232, 210  
251, 246, 224, 239, 220, 211

- (a) Calcule a média, mediana, desvio-padrão e os percentis de ordem 25 e 75.
- (b) Estes dados foram gerados de uma população gaussiana? Justifique sua resposta. (Sugestão: use o resultado do exercício anterior).
10. A distribuição da pressão sistólica por faixa etária em uma população masculina é apresentada na tabela abaixo. Considerando o modelo gaussiano adequado para a descrição dessa distribuição, as faixas de referência estão abaixo. Comente as repercussões da adoção dos valores encontrados como limites para classificação de hipertensão.

Idade (anos)	Média	Desvio-padrão	Faixa de referência
16	118,4	12,17	[94,06; 142,74]
17	121,0	12,88	[95,24; 146,76]
18	119,8	11,95	[95,90; 143,70]
19	121,8	14,99	[91,82; 151,78]
20 $\leftarrow$ 25	123,9	13,74	[96,42; 151,38]
25 $\leftarrow$ 30	125,1	12,58	[99,94; 150,26]
30 $\leftarrow$ 35	126,1	13,61	[98,88; 153,32]
35 $\leftarrow$ 40	127,1	14,20	[98,70; 155,50]
40 $\leftarrow$ 45	129,0	15,07	[98,86; 159,14]
45 $\leftarrow$ 55	132,3	18,11	[96,08; 168,52]
55 $\leftarrow$ 65	139,8	19,99	[99,82; 179,78]

11. É sabido que, para os adultos do sexo masculino, gozando de boa saúde, em uma certa população, a temperatura corporal segue distribuição gaussiana com média de 36,8 graus e desvio-padrão de 0,15 graus.

- (a) Se considerarmos 1000 dessas pessoas, quantas se esperariam com temperatura entre 36,8 e 37,2 graus?
- (b) Em qual intervalo de temperaturas estão 98% dos adultos masculinos saudáveis desta população?

12. Um pesquisador deseja criar um padrão para identificar presença de infecção bacteriana (*Pseudomonas sp*) no trato respiratório através de cultura de escarro. Para isto, coletaram-se dados de pessoas sabidamente saudáveis e determinou-se o número de colônias encontradas em cada cultura. Foram encontrados os seguintes resultados:

17	22	23	23	23	23	24	24	24
24	24	24	25	25	25	25	25	25
25	26	28	28	29	30	30	31	31
35	35	35	36	40	41	41	41	42
51	54	56	56	56	58	60	68	79

- (a) Determine uma faixa de normalidade de 95% para o número de colônias de bactérias no trato respiratório de pessoas saudáveis, usando o método dos percentis e o método da curva de Gauss.

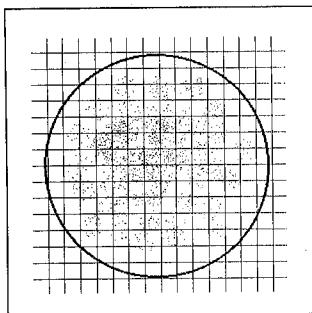
- (b) Qual dos dois métodos é o mais indicado neste caso. Justifique.

13. Para a população masculina nos Estados Unidos (1976-1980) com idade entre 18 e 74 anos, a pressão sistólica tem distribuição aproximadamente gaussiana com média 129 mmHg e desvio-padrão 19,8 mmHg. Recomendações do *Joint National Committee of Hypertension* e *American Heart Association*, 1993 consideram níveis pressóricos normais menores que 130/85 mmHg.

- (a) Qual a probabilidade de um homem dessa população possuir pressão sistólica normal?

- (b) Selecionando-se ao acaso 1000 homens dessa população, quantos seriam diagnosticados com hipertensão moderada (pressão sistólica entre 160 e 179 mmHg)?

14. A figura a seguir representa colônias de bactérias (pontos negros) que se desenvolveram em uma amostra de sangue colocada em uma placa de Petri dividida em pequenos quadrados em condições ideais.



- (a) Realizaram-se oito experimentos com oito diferentes tipos de bactérias. O número de quadrados em que foram observados exatamente  $x$  pontos negros (colônias) está apresentado na tabela a seguir. Comente os resultados encontrados em um dos experimentos.

Pontos negros	Experimento							
	A	B	C	D	E	F	G	H
0	5	26	59	83	8	7	3	60
1	19	40	86	134	16	11	7	80
2	26	38	49	135	18	11	14	45
3	26	17	30	101	15	11	21	16
4	21	7	20	40	9	7	20	9
5	13	0	0	16	7	8	19	0
6	8	0	0	7	0	0	7	0
7	0	0	0	0	0	0	9	0
Total								
quadrados	118	128	244	516	73	55	100	210
colônias	346	195	354	989	168	134	378	254

- (b) A média da distribuição de Poisson ( $\lambda$ ) para cada experimento é mostrada a seguir. Comente os resultados da

tabela e dê o significado do parâmetro do experimento A.

Experimento	Média	Experimento	Média
A	2,93	E	2,30
B	1,52	F	2,44
C	1,45	G	3,78
D	1,92	H	1,21

- (c) Podemos calcular probabilidades associadas ao modelo de Poisson para cada um dos parâmetros apresentados na tabela anterior e então obter os números esperados de quadrados contendo  $x$  colônias. Os resultados são apresentados na tabela a seguir. Complete a linha correspondente ao experimento A e faça uma avaliação da adequacidade do modelo, comparando os valores observados com os previstos pelo modelo.

Experi- mento	Pontos negros (colônias)							
	0	1	2	3	4	5	6	7
A								
B	28,0	42,5	32,4	16,4	6,2	1,9	0,5	0,1
C	57,2	82,9	60,2	29,1	10,5	3,1	0,8	0,1
D	75,6	145,2	139,4	89,2	42,8	16,4	5,3	1,4
E	7,3	16,8	19,4	14,8	8,5	3,9	1,5	0,5
F	4,8	11,7	14,3	11,6	7,1	3,6	1,4	0,5
G	2,3	8,6	16,3	20,5	19,4	14,7	9,2	5,0
H	62,6	75,8	45,8	18,5	5,6	1,3	0,3	0,0

15. A fosfatase alcalina em uma população de pessoas gozando de boa saúde tem distribuição gaussiana com média  $42 \text{ mU/dl}$  e desvio-padrão  $13 \text{ mU/dl}$ . Calcule a porcentagem de pessoas com fosfatase alcalina entre  $15$  e  $69 \text{ mU/dl}$ .
16. Explique porque, algumas vezes, é necessário usar o método dos percentis na determinação de um padrão de referência.
17. Sabe-se que a distribuição de pressão arterial na criança é aproximadamente gaussiana e os valores encontrados são bem menores que os de adultos. Os critérios usados na definição de normalidade são arbitrários e complexos. Já foram propostos

níveis limites como dois desvios-padrão acima da média, acima dos percentis 90 ou 95 ou limites fixos de 130-140 mmHg para pressão sistólica e 85-90 mmHg para a diastólica (Londe et al., 1971; Rance et al., 1974).

O relatório da *Task Force* de 1987 propôs que fossem consideradas *normais* as crianças com níveis de pressão sistólica e diastólica inferiores ao percentil 90 para a idade; *normais altos* os valores de pressão arterial entre os percentis 90 e 95 e *hipertensos* os valores persistentemente acima do percentil 95. Essas crianças hipertensas eram consideradas portadoras de hipertensão significativa quando os níveis pressóricos estivessem entre o 95º e 99º percentil, e hipertensão grave para os níveis persistentemente acima do percentil 99. Para uma população de crianças entre 10 e 12 anos, o percentil 95 da pressão sistólica corresponde a 126 mmHg e da diastólica a 82 mmHg. Já o percentil 99, 134 e 90 mmHg para pressão sistólica e diastólica, respectivamente.

- (a) Calcule a média e o desvio-padrão da distribuição da pressão sistólica em crianças de 10 a 12 anos.
- (b) Acima de que valor de pressão sistólica encontram-se 80% das crianças da população?
18. No estudo de Framingham relatado em Pagano et al. (1993), mediram-se os níveis séricos de colesterol em um grande número de homens saudáveis que foram acompanhados por dezenas de anos. Ao final do período dividiram-se os homens em dois grupos: aqueles que desenvolveram doença cardíaca coronariana (DCC) e aqueles que não apresentaram este problema. A distribuição dos níveis iniciais de colesterol sérico para cada grupo é aproximadamente gaussiana. Para os indivíduos que eventualmente desenvolveram DCC, tem média  $\mu = 244 \text{ mg/dl}$  e desvio-padrão  $\sigma = 51 \text{ mg/dl}$  e, para aqueles que não desenvolveram,  $\mu' = 219 \text{ mg/dl}$  e  $\sigma' = 41 \text{ mg/dl}$ .

- (a) Suponha que um nível sérico de colesterol de 260 mg/dl ou superior seja usado como prognóstico para desenvolvimento de DCC. Qual a probabilidade de se prever a doença em um homem que nunca a desenvolverá?

(b) Ainda para esse valor de referência, qual a probabilidade de fallha na predição de DCC para um homem que a desenvolverá?

- (c) O que aconteceria com os valores da sensibilidade e especificidade alterando-se o valor de referência para 250 mg/dl?
- (d) Nessa população, o uso dos níveis iniciais de colesterol sérico parecem adequados para previsão de DCC?

19. Para exemplo da Seção 5.7.2 sobre o rastreamento de doadores de sangue.

- (a) Calcule VPP, VPN, PFP e PFN para o novo ponto de corte (1,49) sugerido no final da seção. Compare os resultados com o primeiro ponto de corte considerado (1,45).
- (b) Repita os cálculos anteriores considerando a prevalência igual a 1%. Compare com a prevalência de 12 % usada no exemplo.

20. O teste de conjuntos mede a função mental. É um teste rápido e simples e requer que um indivíduo se recorde de itens de quatro categorias comuns. O teste foi aplicado em 65 voluntários idosos sem disfunção mental e os resultados apresentados na tabela a seguir. A média, desvio-padrão e mediana foram:  $\bar{x} = 30,63$ ,  $s = 7,67$  e  $\tilde{x} = 32$ .

4	11	12	15	15	16	20	21	22	22	23	24	24
25	26	26	26	26	27	27	28	28	29	29	29	30
30	30	30	31	32	32	32	32	33	33	34	34	35
35	35	36	36	37	37	37	37	37	37	37	38	38
39	39	39	39	39	39	39	40	40	40	40	40	40

- (a) Calcule o limite inferior de referência, que seja superado por aproximadamente 85% dos idosos testados.
- (b) Calcule e interprete a especificidade deste teste.
- (c) Neste caso podemos calcular a sensibilidade? Por quê?
- (d) Se calculássemos o valor de referência inferior baseado em 95%, ao invés de 85%, o que ocorreria com a especificidade? Por quê?

# Capítulo 6

## Comparando Dois Grupos

### 6.1 Introdução

Em todas as áreas do conhecimento humano há uma busca contínua e ininterrupta por novos métodos e novos procedimentos que melhorem em algum sentido aqueles já existentes. Na agricultura buscam-se variedades mais adequadas e mais produtivas de cereais e na indústria automobilística procuram-se motores de maior rendimento e de menor ruído. A pesquisa médica tem como objetivo por exemplo a obtenção de drogas mais seguras, com maior poder de cura e maior tolerabilidade.

Em todas essas situações é preciso comparar as técnicas usuais com os métodos alternativos. Esta comparação é, pois, uma questão importante que surge freqüentemente no trabalho de pesquisa e desenvolvimento. Não se trata, entretanto, de tarefa simples como o princípio pode parecer. É necessário coletar informações e fazer inferências a partir de evidências experimentais ou observacionais.

Uma questão importante que surge no trabalho de pesquisa na área médica é a comparação de drogas, de métodos cirúrgicos, de procedimentos de laboratórios, de dietas ou de tratamentos.

[Na comparação de um tratamento novo com um tratamento padrão pode-se levar em consideração, entre outros aspectos, o custo, a toxicidade ou ainda a facilidade de implementação.] O objetivo pode ser verificar a superioridade de um tratamento ou a equivalência entre eles.] Após ter-se estabelecido o critério para se considerar um tratamento mais eficiente do que o outro, a próxima etapa é a escolha

desse tratamento.

Consideremos o caso de duas terapias. Se todos os portadores de determinada doença se comportassem de maneira idêntica em relação aos tratamentos utilizados, bastaria examinar o comportamento de um ou no máximo dois deles, frente às alternativas existentes. A decisão sobre qual é a melhor seria óbvia, e portanto, nenhuma análise estatística seria necessária.

Tal, entretanto, não é o caso. A reação a um tratamento varia de indivíduo para indivíduo e, via de regra, não há um tratamento ótimo para todos. Como, em geral, não se conhece *a priori* a reação de cada indivíduo, prescreve-se o tratamento que apresenta *em média* os melhores resultados. Em outras palavras, a situação ideal da escolha do melhor tratamento para cada indivíduo não é possível na prática. Conseqüentemente, considera-se como o melhor tratamento aquele que produz bons resultados para a grande maioria da população em estudo.

Extrapolações dos resultados obtidos no grupo estudado para uma população de interesse exige cuidados especiais no planejamento tanto de ensaios clínicos quanto de estudos observacionais. Fatores que afetam a resposta sendo avaliada devem ser controlados, tanto na fase de coleta como na análise dos dados. Por exemplo, no estudo da pressão arterial, a idade, a raça e o sexo são fatores que devem ser levados em conta.

O procedimento para determinar qual dos dois tratamentos é em média o mais eficiente envolve geralmente a seleção de duas amostras e a comparação dos resultados obtidos. Neste capítulo discutimos como comparar os efeitos médios de dois tratamentos, quando sua resposta é medida para uma variável contínua ou dicotômica.

Para isto, o primeiro passo é a clara identificação dos grupos a serem comparados. Depois é necessário definir a variável resposta a ser usada como critério de comparação. Nos estudos caso-controle, os grupos são os doentes e os não doentes, e o objetivo pode ser, por exemplo, comparar a proporção daqueles que foram expostos ao fator de risco. Neste caso, a resposta de interesse (presença do fator) é dicotômica e existe um teste estatístico específico que deve ser usado como critério de decisão. Em estudos de coorte, os grupos são formados entre os que foram expostos ao fator de risco e os que não foram expostos. Se a resposta de interesse for uma variável quantitativa

(como por exemplo nível de colesterol e pressão sanguínea) também existe um teste apropriado para este caso. Finalmente, é preciso saber se os dados foram obtidos de forma independente ou pareada, pois os testes são específicos para cada tipo de planejamento.

Freqüentemente, para comparação de dois tratamentos observam-se características ou mede-se o valor da variável resposta de interesse em *pares* de indivíduos. No planejamento do estudo, cuida-se para que as características importantes dos indivíduos que integram o mesmo par sejam tão semelhantes quanto possível. O tratamento é administrado a um dos elementos do par e o outro é mantido como controle. A vantagem do procedimento é clara. Os indivíduos no par são semelhantes, exceto no que se refere ao tratamento recebido. Em algumas situações o par é constituído do mesmo indivíduo em duas ocasiões diferentes.

Em resumo, [o teste estatístico a ser usado depende do tipo de variável e do tipo de planejamento]. A seguir serão apresentados os testes mais comuns para quatro situações: [variável dicotômica (amostras independentes e pareadas) e variável contínua (amostras independentes e pareadas)].

## 6.2 Conceitos fundamentais

O procedimento estatístico denominado *Teste de Hipóteses* é usado amplamente nas áreas do conhecimento humano em que as variáveis envolvidas estão sujeitas a variabilidade. Neste texto, entretanto, estamos interessados no problema específico de comparação entre dois grupos. Assim, embora as idéias básicas desta seção sejam de aplicação geral, serão ilustradas apenas no contexto da comparação de dois tratamentos médicos.

### Exemplo 6.1: Eficácia do AZT

Fischl et al. (1987) publicaram o primeiro relato de um ensaio clínico que comprovou a eficácia de zidovudina (AZT) para prolongar a vida de pacientes com AIDS. Os dados centrais do trabalho estão na Tabela 6.1.

Tabela 6.1: Número de sobreviventes tratados com AZT ou placebo

Grupo	Situação		Total
	Vivo	Morto	
AZT	144	1	145
Placebo	121	16	137
Total	265	17	282

Esse experimento foi cercado de muitos cuidados e, embora a análise estatística dos dados seja fundamental, a decisão final de liberação do AZT foi tomada levando-se em consideração muitos outros resultados fornecidos pelo estudo, como aqueles referentes a efeitos colaterais.

A análise dos dados da Tabela 6.1 consiste basicamente na comparação de duas proporções. Entre os pacientes que receberam o AZT, a proporção dos que estavam vivos depois de 24 semanas de tratamento foi de  $\frac{144}{145} = 0,993$ , enquanto a mesma proporção para o grupo placebo foi de  $\frac{121}{137} = 0,883$ .

Como a alocação dos pacientes aos grupos foi feita de forma aleatória, a diferença entre essas duas proporções parece indicar que em pacientes com AIDS o AZT tem o efeito de prolongar a vida. Antes de se aceitar esta conclusão, entretanto, é preciso afastar o acaso como explicação alternativa. Ou seja, deve-se responder à pergunta: será que este resultado ocorreu por mero acaso ou por ser o AZT de fato uma droga efetiva?

### 6.2.1 Hipóteses a serem testadas

Geralmente podemos formular os problemas, como o do AZT, através das hipóteses seguintes.

#### Hipótese nula

No problema de comparação de dois tratamentos é usual fixar como hipótese de interesse a inexistência de diferença entre os dois tratamentos comparados. Como freqüentemente a comparação é feita entre um tratamento padrão e um tratamento novo, esta opção implica colocar o ônus da prova de efetividade no tratamento novo,

uma opção conservadora mas prudente. Por estas razões, a hipótese a ser testada é usualmente chamada hipótese nula ( $H_0$ ), nome que se generalizou mesmo para situações em que o problema não é mais de comparação entre dois tratamentos.

### Hipótese alternativa

A hipótese nula deve ser comparada com uma hipótese alternativa, denominada  $H_1$ . Para cada situação existem muitas hipóteses alternativas adequadas. Neste texto, entretanto, seguindo convenção estabelecida pelos editores de revistas científicas na área médica, a hipótese alternativa será sempre a inexistência de igualdade entre os tratamentos.

No Exemplo 6.1 a hipótese nula é

$$H_0: p_C = p_T \quad (6.1)$$

onde  $p_C$  e  $p_T$  são respectivamente as probabilidades de se observar a resposta de interesse entre os controles e entre os pacientes do grupo tratamento.

De acordo com a opção feita acima, a hipótese alternativa será

$$H_1: p_C \neq p_T \quad (6.2)$$

### 6.2.2 Critério de decisão

Decididas as hipóteses a serem testadas, o próximo passo é construir um critério baseado no qual a hipótese  $H_0$  será julgada. Embora qualquer critério possa em princípio ser considerado, a Teoria Estatística já desenvolveu para um grande número de situações testes que se revelaram as melhores opções e que são apresentados com detalhes em livros textos.

O critério de decisão é baseado na *estatística de teste*. De uma forma bem genérica e intuitiva podemos dizer que a estatística do teste mede a discrepância entre o que foi observado na amostra e o que seria esperado se a hipótese nula fosse verdadeira. Uma grande distância medida pela distribuição de probabilidade é indicação de que  $H_0$  não é verdadeira, devendo portanto ser rejeitada.

Resumindo, rejeita-se a hipótese nula se o valor da estatística do teste é “grande”. Esse valor deve portanto ser comparado a alguma distribuição de probabilidade que depende de cada caso, como veremos nas seções a seguir.

### 6.2.3 Erros do tipo I e II, nível de significância e poder do teste

No exemplo do AZT havia a possibilidade de se rejeitar a hipótese de igualdade entre o AZT e o placebo, mesmo se de fato eles fossem iguais.

Em termos técnicos, a decisão de rejeitar  $H_0$  quando de fato  $H_0$  é verdadeira é chamada de *erro do tipo I*. Para evitá-lo, escolhemos um critério de decisão (que corresponde a um percentil de uma distribuição de probabilidade) que torna este erro pouco provável. Na literatura, a probabilidade de cometer este erro recebe o nome de *nível de significância* do teste, sendo usualmente representado pela letra grega  $\alpha$  (lê-se alfa).

Há, no entanto, um segundo tipo de erro. No exemplo do AZT ele consiste em não rejeitar a hipótese de igualdade entre o AZT e o placebo quando de fato estes dois tratamentos são diferentes. Isto implicaria na não liberação do novo tratamento, cujo efeito real não estaria sendo percebido.

Para um tamanho fixo da amostra, não há como controlar simultaneamente ambos os erros. Convencionou-se que o erro mais sério seria o erro do tipo I. Em um segundo momento, calcula-se o tamanho da amostra que reduza a probabilidade do erro do tipo II, usualmente representado pela letra grega  $\beta$  (lê-se beta), a níveis aceitáveis.

A Tabela 6.2 sintetiza os erros possíveis associados a cada decisão tomada em um teste de hipóteses.

Tabela 6.2: Erros possíveis associados a teste de hipóteses

Conclusão do teste	Situação real	
	$H_0$ verdadeira	$H_0$ falsa
Não rejeitar $H_0$	decisão correta	erro tipo II
Rejeitar $H_0$	erro tipo I	decisão correta

A capacidade de um teste identificar diferenças que realmente existem, ou seja, de rejeitar  $H_0$  quando é realmente falsa, é denominada *poder do teste* e é definida como  $1 - \beta$ .

#### 6.2.4 Probabilidade de significância (valor-p)

Existem duas opções para expressar a conclusão final de um teste de hipóteses. A primeira consiste em comparar o valor da estatística de teste com o valor obtido a partir da distribuição teórica, específica para o teste, para um valor prefixado do nível de significância (por exemplo 5% ou 1%).

Na segunda abordagem, atualmente mais usada, o interesse é quantificar a chance do que foi observado ou resultados mais extremos, sob a hipótese de igualdade dos grupos. Assim, essa opção, baseia-se na probabilidade de ocorrência de valores iguais ou superiores ao assumido pela estatística de teste, sob a hipótese de que  $H_0$  é verdadeira.

Este número é chamado de *probabilidade de significância* ou *valor-p* e freqüentemente é indicado apenas por  $p$ . Como o valor-p é calculado supondo-se que  $H_0$  é verdadeira, pode-se fazer duas conjecturas quando se obtém um valor muito pequeno. Um evento que é extremamente raro pode ter ocorrido ou a hipótese  $H_0$  não deve ser verdadeira, isto é, a conjectura inicial e conservadora não parece plausível.

Portanto, quanto menor o valor-p maior a evidência para se rejeitar  $H_0$ . De modo geral, na área médica, considera-se que valor-p menor ou igual a 0,05 indica que há diferenças significativas entre os grupos.

Para cada um dos testes das seções a seguir vamos mostrar como calcular e interpretar o valor-p. Por exemplo, baseado no valor-p pode-se avaliar se o fator sob investigação é um possível fator de risco ou de proteção.

Todas as idéias aqui apresentadas ficam mais claras após a discussão de situações específicas. Apresentamos também comentários adicionais sobre valor-p na Seção 6.9.

### 6.3 Resposta dicotômica: amostras independentes

Comparar dois grupos através do resultado observado em uma variável dicotômica é um problema comum na pesquisa médica, aparecendo com freqüência em todos os tipos de estudos clínicos.

A variável de interesse é a ocorrência de um evento, como o desenvolvimento de uma doença, ou a presença de certo atributo, por exemplo, albinismo. O problema de comparação das probabilidades de ocorrência do evento ou do atributo nos dois grupos ( $p_1$  e  $p_2$ ) é formulado através das hipóteses

$$H_0: p_1 = p_2 \text{ versus } H_1: p_1 \neq p_2 \quad (6.3)$$

#### 6.3.1 Teste qui-quadrado ( $\chi^2$ )

A Tabela 6.3 apresenta dados genéricos de uma situação envolvendo a comparação de dois grupos e que a resposta de interesse é dicotômica: a ocorrência ou não de um evento.

Tabela 6.3: Distribuição quanto à ocorrência de um evento

Grupo	Ocorrência do evento		Total
	Sim	Não	
I	$a$	$b$	$a + b = n_1$
II	$c$	$d$	$c + d = n_2$
Total	$m_1 = a + c$	$m_2 = b + d$	$n_1 + n_2 = N$

Se não há diferença entre as proporções de ocorrência do evento nos dois grupos, então

$$\frac{a}{n_1} = \frac{c}{n_2} = \frac{a+c}{m_1} = \frac{m_1}{N} \quad (6.4)$$

A partir dessas igualdades podemos escrever

$$a = \frac{m_1 \times n_1}{N} \quad b = \frac{m_2 \times n_1}{N} \quad c = \frac{m_1 \times n_2}{N} \quad d = \frac{m_2 \times n_2}{N} \quad (6.5)$$

Temos, portanto, dois conjuntos de valores: os *observados* ( $O_i$ ) que são denotados por  $a, b, c$  e  $d$  e os *esperados* ( $E_i$ ) calculados sob a hipótese de igualdade das proporções de sucesso entre os grupos e dados pelas expressões anteriores.

Se as proporções de sobrevida são iguais nos dois grupos, a discrepância entre os dois conjuntos de números acima não deve ser grande. Pearson, importante estatístico do início do século, propôs medir a discrepancia entre os valores observados e esperados das quatro entradas de uma tabela  $2 \times 2$  através da expressão

$$X^2 = \sum_{i=1}^4 \frac{(O_i - E_i)^2}{E_i} \quad (6.6)$$

É possível mostrar que o valor do  $X^2$  obtido através de 6.6 pode ser calculado de maneira fácil e algebricamente semelhante usando-se as entradas da tabela  $2 \times 2$  através da expressão:

$$X^2 = \frac{N(ad - bc)^2}{m_1 m_2 n_1 n_2} \quad (6.7)$$

### Exemplo 6.2: Eficácia do AZT (continuação)

Supondo que o AZT e o placebo são equivalentes, teríamos os valores esperados mostrados na Tabela 6.4.

