

Alexandre Sartoris



Estatística e introdução à econometria

2^a edição

Estatística e introdução à econometria

2^a edição

www.saraivauni.com.br

Alexandre Sartoris

Estatística e introdução à econometria

2^a edição



www.saraivauni.com.br



Rua Henrique Schaumann, 270
 Pinheiros – São Paulo – SP – CEP: 05413-010
 Fone PABX: (11) 3613-3000 Fax: (11) 3611-3308
 Televendas: (11) 3613-3344 • Fax vendas: (11) 3268-3268
 Site: <http://www.saraivauni.com.br>

ISBN 978-85-02-19983-5

**CIP-BRASIL. CATALOGAÇÃO NA FONTE
 SINDICATO NACIONAL DOS EDITORES DE LIVROS, RJ.**

Filiais

AMAZONAS/RONDÔNIA/RORAIMA/ACRE

Rua Costa Azevedo, 56 – Centro
 Fone/Fax: (92) 3633-4227 / 3633-4782 – Manaus

BAHIA/SERGIPE

Rua Agripino Dórea, 23 – Brotas
 Fone: (71) 3381-5854 / 3381-5895 / 3381-0959 – Salvador

BAURU/SÃO PAULO (sala dos professores)

Rua Monsenhor Claro, 2-55/2-57 – Centro
 Fone: (14) 3234-5643 – 3234-7401 – Bauru

CAMPINAS/SÃO PAULO (sala dos professores)

Rua Camargo Pimentel, 660 – Jd. Guanabara
 Fone: (19) 3243-8004 / 3243-8259 – Campinas

CEARÁ/PIAUÍ/MARANHÃO

Av. Filomeno Comes, 670 – Jacarecanga
 Fone: (85) 3238-2323 / 3238-1331 – Fortaleza

DISTRITO FEDERAL

SIA/SUL Trecho 2, Lote 850 – Setor de Indústria e Abastecimento
 Fone: (61) 3344-2920 / 3344-2951 / 3344-1709 – Brasília

GOIÁS/TOCANTINS

Av. Independência, 5330 – Setor Aeroporto
 Fone: (62) 3225-2882 / 3212-2806 / 3224-3016 – Goiânia

MATO GROSSO DO SUL/MATO GROSSO

Rua 14 de Julho, 3148 – Centro
 Fone: (67) 3382-3682 / 3382-0112 – Campo Grande

MINAS GERAIS

Rua Além Parába, 449 – Lagoinha
 Fone: (31) 3429-8300 – Belo Horizonte

PARÁ/AMAPÁ

Travessa Apinagés, 186 – Batista Campos
 Fone: (91) 3222-9034 / 3224-9038 / 3241-0499 – Belém

PARANÁ/SANTA CATARINA

Rua Conselheiro Laurindo, 2895 – Prado Velho
 Fone: (41) 3332-4894 – Curitiba

PERNAMBUCO/ALAGOAS/ PARAÍBA/ R. G. DO NORTE

Rua Corredor do Bispo, 185 – Boa Vista
 Fone: (81) 3421-4246 / 3421-4510 – Recife

RIBEIRÃO PRETO/SÃO PAULO

Av. Francisco Junqueira, 1255 – Centro
 Fone: (16) 3610-5843 / 3610-8284 – Ribeirão Preto

RIO DE JANEIRO/ESPÍRITO SANTO

Rua Visconde de Santa Isabel, 113 a 119 – Vila Isabel
 Fone: (21) 2577-9494 / 2577-8867 / 2577-9565 – Rio de Janeiro

RIO GRANDE DO SUL

Av. A. J. Renner, 231 – Farrapos
 Fone: (51) 3371-4001 / 3371-1467 / 3371-1567 – Porto Alegre

SÃO JOSÉ DO RIO PRETO/SÃO PAULO (sala dos professores)

Av. Brig. Faria Lima, 6363 – Rio Preto Shopping Center – V. São José
 Fone: (17) 3227-3819 / 3227-0982 / 3227-5249 – São José do Rio Preto

SÃO JOSÉ DOS CAMPOS/SÃO PAULO (sala dos professores)

Rua Santa Luzia, 106 – Jd. Santa Madalena
 Fone: (12) 3921-0732 – São José dos Campos

SÃO PAULO

Av. Antártica, 92 – Barra Funda
 Fone PABX: (11) 3613-3666 – São Paulo

S26e

2.ed.