Tabela 6.4: Valores esperados de sobreviventes sob a hipótese de equivalência entre AZT e placebo

Grupo	Situação		
	Vivo	Morto	Total
AZT	136,26	8,74	145
Placebo	128,74	8,26	137
Total	265	17	282

Os cálculos necessários para a obtenção do valor da discrepancia proposta por Pearson são mostrados na Tabela 6.5.

Tabela 6.5: Cálculos necessários para a construção do teste  $\chi^2$

$i$	$O_i$	$E_i$	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
1	144	136,26	7,74	59,91	0,44
2	121	128,74	-7,74	59,91	0,47
3	1	8,74	7,74	59,91	6,85
4	16	8,26	7,74	59,91	7,25
Total	282	282	0	239,64	15,01

Portanto,  $X^2 = 0,44 + 0,47 + 6,85 + 7,25 = 15,01$ .

Calculado o valor da expressão  $X^2$ , é preciso decidir se este é ou não um valor “grande”. Devido ao efeito amostral, mesmo se na realidade as proporções fossem exatamente iguais, é possível obter qualquer valor para  $X^2$ . Entretanto, valendo a hipótese de igualdade entre as duas proporções, valores “grandes” ocorrerão pouco frequentemente.

Assim, para se tomar uma decisão sobre a igualdade ou não das duas proporções, é preciso conhecer o comportamento, isto é, a distribuição estatística dos valores de  $X^2$  quando as proporções são iguais. Esta distribuição foi obtida e recebeu o nome de qui-quadrado com 1 grau de liberdade, é indicada por  $\chi_1^2$  e está sintetizada em tabelas de fácil utilização (Tabela A4 do apêndice).

O valor da estatística de teste foi de 15,01. Como este valor é maior que 3,84, valor obtido da distribuição do  $\chi_1^2$  para um nível de significância de 0,05, rejeitamos a hipótese de igualdade entre os grupos de tratamento e controle. Em outras palavras, decidimos com 95% de certeza que há evidência de efeito do AZT.

Para obtermos a probabilidade de significância devemos calcular a probabilidade de encontrar valores maiores que 15,01, isto é,  $\Pr[\chi_1^2 \geq 15,01]$ , sendo verdadeira a hipótese de igualdade das proporções. Da tabela da distribuição do  $\chi_1^2$ , vemos que este valor é aproximadamente 0,0001, ou seja o valor-p é 0,0001.

Baseado nesse estudo podemos dizer com grande certeza que o AZT tem efeito de prolongar a vida de pacientes com AIDS, primeira evidência necessária para a liberação do medicamento.

### Exemplo 6.3: Fatores de risco para AVC

Welin et al. (1987) relataram os resultados de um estudo realizado com o objetivo de determinar os fatores de risco para acidente vascular cerebral (AVC) em homens de meia idade. Todos os homens nascidos em Gotemburgo, em 1913, em dias múltiplos de três, num total de 973 homens, foram convidados a participar do estudo. Destes, 789 aceitaram participar. No período de duração do estudo (de 1967 a 1986), 57 homens sofreram AVC, enquanto 732 não foram acometidos.

A Tabela 6.6, reproduzida do artigo original, apresenta as proporções de presença de alguns fatores de risco nos indivíduos que sofreram AVC e naqueles que não foram acometidos pela doença. Os valores-p apresentados na tabela têm interpretação conforme descrito anteriormente. Por exemplo, a ocorrência de AVC na mãe como causa principal de morte pode ser considerada um possível fator de risco para o aparecimento de AVC no filho, uma vez que a diferença observada foi altamente significativa (valor-p = 0,0005).

Tabela 6.6: Percentuais e valor-p para a comparação do grupo que sofreu com o que não sofreu AVC

Fator	AVC			Valor-p
	Sim	Não		
AVC como causa principal da morte				
na mãe	29,8	11,2		0,0005
no pai	7,0	7,5		0,72
AVC como causa principal ou contribuindo para a causa da morte				
na mãe	29,8	14,2		0,002
no pai	7,0	9,2		0,59
Doença coronariana	7,0	6,3		0,83
Sinais eletrocardiográficos				
hipertrofia do ventrículo esquerdo	3,5	1,0		0,08
doença coronariana	10,5	6,5		0,23

### 6.3.2 Teste qui-quadrado com correção de continuidade

Yates (1934) demonstrou que, se usarmos

$$\chi^2_c = \frac{N(|ad - bc| - \frac{N}{2})^2}{m_1 m_2 n_1 n_2} \quad (6.8)$$

ao invés da expressão (6.7), obtemos um valor mais apropriado a ser comparado com a distribuição do  $\chi^2_1$  e é conhecida como qui quadrado com a *correção de continuidade ou correção de Yates*. A justificativa é que a distribuição de frequências observadas, que é discreta, está sendo aproximada pela distribuição qui-quadrado, que é contínua. Note a semelhança das duas expressões: a diferença é apenas o fator de correção de continuidade. Recomendamos que o teste qui-quadrado seja sempre feito usando-se essa expressão.

### Exemplo 6.4: Contraceptivos orais e infarto do miocárdio

Em um estudo sobre a associação entre o uso corrente de contraceptivos e o infarto de miocárdio, Shapiro et al. (1979) observaram os resultados entre pacientes com idade entre 30 e 34 anos mostrados na Tabela 6.7.

Tabela 6.7: Distribuição de uso de contraceptivo oral segundo grupo que sofreu ou não infarto do miocárdio

Grupo	Uso recente			Total
	Sim	Não		
Casos	9	12		21
Controles	33	390		423
Total	42	402		444

Entre os casos, ou seja, entre as pacientes que tiveram um infarto do miocárdio, a proporção de uso recente de contraceptivos é  $\frac{9}{21} = 0,43$  e entre os controles,  $\frac{33}{423} = 0,08$ . A diferença entre estas duas proporções (0,35) parece indicar que o uso de contraceptivos orais é mais frequente entre os casos. Mas será que este resultado não ocorreu por mero acaso?

Usando a notação da Tabela 6.3,  $a = 9$ ,  $b = 12$ ,  $c = 33$ ,  $d = 390$  e  $N = 444$ . A expressão do  $X^2$  com correção de continuidade nos dá 24,76. A tabela do  $\chi^2$  nos informa que diferenças iguais ou maiores que a observada só ocorrem caso não haja associação entre infarto e uso recente de contraceptivo com freqüência menor que 0,001, ou seja,  $p < 0,001$ . Podemos, portanto, com alto grau de certeza afirmar que existe associação entre o uso de contraceptivos orais e infarto do miocárdio para pacientes entre 30 e 34 anos.

### 6.3.3 Teste exato de Fisher

Há uma dificuldade técnica na aplicação do teste qui-quadrado, com ou sem correção de continuidade, quando o valor esperado em alguma casela na tabela  $2 \times 2$  é menor que 5. Neste caso, o uso da distribuição  $\chi^2$  não é mais completamente apropriado. Ou seja, o grau de certeza na decisão tomada não é exatamente aquele fornecido pela distribuição  $\chi^2$ .

O uso da correção de Yates não remove a exigência de tamanho adequado das freqüências esperadas. A alternativa é usar o teste exato de Fisher, que é a versão exata do teste do qui-quadrado. A maioria dos softwares de análise estatística de dados fornece como opção o teste de Fisher. Não discutiremos aqui, entretanto, as etapas de sua construção. Embora simples, são muito técnicas para o nível pretendido neste texto. Entretanto, apresentaremos o resultado do teste aplicado a uma situação em que seu uso é apropriado.

### Exemplo 6.5: Tratamento de pneumonia bacteriana

Pasternak et al. (1992) avaliaram a eficácia e segurança de dois antibióticos no tratamento de pneumonia bacteriana de origem comunitária em adultos. Foram estudados 63 pacientes, sendo 32 doentes tratados com cefadroxil e 31 com cefalexina. A avaliação da resposta terapêutica foi baseada na evolução do quadro clínico e do exame radiológico do tórax feito na admissão ao estudo e no 10º dia de tratamento.

Dos pacientes avaliados, a cura completa ocorreu em 31 dos 32 (96,9%) pacientes do grupo cefadroxil e em 28 dos 31 (90,3%) pacientes do grupo cefalexina.

Foram observados efeitos adversos em quatro casos: um (3,1%) no grupo que recebeu cefadroxil, com ocorrência de náuseas que desapareceram com o tratamento sintomático instituído; todos os três casos (9,7%) do grupo cefalexina tiveram diarréia, dois se recuperaram espontaneamente e um teve que interromper o tratamento.

Para comparar os percentuais de cura e de ocorrência de efeitos adversos, por se tratar de amostras pequenas, utilizou-se o teste exato de Fisher. Os valores-p são respectivamente  $p = 0,5886$  e  $p = 0,5898$ . Não há portanto evidência de diferença entre os dois tratamentos tanto em termos de cura como na ocorrência de efeitos adversos. Os achados deste trabalho estão resumidos na Tabela 6.8.

Tabela 6.8: Percentuais de curas, de efeitos colaterais e valor-p na comparação de cefadroxil e cefalexina

Quadro	Cefadroxil	Cefalexina	Valor-p
Cura	96,9%	90,3%	0,5886
Efeitos adversos	3,1%	9,7%	0,5898

### 6.3.4 Teste Z para comparação de proporções

Denotamos os dois resultados possíveis da variável dicotômica por sucesso e fracasso. Sejam  $p_1$  e  $p_2$  as proporções de sucesso referentes aos tratamentos a serem comparados, que são estimadas por  $\hat{p}_1$  e  $\hat{p}_2$ , as proporções amostrais baseadas em amostras de tamanhos  $n_1$  e  $n_2$ , respectivamente. Queremos testar as hipóteses:  $H_0: p_1 = p_2$  versus  $H_1: p_1 \neq p_2$ , isto é, estamos investigando a equivalência entre os dois tratamentos.

Apresentaremos uma alternativa para o teste qui-quadrado para a comparação de proporções. Trata-se de um teste aproximado que requer amostras grandes para sua aplicação. Um critério é exigir que  $n_1\hat{p}_1$  e  $n_2\hat{p}_2$  excedam o valor 5.

O teste é baseado em

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} \quad (6.9)$$

Alternativamente, pode-se primeiramente estimar a probabilidade de sucesso nas duas amostras combinadas, denotada por  $\hat{p}$  e calculada como  $\hat{p} = \frac{m_1 + m_2}{n_1 + n_2}$ , onde  $m_1$  e  $m_2$  são respectivamente o número de sucessos de cada amostra. Neste caso, calculamos

$$Z^* = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}} \quad (6.10)$$

O valor  $Z$  (ou  $Z^*$ ) deve ser comparado ao percentil da distribuição gaussiana padrão e rejeita-se  $H_0$  se  $|Z| > z_{\alpha/2}$  (ou  $|Z^*| > z_{\alpha/2}$ ), onde  $z_{\alpha/2}$  é o percentil de ordem  $(1 - \alpha/2)$  da distribuição  $N(0, 1)$ .

Pode-se mostrar que  $Z^2 = X^2$ , isto é, se o valor da estatística (6.9) for elevado ao quadrado obteremos o valor da estatística (6.7). Isto mostra a equivalência dos dois testes para amostras grandes.

### Exemplo 6.6: Comparação de drogas contra náusea

Com o objetivo de comparar a eficácia de dois preventivos contra náusea, dividiu-se aleatoriamente uma amostra de 400 marinheiros em dois grupos de 200. Um grupo recebeu a pílula A e o outro a pílula B, sendo que no primeiro grupo 152 não enjoaram durante uma tempestade e no outro grupo apenas 132. Há indicações de que a eficácia das pílulas A e B é a mesma?

Sejam  $p_A$  e  $p_B$  as proporções de marinheiros que não enjoam, respectivamente para as pílulas A e B. Temos que  $n_A = 200$ ,  $n_B = 200$ ,  $\hat{p}_A = \frac{152}{200} = 0,76$ ,  $\hat{p}_B = \frac{132}{200} = 0,66$ ,  $\hat{p} = \frac{152+132}{200+200} = 0,71$ . Como mostrado abaixo, os valores das duas estatísticas ( $Z$  e  $Z^*$ ) são quase iguais.

$$Z = \frac{0,76 - 0,66}{\sqrt{\frac{0,76(1-0,76)}{200} + \frac{0,66(1-0,66)}{200}}} = \frac{0,10}{0,05} = 2,22$$

$$Z^* = \frac{0,76 - 0,66}{\sqrt{0,71(1-0,71)(\frac{1}{200} + \frac{1}{200})}} = \frac{0,10}{0,05} = 2,20$$

Fixando-se o nível de significância em 5%, rejeita-se  $H_0$ . Os valores-p encontrados para  $Z$  e  $Z^*$  são respectivamente 0,026 e 0,028. Portanto, pode-se concluir que as duas pílulas não são igualmente eficazes. Há indicações de que a pílula A oferece maior proteção contra

náusea comparada à pílula B. O próximo passo é quantificar a diferença dos efeitos das duas pílulas através da construção de um intervalo de confiança, assunto tratado no próximo capítulo.

### 6.4 Resposta dicotômica: amostras pareadas

O exemplo a seguir ilustra a necessidade de desenvolvimento de um teste específico para a situação de dados pareados em que a resposta é dicotômica.

Suponhamos que dois patologistas examinaram, separadamente, o material de 100 tumores e os classificaram como benignos ou malignos. A questão de interesse é saber se os patologistas diferem nos seus critérios de decisão.

Neste caso, a forma adequada de apresentação dos dados é mostrada na Tabela 6.9.

Tabela 6.9: Classificação de dois patologistas (A e B) quanto à malignidade de tumores

Diagnóstico de B	Diagnóstico de A		Total
	Malignos	Benignos	
Malignos	9	1	10
Benignos	9	81	90
Total	18	82	100

É importante observar que a unidade de análise aqui é o *tumor*, avaliado por dois patologistas. Embora tenham sido feitas 200 análises, o total de tumores é, na realidade, apenas 100.

Além disto, alguns tumores serão claramente mais malignos do que outros e, portanto, a hipótese fundamental na construção do teste de probabilidade constante de malignidade não é razoável aqui. Isto explica a necessidade de desenvolvimento de teste específico, isto é, para dados pareados.

Apresentamos a seguir o teste de McNemar, que aparece com mais freqüência na literatura, embora existam alternativas mais interessantes.

#### 6.4.1 Teste de McNemar

Os dados a serem analisados no processo de comparação podem ser resumidos no formato da Tabela 6.10. Usamos a nomenclatura sucesso e fracasso para a ocorrência ou não do evento de interesse.

Tabela 6.10: Apresentação de dados obtidos em uma classificação de dados pareados

Controle	Tratamento		Total
	Sucesso	Fracasso	
Sucesso	$k$	$r$	$n_1$
Fracasso	$s$	$l$	$n_2$
Total	$n_1$	$n_2$	$N$

Se  $p_1$  e  $p_2$  são as probabilidades de sucesso nos grupos controle e tratamento, respectivamente, a hipótese de interesse é

$$H_0: p_1 = p_2 \quad \text{versus} \quad H_1: p_1 \neq p_2 \quad (6.11)$$

Os pares que produziram sucesso ou fracasso nos dois elementos do par não contêm informação para discriminar  $p_1$  de  $p_2$ . Se  $H_0$  é verdadeira, ou seja, os dois grupos são equivalentes, as discordâncias observadas são fruto do acaso. Noutras palavras, se  $r$  e  $s$  têm valores semelhantes, sob  $H_0$  espera-se a metade do número de discordâncias ( $\frac{r+s}{2}$ ). A hipótese  $H_0$  deve, portanto, ser rejeitada se a distância entre os valores observados e esperados for grande, de forma análoga ao que foi visto para amostras independentes.

Usando a correção de continuidade, a estatística do teste é

$$\chi^2_{McN} = \frac{(|r - \frac{r+s}{2}| - \frac{1}{2})^2}{\frac{r+s}{2}} + \frac{(|s - \frac{r+s}{2}| - \frac{1}{2})^2}{\frac{r+s}{2}} = \frac{(|r-s| - 1)^2}{r+s} \quad (6.12)$$

Finalmente, o teste consiste em se rejeitar  $H_0$  quando

$$\chi^2_{McN} = \frac{(|r-s| - 1)^2}{r+s} > \chi^2_{1,1-\alpha} \quad (6.13)$$

onde  $\chi^2_{1,1-\alpha}$  é o percentil de ordem  $1-\alpha$  da distribuição qui-quadrado com 1 grau de liberdade.

#### Exemplo 6.7: Amigdalectomia e doença de Hodgkin

Johnson & Johnson (1972), ao analisarem retrospectivamente a história clínica de pacientes de doença de Hodgkin, um tipo de câncer no tecido linfóide, não encontraram evidências que sustentassem a hipótese de que a amigdalectomia aumenta a suscetibilidade à doença, pela remoção da “barreira linfática” representada pela amígdala. Neste estudo, usaram um planejamento do tipo caso-controle pareado. A conclusão relatada no artigo foi feita, entretanto, com uma análise apropriada para dados provenientes de grupos independentes. Logo após a publicação, vários autores observaram a impropriedade da análise e, utilizando as informações no artigo, fizeram a análise correta, que é baseada nos dados da Tabela 6.11.

Tabela 6.11: Distribuição de pacientes com e sem doença de Hodgkin em um estudo caso-controle pareado segundo à amigdalectomia

Doença de Hodgkin	Controle		
	Operados	Não operados	Total
Operados	26	15	41
Não operados	7	37	44
Total	33	52	85

O valor da estatística do teste de McNemar é

$$\chi^2_{McN} = \frac{(|15 - 7| - 1)^2}{15 + 7} = 2,23$$

Este valor deve ser comparado com 3,84 para um nível de significância de 5%. Ou seja, com uma confiança de 95% acreditamos que não há associação entre a doença de Hodgkin e a amigdalectomia. Esta conclusão não está de acordo com trabalhos anteriores, como o de Vianna et al. (1971).

## 6.5 Resposta contínua: amostras independentes

Nesta seção apresentamos a metodologia para comparar dois grupos de pacientes (por exemplo, docentes versus não docentes) em relação a uma resposta contínua, por exemplo, pressão sistólica. Testase, neste caso, a igualdade das médias das respostas de dois tratamentos.

Sejam  $\mu_1$  e  $\mu_2$  as médias da variável estudada para os dois grupos, respectivamente. As hipóteses a serem testadas são

$$H_0: \mu_1 = \mu_2 \quad \text{versus} \quad H_1: \mu_1 \neq \mu_2 \quad (6.14)$$

### 6.5.1 Teste t

Vamos apresentar aqui o teste mais conhecido, o teste  $t$  para duas amostras, adequado para situações em que as respostas aos dois tratamentos são variáveis quantitativas com distribuição gaussiana com parâmetros  $(\mu_1, \sigma)$  e  $(\mu_2, \sigma)$ , respectivamente. Assim, as suposições para se usar este teste são que as variáveis estudadas têm distribuições gaussianas com o mesmo desvio-padrão. Note que  $\mu_1$ ,  $\mu_2$  e  $\sigma$  são parâmetros populacionais e, portanto constantes desconhecidas.

Para testar a hipótese 6.14, coletamos um amostra de tamanho  $n_1$  no grupo 1 ( $x_{11}, x_{12}, \dots, x_{1n_1}$ ) e uma amostra de tamanho  $n_2$  no grupo 2 ( $x_{21}, x_{22}, \dots, x_{2n_2}$ ). A partir desses dados, calculamos as médias ( $\bar{x}_1$  e  $\bar{x}_2$ ) e os desvios-padrão ( $s_1$  e  $s_2$ ) dos dois grupos.

O critério de decisão para se testar a hipótese nula acima consiste em rejeitar  $H_0$  se

$$T = \frac{\bar{X}_1 - \bar{X}_2}{DP(\bar{X}_1 - \bar{X}_2)} \quad (6.15)$$

é “grande”], sendo  $\bar{X}_1$  e  $\bar{X}_2$  as médias das respostas dos indivíduos do primeiro e do segundo grupos e o denominador é o desvio-padrão da diferença destas medidas.

A decisão baseada nesta estatística de teste é intuitiva, pois quanto maior for a diferença entre as médias amostrais, maior a chance de estarmos diante de dois grupos realmente diferentes.] Isto é captado

pelo numerador da estatística  $T$ . Por outro lado, quanto maior for a variabilidade das respostas, maior será a dificuldade de se detectar diferenças entre os efeitos médios. O desvio-padrão, no denominador da expressão acima, reflete esse fato, fornecendo a unidade na qual as diferenças observadas devem ser medidas. Em outras palavras, com a estatística  $T$  estamos medindo a diferença entre as médias em termos de desvios-padrão.

Devemos rejeitar  $H_0$  se  $T$  for “grande” em valor absoluto. Para tomarmos esta decisão, usamos a distribuição t de Student (Student, 1908). Essa distribuição é caracterizada por um parâmetro que assume apenas valores inteiros positivos que são chamados de graus de liberdade. É muito parecida com a distribuição gaussiana e sua forma típica é mostrada na Figura 6.1.

Os percentis da distribuição t de Student para vários graus de liberdade são mostrados na Tabela A5 do apêndice. Para graus de liberdade muito grandes, seus percentis são praticamente iguais aos da distribuição gaussiana.

[Na comparação de médias de amostras independentes, os graus de liberdade são  $n_1 + n_2 - 2$ , o número total da amostra menos 2.]

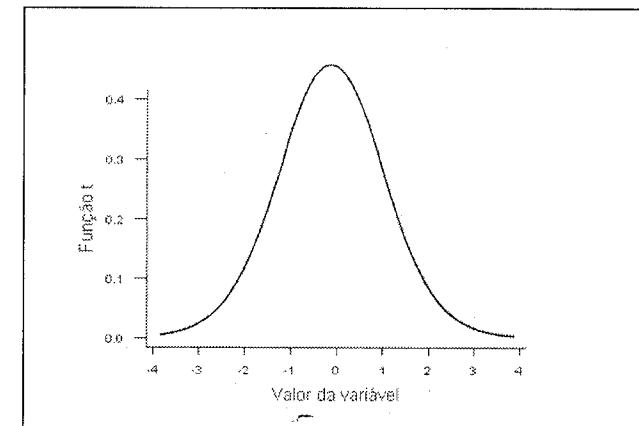


Figura 6.1: Distribuição t de Student

Para efetuarmos os cálculos necessários ao teste, é preciso conhecer a expressão para a variância de  $\bar{X}_1 - \bar{X}_2$ . Para isto usamos o resultado demonstrado pela Estatística Matemática

$$Var(\bar{X}_1 - \bar{X}_2) = \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} = \sigma^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right) \quad (6.16)$$

A partir da primeira amostra estimamos  $\sigma^2$  por

$$s_1^2 = \frac{\sum_{i=1}^{n_1} (x_{1i} - \bar{x}_1)^2}{n_1 - 1} \quad (6.17)$$

e a partir da segunda por

$$s_2^2 = \frac{\sum_{i=1}^{n_2} (x_{2i} - \bar{x}_2)^2}{n_2 - 1} \quad (6.18)$$

Como a suposição é que as variâncias dos dois grupos são iguais, podemos usar tanto a expressão (6.17) ou (6.18) para estimar  $\sigma^2$ . Entretanto, pode-se mostrar que a melhor estimativa para  $\sigma^2$  é a *média ponderada*, com pesos proporcionais aos tamanhos das amostras, ou seja,

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \quad (6.19)$$

Finalmente, a estimativa do desvio-padrão de  $\bar{X} - \bar{Y}$  é dada por

$$\sqrt{s_p^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} \quad (6.20)$$

O teste  $t$  para comparação de duas amostras consiste em se rejeitar  $H_0$  em favor de  $H_1$  ao nível  $\alpha$  de significância, se

$$\left| \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} \right| > t_{n_1+n_2-2;1-\frac{\alpha}{2}} \quad (6.21)$$

onde  $t_{n_1+n_2-2;1-\frac{\alpha}{2}}$  é o percentil de ordem  $1 - \frac{\alpha}{2}$  da distribuição  $t$  com  $n_1 + n_2 - 2$  graus de liberdade e é facilmente obtido consultando-se tabelas, como a Tabela A5 do apêndice.

Embora o valor de  $t_{n_1+n_2-2;1-\frac{\alpha}{2}}$  varie para cada situação, uma consulta às tabelas da distribuição  $t$  mostra que se  $\alpha = 0,05$  então

$t_{n_1+n_2-2;1-\frac{\alpha}{2}} \geq 2,00$ . Ou seja, se a razão  $T$ , em valor absoluto, for maior que 2 estamos diante de uma diferença significativa ao nível de 5%. No entanto, deve-se preferir os valores exatos em trabalhos científicos. Com as facilidades de cálculo trazidas pelos programas de computador especializados em cálculos estatísticos, é cada vez mais frequente o uso do valor- $p$ , ao invés de se comparar o valor de  $T$  com um valor tabelado.

### Exemplo 6.8: Níveis séricos de frutosamina

A frutosamina é um índice do controle metabólico no diabetes mellitus podendo refletir as variações da glicemia nas últimas duas a três semanas. Representa um conjunto de proteínas glicosadas, cuja fração principal é a albumina. Com o objetivo de estabelecer os valores normais da frutosamina em homens, mulheres gestantes ou não, Camargo et al. (1994) realizaram um estudo que incluía um grupo composto de 42 indivíduos normais, 21 mulheres (22 a 71 anos) e 21 homens (24 a 81 anos). O outro grupo era constituído de 36 gestantes (17 a 37 anos) atendidas no ambulatório de Ginecologia e Obstetrícia do Hospital de Clínicas de Porto Alegre. A idade gestacional variou de 17 a 37 semanas.

Como recomendado, os autores apresentam as médias e desvios-padrão de todas as variáveis consideradas para cada grupo (tabelas 6.12 e 6.13), de forma que é possível repetir os cálculos de cada valor- $p$  apresentado.

Tabela 6.12: Média ( $\bar{x}$ ) e desvio-padrão ( $s$ ) de variáveis para mulheres não gestantes e homens

Variável	Mulheres		Homens	
	$\bar{x}_1$	$s_1$	$\bar{x}_2$	$s_2$
Idade (anos)	39	14	43	14
Glicose (mg/dl)	95	20	84	24
Frutosamina (mmol/l)	2,70	0,26	2,99	0,32
Triglicerídos (mg/dl)	145	135	124	62

Tabela 6.13: Média ( $\bar{x}$ ) e desvio-padrão ( $s$ ) de variáveis para gestantes

Variável	$\bar{x}$	$s$
Idade (anos)	25	7
Idade gestacional (semanas)	28	5
Glicose (mg/dl)	103	17
Frutosamina (mmol/l)	2,40	0,22
Albumina (mg/dl)	4,1	0,6

Apresentamos a seguir alguns dos resultados obtidos neste estudo em que o teste t de amostras independentes foi aplicado.

As médias das idades dos grupos de mulheres e homens não apresentaram diferença estatisticamente significante ( $p = 0,30$ ). A glicemia de jejum e a de 2 horas após a sobrecarga oral de glicose não foram diferentes entre homens e mulheres ( $p = 0,08$  e  $p = 0,40$ , respectivamente). A diferença entre sexos para frutosamina é estatisticamente significante ( $p = 0,002$ ). Como a distribuição é gaussiana, a faixa de normalidade obtida foi calculada como média  $\pm 2$  desvios-padrão, sendo portanto 2,35 a 3,63 mmol/l para homens e 2,18 a 3,22 mmol/l para mulheres.

No grupo das gestantes, os valores de frutosamina também apresentaram distribuição gaussiana e a faixa de normalidade neste grupo foi de 1,96 a 2,84 mmol/l.

Os achados principais indicam que devem ser considerados o sexo e a presença ou não de gravidez para se definir os limites normais dos valores da frutosamina sérica.

#### Exemplo 6.9: Comparação da tianeptina com placebo

A tianeptina é um fármaco antidepressivo do grupo dos tricíclicos. Sua ação antidepressiva potencial foi demonstrada em estudos pré-clínicos através de testes em animais. Rocha (1995) relata os resultados de um ensaio clínico aleatorizado, duplo-cego, realizado com o objetivo de comparar a tianeptina com o placebo. Participam deste ensaio pacientes de Belo Horizonte, Rio de Janeiro e Campinas.

Succintamente, o ensaio consistiu em administrar a droga a dois

grupos de pacientes, compostos de forma aleatória, e quantificar a depressão através da escala de Montgomery-Asberg (MADRS), em que os valores maiores indicam maior gravidade da depressão. O escore foi obtido para cada paciente 7, 14, 21, 28 e 42 dias após o início do ensaio.

Pelo planejamento adotado, os dois grupos não diferiam em termos de depressão no início do ensaio. Assim, uma evidência sobre o efeito da tianeptina é obtida comparando-se os dois grupos ao fim de 42 dias.

A Tabela 6.14 apresenta os escores finais dos pacientes dos dois grupos admitidos em Belo Horizonte.

Tabela 6.14: Escore final na escala MADRS de pacientes dos dois grupos admitidos em Belo Horizonte

Grupo	Escores							
	6	33	21	26	10	29	33	29
Placebo	37	15	2	21	7	26	13	
Tianeptina	10	8	17	4	17	14	9	4
	21	3	7	10	29	13	14	2

Para se efetuar o teste t é preciso usar as seguintes informações:

$$\begin{array}{lll} n_1 = 15 & \bar{x}_1 = 20,53 & s_1 = 11,09 \\ n_2 = 16 & \bar{x}_2 = 11,37 & s_2 = 7,26 \end{array}$$

Com estes valores temos que a estimativa de  $\sigma^2$  é dada por

$$s_p^2 = \frac{14(11,09)^2 + 15(7,26)^2}{15 + 16 - 2} = (9,31)^2$$

Finalmente,

$$T = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2(\frac{1}{n_1} + \frac{1}{n_2})}} \right| = \left| \frac{20,53 - 11,37}{9,31 \sqrt{(\frac{1}{15} + \frac{1}{16})}} \right| = 2,74$$

que comparado com o valor de  $t_{29;0,975} = 2,045$  leva à rejeição da igualdade entre os dois grupos no nível de 5%. O valor- $p$  é 0,0104.

### Exemplo 6.10: Fatores de risco para AVC (continuação)

No estudo de Gotemburgo, citado no Exemplo 6.3, além do hábito de fumar (escore) foram consideradas diversas variáveis quantitativas (contínuas) como possíveis fatores de risco para AVC: pressão sistólica ( $mmHg$ ) e diastólica ( $mmHg$ ), peso ( $kg$ ), altura ( $m$ ), índice de massa corporal definido como peso/(altura) $^2$  ( $Kg/m^2$ ), medida da cintura ( $cm$ ), medida do quadril ( $cm$ ), índice de obesidade abdominal definido como cintura/quadril, nível sérico de colesterol ( $mg/dl$ ), glicose no sangue ( $mg/dl$ ), fibrinogênio no plasma ( $g/l$ ), níveis de hemoglobina ( $g/dl$ ) e de hematócrito (%), e capacidade vital ( $l$ ).

Na Tabela 6.15 reproduzimos do artigo original os valores médios das diversas variáveis para os indivíduos do estudo, divididos em dois grupos: os que sofreram AVC e os que não sofreram AVC.

Tabela 6.15: Valores médios dos fatores estudados no início do pesquisa e valor- $p$  obtido através do teste t de comparação de médias

Fator	AVC		
	Sim	Não	Valor-p
Fumo	2,9	2,6	0,16
Pressão sistólica	151,9	143,2	0,004
Pressão diastólica	96,8	90,1	0,0001
Índice massa corporal	25,7	24,9	0,08
Medida do quadril	94,8	93,6	0,23
Índice obesidade abdominal	0,95	0,93	0,0004
Colesterol	275	269	0,40
Glicose	64,8	67,9	0,60
Fibrinogênio	3,6	3,3	0,01
Hemoglobina	14,9	14,8	0,40
Hematócrito	44,3	44,0	0,46
Capacidade vital	4,3	4,5	0,03

Como conclusão da leitura desta tabela, podemos dizer que a

pressão sistólica, a pressão diastólica, o índice de obesidade abdominal, fibrinogênio e capacidade vital são *possíveis* fatores de risco para o AVC em homens. A verificação final da influência destes fatores na ocorrência de AVC foi feita usando-se análises multivariadas.

### 6.5.2 Teste Z para comparação de médias

Um pressuposto importante para aplicar o teste t da seção anterior é que os dois grupos comparados tenham a mesma variabilidade, o que nem sempre acontece na prática. No caso de amostras grandes ( $n_1$  e  $n_2 \geq 30$ ) dispomos de um teste em que não é necessário qualquer suposição adicional sobre  $\sigma_1^2$  e  $\sigma_2^2$ , ou seja, as variâncias podem ser iguais ou diferentes.

Para fazermos inferência sobre a diferença de médias ( $\mu_1 - \mu_2$ ) é natural utilizarmos a informação amostral contida em  $\bar{X}_1 - \bar{X}_2$ , cuja variância é estimada por  $s_1^2/n_1 + s_2^2/n_2$ . A estatística do teste para testar a hipótese de igualdade de médias ( $H_0: \mu_1 = \mu_2$ ) é

$$Z = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} \quad (6.22)$$

que é aproximadamente  $N(0, 1)$ , sob  $H_0$ , isto é, a hipótese nula é rejeitada se o valor absoluto da estatística (6.22) for maior que  $z_{1-\alpha/2}$  o percentil de ordem  $(1 - \alpha/2)$  da distribuição  $N(0, 1)$ .

### Exemplo 6.11: Efeito do halotano em cirurgias cardíacas

O halotano é uma droga bastante utilizada para induzir anestesia geral. Trata-se de um poderoso anestésico de inalação, não inflamável e não explosivo, com um odor relativamente agradável. Pode ser administrado ao paciente com o mesmo equipamento usado para sua oxigenação. Após a inalação, a substância chega aos pulmões tornando possível a passagem para o estado anestésico mais rapidamente do que seria possível com drogas administradas de forma intravenosa.

Entretanto, os efeitos colaterais incluem a depressão do sistema respiratório e cardiovascular, sensibilização a arritmias produzidas por adrenalina e eventualmente o desenvolvimento de lesão hepática. Alguns anestesiologistas acreditam que esses efeitos podem causar complicações em pacientes com problemas cardíacos e sugerem o uso do

morfina como um agente anestésico nesses pacientes devido ao seu pequeno efeito na atividade cardíaca.

Conahan et al. (1973) compararam esses dois agentes anestésicos em um grande número de pacientes submetidos a uma cirurgia de rotina para reparo ou substituição da válvula cardíaca. Para obter duas amostras comparáveis, os pacientes foram alocados aleatoriamente a cada tipo de anestesia.

A fim de estudar o efeito desses dois tipos de anestesia, foram registradas variáveis hemodinâmicas, como pressão sanguínea antes da indução da anestesia, após a anestesia mas antes da incisão, e em outros períodos importantes durante a operação. A questão que surge é se o efeito do halotano e da morfina na pressão sanguínea é o mesmo. Para comparar os dois grupos, necessitamos dos resultados apresentados na Tabela 6.16.

Tabela 6.16: Média e desvio-padrão da pressão sanguínea ( $mmHg$ ) segundo o tipo de anestesia

Informações sobre a amostra	Anestesia	
	Halotano	Morfina
Média	66,9	73,2
Desvio-padrão	12,2	14,4
$n$	61	61

Nas condições do problema, as hipóteses são  $H_0: \mu_1 = \mu_2$  e  $H_1: \mu_1 \neq \mu_2$ , isto é, devemos testar a diferença entre as pressões sanguíneas médias de indivíduos anestesiados com halotano ou morfina.

Como as amostras são grandes, podemos usar o teste  $Z$ , cujo valor da estatística do teste é

$$\frac{66,9 - 73,2}{\sqrt{(12,2)^2/61 + (14,4)^2/61}} = \frac{-6,30}{\sqrt{5,84}} = -2,61$$

Adotando um nível de significância de 5%, o resultado é estatisticamente significativo, já que  $|-2,61| > 1,96$ . Além disso, o valor- $p$  é 0,009, que é menor que o valor de  $\alpha$  estipulado, indicando que os dois anestésicos não são equivalentes.

## 6.6 Resposta contínua: amostras pareadas

Consideremos a seguinte pergunta: será que o armazenamento da amostra do sangue influencia o valor da medida do colesterol e do triglicérides? Se não há mudança, o armazenamento de amostras poderia ser usado em estudos tipo coorte, em que pacientes são acompanhados ao longo do tempo, o que geraria várias facilidades para o pesquisador clínico.

Neste caso, o problema de interesse é uma *comparação* entre dois grupos de medidas: de triglicérides, por exemplo. É razoável supor, e existem evidências empíricas neste sentido, que a distribuição estatística do nível de triglicérides é gaussiana. Se o possível efeito do armazenamento se dá apenas no aumento ou decréscimo na média da distribuição, não na sua variabilidade, então as hipóteses a serem testadas são

$$H_0: \mu_1 = \mu_2 \quad \text{versus} \quad H_1: \mu_1 \neq \mu_2 \quad (6.23)$$

onde  $\mu_1$  e  $\mu_2$  são as médias antes e depois do armazenamento.

O erro do tipo I, neste caso, é dizer que o armazenamento tem efeito, quando, na realidade não tem, enquanto o do tipo II, que não há efeito quando, de fato, os resultados de medições feitas em diferentes momentos são diferentes. A escolha de  $H_0$  implica que, na ausência de outras evidências, consideraremos que o armazenamento não tem efeito.

Intuitivamente, o critério de decisão, a ser utilizado para testar  $H_0$ , deve ser baseado nas diferenças entre os valores do triglicérides nas duas ocasiões das medidas. Se houver influência do armazenamento, então essas diferenças devem ser diferentes de zero.

O problema de escolha de um critério de decisão reduz-se a escolher uma forma de verificar se as diferenças são provenientes de uma distribuição com média zero.

De uma maneira geral, considerando o par formado por um indivíduo que recebe o tratamento e outro do grupo controle, os dados são  $n$  pares de observações:  $(x_{11}, x_{21}), (x_{12}, x_{22}), \dots, (x_{1n}, x_{2n})$ . Para cada par tomamos a diferença das duas observações, isto é,  $d_1 = x_{11} - x_{21}$ ,  $d_2 = x_{12} - x_{22}, \dots, d_n = x_{1n} - x_{2n}$ .

A partir dessas diferenças calculamos a média

$$\bar{d} = \frac{\sum_{i=1}^n d_i}{n} \quad (6.24)$$

e o desvio-padrão

$$s_d = \sqrt{\frac{\sum_{i=1}^n (d_i - \bar{d})^2}{n-1}} \quad (6.25)$$

Um critério razoável é a distância entre a média das diferenças ( $\bar{d}$ ) e zero. Como no problema da seção anterior, esta diferença deve ser medida em termos de desvios-padrão. Usando-se Estatística Matemática pode-se mostrar que, para  $n$  pares de observações, o desvio-padrão de  $\bar{D}$  é  $\frac{\sigma_d}{\sqrt{n}}$ . Como o desvio-padrão  $\sigma_d$  é usualmente desconhecido, estima-se este parâmetro por  $s_d$ , o desvio-padrão das diferenças das duas medidas.

Portanto, o teste consiste em rejeitar  $H_0$  se

$$T_p = \frac{\bar{d} - 0}{\frac{s_d}{\sqrt{n}}} = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}} \quad (6.26)$$

ou seja, se a distância entre a média das diferenças e zero, medida em desvios-padrão tem um valor “grande”.

A decisão de quão “grande” o valor de  $T_p$  deve ser exige o conhecimento de sua distribuição. Aqui também  $T_p$  tem distribuição  $t$ , agora com  $n-1$  graus de liberdade, quando a distribuição da diferença é gaussiana.

A regra do teste é então rejeitar  $H_0$  se

$$|T_p| \geq t_{n-1, 1-\frac{\alpha}{2}} \quad (6.27)$$

onde  $t_{n-1, 1-\frac{\alpha}{2}}$  é o percentil de ordem  $(1 - \frac{\alpha}{2})$  da distribuição  $t$  de Student com  $n-1$  graus de liberdade.

### Exemplo 6.12: Programa para redução do nível de colesterol

Realizou-se um estudo com o objetivo de avaliar a efetividade de uma dieta combinada com um programa de exercícios físicos na redução do nível de colesterol.

A Tabela 6.17 mostra os níveis de colesterol de 12 participantes no início e no final do programa.

Tabela 6.17: Níveis de colesterol no início e no final do programa

Início ( $x_1$ )	Final ( $x_2$ )	Programa	Diferença	Desvio	Desvio ao quadrado
			$d = x_1 - x_2$	$d - \bar{d}$	$(d - \bar{d})^2$
201	200		1	-19,16	367,36
231	236		-5	-25,16	633,36
221	216		+5	-15,16	230,03
260	233		+27	6,83	46,69
228	224		+4	-16,16	261,36
237	216		+21	0,83	0,69
326	296		+30	9,83	96,69
235	195		+40	19,83	393,36
240	207		+33	12,83	164,69
267	247		+20	-0,16	0,03
284	210		+74	53,83	2898,03
201	209		-8	-28,16	793,36

Quanto maior o valor  $\bar{d}$  maior a evidência de que o programa reduz o nível de colesterol; quanto menor a variabilidade das diferenças individuais maior a chance de se detectar um efeito médio significativo, isto é, uma redução significativa do colesterol devido à ação do programa e não ao acaso. Estes aspectos podem ser avaliados através do teste  $t$ .

Sejam  $\mu_a$  e  $\mu_d$  respectivamente as médias dos níveis de colesterol antes e depois do programa. Para testar a hipótese de que o programa altera o nível de colesterol ( $H_0: \mu_a = \mu_d$  versus  $H_1: \mu_a \neq \mu_d$ ) será aplicado o teste  $t$  (11 graus de liberdade).

Apenas dois participantes tiveram o nível de colesterol aumentado após o programa, mas por pequenas quantidades (5 e 8 mg/dl). As médias antes e depois do programa são respectivamente 244,25 e 224,08, correspondente a uma redução média 20,12 mg/dl ( $\bar{d} = 20,17$ ). Além disso,  $s_d = \sqrt{\frac{5885,7}{11}} = 23,13$  e  $T_p = 3,02$  ( $p = 0,012$ ), isto é, há evidência de que, em média, o programa altera o nível de colesterol. No próximo capítulo vamos quantificar a redução que o programa proporcionou.

## 6.7 Testes não-paramétricos

Os testes apresentados nas seções 6.5 e 6.6 foram derivados para situações em que o efeito do tratamento é descrito por variável contínua com distribuição gaussiana. Estes testes são denominados *paramétricos*, pois suas estatísticas de teste usam as estimativas dos parâmetros da distribuição gaussiana.

Existem testes paramétricos para comparação de dois grupos desenvolvidos para outras distribuições, por exemplo Poisson. Entretanto geralmente não se encontram disponíveis em textos elementares e em softwares estatísticos.

Nesta seção vamos apresentar dois *testes não-paramétricos*, desenvolvidos para comparar grupos em situações em que a distribuição da variável de interesse não é conhecida ou tem comportamento não gaussiano.

Esses testes são construídos usando-se os postos das observações. O posto de uma observação é o número de ordem da observação, estando as observações ordenadas. Quando há empates, toma-se como posto de cada observação a média dos postos que seriam atribuídos às observações, caso os empates não existissem.

Os testes não-paramétricos são boas opções para situações em que ocorrem violações dos pressupostos básicos necessários para a aplicação de um teste paramétrico. Por exemplo, para testar a diferença de dois grupos quando a distribuição subjacente é assimétrica ou os dados foram coletados em uma escala ordinal.

Vamos apresentar dois testes para variáveis contínuas, sendo o primeiro para amostra independentes e o segundo para amostras pareadas. Os cálculos necessários são simples mas trabalhosos e por isso recomendamos o uso de computador.

### 6.7.1 Teste de Mann-Whitney

O teste de Mann-Whitney é usado para a comparação de dois grupos independentes. Para sua construção, primeiramente, obtemos os postos a todas as observações como se elas fossem uma única amostra. A estatística do teste de é

$$MW = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - T \quad (6.28)$$

onde  $n_1$  e  $n_2$  são os tamanhos das amostras dos dois grupos e  $T$  é o total dos postos do grupo menor.

Para a tomada de decisão o valor da estatística  $MW$  deve ser comparado com o percentil de uma distribuição especial (não apresentada nesse texto).

O teste pode ser baseado na estatística  $T$  que aparece em (6.28). Para amostras grandes (pelo menos 10 em cada grupo),  $T$  tem aproximadamente distribuição gaussiana com média  $\mu_T = \frac{n_P(n_P+n_G+1)}{2}$  e desvio-padrão  $\sigma_T = \sqrt{\frac{n_G \cdot \mu_T}{6}}$ , onde  $n_P$  e  $n_G$  são os tamanhos da amostra do grupo menor e maior, respectivamente. Neste caso, o valor da estatística  $Z = \frac{T - \mu_T}{\sigma_T}$  deve ser comparado com o percentil da distribuição gaussiana padrão.

#### Exemplo 6.13: Comparação da tianeptina com placebo (continuação)

Para os dados do Exemplo 6.9, vamos comparar o grupo que recebeu tianeptina com o que recebeu placebo através do teste de Mann-Whitney. Os dados ordenados e os postos correspondentes estão apresentados na Tabela 6.18.

O valor-p é 0,0266 indicando que há diferença entre os dois grupos comparados. É mais do dobro do que o valor-p obtido pelo teste t de Student ( $p = 0,0104$ ). Observe que o uso de um teste não paramétrico é mais apropriado já que a escala usada para medir a pressão é ordinal.

Tabela 6.18: Postos atribuídos aos dados agregados dos dois grupos (placebo e tianeptina)

Placebo		Tianeptina	
Escore	Posto	Escore	Posto
2	1,5	2	1,5
		3	3
		4	4,5
		4	4,5
6	6		
7	7,5	7	7,5
		8	9
		9	10
		10	12
10	12	10	12
13	14,5	13	14,5
		14	16,5
		14	16,5
15	18		
		17	19,5
		17	19,5
21	21,5	21	21,5
26	23,5		
26	23,5		
29	25,5	29	25,5
33	27,5		
33	27,5		
39	29		

## 6.7.2 Teste de Wilcoxon

O teste de Wilcoxon é usado para comparar dois tratamentos quando os dados são obtidos através do esquema de pareamento.

Os seguintes passos devem ser seguidos na sua construção:

1. Calcular a diferença entre as observações para cada par.
2. Ignorar os sinais das diferenças e atribuir postos a elas.
3. Calcular a soma dos postos ( $S$ ) de todas as diferenças negativas (ou positivas).

Para amostras pequenas (até 25 pares), o valor-p deve ser obtido através de uma tabela especial (não incluída nesse texto).

Para amostras grandes, a estatística do teste ( $S$ ) tem aproximadamente distribuição gaussiana com média  $\mu_S = \frac{n(n+1)}{4}$  e desvio-padrão  $\sigma_S = \sqrt{\frac{n(n+1)(2n+1)}{24}}$ . Assim, o valor de  $Z = \frac{S - \mu_S}{\sigma_S}$  deve ser comparado ao valor do percentil da distribuição gaussiana.

Embora não haja nenhuma suposição sobre a forma da distribuição das observações, o teste é mais poderoso para distribuições simétricas.

### Exemplo 6.14: Evolução do tratamento com tianeptina

Rocha (1995), além da análise sintetizada na seção anterior, verificou se houve diminuição do escore de depressão entre os pacientes de um dos grupos durante o desenvolvimento do estudo.

A Tabela 6.19 mostra os escores dos pacientes do grupo tianeptina que foram admitidos em Belo Horizonte no primeiro ( $x_1$ ) e no último dia ( $x_{42}$ ).

Como o escore MADRS é, na realidade, uma medida em escala ordinal, o uso de teste não-paramétrico é mais apropriado. O uso do teste  $t$ , só se justifica como aproximação.

Dois pacientes (nímeros 14 e 15) não completaram o estudo. Assim, não fornecem informações suficientes para a comparação pretendida e, por isto, foram excluídos da análise.

As diferenças em valor absoluto,  $|d|$ , e os respectivos postos estão mostrados na Tabela 6.20. O valor-p é 0,0225 que é maior que o obtido através do teste  $t$  ( $p = 0,0013$ ). Há portanto indicação de alteração dos níveis de depressão para pacientes que fizeram uso de tianeptina.

Tabela 6.19: Escores dos pacientes do grupo tianeptina no primeiro ( $x_1$ ) e último dia ( $x_{42}$ )

Nº	$x_1$	$x_{42}$	$d = x_{42} - x_1$	Nº	$x_1$	$x_{42}$	$d = x_{42} - x_1$
1	24	6	-18	9	35	37	+2
2	46	33	-13	10	30	15	-15
3	26	21	-5	11	38	2	-36
4	44	26	-18	12	38	21	-17
5	27	10	-17	13	31	7	-24
6	34	29	-5	14	27	*	*
7	33	33	0	15	34	*	*
8	25	29	+4	16	32	26	-6

Tabela 6.20: Posto atribuído ao valor absoluto da diferença  $|d|$

$ d $	Posto	Nº do paciente	$ d $	Posto	Nº do paciente
0	1	7	15	8	10
2	2	9	17	9,5	5
4	3	8	17	9,5	12
5	4,5	3	18	11,5	1
5	4,5	6	18	11,5	4
6	6	16	24	13	13
13	7	2	36	14	11

## 6.8 Exemplo comentado: Imunogenicidade de uma vacina

A infecção por vírus influenza ocorre através da contaminação de indivíduo suscetível por secreções respiratórias de indivíduo infectado. A doença respiratória febril aguda resultante pode se apresentar como síndromes clínicas diversas: resfriados comuns, faringite, traqueobronquite, pneumonias virais e até mesmo pneumonias bacterianas secundárias. Geralmente auto-limitada, responde por alta morbidade e mortalidade na população idosa, especialmente em virtude do maior risco de complicações pulmonares. Ressalta-se ainda

a ocorrência em surtos e as alterações de antigenicidade.

A principal medida profilática é a utilização de vacina trivalente, contendo um ou mais subtipos do vírus influenza A e B com revisão antigenica anual. No entanto, as vacinas disponíveis contra influenza têm várias limitações. Destaca-se a não obtenção consistente de níveis protetores de anticorpos na população idosa após dose única da vacina, restando uma proporção da população imunizada sob risco.

Em uma tentativa de produzir uma nova vacina com maior imunogenicidade, Gluck et al. (1994) desenvolveram uma vacina virossômica trivalente de influenza inserindo hemaglutinina purificada do vírus na membrana de lipossomos (virossomas).