Sartoris, Alexandre

Estatística e introdução à econometria / Alexandre Sartoris. –
 2. ed. – São Paulo : Saraiva, 2013.

ISBN 978-85-02-19983-5

1. Econometria. 2. Estatística matemática. I. Título.

13-0421

CDD: 330.015195

CDU: 330.43

21.01.12 23.01.13

042282

Editado como livro impresso em 2013

Copyright © Alexandre Sartoris.

2013 Editora Saraiva

Todos os direitos reservados.

Direção editorial Flávia Alves Bravin

Coordenação editorial Rita de Cássia da Silva

Editora – Aquisições Ana Paula Matos

Editora – Universitário Luciana Cruz

Editora – Técnico Alessandra Borges

Editora – Negócios Gisele Folha Mós

Produção editorial Daniela Nogueira Secondo
 Rosana Peroni Fazolari

Produção digital Nathalia Setrini Luiz

Suporte editorial Najla Cruz Silva

Arte e produção Conexão Editorial

Capa Conexão Editorial

Produção gráfica Liliane Cristina Gomes

Impressão e acabamento xxx

Contato com o editorial

editorialuniversitario@editorasaraiva.com.br

2ª edição



Nenhuma parte desta publicação poderá ser
 reproduzida por qualquer meio ou forma
 sem a prévia autorização da Editora Saraiva.
 A violação dos direitos autorais é crime
 estabelecido na lei nº 9.610/98 e punido
 pelo artigo 184 do Código Penal.

Agradecimentos

Este livro é resultado das aulas que ministrei como professor do módulo de Estatística do Curso de Reciclagem em Economia da Ordem dos Economistas de São Paulo e da disciplina Estatística Econômica e Introdução à Econometria na Faculdade de Ciências Econômicas de São Paulo (FACESP), onde lecionei por dois anos. Assim sendo, agradeço aos coordenadores de ambos os cursos, professor Marco Antonio Vasconcellos e professora Maria Sylvia Saes, respectivamente. Ao professor Vasconcellos cabe um agradecimento especial, já que o livro não teria saído se não fosse por ele.

Agradeço também aos alunos da FACESP, em particular à turma de Economia que se formou em 2002, que foram “cobaias” de boa parte deste texto. À minha ex-aluna e sempre amiga Fernanda Brollo, que leu e deu algumas sugestões em alguns capítulos.

Também seria preciso citar vários colegas cujas sugestões contribuíram para o texto final. Não farei isso para não cometer injustiças, pois certamente poderia esquecer alguém. Cito apenas Marislei Nishijima, que, como professora de Econometria na FACESP, fez preciosas sugestões ao curso que se refletiram depois no livro.

Claro que tenho de incluir minha família: pai, mãe, irmã, duas avós, entre outros; minha ex-orientadora, professora Vera Lucia Fava, que se tornou um pouco parente também, e teve muita influência em várias coisas que escrevi; certamente teve influência indireta (no que há de bom, é claro) neste texto.

Incluo um agradecimento à Editora Saraiva, principalmente à Flávia Helena Dante Alves Bravin, pela paciência com minha demora em entregar os capítulos.

O Autor

Sobre o autor

Alexandre Sartoris

É bacharel, mestre e doutor em Economia pela Universidade de São Paulo (USP) e licenciado em Física pela mesma instituição. Foi professor no Mackenzie, Faculdades Oswaldo Cruz e Fundação Escola do Comércio Álvares Penteado (FECAP). Tem experiência em MBAs, cursos de pós-graduação *lato sensu* e no curso PROANPEC, preparatório ao exame da ANPEC. Atualmente, é professor do departamento de Economia da Universidade Estadual Paulista (UNESP, Câmpus de Araraquara), onde leciona na graduação e pós-graduação.

Contato com o autor:

sartoris@editorasaraiva.com.br

Sobre a segunda edição

Agradeço as inúmeras sugestões, críticas e correções feitas por alunos e colegas, alguns que, infelizmente, não tive a oportunidade de conhecer pessoalmente. Não os cito nominalmente porque poderia cometer injustiças.

Os erros que, por ventura, ainda permanecem são, no entanto, responsabilidade exclusiva do autor.