Para verificar a segurança e imunogenicidade, conduziram um estudo em uma casa de repouso para idosos em Cagliari, Itália. Os residentes foram excluídos na presença de evidências de doença aguda ou crônica no momento da vacinação, história de influenza nos últimos 12 meses, alergia a qualquer dos componentes da vacina ou proteínas de aves, uma reação grave prévia a essa vacina ou estar participando de outro estudo.

Cento e vinte e seis residentes selecionados receberam, aleatoriamente, uma das três vacinas: vírus inteiros, subunidade viral e virossômica. As reações adversas foram diariamente registradas e duas amostras de sangue, coletadas, antes e 28 dias após a imunização.

A idade média do grupo estudado foi de 78 anos, sendo 70% dos participantes mulheres. Além disso, 12% haviam recebido a vacina contra esse vírus no ano anterior. Não foram relatadas reações locais ou sistêmicas nos grupos. As características dos três grupos estudados estão resumidas na Tabela 6.21.

Tabela 6.21: Informação sobre idade dos pacientes e distribuição quanto ao sexo e à vacinação contra influenza nos últimos 24 meses

Vacina	Idade $\bar{x}$ (faixa)	Nº de idosos			Vacinados
		Homens	Mulheres	Total	
Vírus inteiros	78 (64-88)	9	23	32	1
Virossômica	78 (63-102)	22	41	63	10
Subunidade	79 (63-102)	7	24	31	4
Total	78 (63-102)	38	88	126	15

A Tabela 6.22 mostra a resposta imune às três vacinas. Através do teste t pareado constatou-se elevação significativa ( $p < 0,01$ ) da média geométrica dos títulos em relação aos títulos basais para as três cepas virais. O uso da média geométrica justifica-se pelo tipo de variável (titulação); neste caso, trabalha-se com o logaritmo da titulação e, então, o logaritmo da média geométrica é igual à média aritmética na escala logarítmica.

Tabela 6.22: Média geométrica e faixa dos títulos de Ac anti-Ha antes e após vacinação com os três preparados vacinais referentes às três cepas: H1N1 Singapore (Cepa 1), H3N2 Beijing (Cepa 2) e B/Yamagata (Cepa 3)

Tipo de vacina	Cepa 1		Cepa 2		Cepa 3	
	Pré	Pós	Pré	Pós	Pré	Pós
Vírus inteiros	2,9 (0-80)	29,1 (0-320)	25,2 (0-160)	130 (10-640)	3,4 (0-20)	16,3 (0-640)
Virossômica	2,4 (0-80)	44,2 (0-640)	20,6 (0-320)	142 (20-2560)	2,4 (0-40)	17,5 (0-320)
Subunidade	1,8 (0-40)	14 (0-320)	26,4 (0-160)	121 (10-5120)	2,9 (0-40)	11,4 (0-160)

A eficácia vacinal foi avaliada através da proporção de imunizados que adquiriram níveis protetores de anticorpos. A Tabela 6.23 mostra os valores-p obtidos através do teste t. A vacina virossômica levou à proporção mais alta de pessoas com títulos protetores de anticorpos contra duas de três cepas virais quando comparada à outra vacina. Para a cepa H3N2 não houve diferença provavelmente pela alta porcentagem de indivíduos com níveis basais de anticorpos protetores.

Tabela 6.23: Valor-p para a avaliação de vacinas

Tipo de vacina	H1N1 Singapore	B/Yamagata
Vírus inteiros	0,022	0,035
Virossômica	0,007	0,002

## 6.9 Considerações práticas sobre testes de hipóteses

Reunimos nesta seção algumas considerações essenciais para o uso crítico dos testes de hipóteses na prática. Como atualmente os cálculos para a realização de qualquer teste estão disponíveis em programas de computador, várias observações desta seção são potencialmente mais importantes que aquelas das seções anteriores, em que se explica basicamente a mecânica de cálculo das estatísticas de alguns testes de hipóteses.

### 6.9.1 Valor-p

A decisão final sobre a hipótese nula é tomada comparando-se o valor-p com um valor pré-fixado, usualmente 0,05. Quando o valor-p é menor que este ponto de corte, o resultado é chamado *estatisticamente significante*, e, *altamente significante*, quando ele é menor que um ponto de corte ainda menor (digamos 0,01). Nas outras situações o teste é dito *não significante*. Por esta razão os testes são denominados *testes de significância*.

Esses pontos de corte são arbitrários e não se deve dar uma importância exagerada a eles. É inaceitável que os resultados de dois estudos em que os valores-p sejam 0,045 e 0,055 sejam interpretados de forma diferente para  $\alpha = 0,05$ . Esses valores devem levar a conclusões muito parecidas, não a conclusões diametralmente opostas. Esse raciocínio ficará mais claro no próximo capítulo, quando estudarmos o conceito de intervalo de confiança.

### 6.9.2 Hipóteses bilaterais versus unilaterais

As hipóteses alternativas, respectivamente para o teste de comparação de proporções e de médias são:  $H_1: p_1 \neq p_2$  e  $H_1: \mu_1 \neq \mu_2$ . Podem ser desmembradas como:  $H_1: p_1 > p_2$  ou  $p_1 < p_2$  e  $H_1: \mu_1 > \mu_2$  ou  $H_1: \mu_1 < \mu_2$ .

Estas hipóteses assumem, portanto, que qualquer um dos dois grupos pode ter uma proporção ou uma média maior do que o outro. Por isto este tipo de hipótese é denominada *bilateral*. O valor-p bilateral é a probabilidade de se obter em *qualquer direção* uma diferença igual ou mais extrema do que a observada.

Existe também a possibilidade de se formular hipóteses alternativas ( $H_1$ ) unilaterais, como mostra o quadro abaixo.

Situação	Proporções	Médias
(1)	$H_1: p_1 > p_2$	$H_1: \mu_1 > \mu_2$
(2)	$H_1: p_1 < p_2$	$H_1: \mu_1 < \mu_2$

Nestes casos, as comparações são estabelecidas em uma determinada direção. Assim, por exemplo, se se comparar um procedimento novo com um padrão, estamos avaliando se a inovação deve ser recomendada. Portanto, a escolha de hipóteses unilaterais ou bilaterais influencia decisivamente a interpretação dos resultados da análise estatística.

Para se testar hipóteses unilaterais baseadas na distribuição t de Student ou na distribuição gaussiana, utiliza-se apenas uma cauda da distribuição. Por exemplo, fixando-se um nível de significância de 5%, o percentil utilizado para testes baseados na distribuição gaussiana é 1,64 e não 1,96. Embora a diferença não seja grande, a interpretação pode ser diferente para um nível de significância fixado.

As duas opções de testes (unilateral ou bilateral) estão disponíveis em programas de computador. Em geral, o valor-p para teste bilateral é o dobro do valor-p correspondente à hipótese unilateral.

Há circunstâncias em que hipótese unilateral é a melhor forma de se formular a questão de interesse. Em estudos comparando uma inovação de um procedimento padrão, a hipótese alternativa mais interessante é que a inovação é superior, um apelo à hipótese unilateral. Em um estudo sobre a qualidade da água, é importante verificar se as contagens bacteriológicas das amostras analisadas estão acima do padrão considerado aceitável. Obviamente, não há nenhuma preocupação quando a água estiver mais pura do que o padrão estabelecido, o que justificaria a hipótese unilateral.

Um teste unilateral também pode ser justificado quando se pode afirmar que uma das direções contempladas pela hipótese bilateral é completamente inconcebível, como a situação descrita por Luthra et al. (1982).

O argumento mais forte contra o uso de hipóteses unilaterais é que por maior que seja a evidência de que um tratamento seja superior ao outro, nunca se tem certeza absoluta do que realmente pode

acontecer. Embora exista uma boa razão para esperar que novas drogas ou novos procedimentos sejam melhores que os do grupo controle (caso contrário o estudo não estaria sendo realizado), ainda assim existe a possibilidade, mesmo que remota, de que seus resultados sejam piores. Na comparação de uma droga com o placebo, não se pode descartar a possibilidade de que a droga tenha um efeito deletério, e portanto não deva ser recomendada.

Na escolha entre hipótese bilateral e unilateral os seguintes aspectos devem ser considerados:

1. O tipo de hipótese adotado deve preceder a análise dos dados, isto é, a escolha não deve ser influenciada pelo resultado da amostra.
2. O teste bilateral é mais conservador do que o unilateral. Na maioria dos casos, testes unilaterais são vistos como uma maneira de se exagerar a força dos achados. Se houver qualquer dúvida, deve-se optar pelo teste bilateral.
3. Segundo alguns autores, a distinção entre testes unilaterais e bilaterais não é de fundamental importância na interpretação dos resultados, desde que fique claro qual foi usado.
4. Mesmo quando o teste unilateral pode ser justificado, pode não encontrar resistência editorial para publicar tais achados.
5. Teste bilateral é a forma padrão usada em periódicos médicos.

Alguns estatísticos e editores de jornais acreditam que o valor p unilateral nunca deve ser usado. O primeiro argumento é a uniformidade de apresentação dos resultados, tal que um determinado valor p tenha um mesmo significado em todos os artigos. Segundo, acreditam que situações que justificam o uso de testes unilaterais são extremamente raras. Terceiro, em estudos sobre importantes questões, como a regulamentação de uma droga, o valor-p é apenas um fator usado na tomada de decisões. O critério de que o valor-p seja menor do que 0,05 em geral é insuficiente para estabelecer eficiência e pelo menos o teste bilateral é mais conservador.

Aceitando estes argumentos e considerando a padronização já existente na maioria dos periódicos médicos, recomendamos o uso rotineiro de hipóteses bilaterais.

### 6.9.3 Conclusões a partir de testes de hipóteses

A partir de testes bilaterais não se pode concluir que um tratamento é superior ao outro e sim que há diferença entre os tratamentos comparados. Uma maneira apropriada para avaliar qual é o mais eficiente é através da quantificação dos “efeitos” dos tratamentos, assunto tratado no próximo capítulo.

A evidência fornecida pelo valor-p é uma evidência probabilística. Assim em um ensaio clínico, se o valor-p é “grande” (digamos maior que 0,20), podemos dizer que a evidência fornecida sugere que os tratamentos são equivalentes. Não podemos excluir, entretanto, que os tratamentos sejam diferentes, apenas estamos mais seguros que são equivalentes.

Por outro lado, se o valor-p é pequeno (digamos menor que 0,001), então a igualdade dos dois tratamentos não parece ser plausível. A evidência sugere que um dos tratamentos é superior.

Conforme comentado na Seção 6.2.4, o valor-p igual a 0,05 é usado como referência para tomada de decisão em pesquisas médicas. Isto significa um resultado que pode ocorrer menos que uma vez em vinte quando a hipótese nula é verdadeira. Nesta formulação, quando se rejeita a hipótese nula, aceita-se a hipótese alternativa complementar. Se o valor-p não excede o ponto de corte, não se rejeita a hipótese nula. Entretanto, não se pode dizer que a hipótese nula é verdadeira, mas somente que não há evidência suficiente para que ela seja rejeitada. Esta é uma distinção útil mas importante.

Concluindo, não se deve fixar no critério de  $p < 0,05$ , mas sim interpretar corretamente o resultado do teste através do valor-p.

### 6.9.4 Significância estatística e significância clínica

Dentro do contexto de teste de hipóteses, a palavra significante tem sido muitas vezes usada de forma incorreta. Por essa razão não é recomendada em alguns periódicos médicos. Pode-se usar corretamente o conceito de significância estatística e na prática interpretá-lo de forma indevida. Portanto, é importante entender a diferença entre significância estatística e significância clínica.

Um resultado estatisticamente significante não implica necessariamente em uma importância clínica. Por exemplo, Gould et al. (1985) relatam um estudo em que uma pequena diferença de 1 mmHg na

pressão sanguínea foi estatisticamente significante, mas sem qualquer importância clínica. Por outro lado, pode ocorrer que o teste não tenha sido estatisticamente significante, mas os dados podem ter mostrado uma diferença importante na prática. Esses efeitos podem estar ligados ao tamanho da amostra (Lemeshow et al., 1990).

### 6.9.5 Fatores de confusão

Nem sempre a diferença observada entre as respostas, através da aplicação do teste t ou teste qui-quadrado, pode ser atribuída ao fator usado para dividir os pacientes nos dois grupos de comparação. Como visto no Capítulo 2, eventuais diferenças podem ser atribuídas a um outro fator que recebe neste caso, a denominação de fator de confusão.

Como qualquer realidade clínica é complexa, a possibilidade de existência de fatores de confusão deve ser sempre considerada. De fato, qualquer resposta clínica, como sobrevida, remissão, recidiva ou desaparecimento dos sintomas, é influenciada por vários fatores.

Para se fazer comparações válidas, na presença de fatores de confusão, foram desenvolvidos vários métodos estatísticos que recebem na literatura médica o nome genérico de métodos multivariados.

### 6.9.6 Apresentação dos resultados

Em análises que envolvem testes de hipóteses, é boa prática apresentar os valores necessários para que o valor-p possa ser recalculado. Por exemplo, em testes de comparação de médias para amostras independentes, as médias e os desvios-padrão para cada grupo devem ser apresentados.

Deve-se evitar a apresentação do valor-p em forma de intervalos ( $p > 0,05$ ;  $0,01 < p < 0,05$ ). É recomendado apresentar o valor-p exato obtido (por exemplo,  $p = 0,012$ ), que em geral é a saída padrão de programas de computador. Em geral duas ou três casas decimais são suficientes.

### 6.9.7 Pressupostos dos testes

Como mencionado nas seções em que os testes foram apresentados, cada um deles foi desenvolvido a partir de determinados pressupostos. Após a escolha do teste a ser utilizado, as seguintes questões

relevantes podem ser levantadas: Os dados analisados satisfazem os pressupostos? Se eles não forem completamente satisfeitos, quais são as consequências? Elas são graves?

Existem várias formas para a verificação dos pressupostos: análise descritiva dos dados, gráficos específicos e os chamados testes de aderência. Cada suposição é avaliada de uma forma específica. Por exemplo, no Capítulo 5 discutimos formas de avaliar a suposição de que a variável tem distribuição gaussiana.

Em geral, a violação da suposição altera o nível de significância. Por exemplo, alguém pode estar pensando que está testando a um nível de significância de 5% e realmente o nível ser maior, digamos 8%. Outros efeitos também podem ocorrer, dependendo da situação e do tipo de teste.

A seguir apresentamos alguns comentários sobre os pressupostos de dois testes discutidos neste capítulo: teste qui-quadrado e teste t.

### Teste qui-quadrado

O teste qui-quadrado foi desenvolvido para situações em que são válidas as seguintes suposições:

1. A probabilidade de sucesso não varia de indivíduo para indivíduo dentro dos grupos de controle e tratamento.
2. O resultado que ocorre para um paciente de qualquer dos grupos não influencia de nenhuma maneira o resultado de outros indivíduos.

Quando a condição 1 não for válida, é necessário usar uma outra análise, por exemplo, combinação de tabelas  $2 \times 2$  (Fleiss, 1981).

Quando a segunda hipótese, de independência, não é verdadeira, há maior dificuldade na análise. Métodos adequados de análise para esta situação só foram desenvolvidos muito recentemente. Estão agrupados no tópico Análise de Dados Dependentes.

### Teste t

Os seguintes pressupostos são feitos para aplicar o teste t descrito na Seção 6.5.1 para a comparação de médias quando as amostras são independentes: a variável tem distribuição gaussiana e as variâncias

são iguais. Com essas duas suposições, a hipótese de igualdade da média é mais geral do que a princípio parece, pois é equivalente à hipótese de que os dois tratamentos produzem o mesmo efeito, ou seja, não apenas que os tratamentos sejam em médias comparáveis. Esse é o tipo de comparação que mais interessa na prática.

Estudos mostram que o teste t é robusto com relação à suposição da distribuição, isto é, o teste é apropriado mesmo para variáveis que sejam um pouco assimétricas ou que efetivamente não tenham distribuição gaussiana (por exemplo escalas). Entretanto, a suposição de igualdade de variância, denominada *homocedasticidade*, é um importante aspecto a ser avaliado, pois sua violação pode resultar em conclusões incorretas.

Outros estudos demonstraram que se o tamanho das amostras dos dois grupos é igual ( $n_1 = n_2$ ) e as amostras são grandes, o nível de significância calculado segundo a suposição de que a distribuição subjacente é gaussiana e há igualdade de desvio-padrão é válido, mesmo se essas suposições forem violadas. Entretanto, se  $n_1 \neq n_2$  e os desvios padrão não são iguais, o nível de significância pode ser alterado por uma quantidade que depende de quantas vezes um desvio-padrão é maior que o outro.

A homocedasticidade pode ser avaliada por um teste de igualdade de variância. A *heterocedasticidade*, variâncias diferentes, pode ser corrigida pelo uso de transformação da variável; alternativamente pode-se usar testes desenvolvidos para esta situação.

Além do aspecto técnico apresentado sobre o desvio de homocedasticidade, é importante estar atento ao significado das hipóteses no caso de heterocedasticidade. Nesse caso, o teste de comparação de médias pode não ter muito interesse na prática. Dois tratamentos que tenham a mesma média podem ter efeitos completamente diferentes, e portanto não serem considerados equivalentes.

## 6.10 Considerações finais

### 6.10.1 Denominações de testes

Nas seções 6.3.1 e 6.5.1, usamos as denominações teste qui-quadrado e teste t. Essas são as formas populares como esses testes são conhecidos. Entretanto, existe uma pequena imprecisão nessas

denominação pois na verdade existem inúmeros testes baseados nas distribuições qui-quadrado e t de Student.

### 6.10.2 Amostras pareadas e amostras independentes

Em geral, o planejamento de amostras pareadas melhora a precisão das estimativas das diferenças entre as médias dos tratamentos. Entretanto, quando os indivíduos já forem homogêneos ou quando a heterogeneidade não puder ser atribuída a fatores identificáveis, o pareamento arbitrário nem sempre contribui para uma redução da variabilidade, ou seja, para a melhoria da precisão. Neste caso, o planejamento de amostras independentes seria recomendado (ver seção 9.5 de Bhattacharyya & Johnson, 1977).

### 6.10.3 Sobre pareamentos

#### Tipos de pareamentos

Podemos distinguir três tipos de pareamento: auto-pareamento, pareamento natural e pareamento artificial.

O *auto-pareamento* ocorre quando o indivíduo serve como seu próprio controle, como na situação em que um indivíduo recebe duas drogas administradas em ocasiões diferentes. Outra situação é a que foi apresentada no Exemplo 6.12, em que o nível de colesterol foi medido antes e depois do programa. Finalmente, a comparação de dois órgãos no mesmo indivíduo, como braços, pernas, olhos, narinas, segundo alguma característica estudada também constitui um auto-pareamento.

O *pareamento natural* consiste em formar pares tão homogêneos quanto possível, controlando os fatores que possam interferir na resposta, sendo que o pareamento aparece de forma natural. Por exemplo, em experimentos de laboratório pode-se formar pares de cobaias selecionadas da mesma ninhada; em investigações clínicas, gêmeos univitelinos são muitos usados.

No *pareamento artificial* escolhe-se indivíduos com características semelhantes, tais como, idade, sexo, nível sócio-econômico, estado de saúde ou, em geral, fatores que podem influenciar de maneira relevante a variável resposta.

### Dificuldades práticas na formação de pares homogêneos

Na prática podem existir dificuldades no conhecimento das características que devem ser controladas e mesmo conhecendo-as pode ser difícil formar pares homogêneos, como por exemplo no caso em que o número de fatores é muito grande.

Em muitas situações, embora desejável, torna-se difícil ou mesmo impossível a implementação do planejamento com amostras pareadas.

### 6.10.4 Amostragem

Uma questão essencial em qualquer pesquisa é sobre a amostra a ser coletada ou pesquisada. Além da necessidade de se ter uma amostra com uma boa representatividade da população, que pode ser obtida através de um planejamento adequado, (a quantidade de indivíduos é crucial para a validade das conclusões) A chave de uma pesquisa é portanto, *como e quanto amostrar*. Um dos problemas que pode acompanhar a não detecção de diferença significativa em um estudo é o tamanho de amostra insuficiente (Freiman et al., 1978).

Os principais esquemas amostrais que podem ser implementados são: amostragem aleatória simples, estratificada, por conglomerados e sistemática ou combinações entre eles.

Os critérios usuais na determinação do tamanho da amostra são: precisão, erros associados a testes de hipóteses, poder estatístico, custos, ou ainda a combinação de alguns destes critérios. Existem fórmulas desenvolvidas para dimensionamento de amostras que consideram o esquema amostral e/ou o planejamento do estudo. Entretanto, algumas fórmulas populares de tamanho de amostra nem sempre são apropriadas (Kupper & Hafner, 1989).

Gail (1976), Kraemer & Thiemann (1987), Lemeshow et al. (1990), Levi & Lemeshow (1991), Lwanga & Lemeshow (1991) e Silva (1998) são boas referências para esses tópicos.

### 6.10.5 Extensões

#### Hipóteses mais gerais

Neste capítulo as hipóteses testadas referem-se a igualdade de médias e igualdade de proporções dos tratamentos comparados. Em algumas situações pode haver o interesse de testar hipóteses mais

gerais, isto é, que a diferença entre as médias ou as proporções de sucesso dos tratamentos é uma certa quantidade. Por exemplo, podemos testar se a diferença entre as médias da pressão sanguínea antes e depois de um exercício físico é 20 mmHg. Poderíamos ainda testar se a preferência para o partido político A está 10 pontos percentuais acima do partido B.

Para estas situações, testamos as seguintes hipóteses de comparação de médias:  $H_0: \mu_1 - \mu_2 = \mu_0$  versus  $H_1: \mu_1 - \mu_2 \neq \mu_0$  e as seguintes hipóteses de comparação de proporções:  $H_0: p_1 - p_2 = p_0$  versus  $H_1: p_1 - p_2 \neq p_0$ . Note que para o teste de igualdade de médias  $\mu_0 = 0$  e para o teste de igualdade de proporções  $p_0 = 0$ , isto é, estes testes são um caso particular do teste mais geral.

As estatísticas para os testes de média, para amostras independentes ou emparelhadas, e para o teste Z para comparação de proporções são semelhantes àquelas apresentadas nas seções anteriores; entretanto, no numerador da estatística do teste, a diferença das estimativas dos parâmetros é comparada com o valor especificado na hipótese nula. Por exemplo, no teste Z de comparação de proporções devemos calcular

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - p_0}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$$

onde  $p_0$  é o valor especificado no problema (no exemplo da preferência política  $p_0 = 0,10$ ).

### Mais de dois grupos

Este capítulo trata do caso de comparação de dois grupos, mas frequentemente o objetivo de uma pesquisa é a comparação de mais de dois grupos. Não é apropriado fazer-se comparações duas a duas sem qualquer ajuste. Torna-se, portanto, necessária a extensão dos testes apresentados.

A comparação de médias de mais de dois tratamentos é feita através de análise de variância e comparações múltiplas. Para a comparação de mais de duas proporções podemos utilizar o teste qui-quadrado, com as devidas adaptações. Altman (1991) é uma referência para esses dois tópicos.

## 6.11 Exercícios

- Observou-se nos anos 60 que a ocorrência de abortos espontâneos nas gestações de médicas anestesiadoras era mais elevada do que o normal. Para verificar se esta observação refletia ou não uma condição geral, realizou-se em 1970 um estudo no Hospital da Universidade de Stanford, Califórnia. Foram encontrados os seguintes resultados:

Quadro	Especialidade		Total
	Anestesiista	Outra	
Gestação normal	23	52	75
Aborto espontâneo	14	6	20
Total	37	58	95

- Escolha a hipótese nula e a alternativa que são razoáveis nesta situação.
- Faça o teste adequado considerando uma margem de erro provável inferior a 5%. Calcule o valor-p. Qual é a sua conclusão?
- Para os dados de glicose e triglicerídeos do Exemplo 6.8, calcule o valor-p para o teste de comparação de mulheres e homens. Comente.
- Frick et al.(1987) realizaram em Helsinque um ensaio clínico aleatorizado com o objetivo de verificar se o fármaco gemfibrozil, que reduz o nível de colesterol, era eficiente para a prevenção primária de doença coronariana em homens de meia-idade assintomáticos com dislipidemia. Para isto, compararam o grupo que durante o ensaio tomou gemfibrozil (grupo 1), com o grupo que usou placebo (grupo 2). Entre as várias análises, verificaram o nível de colesterol (mg/dl) observando os seguintes resultados:

$n_1 = 1973$	$\bar{x}_1 = 244,7$	$s_1 = 33,7$
$n_2 = 1958$	$\bar{x}_2 = 272,5$	$s_2 = 31,4$

Com base nos resultados acima pode-se afirmar que, ao nível de significância de 5%, o gemfibrozil atua sobre o nível de colesterol?

4. Transcrevemos a seguir resultados do artigo de Rothenberg et al. (1987), onde são descritas as características dos primeiros 5833 casos de AIDS em Nova York. O termo sobrevida zero é usado para indicar que o paciente morreu no primeiro mês após o diagnóstico de AIDS.

Sexo	Número de pacientes	Sobrevida		
		Média (dias)	Mediana (dias)	Zero (%)
Masculino	5281	374	357	10,9
Feminino	552	298	263	16,3

- (a) Complete o quadro:

Sexo	Sobrevida zero		
	Sim	Não	Total
Masculino			
Feminino			
Total			

- (b) Teste a hipótese de que a proporção de homens com sobrevida zero é igual à proporção de mulheres com sobrevida zero (use  $\alpha = 0,05$ ). Calcule o valor-p.

5. Realizou-se um estudo tipo coorte para investigar a associação entre o nível de catecolamina no sangue e a incidência posterior de doença coronariana. Para isto, 609 homens com idade entre 40 e 76 anos foram acompanhados no período de 1960 a 1989. O nível de catecolamina foi determinado em 1960 e classificado como alto ou baixo. Os resultados são apresentados na tabela abaixo:

Doença coronariana	Nível inicial de catecolamina		
	Alto	Baixo	Total
Presente	27	44	71
Ausente	95	443	538
Total	122	487	609

Teste a hipótese de associação entre o nível de catecolamina no sangue e a incidência de doença coronariana, ao nível de significância de 1%.

6. Nove indivíduos do sexo masculino, sadios, com idade média de 21 anos participaram, voluntariamente, de uma pesquisa cujo objetivo era verificar se a alcalose respiratória, induzida por hiperventilação voluntária, aumenta a capacidade física avaliada pelo tempo de corrida de 800 metros. Neste estudo, os nove indivíduos participaram da corrida de 800 metros em 2 momentos: um deles em condições normais (sem hiperventilação) e no outro após a hiperventilação voluntária. Os dados obtidos (em minutos) estão apresentados a seguir:

Indivíduo	Hiperventilação	
	Com	Sem
1	155	151
2	150	152
3	174	172
4	158	150
5	155	152
6	148	150
7	142	144
8	164	169
9	143	144

- (a) Expresse em termos estatísticos as hipóteses de interesse a serem testadas pelo pesquisador.  
 (b) Teste a hipótese estabelecida em (a), ao nível de significância de 5%. Calcule o valor-p e apresente suas conclusões.

7. O artigo transcrito a seguir foi publicado, no primeiro semestre de 1990, pela Folha de São Paulo:

#### ENZIMA INDICA PROPENSÃO À RECIDIVA

Boston, Massachusetts - Uma enzima denominada catepsina D pode indicar quais as pacientes com câncer de mama são mais suscetíveis a recidivas, indica estudo da Universidade do Texas. De acordo com os pesquisadores, mulheres com altos níveis de

catepsina D parecem ter mais do dobro de chances de recidiva cinco anos após a cirurgia, devendo por isso ser submetidas a tratamentos específicos. Uma das funções da catepsina D é decompor proteínas encontradas nos alimentos. O trabalho acompanhou 199 mulheres cujo câncer não havia se estendido aos nódulos linfáticos. Um terço apresentou altos níveis de catepsina D; dessas, 60% sofreram recidivas. Nas mulheres com baixas taxas da enzima, o percentual caiu para 29%, equivalente ao da maioria das mulheres americanas que tiveram a doença diagnosticada em estágios iniciais. Os resultados do estudo indicam que pacientes precocemente diagnosticadas que tenham alto nível da enzima devem ser submetidas à quimioterapia ou à terapia hormonal para reduzir os riscos de recidiva. Cerca de 140 mil americanas morrem anualmente nos Estados Unidos em consequência da doença. Metade dos casos é diagnosticada em estágios iniciais.

- (a) Que forma de pesquisa médica foi utilizada neste estudo? Justifique, identificando os grupos e a resposta de interesse.
- (b) Utilizando os dados fornecidos no artigo, faça o teste adequado para verificar se os resultados obtidos pelos pesquisadores indicam que a enzima catepsina D está associada à recidiva nos casos de câncer de mama (use  $\alpha = 0,05$ ). Calcule o valor-p.
- Para determinar como uma dose experimental de anestesia afeta homens e mulheres, uma amostra de 15 homens e 17 mulheres foi selecionada aleatoriamente em uma clínica odontológica e seus tempos de reação (em minutos), registrados. As seguintes estatísticas sumarizam os dados:

Estatística	Homens	Mulheres
Média	4,8	4,4
Desvio-padrão	0,8	0,9

- (a) Usando  $\alpha = 0,05$ , teste se existe diferença significativa entre os tempos de reação de homens e mulheres.

- (b) Qual é a relevância do fato de os pacientes terem sido escolhidos aleatoriamente?

9. Cinquenta amostras de saliva sabidamente positivas para o bacilo da tuberculose foram colocadas em duas diferentes culturas (A e B). O objetivo do experimento era a comparação destes meios na detecção do bacilo. Os resultados estão resumidos na tabela abaixo:

Meio A	Meio B	
	Detectou	Não detectou
Detectou	20	12
Não detectou	2	16

Existe evidência de que os métodos sejam diferentes? (use  $\alpha = 0,05$ ).

10. Numa pesquisa de opinião pública 1000 homens e 1000 mulheres foram entrevistados sobre a posição acerca do aborto. Entre as mulheres 356 manifestaram-se contra a legalização do aborto, enquanto que 515 homens tiveram a mesma posição. Existe diferença significativa entre os dois sexos quanto à opinião sobre a legalização do aborto? (use  $\alpha = 0,05$ ). Calcule o valor-p.
11. O *Canadian Medical Association Journal* publicou em novembro de 1972 os seguintes resultados de um experimento que procurou investigar se o uso de vitamina C previne resfriados. Por um determinado período de tempo, 407 indivíduos tomaram fortes doses de vitamina C e 411 receberam placebo. No grupo da vitamina, 105 participantes ficaram livres de doenças do tipo respiratório enquanto que no grupo dos pseudomedicados, apenas 76. O que os pesquisadores puderam concluir? (use  $\alpha = 0,05$ ). Calcule o valor-p.
12. Para verificar se, na faixa etária de 7 a 10 anos, crianças do sexo masculino têm desempenho em uma corrida de 50 metros diferente de crianças do sexo feminino, obteve-se uma amostra de 30 meninos e 30 meninas escolhidos aleatoriamente entre os 804 alunos do Colégio Pitágoras Pampulha em Belo Horizonte

que participaram da Mini-Olimpíada de 1989. A partir dos dados referentes ao tempo gasto para completar o percurso foram calculadas as seguintes estatísticas:

Estatística	Meninos	Meninas
$\bar{x}$	8,75	9,08
s	1,56	1,68

- (a) Estabeleça a hipótese a ser testada.
  - (b) Teste-a ao nível de significância de 5%.
  - (c) Existe algum fator de confusão possível neste caso?
  - (d) A escolha aleatória dos alunos resolve a dificuldade trazida por um eventual fator de confusão?
13. A seguir apresentamos o resumo do artigo intitulado *Prevalence of Low Birth Weight and Preterm Delivery in Relation to the Interval Between Pregnancies among White and Black Women*, de autoria de Rawlings, J. S. et al., publicado no *The New England Journal of Medicine* em 12 de janeiro de 1995.

A taxa de mortalidade infantil nos Estados Unidos é mais alta entre os negros do que entre as outras raças. Isto pode ser devido a uma maior freqüência de crianças nascidas prematuramente ou com baixo peso entre os negros, fatores que, por sua vez, estão associados a pequenos intervalos entre as gestações. Assim, uma possível explicação para a diferença de mortalidade infantil entre as raças seria o menor intervalo entre gestações para as mulheres da raça negra.

Para verificar essa hipótese, os autores fizeram um estudo retrospectivo envolvendo 1624 mulheres brancas e 298 negras, todas com acesso a um serviço de saúde gratuito de alta qualidade. Só foram incluídas mulheres que tiveram duas gestações simples com o mesmo parceiro.

Alguns resultados estão sintetizados na tabela a seguir. Para as variáveis idade e paridade, apresentamos as médias e os desvios-padrão e para as variáveis solteira e pré-natal tardio, as freqüências com as respectivas porcentagens entre parênteses. Por pré-natal tardio entende-se que o acompanhamento

médico iniciou-se após o primeiro trimestre da gravidez ter-se completado.

Informações	Intervalo Intergenésico		Valor-p
	< 6 meses	≥ 6 meses	
<b>Brancas</b>			
Freqüência	286	1338	
Idade	$23,5 \pm 4,2$	$25,8 \pm 4,6$	<0,001
Paridade	$2,7 \pm 1,1$	$2,7 \pm 1,0$	0,924
Solteira	11(3,8)	45(3,4)	0,685
Pré-natal tardio	11(3,8)	24(1,8)	0,030
<b>Negras</b>			
Freqüência	89	209	
Idade	$24,3 \pm 4,5$	$25,1 \pm 4,4$	0,183
Paridade	$2,9 \pm 1,1$	$2,7 \pm 1,2$	0,381
Solteira	9(10,1)	19(9,1)	0,782
Pré-natal tardio	4(4,5)	5(2,4)	0,332

Responda às questões usando  $\alpha = 0,05$ :

- (a) Os dados evidenciam diferença entre mulheres negras e brancas quanto ao intervalo entre gestações? Faça o teste que julgar apropriado, explicitando a hipótese, a estatística do teste, o valor-p e a sua conclusão.
  - (b) O estado civil das mulheres negras influencia no intervalo entre gestações?
  - (c) Verifique se a idade média das mulheres no grupo  $< 6$  meses é estatisticamente diferente da idade média no grupo  $\geq 6$  meses. Interprete os resultados.
    - i. Considerando as brancas.
    - ii. Considerando as negras.
14. Considere as seguintes informações adicionais do estudo descrito no Exemplo 6.11 sobre a comparação do efeito do halotano e

da morfina na pressão sanguínea: dos 61 pacientes anestesiados com halotano 8 (13,1%) morreram, assim como 10 dos 67 (14,9%) pacientes anestesiados com morfina.

- (a) Compare os valores das estatísticas (6.9) e (6.10).
  - (b) Ao nível de significância de 5% use o teste Z e o teste qui-quadrado para avaliar em termos de mortalidade se os dois anestésicos são equivalentes.
15. Dezenove crianças com diagnóstico de AIDS foram separadas em dois grupos de acordo com a susceptibilidade à droga. Considera-se suscetível quando o vírus HIV é inibido por concentração de zidovudina (AZT) menor do que  $0,1 \mu\text{g/l}$  e resistente quando a inibição exige nível acima de  $10 \mu\text{g/l}$ . A duração em meses da terapia com AZT relatada por Ogino (1993) é mostrada a seguir.
- | Suscetíveis<br>(n = 10) | Resistentes<br>(n = 9) |
|-------------------------|------------------------|
| 9                       | 12                     |
| 7                       | 14                     |
| 12                      | 5                      |
| 18                      | 14                     |
| 10                      | 15                     |
| 10                      | 17                     |
| 10                      | 13                     |
| 13                      | 13                     |
| 13                      | 12                     |
| 9                       |                        |
- (a) Para cada grupo calcule a média e o desvio-padrão do tempo de tratamento com AZT.
  - (b) Existe diferença significativa entre os dois grupos no que diz respeito ao tempo de terapia com AZT? Formule a hipótese de interesse e teste-a ao nível de significância de 5%. Calcule o valor-p e interprete-o.
16. Um estudo realizado em Recife (Linhares et al., 1986) comparou crianças entre 7 e 17 anos de classes sócio-econômicas elevadas

(grupo controle ou grupo 1) com crianças desprivilegiadas, de camadas sociais mais baixas (grupo 2). As crianças do grupo controle vinham de 15 escolas particulares da cidade, enquanto as do outro grupo, de orfanatos e casas de caridade ou de outras áreas da cidade, onde cheiravam cola, roubavam ou vendiam frutas, chocolates e doces.

Amostras de sangue das crianças foram colhidas e um dos aspectos estudados foi a concentração total de proteínas plasmáticas e também níveis de albumina e globulina, responsáveis, entre outras coisas, pela produção de pressão osmótica a nível de membrana capilar e pela imunidade natural e adquirida que o organismo possui, respectivamente. Os resultados obtidos, em  $\text{g/l}$ , estão sintetizados na tabela a seguir.

Informações	Proteína total	Albumina	Globulina
Grupo 1			
Média	69,2	32,4	15,4
Desvio-padrão	5,1	4,0	3,1
n	412	412	412
Grupo 2			
Média	70,0	30,3	18,9
Desvio-padrão	5,9	4,0	4,3
n	196	196	196

- (a) Verifique se existe diferença significativa na concentração total de proteínas entre os dois grupos. Escreva as hipóteses de interesse e a conclusão.
  - (b) A partir unicamente do resultado obtido em (a), pode-se concluir a presença ou não de desnutrição nas crianças desprivilegiadas? (Sugestão: análise do significado das proteínas albumina e globulina).
17. Uma das questões mais controversas da Medicina contemporânea diz respeito à eutanásia e suicídio assistido por médicos em pacientes com doenças terminais. Além do "direito de morrer", outros aspectos vêm sendo questionados, desde a durabilidade

do desejo pela morte até fatores que o influenciem, especialmente aqueles remediáveis com intervenção médica e social. Existe hoje uma preocupação de que o desejo pela morte, mesmo entre doentes terminais, pode ser um indicativo de desordem psiquiátrica potencialmente tratável.

Chochinov et al. (1995) entrevistaram 199 pacientes de dois hospitais em Winnipeg, Canadá, com diagnóstico primário de câncer terminal, admitidos no período de coletas de dados. Uma das preocupações era verificar a existência de relação entre o desejo de morte relatado por alguns deles e a depressão diagnosticada em alguns casos. A tabela abaixo mostra os dados obtidos para essa variável (diagnóstico de depressão).

Desejo pela morte	Depressão		
	Presente	Ausente	Total
Presente	10	7	17
Ausente	14	168	182
Total	24	175	199

- (a) Formule a hipótese de interesse e teste-a ao nível de significância de 5%.
- (b) Calcule o valor-p.
- (c) Qual é a relevância do resultado para a discussão atual sobre eutanásia?
18. O consumo de pequenas quantidades de álcool tem sido recomendado a mulheres que amamentam a fim de ajudar à lactação. Entretanto, não há evidências científicas para tanto. As crianças são expostas ao álcool do leite de suas mães, que representam uma pequena fração do total ingerido, sendo mínimos os efeitos a longo prazo. A ocorrência de alteração nas características sensoriais do leite ou no comportamento dos lactentes foi objeto de estudo de Mennella & Beauchamp (1991).

Durante dois dias, com intervalo de uma semana, doze mulheres consumiram suco de laranja puro ou contendo etanol (0,3 g/kg de massa corporal) antes de amamentarem suas crianças e meia-hora após ter sido obtida uma amostra do leite. Metade das

mulheres bebeu o etanol no suco de laranja no primeiro dia de teste e o suco de laranja puro no segundo. A ordem foi invertida para as seis mulheres remanescentes. Amostras de leite para o estudo foram obtidas 30 minutos, 1, 2 e 3 horas após o consumo da bebida, de modo a interromper o mamar possível a amamentação.

Os bebês alimentaram-se livremente e o total de leite ingerido foi estimado a partir de cálculos sobre o peso da criança e a densidade do leite humano. Já a duração da amamentação (tempo) foi obtida através de gravações em vídeo.

Esses dados encontram-se expostos na tabela a seguir, juntamente com os resultados do teste t pareado (11 graus de liberdade) que compara o comportamento das crianças de acordo com a bebida tomada por suas mães.

Variável	Bebida alcóolica		Teste t	Valor p
	Sim	Não		
Leite ingerido (ml)	156,4 ± 8,2	120,4 ± 9,5	4,69	< 0,001
Tempo (min)	28,6 ± 7,7	28,2 ± 7,3	0,15	0,88

- (a) Para cada variável, formule as hipóteses que foram testadas.
- (b) Existe associação entre a presença ou não de álcool na bebida da mãe e o total de leite ingerido pelo filho? Justifique.
- (c) Interprete o valor-p para a variável duração da amamentação.
19. A existência de relação entre criatividade e certos distúrbios mentais sempre foi alvo de especulações. Relatos mais antigos já indicam taxas mais altas de esquizofrenia, psiconeuroses e, mais modernamente, distúrbios afetivos, depressão e alcoolismo. Estudos clínicos e científicos ao longo dos anos não foram capazes de resolver essas controvérsias. Diferenças na conceituação e medida da criatividade, nos critérios de diagnóstico psiquiátrico, na metodologia e interpretação fazem com que comparações entre esses estudos sejam impraticáveis.

Um dos trabalhos mais importantes para essa discussão foi publicado por Ludwig em 1992, no qual ele compara o comportamento criativo e psicopatologia entre diversos grupos profissionais, a partir de pelo menos uma fonte biográfica bem documentada e bem estudada.

Mais de 1000 personalidades eminentes, já falecidas e que viveram pelo menos uma parte de suas vidas no século XX foram divididas em dezenove categorias profissionais e observações comportamentais e julgamentos sobre eles, realizadas com base em evidências científicas. Um dos aspectos abordados no estudo envolveu a comparação entre dois grandes grupos, um com os integrantes da chamada “arte criativa” (arquitetos, pintores, músicos, poetas, escritores, entre outros) e outro com aqueles de outras profissões (por exemplo, banqueiros, atletas, filósofos e cientistas).

- (a) Sabendo que 38,7% dos membros do grupo de “arte criativa” ( $n = 555$ ) e 16,4% dos membros de outras profissões ( $n = 450$ ) experimentaram depressão antes dos 40 anos, avalie a existência de diferença significativa entre esses grupos.
- (b) Outro objeto de estudo foi a existência de diferenças na psicopatologia para integrantes do grupo artístico. A tabela a seguir mostra a incidência de suicídio como causa de morte para poetas e escritores de ficção.

Causa da morte	Poetas	Escritores de ficção	Total
Suicídio	10	7	17
Outras	43	173	216
Total	53	180	233

Formule a hipótese de interesse e teste-a ao nível de significância de 5%. Calcule o valor-p.

20. Em um estudo sobre o efeito de condições ambientais sobre a qualidade do sono de pacientes, Pimentel-Souza et al. (1996) compararam dois hospitais de Belo Horizonte: o Hospital das

Clínicas da UFMG ( $n = 13$ ), representando um hospital localizado em uma região barulhenta, e o Hospital da Baleia ( $n = 22$ ), considerado um hospital mais silencioso. Os percentuais de queixas dos pacientes quanto a fatores que podem atrapalhar o sono observados nos dois hospitais e os valores-p do teste exato de Fisher encontram-se na tabela abaixo.

Condição ambiental	Hospital (%)		Valor-p
	Baleia	Clínicas	
Barulho	31,82	69,23	0,043
Acomodação ruim	9,09	7,69	1,000
Temperatura	27,2	15,38	0,680
Presença ou atenção a outra pessoa	9,09	38,46	0,075
Umidade, mofo, cheiro, pó	0,00	7,69	0,371
Luz	22,73	46,15	0,258
Silêncio excessivo	4,55	0,00	1,000
Desadaptação a nova situação	9,09	15,38	0,618
Outros	0,00	7,69	0,371

- (a) Justifique o uso do teste exato de Fisher.
- (b) Comente os resultados mostrados na tabela.

## Capítulo 7

# Medida do Efeito de uma Intervenção ou Exposição

### 7.1 Introdução

Nos estudos do tipo caso-controle, estamos interessados em verificar se o grupo de pacientes com a patologia de interesse foi mais (ou menos) exposto ao fator de risco em análise. Ao verificarmos, através de testes de hipóteses, a *existência* de efeito da exposição, demos uma resposta parcial a essa questão.

Neste capítulo vamos responder à mesma pergunta *medindo*, entretanto, o efeito da exposição. Essa abordagem é, em muitos aspectos, mais completa e útil do que a apresentada no capítulo anterior. Por isto tornou-se a forma padrão de análise da associação entre fator de risco e doença nas revistas científicas da área médica. Sua utilidade se estende, naturalmente, aos outros tipos de estudos clínicos, entre eles estudos de coorte e ensaios clínicos.

Como fizemos no Capítulo 6, vamos dividir a apresentação do tema de acordo com o tipo da resposta de interesse (contínua ou dicotómica) e da forma de coleta de dados utilizada (grupos independentes ou pareamento). No entanto, para tratar destes assuntos, é preciso conhecer alguns resultados gerais, que serão introduzidos nas seções 7.2 e 7.3.

### 7.2 Conceitos fundamentais

Nesta seção vamos apresentar os conceitos básicos necessários para a compreensão do capítulo. Utilizaremos para isto o contexto de uma análise de dados em que existe apenas uma variável resposta de interesse, descrita pela distribuição gaussiana.

#### Exemplo 7.1: Níveis plasmáticos de vitamina A

Moura (1990) avaliou os níveis plasmáticos de vitamina A em um grupo de 47 crianças diabéticas com idade até 12 anos. Um de seus interesses era conhecer o nível sanguíneo da vitamina A neste grupo, composto de pacientes típicos dos atendidos pelo setor de Endocrinologia Pediátrica da Faculdade de Medicina da UFMG.

#### 7.2.1 Parâmetro de interesse

Em termos estatísticos o objetivo do estudo descrito no Exemplo 7.1 é conhecer o nível médio,  $\mu$ , da distribuição do nível sanguíneo de vitamina A em crianças diabéticas. Neste caso,  $\mu$  é o *parâmetro de interesse*.

Em geral, para a análise estatística dos dados de um problema clínico é preciso, primeiramente, identificar os parâmetros de interesse, isto é, aqueles cujo conhecimento viabiliza a solução da questão colocada. Esta identificação exige de quem analisa os dados experiência clínica relevante, intuição e conhecimentos estatísticos.

A tomada de qualquer decisão deve ser baseada no valor do parâmetro de interesse que, entretanto, não é conhecido na prática. Para atender a esta necessidade foram desenvolvidos métodos agrupados na chamada Teoria da Estimação, uma grande e importante área da Estatística.

Aí se aprende como, a partir de dados amostrais, podemos obter valores que se aproximam do verdadeiro valor do parâmetro. Também pode-se entender exatamente como estas aproximações devem ser construídas e de que forma estão próximas do valor do parâmetro. Este capítulo trata dos aspectos mais simples e sempre tendo como referência as necessidades dos estudos clínicos.

## 7.2.2 Estimadores

É razoável admitir que o nível sanguíneo de vitamina A tem distribuição gaussiana. Com esta especificação, o problema do Exemplo 7.1 se reduz a encontrar estimativas para os dois parâmetros da distribuição ( $\mu$  e  $\sigma$ ).

A Teoria Estatística tem várias soluções para este problema. A mais comum é dada pelo *método da máxima verossimilhança*. Em termos intuitivos este método baseia-se no seguinte princípio: entre todas as amostras possíveis, aquela observada é a que tem maior probabilidade.

Assumindo que os valores dos níveis da vitamina A são representados por  $X_1, X_2, \dots, X_n$  ( $n = 47$ ), e representando  $\hat{\mu}$  e  $\hat{\sigma}^2$  os estimadores fornecidos pelo método da máxima verossimilhança, temos que

$$\hat{\mu} = \bar{X} \quad (7.1)$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} \quad (7.2)$$

Um grande número de autores recomenda, entretanto, para  $\sigma^2$  o seguinte estimador

$$\hat{S}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} \quad (7.3)$$

Existem motivos teóricos e práticos para esta indicação. Sem explicitar as justificativas, cujo entendimento exige conhecimentos que ultrapassam o nível teórico pretendido para este texto, vamos adotá-la.

Estes dois estimadores são chamados de pontuais, pois fornecem apenas o valor da estimativa do parâmetro. A questão da sua variabilidade é tratada pela construção do intervalo de confiança, assunto da próxima seção.

## 7.2.3 Intervalo de confiança

Com a construção de intervalos de confiança agregamos ao estimador pontual informação sobre sua variabilidade. Isto é feito escolhendo-se um limite inferior e outro superior para a estimativa. Descutimos a seguir o sentido preciso destes limites e a forma de sua obtenção.

Para construirmos o intervalo de confiança para  $\mu$ , usaremos o seguinte resultado da Teoria Estatística, que é justificado intuitivamente na próxima seção

$$\frac{\bar{X} - \mu}{\frac{S_x}{\sqrt{n}}} \sim t_{n-1} \quad (7.4)$$

onde o símbolo  $\sim$  significa “ter distribuição”,  $S_x = \sqrt{\frac{\sum(X_i - \bar{X})^2}{n-1}}$  e  $t_{n-1}$  representa a distribuição  $t$  com  $n-1$  graus de liberdade, cuja tabela está disponível no apêndice A5. Ou seja, mesmo não se conhecendo  $\mu$ , podemos calcular probabilidades envolvendo a estatística

$$T = \frac{\bar{X} - \mu}{\frac{S_x}{\sqrt{n}}}, \quad (7.5)$$

que nos permite construir o *intervalo de confiança* para  $\mu$ .

Para isto, através da tabela da distribuição  $t_{n-1}$  obtemos um valor  $t^*$  tal que

$$Pr[-t^* \leq t_{n-1} \leq t^*] = 1 - \alpha$$

ou seja,

$$Pr\left[-t^* \leq \frac{\bar{X} - \mu}{\frac{S_x}{\sqrt{n}}} \leq t^*\right] = 1 - \alpha$$

expressão que pode ser reescrita como

$$Pr\left[\bar{X} - t^* \frac{S_x}{\sqrt{n}} \leq \mu \leq \bar{X} + t^* \frac{S_x}{\sqrt{n}}\right] = 1 - \alpha$$

Assim, um *intervalo confiança* para  $\mu$  com nível de confiança de  $100(1 - \alpha)\%$  é dado por

$$\left[ \bar{X} - t^* \frac{S_x}{\sqrt{n}} ; \bar{X} + t^* \frac{S_x}{\sqrt{n}} \right] \quad (7.6)$$

Veremos na próxima seção que  $\frac{S_x}{\sqrt{n}}$  é uma estimativa do desvio-padrão de  $\bar{X}$ . Assim, o intervalo de confiança para  $\mu$  é centrado na estimativa do efeito, e varia de uma quantidade  $t^*$  desvios-padrão para baixo até o mesmo número de desvios-padrão para cima.