Sumário

INTRODUÇÃO.....	XV
CAPÍTULO 1 – PROBABILIDADE	17
1.1 Conceito	17
1.2 Probabilidade subjetiva	20
1.3 Probabilidade do e e do ou	21
1.4 Probabilidade condicional	25
1.5 Regra de Bayes	32
Exercícios	33
Apêndice 1.A – Revisão de análise combinatória	39
Apêndice 1.B – Definição axiomática de probabilidade	42
CAPÍTULO 2 – MEDIDAS DE POSIÇÃO E DISPERSÃO.....	45
2.1 Variável aleatória	45
2.2 Medidas de posição central	45
2.3 Medidas de dispersão	52
Exercícios	63
Apêndice 2.B – Demonstrações	66
CAPÍTULO 3 – DISTRIBUIÇÃO DE PROBABILIDADE.....	69
3.1 Esperança matemática	69
3.2 Algumas distribuições discretas especiais	73
Exercícios	84
Apêndice 3.A – Progressão geométrica.....	87
Apêndice 3.B – Tópicos adicionais em distribuições de probabilidade.....	89
CAPÍTULO 4 – DISTRIBUIÇÕES CONTÍNUAS E TEOREMA DE TCHEBICHEV.....	91
4.1 Distribuições contínuas	91
4.2 Função de distribuição de variáveis contínuas	99
4.3 Esperança e variância de variáveis aleatórias contínuas	100
4.4 A distribuição normal	103
4.5 Transformações de variáveis	110
4.6 Teorema de Tchebichev	112
4.7 Momentos de uma distribuição	113
Exercícios	114
Apêndice 4.A – Cálculo diferencial e integral	118
Apêndice 4.B – Demonstração dos teoremas e momentos de uma distribuição	120

CAPÍTULO 5 – DISTRIBUIÇÃO DE PROBABILIDADE CONJUNTA.....	124
5.1 Distribuição conjunta de variáveis discretas	124
5.2 Distribuição conjunta de variáveis contínuas	134
Exercícios	145
Apêndice 5.B – Tópicos adicionais em distribuição conjunta	150
CAPÍTULO 6 – ESTIMAÇÃO.....	156
6.1 O que é inferência estatística?	156
6.2 Estimadores	156
6.3 Estimadores não viesados	158
6.4 Variância de estimadores: estimadores eficientes	162
6.5 Estimador para a variância: variância amostral	168
6.6 Melhor estimador linear não viesado	171
6.7 Propriedades assintóticas: estimadores assintoticamente não viesados	172
6.8 Estimadores consistentes	173
6.9 Lei dos grandes números	176
6.10 Teorema do limite central	176
6.11 População finita	180
6.12 Métodos de estimação	182
Exercícios	187
Apêndice 6.B – Convergências e mais propriedades de estimadores	191
CAPÍTULO 7 – INTERVALO DE CONFIANÇA E TESTES DE HIPÓTESES.....	193
7.1 Intervalo de confiança	193
7.2 Testes de hipóteses	199
7.3 Testando a variância	208
7.4 Testando a média quando a variância é desconhecida e.....	213
7.5 Comparação de variâncias	217
7.6 Erros e poder de um teste.....	220
7.7 O valor-p de um teste	222
Exercícios	224
Apêndice 7.B – Propriedades e conceitos adicionais de testes de hipóteses	227
CAPÍTULO 8 – REGRESSÃO LINEAR.....	230
8.1 Regressão linear simples	230
8.2 Método dos mínimos quadrados	232
8.3 A hipótese de normalidade	246
8.4 Propriedades dos estimadores de mínimos quadrados	247

8.5 Modelos não lineares nas variáveis	248
8.6 Regressão múltipla	252
8.7 Variáveis <i>dummy</i>	258
8.8 Seleção de modelos	261
Exercícios	265
Apêndice 8.A — Matrizes	272
Apêndice 8.B — Mais sobre regressão linear	275
CAPÍTULO 9 – VIOLANDO AS HIPÓTESES BÁSICAS.....	282
9.1 Violando a hipótese VI: a multicolinearidade	283
9.2 Violando a hipótese V: a autocorrelação	289
9.3 Violando a hipótese IV: a heterocedasticidade	299
9.4 Violando a hipótese III: o problema da simultaneidade	309
Exercícios	318
Apêndice 9.B – O método dos mínimos quadrados generalizados	323
CAPÍTULO 10 – SÉRIES DE TEMPO.....	325
10.1 Métodos ingênuos de previsão	325
10.2 Séries estacionárias e regressão espúria	326
10.3 Procedimento de Box e Jenkins [modelos ARIMA]	330
10.4 Testes de raízes unitárias	349
10.5 Cointegração	351
Exercícios	353
CAPÍTULO 11 – NÚMEROS-ÍNDICES	355
11.1 Construindo números-índices	355
11.2 Índices de preços	358
11.3 Índices de quantidades e de valor	371
11.4 Valores nominais e reais – deflacionamento de séries	373
11.5 Tipos de índices de preços	375
Exercícios	376
Tabelas estatísticas	381
Referências bibliográficas	391
Índice remissivo	393