#### Exemplo 7.2: Níveis plasmáticos de vitamina A (continuação)

Para os dados do Exemplo 7.1,  $\bar{x} = 25,5 \text{ mcg/dl}$  e  $s = 8,5 \text{ mcg/dl}$ . Para um nível de confiança de 95%, o percentil da distribuição com 46 graus de liberdade é  $t_{0,975}^* = 2,0129$ . Portanto, o intervalo de confiança para a média do nível plasmático de vitamina A é [23,00; 28,00]. Com 95% de confiança podemos afirmar que o nível plasmático de vitamina A em crianças diabéticas com idade até 12 anos varia entre 23  $\text{mcg/dl}$  e 28  $\text{mcg/dl}$ .

#### 7.2.4 Distribuição da média amostral

O objetivo desta seção é apresentar os resultados da Teoria Estatística, usados na construção de intervalos de confiança. Esta seção apresenta resultados teóricos e pode ser omitida em uma primeira leitura. É, entretanto, fundamental para um entendimento mais profundo da forma de construção dos intervalos de confiança.

São três os resultados a serem apresentados: a fórmula do desvio-padrão da média, a distribuição da média amostral e sua distribuição padronizada pelo desvio-padrão.

Consideremos que a variável aleatória nível da hemoglobina em mulheres jovens gozando de boa saúde tenha distribuição gaussiana com média 12 e desvio-padrão 1. Vamos tirar sucessivas amostras de tamanho 6 desta população e para cada amostra, calcular sua média. Para 50 repetições, obtivemos os resultados da Tabela 7.1.

Tabela 7.1: Amostras de tamanho 6 e suas médias

Amostra	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	Média
1	11,7	11,3	11,3	11,8	11,7	12,7	11,8
2	12,9	12,5	9,9	11,7	12,9	12,3	12,0
3	12,3	13,5	12,8	12,6	10,9	12,8	12,5
4	13,1	11,4	11,9	13,5	12,6	13,3	12,6
5	12,4	14,0	12,0	12,5	12,1	11,4	12,4
6	12,0	12,0	12,5	13,5	11,1	12,3	12,2
7	12,6	12,7	10,9	10,7	10,9	13,8	11,9
8	11,6	12,3	11,8	13,2	13,5	13,5	12,7
9	12,7	12,8	13,4	11,4	11,2	12,9	12,4
10	9,7	13,4	11,2	12,0	12,6	11,4	11,7
:	:	:	:	:	:	:	:
50	11,2	11,4	12,1	13,4	11,9	11,5	11,9

Na Figura 7.1 apresentamos o histograma da coluna de médias, que tem média 12,072 e o desvio-padrão 0,359.

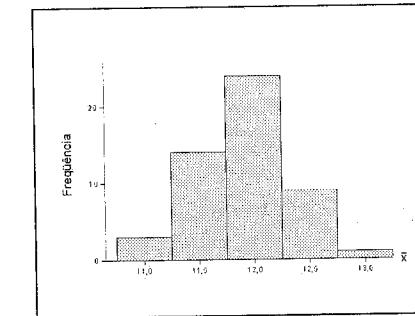


Figura 7.1: Histograma das médias da Tabela 7.1

Os dois primeiros resultados que devemos explicar referem-se aos valores teóricos da média e do desvio-padrão da população de médias, uma amostra da qual está apresentada na última coluna da Tabela 7.1.

O primeiro resultado diz que

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} \quad (7.7)$$

ou seja, que o desvio-padrão da média é dado pelo desvio-padrão da população dividido por  $\sqrt{n}$ .

Note o leitor que o desvio-padrão da média, também chamado de *erro padrão*, é sempre menor que o desvio-padrão da população para  $n \geq 2$ . Intuitivamente este resultado é esperado já que a média de um conjunto de valores, de uma mesma população, deve variar menos que os próprios valores. Além disso, o desvio-padrão da média diminui à medida que cresce o tamanho da amostra, mas precisamente à medida que cresce a raiz quadrada do tamanho da amostra.

Neste exemplo deveríamos obter  $\frac{1}{\sqrt{6}} = 0,41$ . A diferença entre o valor teórico de 0,41 e o valor observado de 0,359 advém do fato de estarmos usando uma amostra de tamanho 50.

O segundo resultado, *Teorema Central do Limite*, diz que a média tem aproximadamente distribuição gaussiana centrada na média da população e desvio-padrão dado pela expressão (7.7), isto é,

$$\bar{X} \sim N \left( \mu, \frac{\sigma}{\sqrt{n}} \right) \quad (7.8)$$

O histograma da Figura 7.1 sugere a validade teórica deste resultado. Repetindo o processo de geração de médias para 1.000 amostras, obtemos a Figura 7.2, cujo aspecto gaussiano é mais claro.

O terceiro resultado é devido a W. Gosset que descobriu, no princípio do século, que a razão

$$T = \frac{\bar{X} - \mu}{\frac{s_x}{\sqrt{n}}} \quad (7.9)$$

tem distribuição que não depende de  $\mu$  e  $\sigma^2$ , mas apenas do tamanho da amostra. Produziu as primeiras tabelas para esta nova distribuição que recebeu o nome de *t de Student*, pseudônimo utilizado por Gosset para publicação do resultado.

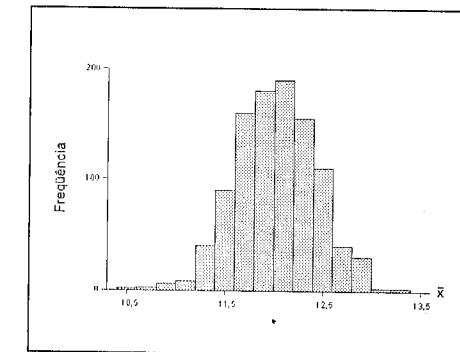


Figura 7.2: Histograma das médias de 1.000 amostras de tamanho 6

### 7.3 Medida do efeito: resposta contínua

O objetivo desta seção é mostrar como o efeito pode ser estimado a partir dos dados e como deve ser tratada a questão da variabilidade desta estimativa para os dois tipos de coleta de dados: amostras emparelhadas e independentes.

Consideremos uma situação em que os pacientes do estudo estão divididos em dois grupos e a resposta de interesse é contínua. Um exemplo típico é um ensaio clínico para verificar se uma droga tem efeito hipotensor. Nesta circunstância, o comportamento global em cada grupo é sintetizado pela média aritmética dos valores da variável resposta. O *efeito da intervenção* é definido como a diferença das médias das respostas dos pacientes nos dois grupos.

#### 7.3.1 Amostras pareadas

A variável resposta é denotada por  $X_1$  e  $X_2$ , respectivamente para os dois grupos a serem comparados. Os dados são pares de observações:  $(x_{11}, x_{12}), (x_{21}, x_{22}), \dots, (x_{n1}, x_{n2})$ . É bastante intuitivo pensar que o efeito de intervenção para cada par é medido pelas diferenças  $d_1 = x_{11} - x_{12}, d_2 = x_{21} - x_{22}, \dots, d_n = x_{n1} - x_{n2}$  e que a estimativa do efeito  $\mu_1 - \mu_2 = \mu_d$  é dada por

$$\bar{d} = \bar{x}_1 - \bar{x}_2 \quad (7.10)$$

Obtida esta estimativa, é natural perguntar-se sobre a variabilidade associada a este valor. A resposta é dada pelo *intervalo de confiança*, que neste caso é

$$\left[ \bar{d} - t^* \frac{s_d}{\sqrt{n}}, \quad \bar{d} + t^* \frac{s_d}{\sqrt{n}} \right] \quad (7.11)$$

onde  $t^*_{1-\alpha/2}$  é o percentil de ordem  $1 - \alpha/2$  da distribuição  $t$  com  $n-1$  graus de liberdade e  $s_d$  é o desvio-padrão das diferenças, isto é,  
 $s_d = \sqrt{\frac{\sum(d_i - \bar{d})^2}{n-1}}$ .

### Exemplo 7.3: Avaliação de redução da pressão intraocular

Em estudo realizado no Hospital São Geraldo da UFMG, Cronemberg & Calixto (1991) estudaram a capacidade de redução da pressão intraocular das drogas timolol, betaxolol e levobunolol. Para isto, medicaram 10 pacientes com as três drogas e compararam os resultados com aqueles obtidos quando os pacientes receberam um placebo. As medidas da pressão intraocular, expressas em  $mmHg$ , obtida às 6 horas da manhã com os pacientes usando timolol e placebo são apresentadas na Tabela 7.2.

Tabela 7.2: Pressão ocular de pacientes em uso de placebo e timolol

Placebo	22	25	23	18	24	24	17	23	22	23
Timolol	18	20	20	17	16	20	20	20	20	24
Efeito	-4	-5	-3	-1	-8	-4	3	-3	-2	1

Por exemplo, o efeito do timolol para o primeiro paciente foi de reduzir a pressão ocular em  $4 mmHg$ . Para descrever o efeito global dos 10 pacientes considerados, usamos  $\bar{d}$ , a média do efeito em cada um dos 10 pares. Neste caso  $\bar{d} = -2,6 mmHg$  é uma estimativa do efeito hipotensor do timolol, ou seja, esses 10 pacientes fornecem evidências de que o timolol diminui a pressão ocular em  $2,6 mmHg$ .

Além disso,  $s_d = 3,098$ ,  $\frac{s_d}{\sqrt{n}} = 0,98$ ,  $t^* = 2,262$ . Assim, o intervalo de confiança com nível de 95% para o efeito hipotensor do timolol é

$$[-2,6 - (2,262)(0,98) ; -2,6 + (2,262)(0,98)] = [-4,8 ; -0,4]$$

Este intervalo nos dá uma boa representação da variabilidade com que conseguimos medir o efeito hipotensor do timolol. Podemos afirmar com 95% de confiança que o timolol reduz a pressão ocular por uma quantidade que varia de  $0,4 mmHg$  a  $4,8 mmHg$ . A não inclusão do zero no intervalo indica, ao nível de confiança de 95%, a existência do efeito, mesmo resultado que obteríamos através do uso do teste  $t$  para dados pareados, descrito no Capítulo 6.

### 7.3.2 Amostras independentes

É muito comum em ensaios clínicos a necessidade de realizar comparações entre grupos independentes de pacientes. Chamando as observações do primeiro grupo de  $x_{11}, \dots, x_{n_1}$ , e as do segundo  $x_{12}, \dots, x_{n_2}$ , a *estimativa do efeito* é dada por

$$\bar{x}_1 - \bar{x}_2 \quad (7.12)$$

Como no caso anterior, este valor é apenas a estimativa do verdadeiro valor do efeito  $\mu_1 - \mu_2$ .

No Capítulo 6 usamos o seguinte resultado demonstrado pela Teoria Estatística:

$$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2} \quad (7.13)$$

onde

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \quad (7.14)$$

Obtendo da tabela da distribuição  $t$  com  $n_1 + n_2 - 2$  graus de liberdade um número  $t^*$  tal que  $Pr[-t^* \leq t_{n_1+n_2-2} \leq t^*] = 1 - \alpha$ , onde  $1 - \alpha$  é o nível de confiança escolhido.

Podemos escrever o intervalo de confiança para o efeito  $\mu$  como

$$\left[ (\bar{x}_1 - \bar{x}_2) - t^* s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} ; (\bar{x}_1 - \bar{x}_2) + t^* s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right] \quad (7.15)$$

Um exemplo ilustra que os cálculos necessários são razoavelmente simples.

#### Exemplo 7.4: Comparação da tianeptina com placebo

Utilizando os dados obtidos no Exemplo 6.9 do Capítulo 6, podemos obter uma estimativa do efeito da tianeptina ao fim de 42 dias de uso.

Grupo placebo	$n_1 = 15$	$\bar{x}_1 = 20,53$	$s_1 = 11,09$
Grupo tianeptina	$n_2 = 16$	$\bar{x}_2 = 11,37$	$s_2 = 7,26$

1. Estimativa do efeito da droga:  $20,53 - 11,37 = 9,16$

2. Intervalo de 95% de confiança:

$$(20,53 - 11,37) \pm (2,0452)(9,31) \sqrt{\frac{1}{15} + \frac{1}{16}} = \\ [9,16 - 6,84; 9,16 + 6,84] = [2,32 ; 16,00]$$

Ou seja, a tianeptina tem claro efeito, ainda que medido com grande variabilidade, neste ensaio clínico.

## 7.4 Medida do efeito: resposta dicotômica

A variável resposta pode ser dicotômica em qualquer tipo de estudo clínico. No entanto, é o tipo padrão de resposta em estudos tipo caso-controle e em muitos estudos de coorte.

A Tabela 7.3 apresenta, neste caso, as proporções dos pacientes correspondentes a cada possível classificação. Para exemplificar, vamos supor que  $P_1 = 0,20$ ,  $P_0 = 0,10$  e  $p = 0,4$ . Obtemos a distribuição da Tabela 7.4.

Tabela 7.3: Proporções em estudos caso-controle ou coorte

Doença	Fator		Total
	Exposto	Não-exposto	
Presente	$pP_1$	$qP_0$	$pP_1 + qP_0$
Ausente	$pQ_1$	$qQ_0$	$pQ_1 + qQ_0$
Total	$p$	$q$	1

Tabela 7.4: Exemplo numérico da Tabela 7.3

Doença	Fator		Total
	Exposto	Não-exposto	
Presente	0,08	0,06	0,14
Ausente	0,32	0,54	0,86
Total	0,4	0,6	1

Em um estudo de coortes, as proporções  $p$  e  $q = 1 - p$  indicam simplesmente o tamanho relativo das coortes do estudo. Entre os *pacientes expostos*  $P_1$  é a proporção dos que desenvolveram a doença e  $P_0$  a proporção análoga entre os *pacientes não expostos*.

Os dados amostrais, em estudos tipo caso-controle e de coorte, são usualmente apresentados como na Tabela 7.5.

Tabela 7.5: Freqüências observadas em estudos caso-controle ou coorte

Doença	Fator		Total
	Presente	Ausente	
Presente	$a$	$b$	$a + b$
Ausente	$c$	$d$	$c + d$
Total	$a + c$	$b + d$	$N$

### 7.4.1 Risco relativo

No estudo de coorte é bastante natural pensar na razão de  $P_1$  e  $P_0$ , como medida de efeito da exposição ao fator. Esta razão recebe o nome de *risco relativo*, será denotada por  $RR$  e definida por

$$RR = \frac{P_1}{P_0} \quad (7.16)$$

O risco relativo é, naturalmente, estimado pela razão das duas proporções amostrais  $\hat{P}_1$  e  $\hat{P}_0$ .

$$\hat{RR} = \frac{\hat{P}_1}{\hat{P}_0} = \frac{\frac{a}{a+c}}{\frac{b}{b+d}} = \frac{a/(a+c)}{b/(b+d)} \quad (7.17)$$

O leitor deve se lembrar de que um ensaio clínico pode ser visto como um estudo de coortes, em que os grupos foram criados através de alocação aleatória de pacientes aos grupos. Portanto, o risco relativo pode ser calculado para ensaios clínicos.

#### Exemplo 7.5: Efeito preventivo da aspirina

Com o objetivo de testar se 325 mg de aspirina (65% de um comprimido usual que é de 0,5 g), ingeridas em dias alternados, reduzem a mortalidade devida a doenças cardiovasculares, um ensaio clínico controlado, duplo-cego, envolveu um total de 22.071 médicos americanos, com idades entre 40 e 84 anos, que não tinham história de infarto do miocárdio, acidente vascular cerebral (AVC) ou ataque isquêmico transitório, não usavam regularmente aspirina e nem apresentavam contra-indicações ao seu uso. Destes, 11.037 receberam aspirina ativa e 11.034, placebo. O experimento teve início em 1982 com os resultados finais publicados pelo *Steering Committee of Physicians' Healthy Group* em 1989. Durante o tempo de seguimento (média de 57 meses), foram observados 139 infartos no grupo que tomava aspirina e 239 no grupo placebo. Assim,  $\hat{P}_1 = \frac{139}{11037} = 0,013$  e  $\hat{P}_0 = \frac{239}{11034} = 0,022$ .

Portanto, uma estimativa do risco relativo é:

$$\hat{RR} = \frac{0,013}{0,022} = 0,59 \implies \frac{1}{\hat{RR}} = \frac{0,022}{0,013} = 1,69$$

O risco de quem tomava aspirina regularmente é 59% do risco dos que não o faziam ou ainda, o risco de quem não toma a droga é 1,69 vezes maior do que o dos usuários.

### 7.4.2 Razão das chances

O risco relativo, embora seja uma medida de efeito com muitas qualidades, tem uma grande limitação: não pode ser estimado em estudo de caso-controle, forma mais freqüente de estudo comparativo. Isto porque, neste tipo de estudo, as incidências observadas são meras consequências do número escolhido de casos e controles e não características dos grupos em estudo. Por isto, buscaram-se maneiras alternativas para se definir o efeito da exposição de forma adequada a todos os estudos epidemiológicos. A solução encontrada foi o uso de uma medida denominada *razão das chances*.

Define-se *chance* de se desenvolver a doença entre os expostos como  $\frac{P_1}{Q_1}$  e entre os não expostos como  $\frac{P_0}{Q_0}$ . Portanto, a *razão das chances* é:

$$\psi = \frac{\frac{P_1}{Q_1}}{\frac{P_0}{Q_0}} = \frac{P_1 Q_0}{P_0 Q_1} \quad (7.18)$$

Este é o parâmetro de interesse para análise de tabelas 2 × 2 e é estimado por

$$\hat{\psi} = \frac{\frac{a/(a+c)}{c/(a+c)}}{\frac{b/(b+d)}{d/(b+d)}} = \frac{ad}{bc} \quad (7.19)$$

Por razões teóricas a variação de  $\hat{\psi}$  é mais facilmente calculada na escala logarítmica. Vale o seguinte resultado

$$Var[\ln \hat{\psi}] = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d} \quad (7.20)$$

Como  $\ln \hat{\psi}$  tem aproximadamente distribuição gaussiana construímos intervalos de confiança para  $\ln \hat{\psi}$  da mesma forma da seção anterior, isto é,

$$\left[ \ln \hat{\psi} - z^* \sqrt{Var[\ln \hat{\psi}]} ; \quad \ln \hat{\psi} + z^* \sqrt{Var[\ln \hat{\psi}]} \right] \quad (7.21)$$

onde  $z^*$  é o percentil obtido de uma tabela da gaussiana padrão tal que  $Pr[-z^* < Z < z^*] = 1 - \alpha$  e  $Var[\ln \hat{\psi}]$  é dada por 7.20.

Considerando que as probabilidades associadas aos valores 0,01 e 0,05, correspondendo a um nível de confiança de 99% e 95%, tomados freqüentemente como referência, temos  $z^*$  igual a 2,33 e 1,96.

Se este intervalo contém o número zero (correspondendo ao valor 1 para a razão de chances) então a associação entre a doença e o fator não é estatisticamente significativa.

Para obter o intervalo de confiança para  $\psi$  basta exponenciar os limites inferior e superior do intervalo (7.21), isto é,  $[e^{L_i}; e^{L_s}]$ .

Dois motivos concorrem para adoção da razão das chances como a forma de se medir associação entre fator de risco e doença, como discutido a seguir.

Usualmente as doenças são raras, isto é,  $P_1$  e  $P_0$  são pequenos e portanto  $Q_1 = Q_0 \cong 1$ . Nesta situação,  $\psi = \frac{P_1}{P_0} = RR$ . Noutras palavras, para doenças raras, a razão das chances é próxima do risco relativo e pode portanto ser interpretada da mesma maneira.

O segundo motivo é que  $\psi$  pode ser estimado com dados de qualquer tipo de estudo. Esta segunda vantagem foi percebida e divulgada por Cornfield (1956), em um artigo fundamental para a epidemiologia moderna. A seguir apresentamos uma justificativa intuitiva para este resultado.

Considerando a Tabela 7.3 como proveniente de um estudo caso-controle, temos que

$$p_1 = Pr[\text{exposição} | \text{caso}] = \frac{pP_1}{pP_1 + qP_0}$$

$$p_0 = Pr[\text{exposição} | \text{controle}] = \frac{pQ_1}{pQ_1 + qQ_0}$$

e portanto podemos verificar que:

$$\frac{p_1/q_1}{p_0/q_0} = \frac{p_1q_0}{p_0q_1} = \frac{P_1Q_0}{P_0Q_1} = \psi \quad (7.22)$$

Conseqüentemente, o risco relativo, aproximado pela razão das chances, pode ser estimado diretamente de um estudo tipo caso-controle, ainda que esta razão não forneça nenhuma informação sobre o valor absoluto das taxas de incidência ( $P_0$  e  $P_1$ ) nos grupos expostos e não-expostos. Esta é uma forte justificativa para o uso da razão das chances como medida de efeito quando a variável resposta é dicotômica.

Para se apreciar completamente a importância deste resultado, alguns outros fatos são necessários. O leitor interessado é convidado a consultar Breslow e Day (1980), um livro clássico da área.

Em estudos pareados com resposta dicotômica é possível estimar a razão das chances. A forma mais simples de se fazer isto, entretanto, é via a combinação de tabelas  $2 \times 2$  (Fleiss, 1981).

#### Exemplo 7.6: Amamentação na infância e câncer de mama

Para verificar se o fato de ter sido amamentado pela mãe é um fator de proteção para o câncer de mama, Freudenheim et al. (1994) realizaram estudo do tipo caso-controle nos condados de Erie e Niagara situados na parte oeste do estado de Nova York (EUA).

As pacientes tomadas como controle foram escolhidas na população da região, não havendo emparelhamento. Os dados obtidos estão apresentados na Tabela 7.6.

Tabela 7.6: Distribuição de casos e controles segundo a amamentação

Grupo	Amamentação	
	Sim	Não
Casos	353	175
Controles	449	153

O risco de desenvolver câncer de mama entre mulheres amamentadas pela mãe, aproximado pela razão de chances, é estimado por

$$\hat{\psi} = \frac{(353)(153)}{(175)(449)} = 0,69$$

ou seja, nominalmente, o risco do grupo amamentado é apenas 69% do risco do grupo não amamentado. Para obtermos um intervalo de confiança para  $\psi$  temos que calcular:

$$\ln \hat{\psi} = \ln(0,69) = -0,35$$

$$Var[\ln \hat{\psi}] = \frac{1}{353} + \frac{1}{175} + \frac{1}{449} + \frac{1}{153} = 0,02$$

ou seja, a estimativa do desvio-padrão de  $\ln \hat{\psi}$  é  $\sqrt{0,02} = 0,14$ . Um intervalo de 95% de confiança para  $\ln \hat{\psi}$  é portanto

$$[-0,37 - 1,96 \times 0,14 ; -0,37 + 1,96 \times 0,14] = [-0,64 ; -0,10]$$

O intervalo para  $\psi$  é então obtido tomando-se a exponencial dos limites inferior e superior deste intervalo, ou seja,  $[0,53 ; 0,90]$ , indicando uma associação significativa entre ter sido amamentada e câncer de mama.

Esse resultado deve ser interpretado com cuidado, uma vez que não foram considerados fatores importantes, como história familiar e idade na primeira gestação. De fato, ao ajustar o modelo incorporando essas variáveis, a associação deixou de ser significativa.

#### Exemplo 7.7: Efeito preventivo da aspirina (continuação)

Os resultados finais do ensaio clínico sobre a eficácia da aspirina na prevenção de doenças coronarianas, são dados pela Tabela 7.7. Nesta tabela a estimativa do risco relativo e intervalo de confiança foram obtidos por métodos, não descritos aqui, que consideram o tempo de acompanhamento de cada indivíduo.

Quem tomou aspirina teve uma redução de 44% na chance de ter infarto do miocárdio (fatal e não fatal) em relação ao grupo que não tomou. Este resultado é estatisticamente significativo já que o intervalo de confiança não contém o valor 1. Quanto ao AVC (fatal e não fatal), quem tomou aspirina teve sua chance aumentada de 22%, em relação aos que não receberam a aspirina. Este valor não é, entretanto, estatisticamente significativo.

Tabela 7.7: Freqüências e riscos relativos referentes a infarto do miocárdio (IM) e acidente vascular cerebral (AVC) nos grupos que receberam aspirina ou placebo

Resultados finais	Grupo		Risco relativo ( $\hat{RR}$ )	I.C. para $RR$ (95%)
	Aspirina	Placebo		
<b>IM</b>				
Fatal	10	26	0,34	0,15 - 0,75
Não fatal	129	213	0,59	0,47 - 0,74
Total	139	239	0,56	0,45 - 0,70
<b>AVC</b>				
Fatal	9	6	1,51	0,54 - 4,28
Não fatal	110	92	1,20	0,91 - 1,59
Total	119	98	1,22	0,92 - 1,60

## 7.5 Exemplo comentado: Desnutrição hospitalar

A desnutrição, um estado mórbido secundário a uma deficiência ou excesso, relativo ou absoluto, de um ou mais nutrientes tem alta prevalência em pacientes internados, especialmente cirúrgicos. Pode relacionar-se à doença de base (câncer, AIDS e doenças inflamatórias intestinais, dentre tantas outras), a componente iatrogênico (jejum prolongado para realização de propedêutica, por exemplo) e também a alterações metabólicas induzidas no trauma operatório. Pacientes cirúrgicos desnutridos têm maior incidência de complicações e índices superiores de mortalidade.

Na avaliação do estado nutricional podem ser empregadas várias abordagens. São utilizadas, por exemplo, avaliação subjetiva global (história clínica completa), medidas de dados antropométricos (peso, altura, índice de massa corporal:  $IMC = \text{peso}/\text{altura}^2$ , medida de pregas cutâneas), análises bioquímicas (contagem total de linfócitos e albumina) e resposta a testes de sensibilidade cutânea.