Introdução

Este livro foi elaborado para disciplinas de Estatística em cursos de graduação que a utilizam como ferramenta. A minha experiência pessoal é com cursos de Economia, mas creio que seja útil por abranger o programa de Estatística em cursos como Administração, Engenharia e outros.

Os Capítulos 1 a 8 são o que às vezes chamamos de “Estatística Básica”. Envolvem tópicos como probabilidade, distribuição de probabilidade, inferência estatística, testes de hipóteses e regressão linear. Neles são vistos o que há de comum na maioria das disciplinas introdutórias de Estatística.

O Capítulo 9 é uma discussão sobre problemas num modelo de regressão linear, e o 10 é sobre análise de séries temporais, ambos num nível bastante inicial. Podem ser utilizados como uma introdução a estes temas num primeiro momento e depois partir para textos mais avançados, ou mesmo serem utilizados em cursos nos quais este tema seja interessante, mas não fundamental a ponto de necessitar uma análise mais profunda. Creio que, neste último caso, se enquadrem alguns cursos de pós-graduação *lato sensu* (já dei aula em pelo menos um curso que faria bom uso destes capítulos).

O Capítulo 11 é sobre números-índices e índices de preços, o que é de interesse óbvio para cursos de Economia, mas certamente é algo que, pelo menos no Brasil, dado o passado recente, acaba atraindo o interesse de muita gente.

Em alguns capítulos houve a necessidade de revisar brevemente conceitos de Matemática. Em outros, pareceu interessante aprofundar um pouco o tema em questão. Em alguns casos, combinamos as duas coisas. Neste sentido, foi adotada a seguinte convenção: apêndices aos capítulos marcados com a letra “A” se referem a revisões de conceitos matemáticos; os com a letra “B” se referem a discussões um pouco mais aprofundadas da Estatística propriamente dita, seja de conceitos adicionais, ou demonstrações que não são absolutamente necessárias para o entendimento do texto.



Capítulo 1

PROBABILIDADE

1.1 Conceito

O conceito de probabilidade está sempre presente em nosso dia a dia: qual é a probabilidade de meu time ser campeão? Qual é a probabilidade de eu passar naquela disciplina? Qual é a probabilidade de eu ganhar na loteria?

Probabilidade é uma espécie de medida associada a um evento. No caso específico da primeira pergunta do parágrafo anterior, o evento em questão é *meu time ser campeão*. Se esse evento é *impossível* de ocorrer, dizemos que sua probabilidade é *zero*. Se, entretanto, ele ocorrerá *com certeza*, sua probabilidade é igual a um (ou 100%).

Chamando esse evento simplesmente de *A*, então, dizemos que:

Se *A* é impossível de ocorrer, então $P(A) = 0$.

Se *A* ocorre com certeza, então $P(A) = 1$.

Note que a expressão $P(A)$ é lida como *probabilidade de A ocorrer*, ou simplesmente *probabilidade de A*.

A probabilidade de um evento *A* qualquer pode ser definida, de maneira simplificada,¹ como:

$$P(A) = \frac{\text{número de vezes em que } A \text{ ocorre}}{\text{número de vezes em que todos os eventos ocorrem}}$$

Essa definição deve ser vista com ressalvas: não se trata do número de vezes que de fato ocorreriam em um experimento, mas sua proporção teórica. Assim, se jogássemos uma moeda comum três vezes e nas três ela desse cara, isso não significa que a probabilidade de dar cara é igual a 1. Essa dedução nos levaria a pensar que *com certeza* essa moeda dará cara sempre, o que é um absurdo.

O conjunto de todos os eventos possíveis desse experimento (conjunto esse que chamamos de **espaço amostral**) é composto de dois possíveis resultados: cara ou coroa.