Correia (1997) estudou pacientes do Sistema Único de Saúde com idade superior a 18 anos, internados em 25 hospitais da rede pública,

conveniados, filantrópicos e universitários de 12 estados do Brasil e do Distrito Federal entre maio e novembro de 1996.

Neste estudo, que recebeu a denominação Inquérito Brasileiro de Nutrição (IBRANUTRI), foram incluídos 4000 casos, 2072 clínicos e 1928 cirúrgicos. Dos pacientes cirúrgicos, 374 referiam-se a cirurgias do aparelho digestivo e hérnias da parede abdominal.

A Tabela 7.8 evidencia o estado nutricional de toda a amostra do estudo, dos pacientes cirúrgicos e dos pacientes submetidos a cirurgias do aparelho digestivo e hérnias da parede abdominal.

Tabela 7.8: Distribuição do estado nutricional de todos os pacientes do IBRANUTRI e dos pacientes cirúrgicos destacando-se os submetidos a cirurgias do aparelho digestivo e hérnias da parede abdominal (CAD+HPA)

Estado nutricional	Amostra geral	Tipo de Cirurgia		
		Geral	CAD+HPA	
Nutrido	2061 (51,5%)	1233 (64,0%)	188 (45%)	
Desnutrido				
moderado	1407 (35,2%)	543 (28,2%)	135 (36%)	
grave	498 (12,4%)	140 (7,3%)	71 (19%)	
Sem diagnóstico	34 (0,8%)	12 (0,6%)	-	
Total	4000 (100%)	1928 (100%)	374 (100%)	

A Tabela 7.9 mostra a distribuição para algumas características dos pacientes submetidos a cirurgias do aparelho digestivo e hérnias da parede abdominal, a estimativa para a razão de chances e o respectivo intervalo de confiança.

Estimou-se que pacientes portadores de neoplasias malignas têm um risco de desnutrição cerca de 8 vezes maior do que os não portadores.

Pacientes internados por um período superior a 7 dias ( $\geq 8$  dias) apresentaram uma chance de desnutrição 2,73 vezes maior do que aqueles internados há menos tempo.

Tabela 7.9: Distribuição para algumas características de pacientes submetidos a cirurgias do aparelho digestivo e hérnias da parede abdominal, a estimativa para a razão de chances ( $\hat{\psi}$ ) e o respectivo intervalo de 95% de confiança

Característica dos pacientes	Pacientes		$\hat{\psi}$	IC para $\hat{\psi}$ (95%)
	Desnutridos	Nutridos		
Idade				
> 60 anos	85 (64,4%)	47 (35,6%)	1,81	1,14 - 2,87
$\leq 60$ anos	121 (50%)	121 (50%)		
Câncer				
Sim	116 (85,5%)	23 (16,6%)	8,13	4,70 - 14,14
Não	90 (38,3%)	145 (61,7%)		
Infecção				
Sim	92 (61,7%)	57 (31,1%)	1,57	1,01 - 2,15
Não	114 (50,7%)	111 (49,3%)		
Internação				
$\geq 8$ dias	114 (69,9%)	49 (30,1%)	2,73	1,66 - 4,51
< 8 dias	92 (43,6%)	119 (56,4%)		

## 7.6 Considerações finais

### 7.6.1 Significância estatística e significância clínica

De acordo com a convenção científica, um achado é estatisticamente significante se o valor-p é menor que um ponto de corte (geralmente 0,05). Por outro lado, um resultado é clinicamente importante se o intervalo de confiança para o parâmetro de interesse difere do valor especificado na hipótese nula por uma quantidade considerada clinicamente relevante.

A Figura 7.3 ilustra os tipos diferentes de resultados na situação em que a inexistência de efeito corresponde ao valor zero e em que W é a quantidade considerada clinicamente importante. Quando o intervalo de confiança representado pelo segmento de reta inclui o zero, o resultado não é significante. Se o efeito estimado é inferior a W, o efeito não é considerado clinicamente importante.

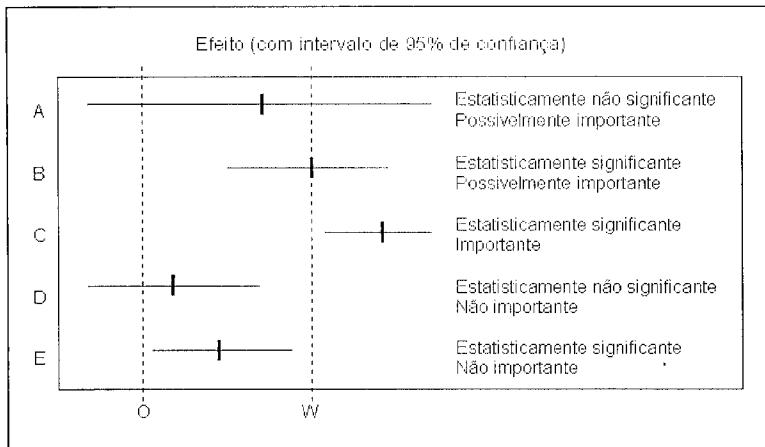


Figura 7.3: Esquema ilustrativo da significância estatística e significância clínica.

### 7.6.2 Equivalência entre teste de hipóteses e intervalo de confiança

Existe equivalência entre teste de hipóteses bilaterais e intervalo de confiança. Pode-se entender que o intervalo de confiança engloba os valores plausíveis do parâmetro para um determinado coeficiente, que é o complemento do nível de significância. Assim, se o intervalo de confiança contém o valor especificado na hipótese nula, não se rejeita a hipótese.

Por exemplo, baseados no intervalo de confiança para a razão de chances ( $\psi$ ) do Exemplo 7.6, podemos testar as hipóteses  $H_0: \psi = 1$  versus  $H_1: \psi \neq 1$ . Como o intervalo de 95% de confiança obtido, [0,53; 0,90], não contém o valor 1, não se rejeita  $H_0$  ao nível de significância de 5%.

### 7.6.3 Diferentes tempos de seguimento

Nas seções anteriores, ao tratar do estudo de coorte, assumimos que tanto os pacientes do grupo exposto quanto os do grupo não exposto foram acompanhados durante o mesmo período de tempo. Isto é mais a exceção do que a regra e portanto, para se descrever o risco relativo em estudo de coortes usa-se muito freqüentemente

a razão de *taxas* de incidência da doença. Neste caso, divide-se o número de doentes pelo número total de unidades de tempo em que foram acompanhados. Alternativamente usam-se modelos estatísticos para o tempo de acompanhamento, ou sobrevivência (Kleinbaum, 1995).

### 7.6.4 O termo *odds-ratio*

Não há consenso em como traduzir o termo *odds-ratio*. Preferimos o nome *razão das chances* que é criticado porque a palavra chance está mais relacionada com probabilidade, conceito diferente daquele captado pela palavra *odds*. Optamos por criar um novo sentido para a palavra chance ao invés de criar um neologismo, como defendido por aqueles que usam termos como *razão de odds*.

## 7.7 Exercícios

- 1.) O Centro de Acompanhamento Pré-Natal, para dependentes de drogas químicas, da Escola de Medicina da Universidade de Northwestern - Chicago acompanhou a gravidez de 55 mulheres dependentes de cocaína. Destas, apesar de todo o esforço do centro, apenas 19 conseguiram parar de usar a droga durante o 1º trimestre. A tabela abaixo apresenta os resultados dos pesos dos recém-nascidos do grupo 1, filhos de mães que usaram cocaína apenas no 1º trimestre de gravidez, e do grupo 2, filhos de mães que usaram cocaína durante toda a gravidez.

Informação	Grupo 1	Grupo 2
Tamanho da amostra	19	36
Média (g)	3160	2829
Desvio-padrão (g)	453	708

- (a) Faça o teste estatístico adequado para verificar a hipótese de que os filhos de mães que usaram cocaína durante toda a gravidez têm peso ao nascer diferente dos filhos de mães que usaram cocaína apenas no primeiro trimestre de gravidez.

- (b) Estime o efeito da cocaína no peso dos recém-nascidos e construa o respectivo intervalo de confiança. Comente os resultados.
2. Trinta pacientes com melanoma foram acompanhados até a morte pelo Centro de Ciências da Saúde de Oklahoma. O objetivo era comparar a imunoterapia C. Parvum (método novo) com o BCG (método padrão) quanto a sua capacidade de prolongar o tempo de sobrevida e o tempo de remissão de indivíduos com a doença. Os tempos de remissão e de sobrevida em meses observados por Calvete (1986) estão sintetizados na tabela abaixo.
- | Tempo de remissão |           | Tempo de sobrevida |           |       |
|-------------------|-----------|--------------------|-----------|-------|
| BCG               | C. Parvum | BCG                | C. Parvum |       |
| <i>n</i>          | 11        | 19                 | 11        | 19    |
| $\bar{x}$         | 13,73     | 11,53              | 17,18     | 14,26 |
| <i>s</i>          | 12,11     | 7,43               | 10,20     | 6,82  |
- (a) Escreva em notação estatística as hipóteses relevantes.  
 (b) Faça os respectivos testes de hipótese.  
 (c) Determine, na precisão possível da tabela do texto, a probabilidade de significância.  
 (d) Determine o efeito do método novo.
3. Freudenheim et al. (1994) publicaram um estudo sobre a influência da amamentação com leite materno no desenvolvimento, anos depois, de câncer de mama. O estudo foi feito no formato de caso-controle. Os casos foram todos recém-diagnosticados e os controles escolhidos na mesma faixa etária e residentes no mesmo local dos casos. A tabela fornece alguns resultados para as pacientes que ainda não tinham entrado na menopausa.
- | Idade | Casos       |     | Controles   |     |
|-------|-------------|-----|-------------|-----|
|       | Amamentação |     | Amamentação |     |
|       | Sim         | Não | Sim         | Não |
| 40-50 | 96          | 108 | 124         | 94  |
| >50   | 16          | 9   | 20          | 8   |
- (a) Para as mulheres com idade entre 40 e 50 anos verifique, usando o teste adequado, se ter sido amamentada pela mãe é fator de proteção para o câncer de mama. Use  $\alpha = 0,05$ .  
 (b) Determine, com a precisão permitida pela tabela, o valor-p associado ao teste feito em (a). Interprete este valor.  
 (c) Para as mulheres com mais de 50 anos, calcule o efeito da amamentação com leite materno e o respectivo intervalo de 95% de confiança. Interprete os resultados.
4. Segundo Grady et al. (1986), a probabilidade de uma mulher casada experimentar uma gravidez indesejada durante o primeiro ano de uso de um dos seguintes métodos de contracepção é estimada por:

Método de contracepção	Probabilidade de gravidez
Nenhum	0,431
Diafragma	0,149
Preservativo	0,106
DIU	0,071
Pílula	0,037

Para cada método listado, calcular o risco de gravidez para mulheres usando o método comparado com o de mulheres não utilizando nenhum tipo de contracepção. Interprete e compare os resultados obtidos.

5. A deficiência de vitamina A, reconhecida por suas complicações oculares, tem mostrado efeitos sistêmicos que contribuem para aumento de mortalidade e morbidade. O sarampo continua sendo uma das doenças infecciosas mais graves em comunidades pobres, e para verificar se a suplementação de vitamina A reduz a morbidade dessa enfermidade, Coutsoudis et al. (1991) conduziram um estudo na África do Sul.

Sessenta crianças de baixas condições sócio-econômicas (idade entre 4 e 24 meses) com sarampo e complicações (pneumonia e diarréia grave), admitidas no *King Edward VIII Hospital* em

Durante foram aleatoriamente divididas em dois grupos. Ambos receberam terapia padrão (antibióticos, antipiréticos e reidratação oral) e monitoramento para as condições apresentadas. O grupo experimental recebeu vitamina A em dose recomendada pela Organização Mundial de Saúde, enquanto o controle recebeu placebo. Vitamina A e placebo foram administrados pela mesma pessoa na admissão e nos dias 2 e 8 (estudo duplo-cego). Não se conhecia o estado nutricional prévio das crianças, mas nenhuma demonstrou sinais de deficiência de vitamina A que exigisse cuidados.

A tabela a seguir mostra alguns resultados para as seguintes variáveis de interesse obtidas a partir do acompanhamento das crianças:  $X_1$ - duração de pneumonia clínica (dias);  $X_2$ - duração da diarréia (dias);  $X_3$ - ganho de peso após 6 semanas ( $Kg$ )  $X_4$ - ganho de peso após 6 meses ( $Kg$ ).

Variável	Placebo	Vitamina A	Valor-p
$X_1$	$5,70 \pm 0,79$ (n=31)	$3,80 \pm 0,40$ (n=29)	0,04
$X_2$	$4,50 \pm 0,35$ (n=31)	$3,20 \pm 0,71$ (n=29)	
$X_3$	$0,90 \pm 0,14$ (n=24)	$1,29 \pm 0,31$ (n=24)	0,04
$X_4$	$2,37 \pm 0,20$ (n=16)	$2,89 \pm 0,23$ (n=20)	

- Calcule o valor-p para as demais variáveis e verifique se existe diferença significativa entre os dois grupos.
  - Estime o efeito da vitamina A com seu respectivo intervalo de 95% de confiança.
6. Trinta e cinco pacientes com doença arterial coronariana, de idade entre 35 e 75 anos, foram submetidos a teste ergométrico (Miller et al., 1990). Os estudos realizaram-se em bicicleta ergométrica, com os exercícios continuando até angina ou fadiga que requeresse a suspensão da atividade. O tempo médio de duração do exercício foi  $716\ s$ , desvio-padrão  $437\ s$ , variando de 240 a 1845 s. A tabela abaixo mostra a pressão sanguínea sistólica em  $mmHg$  dos indivíduos em repouso e no pico do exercício. Calcule o efeito do exercício sobre a pressão sistólica dos participantes da pesquisa e o intervalo de confiança correspondente.

Nº do Paciente	Pressão sanguínea		Nº do Paciente	Pressão sanguínea	
	repouso	exercício		repouso	exercício
1	175	190	19	120	186
2	140	160	20	150	180
3	140	225	21	162	194
4	148	190	22	172	204
5	154	232	23	178	220
6	140	170	24	168	220
7	170	200	25	134	168
8	158	190	26	154	180
9	175	185	27	130	220
10	140	170	28	120	146
11	154	166	29	160	214
12	150	174	30	160	160
13	132	144	31	150	170
14	178	214	32	170	204
15	122	154	33	122	162
16	128	150	34	180	220
17	160	218	35	140	190
18	125	172			

7. Um dos objetivos da pesquisa descrita no exercício 19 do Capítulo 4 era averiguar a associação de alguns fatores de risco para hemorragia peri-intraventricular (HPIV). A partir dos dados da tabela a seguir, estime o risco (*odds ratio*) associado a peso ao nascer inferior a 1500 g e seu respectivo intervalo de confiança.

HPIV	Peso (g)		Total
	< 1500	1500-2000	
Presente	24	15	39
Ausente	32	49	81
Total	56	64	120

8. Baseado nos dados do Exemplo 6.12 do Capítulo 6, estime a redução no nível de colesterol proporcionado pelo programa de exercícios físicos. Construa o intervalo de confiança e interprete.

# Apêndice A

## Tabelas

Tabela A1: Números aleatórios

Linha	Coluna								
	1	2	3	4	5	6	7	8	9
1	6640	5245	2410	5929	9395	2375	9430	7560	9704
2	9199	5866	7747	8516	2092	6931	8586	1494	2947
3	7572	3016	3052	9135	7997	6860	5981	9052	6540
4	6003	5876	8380	7513	5443	3572	7010	5784	6023
5	5232	3748	8135	2104	1983	1781	8599	2345	6241
6	6122	5390	4649	6549	6961	6549	7223	1633	7176
7	4747	1480	4225	4401	6765	6691	2028	6770	9680
8	9195	5493	9353	9430	6236	3120	2689	5482	5634
9	2256	8490	9943	1055	6528	3741	1500	3608	7843
10	2803	7471	2884	4722	7652	5732	3118	5662	9370
11	1616	1914	3103	5856	1470	7273	8376	8479	3771
12	4290	2794	6221	8280	5290	4040	6008	4539	7905
13	2848	8528	5328	2664	4510	7935	9752	9831	5280
14	4030	4827	2277	5001	1227	7892	2646	6564	4700
15	2950	8725	4141	3611	5879	1833	8270	8875	4482
16	6707	3274	4461	5435	5601	5900	2606	1132	5006
17	1762	4679	2700	1729	9674	7035	3150	1836	3260
18	6170	1830	6496	1927	6920	4901	3305	4448	1600

Tabela A2: Distribuição de Poisson:  $\Pr(X = x)$

x	$\lambda$							
	1	2	2,5	3	3,5	5	8	10
0	0,368	0,135	0,082	0,050	0,030	0,007	0,000	0,000
1	0,368	0,271	0,205	0,149	0,106	0,034	0,003	0,000
2	0,184	0,271	0,257	0,224	0,185	0,084	0,011	0,002
3	0,061	0,180	0,214	0,224	0,216	0,14	0,029	0,008
4	0,015	0,090	0,134	0,168	0,189	0,175	0,057	0,019
5	0,003	0,036	0,067	0,101	0,132	0,175	0,092	0,038
6	0,001	0,012	0,028	0,050	0,077	0,146	0,122	0,063
7	0,000	0,003	0,010	0,022	0,039	0,104	0,140	0,090
8	0,000	0,001	0,003	0,008	0,017	0,065	0,140	0,113
9	0,000	0,000	0,001	0,003	0,007	0,036	0,124	0,125
10	0,000	0,000	0,000	0,001	0,002	0,018	0,099	0,125
11	0,000	0,000	0,000	0,000	0,001	0,008	0,072	0,114
12	0,000	0,000	0,000	0,000	0,000	0,003	0,048	0,095
13	0,000	0,000	0,000	0,000	0,000	0,001	0,030	0,073
14	0,000	0,000	0,000	0,000	0,000	0,000	0,017	0,052
15	0,000	0,000	0,000	0,000	0,000	0,000	0,009	0,035
16	0,000	0,000	0,000	0,000	0,000	0,000	0,005	0,013
17	0,000	0,000	0,000	0,000	0,000	0,000	0,007	0,022
18	0,000	0,000	0,000	0,000	0,000	0,000	0,002	0,013
19	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,007
20	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,004
21	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,002
22	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,001

Exemplo:  $\Pr(X = 1) = 0,149$  para  $\lambda = 3$

Tabela A3: Distribuição gaussiana para valores negativos:  
 $\Pr(X \leq x)$

x	Segunda casa decimal de x									
	0	1	2	3	4	5	6	7	8	9
-2,9	0019	0018	0018	0017	0016	0016	0015	0015	0014	0014
-2,8	0026	0025	0024	0023	0023	0022	0021	0021	0020	0019
-2,7	0035	0034	0033	0032	0031	0030	0029	0028	0027	0026
-2,6	0047	0045	0044	0043	0041	0040	0039	0038	0037	0036
-2,5	0062	0060	0059	0057	0055	0054	0052	0051	0049	0048
-2,4	0082	0080	0078	0075	0073	0071	0069	0068	0066	0064
-2,3	0107	0104	0102	0099	0096	0094	0091	0089	0087	0084
-2,2	0139	0136	0132	0129	0125	0122	0119	0116	0113	0110
-2,1	0179	0174	0170	0166	0162	0158	0154	0150	0146	0143
-2,0	0228	0222	0217	0212	0207	0202	0197	0192	0188	0183
-1,9	0287	0281	0274	0268	0262	0256	0250	0244	0239	0233
-1,8	0359	0351	0344	0336	0329	0322	0314	0307	0301	0294
-1,7	0446	0436	0427	0418	0409	0401	0392	0384	0375	0367
-1,6	0548	0537	0526	0516	0505	0495	0485	0475	0465	0455
-1,5	0668	0655	0643	0630	0618	0606	0594	0582	0571	0559
-1,4	0808	0793	0778	0764	0749	0735	0721	0708	0694	0681
-1,3	0968	0951	0934	0918	0901	0885	0869	0853	0838	0823
-1,2	1151	1131	1112	1093	1075	1056	1038	1020	1003	0985
-1,1	1357	1335	1314	1292	1271	1251	1230	1210	1190	1170
-1,0	1587	1562	1539	1515	1492	1469	1446	1423	1401	1379
-0,9	1841	1814	1788	1762	1736	1711	1685	1660	1635	1611
-0,8	2119	2090	2061	2033	2005	1977	1949	1922	1894	1867
-0,7	2420	2389	2358	2327	2296	2266	2236	2206	2177	2148
-0,6	2743	2709	2676	2643	2611	2578	2546	2514	2483	2451
-0,5	3085	3050	3015	2981	2946	2912	2877	2843	2810	2776
-0,4	3446	3409	3372	3336	3300	3264	3228	3192	3156	3121
-0,3	3821	3783	3745	3707	3669	3632	3594	3557	3520	3483
-0,2	4207	4168	4129	4090	4052	4013	3974	3936	3897	3859
-0,1	4602	4562	4522	4483	4443	4404	4364	4325	4286	4247
0,0	5000	4960	4920	4880	4840	4801	4761	4721	4681	4641

A tabela apresenta a parte decimal de  $\Pr[X \leq x]$   
Exemplo:  $\Pr[X \leq -1,96] = 0,0250$

Tabela A3: Distribuição gaussiana para valores positivos:  
 $\Pr(X \leq x)$  (continuação)

x	Segunda casa decimal de x									
	0	1	2	3	4	5	6	7	8	9
0,0	5000	5040	5080	5120	5160	5199	5239	5279	5319	5359
0,1	5398	5438	5478	5517	5557	5596	5636	5675	5714	5753
0,2	5793	5832	5871	5910	5948	5987	6026	6064	6103	6141
0,3	6179	6217	6255	6293	6331	6368	6406	6443	6480	6517
0,4	6554	6591	6628	6664	6700	6736	6772	6808	6844	6879
0,5	6915	6950	6985	7019	7054	7088	7123	7157	7190	7224
0,6	7257	7291	7324	7357	7389	7422	7454	7486	7517	7549
0,7	7580	7611	7642	7673	7704	7734	7764	7794	7823	7852
0,8	7881	7910	7939	7967	7995	8023	8051	8078	8106	8133
0,9	8159	8186	8212	8238	8264	8289	8315	8340	8365	8389
1,0	8413	8438	8461	8485	8508	8531	8554	8577	8599	8621
1,1	8643	8665	8686	8708	8729	8749	8770	8790	8810	8830
1,2	8849	8869	8888	8907	8925	8944	8962	8980	8997	9015
1,3	9032	9049	9066	9082	9099	9115	9131	9147	9162	9177
1,4	9192	9207	9222	9236	9251	9265	9279	9292	9306	9319
1,5	9332	9345	9357	9370	9382	9394	9406	9418	9429	9441
1,6	9452	9463	9474	9484	9495	9505	9515	9525	9535	9545
1,7	9554	9564	9573	9582	9591	9599	9608	9616	9625	9633
1,8	9641	9649	9656	9664	9671	9678	9686	9693	9699	9706
1,9	9713	9719	9726	9732	9738	9744	9750	9756	9761	9767
2,0	9772	9778	9783	9788	9793	9798	9803	9808	9812	9817
2,1	9821	9826	9830	9834	9838	9842	9846	9850	9854	9857
2,2	9861	9864	9868	9871	9875	9878	9881	9884	9887	9890
2,3	9893	9896	9898	9901	9904	9906	9909	9911	9913	9916
2,4	9918	9920	9922	9925	9927	9929	9931	9932	9934	9936
2,5	9938	9940	9941	9943	9945	9946	9948	9949	9951	9952
2,6	9953	9955	9956	9957	9959	9960	9961	9962	9963	9964
2,7	9965	9966	9967	9968	9969	9970	9971	9972	9973	9974
2,8	9974	9975	9976	9977	9978	9979	9979	9980	9981	
2,9	9981	9982	9982	9983	9984	9984	9985	9986	9986	

A tabela apresenta a parte decimal de  $\Pr(X \leq x)$   
Exemplo:  $\Pr(X \leq 1,96) = 0,9750$

Tabela A5: Distribuição  $t$  de Student :  $\Pr(X \leq -x) + \Pr(X \geq x)$ Tabela A4: Distribuição do  $\chi^2$ :  $\Pr(X \geq x)$ 

Graus de liberdade	Pr( $X \geq x$ )						
	0,750	0,250	0,100	0,050	0,025	0,010	0,005
1	0,102	1,323	2,706	3,841	5,024	6,635	7,879
2	0,575	2,773	4,605	5,991	7,378	9,210	10,597
3	1,213	4,108	6,251	7,815	9,348	11,345	12,838
4	1,923	5,385	7,779	9,488	11,143	13,277	14,860
5	2,675	6,626	9,236	11,070	12,833	15,086	16,750
6	3,455	7,841	10,645	12,592	14,449	16,812	18,548
7	4,255	9,037	12,017	14,067	16,013	18,475	20,278
8	5,071	10,219	13,362	15,507	17,535	20,090	21,955
9	5,899	11,389	14,684	16,919	19,023	21,666	23,589
10	6,737	12,549	15,987	18,307	20,483	23,209	25,188
11	7,584	13,701	17,275	19,675	21,920	24,725	26,757
12	8,438	14,845	18,549	21,026	23,337	26,217	28,300
13	9,299	15,984	19,812	22,362	24,736	27,688	29,819
14	10,165	17,117	21,064	23,685	26,119	29,141	31,319
15	11,037	18,245	22,307	24,996	27,488	30,578	32,801
16	11,912	19,369	23,542	26,296	28,845	32,000	34,267
17	12,792	20,489	24,769	27,587	30,191	33,409	35,718
18	13,675	21,605	25,989	28,869	31,526	34,805	37,156
19	14,562	22,718	27,204	30,143	32,852	36,191	38,582
20	15,452	23,828	28,412	31,410	34,170	37,566	39,997
30	24,478	34,800	40,256	43,773	46,979	50,892	53,672
40	33,660	45,616	51,805	55,759	59,342	63,691	66,767
50	42,942	56,334	63,167	67,505	71,420	76,154	79,490
60	52,294	66,982	74,397	79,082	83,298	88,381	91,955

Exemplo: para 1 grau de liberdade,  $\Pr(X \geq 3,841) = 0,050$

Graus de liberdade	Pr( $X \leq -x$ ) + Pr( $X \geq x$ )						
	0,50	0,40	0,20	0,10	0,05	0,02	0,01
1	1,000	1,376	3,078	6,314	12,706	31,821	63,657
2	0,817	1,061	1,886	2,920	4,303	6,965	9,925
3	0,765	0,978	1,638	2,353	3,182	4,541	5,841
4	0,741	0,941	1,533	2,132	2,776	3,747	4,604
5	0,727	0,920	1,476	2,015	2,571	3,365	4,032
6	0,718	0,906	1,440	1,943	2,447	3,143	3,708
7	0,711	0,896	1,415	1,895	2,365	2,998	3,500
8	0,706	0,889	1,397	1,860	2,306	2,897	3,355
9	0,703	0,883	1,383	1,833	2,262	2,821	3,250
10	0,700	0,879	1,372	1,812	2,228	2,764	3,169
11	0,697	0,876	1,363	1,796	2,201	2,718	3,106
12	0,695	0,873	1,356	1,782	2,179	2,681	3,055
13	0,694	0,870	1,350	1,771	2,160	2,650	3,012
14	0,692	0,868	1,345	1,761	2,145	2,625	2,977
15	0,691	0,866	1,341	1,753	2,132	2,603	2,947
16	0,690	0,865	1,337	1,746	2,120	2,584	2,921
17	0,689	0,863	1,333	1,740	2,110	2,567	2,898
18	0,688	0,862	1,330	1,734	2,101	2,552	2,878
19	0,688	0,861	1,328	1,729	2,093	2,540	2,861
20	0,687	0,860	1,325	1,725	2,086	2,528	2,845
30	0,683	0,854	1,310	1,697	2,042	2,457	2,750
40	0,681	0,851	1,303	1,684	2,021	2,423	2,705
60	0,679	0,848	1,296	1,671	2,000	2,390	2,660
120	0,677	0,845	1,289	1,658	1,980	2,358	2,617

A tabela fornece probabilidades acumuladas nas duas caudas.