¹ No Apêndice 1.B é dada uma definição formal de probabilidade.

Considerando que esses dois eventos têm a mesma chance de ocorrer – o que vale dizer que a moeda não está viciada –, teremos:

$$P(\text{cara}) = \frac{\text{número de vezes em que ocorre cara}}{\text{número de vezes em que todos os eventos ocorrem}} = \frac{1}{2} = 0,5$$

Todos os eventos, nesse caso, são dois: cara ou coroa. Desses dois, um deles é o evento em questão (cara). Portanto, a probabilidade de dar cara é igual a 0,5 (ou 50%).

De maneira idêntica, temos a probabilidade para o evento coroa:

$$P(\text{coroa}) = \frac{\text{número de vezes em que ocorre coroa}}{\text{número de vezes em que todos os eventos ocorrem}} = \frac{1}{2} = 0,5$$

Repare que a soma das duas probabilidades é igual a 1. E tinha de ser mesmo. A soma das probabilidades (nesse caso específico) representa a probabilidade de o evento *dar cara ou coroa* ou, generalizando, ocorrer qualquer evento possível, que é algo que ocorrerá com certeza.

Mudemos o jogo: de cara ou coroa para dados. Se jogarmos o dado uma única vez, teremos seis possibilidades, que correspondem aos números inteiros de 1 a 6. A probabilidade de cair um número qualquer (digamos, o 3) será dada por:

$$P(\text{coroa}) = \frac{\text{número de vezes em que ocorre coroa}}{\text{número de vezes em que todos os eventos ocorrem}} = \frac{1}{6} = 0,1666\overline{6}$$

Outra maneira de encontrarmos essas probabilidades seria fazer um experimento (por exemplo, jogar a moeda) um número muito grande de vezes (na verdade, deveriam ser infinitas vezes) e encontrar a proporção entre caras e coroas. Esse experimento foi feito² e os resultados são mostrados na Tabela 1.1.

Tabela 1.1 – Proporção de caras e coroas para um dado jogado infinitas vezes

Número de jogadas	Número de caras	Número de coroas	Proporção de caras	Proporção de coroas
10	6	4	0,6000	0,4000
100	47	53	0,4700	0,5300
1.000	509	491	0,5090	0,4910
10.000	4.957	5.043	0,4957	0,5043
25.000 ²	12.486	12.514	0,4994	0,5006

O experimento evidencia que, à medida que o número de jogadas aumenta, a proporção de caras e de coroas se aproxima do valor 0,5.

Chamando de n o número de vezes que o experimento é feito, uma maneira de definir probabilidade é:

² Na verdade, a moeda não foi realmente jogada 25 mil vezes; os resultados foram obtidos por meio de uma simulação por computador.

$$P(\text{cair } 3) = \frac{\text{número de vezes em que ocorre } 3}{\text{número de vezes em que todos os eventos ocorrem}} = \frac{1}{6}$$

Essa fórmula é chamada de definição de probabilidade por frequência relativa ou, ainda, definição frequentista de probabilidade.

Exemplo 1.1.1

Qual a probabilidade de, jogando um único cartão, acertar a Mega Sena (seis dezenas em um total de 60)?

O acerto exato das seis dezenas é uma única possibilidade entre todas as combinações possíveis (combinações mesmo,³ já que a ordem em que os números são sorteados não é relevante):

$$P(\text{ganhar na sena}) = \frac{1}{C_{60,6}} = \frac{1}{\frac{60!}{54! \times 6!}} = \frac{1}{50.063.860} \cong 0,0000002$$

Portanto, a probabilidade de acertar a Mega Sena com apenas um cartão é de 1 para cada 50.063.860, ou aproximadamente 0,000002%.

Exemplo 1.1.2

Sendo o conjunto X definido por $X = \{x \in IR \mid 0 < x < 2\}$, qual a probabilidade de, ao sortearmos um número qualquer desse conjunto, esse número pertencer ao intervalo $[0,5; 1,5]$? E qual a probabilidade de esse número ser exatamente igual a 1?