Exemplo: para 10 graus de liberdade,  
 $\Pr(X \leq -2,228) + \Pr(X \geq 2,228) = 0,05$ ;  $\Pr(X \geq 2,228) = 0,025$

## Apêndice B

# Respostas dos exercícios

### Capítulo 3

3.  $s_{n-1} = 0,49$ ;  $s_n = 0,46$ .

5. Considerando o menor valor como 17 anos e o maior 21 anos:  
 $6s = 4$  e  $s = 0,67$ .

6. (a) Haverá aumento do valor da média correspondente ao número adicionado. O valor do desvio-padrão não sofrerá alteração.

(b) A média e o desvio-padrão ficarão multiplicados pela constante.

7. (a) A média está abaixo de no mínimo um dos dados. É a mediana que se encontra sempre abaixo da metade dos dados.

(b) A média é uma medida representativa apenas para conjunto de dados simétricos.

(c) A frase é muito ambígua. Se entendermos média no sentido estatístico, sempre haverá alunos com nota abaixo da média, pela própria definição desse conceito. No entanto, a palavra média é usada na linguagem coloquial com sentido de padrão (a média para aprovação, por exemplo, é 60). Nesse sentido, a frase expressa uma preocupação correta.

8. (a)  $\bar{x} = 20,26$ ;  $\tilde{x} = 21,00$ .

(b)  $Q_1 = 16,5$ ;  $Q_3 = 23,0$ .

(c)  $CV_A = 0,200$ ;  $CV_B = 0,225$ ;  $CV_C = 0,260$ . Os resultados mostram maior homogeneidade na Turma A, enquanto a maior variabilidade é encontrada na Turma C.

10.  $Q_1 = 8$ ;  $Q_2 = 14$ ;  $Q_3 = 17,5$ .

11. (a)  $\bar{x}_m = 10,71$ ;  $s_m = 6,82$ ;  $\tilde{x}_m = 9$ ;  $CV_m = 0,64$

$\bar{x}_f = 7,17$ ;  $s_f = 3,67$ ;  $\tilde{x}_f = 7$ ;  $CV_f = 0,51$

(b)  $\bar{x} = 8,94$ ;  $s = 5,70$ ;  $\tilde{x} = 7,5$ ;  $CV = 0,64$

12. O valor para albumina dessa pessoa supera as medidas encontradas em 95% dos indivíduos. Já o valor do colesterol supera apenas 5% dos valores.

13. (b)  $Q_1$ ,  $\tilde{x}$  e  $P_{90}$  estão nas seguintes classes: 51 + 57, 60 + 63 e 69 + 72.

(c) Aproximadamente 53, correspondendo a  $P_{20}$ .

14.

(a)

Taxa de colesterol	Frequência absoluta		Frequência relativa	
	simples	acumulada	simples	acumulada
112 + 132	1	1	0,02	0,02
132 + 152	3	4	0,06	0,08
152 + 172	6	10	0,12	0,20
172 + 192	9	19	0,18	0,38
192 + 212	14	33	0,28	0,66
212 + 232	9	42	0,18	0,84
232 + 252	5	47	0,10	0,94
252 + 272	2	49	0,04	0,98
272 + 292	1	50	0,02	1,00
Total	50	-	1	-

(c)  $P_{90} \cong 244$ .

(d) O grupo A (entre 10 e 19 anos) é mais homogêneo, por apresentar menor coeficiente de variação:  $CV_A = 0,17$ ,  $CV_H = 0,19$ .

(e) Há uma tendência de aumento da taxa de colesterol médio com a idade.

16. O procedimento utilizado foi inadequado porque qualquer variação nos valores da amostra trará alterações indesejáveis ao valor da média. A forma mais apropriada de síntese dos dados, nesse caso, é a mediana, pois seu cálculo independe de valores extremos.

17. (a)  $P_{13}$ .

(b) 1850. Opus 87.

(c) Até 1833 havia composto somente 6% de sua obra, portanto, uma perda de 94% de seu trabalho.

(d) Durante esses dois anos, Schumann compôs 35, 12% de sua obra.

18. (b) A expressão “erro em exames” é inadequada para o caso. O que houve, na verdade, foi detecção de variação para as medidas de colesterol e triglicírides. Como não é fornecido um padrão para comparação, não se pode demonstrar se a variação encontrada é aceitável ou se estão configurados realmente erros nos exames.

21. (c) 73,7%

(d) 94,5%

22. (a) 39 meninas.

(b) 21

## Capítulo 4

1. (a) i. 0,0488 ii. 0,4729 iii. 0,0137 iv. 0,0800  
(b) i. 0,0442 ii. 0,0001 iii. 0,0024 iv. 0,9048  
(d) 49 indivíduos.

2.  $Pr[A | B] = Pr[B | A] = 0$ .

3. (a)  $s = 0,953$ ;  $e = 0,998$ .  
(b)  $VPP = 0,0667$ .

4. Alterando-se a faixa de normalidade de 75 a 85 para 75 a 120, alguns dos indivíduos, antes considerados doentes, passam a apresentar teste negativo, significando diminuição de sensibilidade e aumento de especificidade. Ao variar a faixa para 50 a 80, acontece um pro-

blema inverso. Há uma diminuição da especificidade e aumento da sensibilidade porque alguns testes negativos tornaram-se positivos.

5. (a) Geralmente, ao se aumentar a prevalência, verifica-se aumento significativo do valor da predição positiva, enquanto o valor da predição negativa sofre pequena variação.

6. O valor de predição positiva é dado por  $a/(a + c)$  quando a prevalência populacional é igual à prevalência da tabela, o que raramente ocorre na prática.

7. Para  $p = 0,05$ :  $VPP = 0,139$ ;  $VPN = 0,986$ ;  $PPV = 0,861$ ;  $PFN = 0,014$ .

Para  $p = 0,10$ :  $VPP = 0,254$ ;  $VPN = 0,970$ ;  $PPV = 0,746$ ;  $PFN = 0,030$ .

Para  $p = 0,15$ :  $VPP = 0,351$ ;  $VPN = 0,954$ ;  $PPV = 0,649$ ;  $PFN = 0,046$ .

Conforme o esperado, a um aumento na prevalência segue-se no o aumento no valor de  $VPP$  e a pouca alteração de  $PPV$ .

8. (a)  $s = 0,733$ ;  $e = 0,972$ .

(b) Definindo-se proporção de falso-positivo como  $1 - VPP$  e proporção de falso-negativo como  $1 - VPN$ , obtém-se 0,695 e 0,005 respectivamente.

(c) Se um médico tem 75% de certeza sobre o diagnóstico, o  $VPP$  passa a ser 0,998, aumentando a probabilidade em 24,8%.

9. (a)  $s = 0,841$ ;  $e = 0,982$ .

(b) Probabilidade de teste positivo em indivíduo saudável é 0,302.

(c)  $VPP = 0,698$ ;  $VPN = 0,992$ .

(d) A recomendação dos pesquisadores é precipitada. O teste apresenta valor baixo para VPP, mostrando que um teste positivo só corresponderia a câncer hepático em 70% dos casos. Assim, antes de encaminhar um paciente para uma clínica especializada, outros exames deveriam ser realizados de modo a confirmar esse resultado.

10. (a) Considerando  $p = 0,05$ , um valor razoável para prevalência de diabetes na população,  $VPP = 0,50$ .

(b) Espera-se que haja um aumento significativo no valor de  $VPP$ , uma vez que o paciente apresenta um valor  $p$  maior do que aquele correspondente ao da população em geral.

11. (a)  $s = 0,696$ ;  $e = 0,892$ .

(b)  $VPP = 0,658$ ;  $VPN = 0,907$ .

(c) Considerando  $p = 0,005$ :  $VPP = 0,031$ .

12. (a) O  $VPP$  depende fundamentalmente da prevalência da doença de interesse na população, daí a incoerência da afirmativa. Um teste que produza  $VPP$  baixo em determinada população pode não fazê-lo em outra. E mesmo naquela, dependendo do interesse, poderá ser utilizado satisfatoriamente (um  $VPP$  baixo não influí no  $VPN$ , que uma vez alto, será de grande utilidade para descartar a possibilidade de uma doença).

(b) Especificidade alta significa que grande número de pessoas não doentes terão testes negativos. Para avaliar a probabilidade de um teste negativo corresponder a um estado de ausência de doença, é necessária a utilização do  $VPN$ .

13. A divulgação do teste foi feita de maneira confusa. Mas muito mais importante, é que apenas os dados relevantes ao cálculo da sensibilidade do teste são apresentados. Não existe qualquer menção à capacidade do teste não detectar pessoas saudáveis (especificidade), tão importante quanto a detecção do tumor de próstata em fase inicial.

14. A afirmativa sintetiza de forma correta os significados dos índices  $s$ ,  $e$  (em que a condição prévia dos indivíduos é conhecida) e  $VPP$ ,  $VPN$  (em que há necessidade de se conhecer a condição do paciente a partir do resultado de um teste).

15.  $s = 0,99$ ;  $e = 0,99$ ;  $VPP = 0,50$ ;  $VPN = 0,99$ ;  $p = 0,01$ .  $PFP = 1 - VPP = 0,50$ ;  $PFN = 1 - VPN = 0,01$ . O teste não é confiável nesta população, por apresentar pequeno valor de predição positiva. No caso, apenas 50% das pessoas com teste positivo terão a doença.

16. (a)  $p = 0,167$  (16,7%).

(b)  $s = 0,94$ ;  $e = 0,50$ .

(c)  $VPP = 0,273$ ;  $VPN = 0,977$ .

(d) Sim. Como o  $VPN$  é alto, o teste em questão é útil para excluir a doença; o baixo  $VPP$ , faz com que um resultado positivo não seja confiável para diagnosticar a doença.

17. O clínico que usa como critério de anormalidade concentrações de glicose maiores do que  $L_1$ , faz com que o teste de todos os saudáveis seja negativo mas leva à não detecção da doença em alguns pacientes. O outro clínico, cujo critério é o valor  $L_2$ , observa teste positivo para todos os doentes, mas também, para alguns saudáveis. Desse modo, ao utilizar  $L_1$ , há ganho em especificidade e perda de sensibilidade quando comparado ao uso do valor  $L_2$ .

18. (a)  $s = 0,93$ ;  $e = 0,88$ .

(b)  $s = 0,422$ ;  $e = 0,992$ .

(c) Para o teste 1:  $VPP = 0,931$ ;  $VPN = 0,884$ . Para o teste 2:  $VPP = 0,989$ ;  $VPN = 0,492$ .

19. (a)  $s = 0,658$ ;  $e = 0,612$ ;  $p = 0,319$

(b)  $P_{95} = 1760g$ .

20. Para ruptura do menisco medial:  $s = 0,882$ ;  $e = 0,773$ . Para condromalácia:  $s = 0,379$ ;  $e = 1,00$ .

21. (a)  $s = 0,938$ ;  $e = 1,00$ .

(b) A prevalência na amostra é 0,86, muito menor do que a encontrada na população em geral.

(c)  $p = 0,7$ ;  $VPN = 0,87$ .

22. (a) A sensibilidade e especificidade do teste conjunto são 0,63 e 0,97 independente da ordem.

(b) Suponhamos que o teste seja usado em 11.000 pessoas. Quando o teste A (o de maior sensibilidade) é usado primeiro, 3900 pessoas devem receber ambos os testes. Quando o teste B (o de maior especificidade) é usado primeiro, somente 1.700 pessoas devem receber ambos os testes. Então é economicamente vantajoso usar o teste B primeiro, já que o teste B custa 2,3 vezes ( $3900/1700$ ) menos que o teste A.

## Capítulo 5

1. 6 anos.
2. 21 alunos.
4.  $z_0 = 1,29$ .
5. (a) 31% dos candidatos têm chance de serem aprovados.  
(b) O tempo máximo para merecer o certificado é 57 minutos.
6. (a) 489 estudantes.  
(b) 106 estudantes.
7. (a) Pelo método de percentis: [97,5; 146,5]. Pelo método de Gauss: [98,28; 147,12].
8. Mediana = média =  $\mu$ ;  $P_{75} = \mu + 0,68\sigma$ .
9. (a)  $\bar{x} = 243,92$ ;  $\tilde{x} = 235,5$ ;  $s = 43,04$ ;  $P_{25} = 215,5$ ;  $P_{75} = 248,5$ .  
(b) Não há indicação de que os dados foram gerados a partir de uma população gaussiana, uma vez que os valores observados diferem dos esperados se eles seguissem essa distribuição (média = mediana  $\mu$ ,  $P_{25} = 257,7$  e  $P_{75} = 273,2$ ).
11. (a) 496 pessoas.  
(b) [36,45; 37,15].
12. (a) Método dos percentis: [19,5; 73,5]. Método da curva de Gauss: [6,06; 63,54].  
(b) O método mais indicado nesse caso é o método dos percentis, uma vez que o método da curva de Gauss coloca o limite inferior sem significado para os dados obtidos.
13. (a) 0,52  
(b) 52 homens.

14.

(c)	0	1	2	3	4	5	6	7
	6,3	18,5	27,0	26,3	19,3	11,3	5,5	2,3

15. 96,20% das pessoas apresentam fosfatase alcalina entre 15 e 69  $mU/dl$ .

16. Nem sempre a variável de interesse tem distribuição gaussiana. Além disso, a expressão  $(\bar{x} - 2s, \bar{x} + 2s)$  pode não ser razoável para o problema em questão. Ela força a existência de limites superior e inferior, que em determinadas situações perdem sentido. Outras vezes, essa expressão gera valores que não condizem com os limites de normalidade.

17. (a) Média e desvio-padrão são 106,42 e 11,94.  
(b) 80% das crianças em questão têm pressão sistólica acima de 94,40  $mmHg$ .

18. (a) 0,16  
(b) 0,62  
(c) Aumento de sensibilidade e diminuição de especificidade.

19. (a)  $VPP = 0,957$ ;  $VPN = 0,787$ ;  $PFP = 0,043$ ;  $PFN = 0,213$ .

20. (a)  $P_{15} = 22$ .  
(b)  $e = 0,85$ .  
(c) Não se pode calcular a sensibilidade porque os testes foram realizados somente em idosos sem disfunção mental.  
(d) Ocorrerá aumento da especificidade, já que um número maior de resultados negativos nos indivíduos testados.

## Capítulo 6

1. (a)  $H_0$ : A proporção de abortos espontâneos em médicas anestesiistas é igual à proporção em médicas de outras especialidades.  
 $H_1$ : A proporção de abortos espontâneos em médicas anestesiistas é diferente da proporção encontrada em médicas de outras especialidades.

(b)  $X^2 = 10,27$ ;  $X_c^2 = 8,69$  (com correção); valor  $p < 0,005$ .

Existe diferença na proporção de abortos espontâneos entre médicas anestesistas e de outras especialidades.

2. Em relação aos dados de glicose, valor- $p \cong 0$ ; para triglicérides, valor- $p = 0,52$ .

3. Pode-se afirmar que o gemfibrozil atua sobre o nível de colesterol ( $Z = 26,76$ , valor- $p \cong 0$ ).

4.

(a)

Sexo	Sobrevida zero		
	Sim	Não	Total
Masculino	576	4705	5281
Feminino	90	462	552
Total	666	5167	5833

(b)  $X^2 = 14,39$  (valor- $p = 0,0001$ );  $X_c^2 = 13,87$  (valor- $p = 0,0002$ ).

5.  $X^2 = 16,25$  (valor- $p \cong 0$ );  $X_c^2 = 15,00$  (valor- $p \cong 0$ ).

6. (a) Definindo  $\mu_h$  como a média do tempo de corrida dos indivíduos quando hiperventilados e  $\mu_s$ , média sem hiperventilação, as hipóteses a serem testadas são:  $H_0: \mu_h = \mu_s$  vs  $H_0: \mu_h \neq \mu_s$ .

(b)  $T = 0,42$ ; valor- $p = 0,69$ . Não existe diferença na capacidade física (avaliada por corrida de 800 metros) de indivíduos submetidos ou não à alcalose respiratória, adquirida através de hiperventilação.

7. (a) Estudo de coorte - estudo prospectivo visando verificar a influência dos altos níveis de catepsina D na recidiva de câncer de mama em pacientes submetidas à cirurgia. Acompanhamento de dois grupos, um com altas taxas de catepsina D e outro com baixas.

(b)  $X^2 = 18,03$  (valor- $p \cong 0$ );  $X_c^2 = 16,75$  (valor- $p \cong 0$ ).

8. (a) Não existe diferença significativa entre os tempos de reação de homens e mulheres ( $T = 1,32$ ; valor- $p = 0,20$ ).

9. Teste de McNemar:  $X_{McN}^2 = 5,79$  e valor- $p = 0,016$ .

10.  $X^2 = 51,42$  (valor- $p \cong 0$ );  $X_c^2 = 50,77$  (valor- $p \cong 0$ );  $Z = 7,26$  (valor- $p \cong 0$ ).

11.  $X^2 = 6,34$  (valor- $p = 0,01$ );  $X_c^2 = 5,92$  (valor- $p = 0,02$ );  $Z = 2,67$  (valor- $p = 0,01$ ).

12. (a)  $H_0$ : Não existe diferença na performance em corrida de 50 metros entre meninos e meninas.

(b)  $t = 0,79$ ; valor- $p = 0,43$ .

(c) A diferença de idade entre os grupos, a prática esportiva por algumas crianças, além de características individuais podem ser fatores de confusão.

(d) A escolha aleatória garante representatividade da amostra, mas não a comparabilidade entre grupos, o que não elimina os fatores de confusão.

13. (a)  $H_0$ : Intervalo entre gestações menor que 6 meses ocorre com as mesmas proporções em mulheres negras e brancas.

$X^2 = 24,08$  (valor- $p \cong 0$ );  $X_c^2 = 23,31$  (valor- $p \cong 0$ ). Existe diferença na proporção de mulheres negras que têm intervalo entre gestações menor que 6 meses em relação às mulheres brancas.

(b) Não existe influência do estado civil de mulheres negras no intervalo entre gestações (valor- $p = 0,78$ ).

(c) i. Para mulheres brancas: existe diferença na idade média das mulheres no grupo  $< 6$  meses e no grupo  $\geq 6$  meses ( $T = 7,79$ )

ii. Para mulheres negras: a idade média é igual para os grupos de mulheres negras com intervalo de gestação  $< 6$  meses e  $\geq 6$  meses ( $T = 1,43$ ; valor- $p \cong 0,16$ ).

14. (a)  $Z = 0,294$  (valor- $p = 0,77$ );  $Z^* = 0,292$  (valor- $p = 0,77$ )

(b)  $X^2 = 0,087$  (valor- $p = 0,77$ );  $X_c^2 = 0,0015$  (valor- $p = 0,97$ ).

15. (a) Para o grupo de crianças susceptíveis:  $\bar{x} = 11,1$ ;  $s = 3,1$ . Para o grupo de crianças em que há resistência ao AZT:  $\bar{x} = 12,8$ ;  $s = 3,3$ .

(b)  $T = 3,20$ ; valor- $p = 0,01$ .

16. (a)  $H_0$ : Não existe diferença na concentração plasmática de proteínas totais nos grupos estudados.  $H_1$ : Existe diferença na concentração dessas proteínas nas crianças estudadas.

Não existe diferença significativa nos dois grupos, comparando-se o valor encontrado no teste com o respectivo padrão:  $Z = 1,63$  (valor-p = 0,10);  $T = 1,72$  (valor-p = 0,09).

17. (a)  $H_0$ : Não existe diferença na presença de depressão nos grupos.  $H_1$ : Existe diferença na presença de depressão nos pacientes estudados.

(b)  $X^2 = 38,33$  (valor-p  $\cong 0$ );  $X_c^2 = 33,66$  (valor-p  $\cong 0$ ).

19. (a)  $X^2 = 60,29$  (valor-p  $\cong 0$ );  $X_c^2 = 59,21$  (valor-p  $\cong 0$ ).

(b)  $H_0$ : Não existe diferença na ocorrência de suicídio como causa de morte em poetas e escritores de ficção.  $H_1$ : Existe diferença na ocorrência de suicídio nos grupos estudados.  $X^2 = 13,58$  (valor-p = 0,0002);  $X_c^2 = 11,46$  (valor-p = 0,0007).

## Capítulo 7

1. (a)  $T = 1,84$  (valor-p = 0,07).

(b) Estimativa do efeito = 331 g; intervalo de confiança, I.C. [-20,82; 682,82]. Não existe diferença significativa no peso médio de recém-nascidos cujas mães usaram cocaína apenas no 1º trimestre e daquelas que a utilizaram durante toda a gravidez.

2. (a) Considerando média no tempo de remissão  $\mu_B$  para imunoterapia com BCG e  $\mu_C$  para aquela com C. Parvum, e média no tempo de sobrevida  $\mu'_B$  e  $\mu'_C$  para os dois tratamentos, tem-se que: para tempo de remissão:  $H_0: \mu_B = \mu_C$  e  $H_1: \mu_B \neq \mu_C$ ; para tempo de sobrevivência:  $H_0: \mu'_B = \mu'_C$  e  $H_1: \mu'_B \neq \mu'_C$ .

(b) Tempo de remissão:  $T = 0,62$ . Tempo de sobrevivência:  $T = 0,94$ .

Não há evidência de diferença entre os métodos BCG e C. parvum, comparados em relação aos tempos médios de remissão e de sobrevivência.

(c) Tempo de remissão: valor-p  $\cong 0,10$ . Tempo de sobrevivência: valor-p  $\cong 0,35$ .

- (d) Diminuir o tempo de remissão em 2,20 meses e o tempo de sobrevivência em 2,92 meses.

3. (a)  $X^2 = 4,07$ ;  $X_c^2 = 3,69$ .

(b) Valor-p  $\cong 0,05$ .

(c) Estimativa do efeito =  $\hat{\psi} = 0,71$ ; I.C. para  $\psi \in [0,29, 2,26]$ . Para mulheres acima de 50 anos, o fator ter sido aumentada não exerce influência no desenvolvimento de câncer de mama.

4. Estimativa do risco para os métodos: diafragma = 0,346; preservativo = 0,246; DIU = 0,165; pílula = 0,086.

5. (a) Para  $X_2$  e  $X_4$ , valor-p  $\cong 0$ .

(b)  $X_1$  = Duração da pneumonia (dias): efeito = 1,9; I.C. [1,57; 2,23];  $X_2$  = Duração da diarréia (dias): efeito = 1,3; I.C. [1,02; 1,58];  $X_3$  = Ganhos de peso após 6 semanas (kg): efeito = 0,39; I.C. = [0,25; 0,53];  $X_4$  = Ganhos de peso após 6 meses (kg): efeito = 0,52; I.C. = [0,37; 0,57].

6. Atividade física na pressão arterial: efeito = 36,63 mmHg; I.C. = [29,93; 43,33].

7.  $\psi = 2,45$ ; I.C. para  $\psi$ : [1,04; 5,80].

8. Atividade física no colesterol: efeito = 20,17 mmHg; I.C. = [15,93; 24,41].