O conjunto X é um **conjunto contínuo**, já que contém *todos* os números reais que sejam maiores do que 0 e menores do que 2. Tem, por exemplo, o número 1, o número 0,5 e o número 0,4; mas também tem o 0,45, o 0,475, o 0,46. Dados dois elementos desse conjunto, *sempre* é possível encontrar um número que esteja entre esses dois. Não há saltos ou buracos, daí a ideia de continuidade. Ao contrário do dado, cujos valores possíveis são 1, 2, 3, 4, 5 ou 6 (não existe 1,5 ou 2,1) e, por isso, é um **conjunto discreto**⁴. Nesse caso, a probabilidade de sortearmos qualquer número entre 0,5 e 1,5 (inclusive), que é um intervalo de comprimento igual a 1 ($= 1,5 - 0,5$), de um intervalo possível que tem comprimento igual a 2 ($= 2 - 0$), será dada por:

$$P(0,5 \leq x \leq 1,5) = \frac{1}{2}$$

³ Para uma revisão de análise combinatória, veja o Apêndice 1.A.

⁴ Não há necessidade de que um conjunto discreto seja composto apenas de números inteiros. Numa prova com 20 questões de múltipla escolha, cada uma delas valendo meio ponto, haverá notas variando nesse intervalo, isto é, poderá existir nota 7,0 ou 7,5, mas nunca 7,2 ou 7,3. É, portanto, um conjunto discreto.

E a probabilidade de ser exatamente 1? Ou seja, de sortearmos um único número entre todos os números presentes no conjunto X , o que dá um total de... infinitos números! A probabilidade será dada, então, por:

$$P(x = 1) = \lim_{n \rightarrow \infty} \frac{1}{n} = 0$$

Portanto, embora possível de ocorrer, a probabilidade de sortearmos um número igual a 1 (ou igual a qualquer número) é igual a zero, se estivermos falando de um conjunto contínuo. A probabilidade só será diferente de zero se estivermos falando de um **intervalo** contido nesse conjunto.

Como consequência disso, não faria diferença se o intervalo para o qual encontramos inicialmente a probabilidade (entre 0,5 e 1,5) fosse **fechado** ou **aberto** (isto é, incluísse ou não os extremos), pois a probabilidade de ser exatamente 0,5 ou 1,5 é zero. Portanto, como X é um conjunto contínuo:

$$P(0,5 \leq x \leq 1,5) = P(0,5 < x < 1,5) = \frac{1}{2}$$

1.2 Probabilidade subjetiva

Nos casos exemplificados, assumindo que os dados e as moedas utilizadas não sejam viciados, as probabilidades calculadas são exatas. Nem sempre isso é possível.

Imagine o evento *meu time será campeão*. Não é possível repetir esse experimento (o campeonato) um número muito grande de vezes. Na verdade, esse campeonato, com esses times, com os mesmos jogadores, nas mesmas condições, só é jogado uma única vez. Entretanto, é possível atribuir um valor que represente as chances de o time ganhar o campeonato, mas, evidentemente, esse valor será diferente para cada pessoa que opinar a respeito: um torcedor fanático tenderá a atribuir um valor maior do que um analista frio e imparcial (se é que isso existe).

Qualquer que seja esse valor, entretanto, deve seguir as mesmas regras da probabilidade objetiva, isto é, tem de estar entre 0 e 1, correspondendo 0 à impossibilidade e 1 à certeza de que o time será campeão.

E assim vale para uma série de situações: a probabilidade de que o governo mude a política econômica (é certamente maior em períodos de crise); a probabilidade de chover ou não (é maior ou menor quando a previsão do tempo afirma que vai chover?); a probabilidade de ser assaltado quando se passa por determinada rua etc.

Exemplo 1.2.1

Qual a probabilidade de se acertar os 13 pontos na Loteria Esportiva?

Aí é mais complicado, porque depende da avaliação subjetiva que se faz dos times em cada um dos jogos. Pode-se imaginar que um teste da Loteria Esportiva em que predominem jogos equilibrados será mais difícil de acertar e tenderá a ter menos acertadores do que um teste com muitas *barbadas*.

Por exemplo, Flamengo × Olaria (um jogo teoricamente fácil):

$$P(\text{Flamengo}) = 70\%$$

$$P(\text{empate}) = 20\%$$

$$P(\text{Olaria}) = 10\%$$

Já Corinthians × São Paulo (jogo equilibrado):

$$P(\text{Corinthians}) = 30\%$$

$$P(\text{empate}) = 40\%$$

$$P(\text{São Paulo}) = 30\%$$

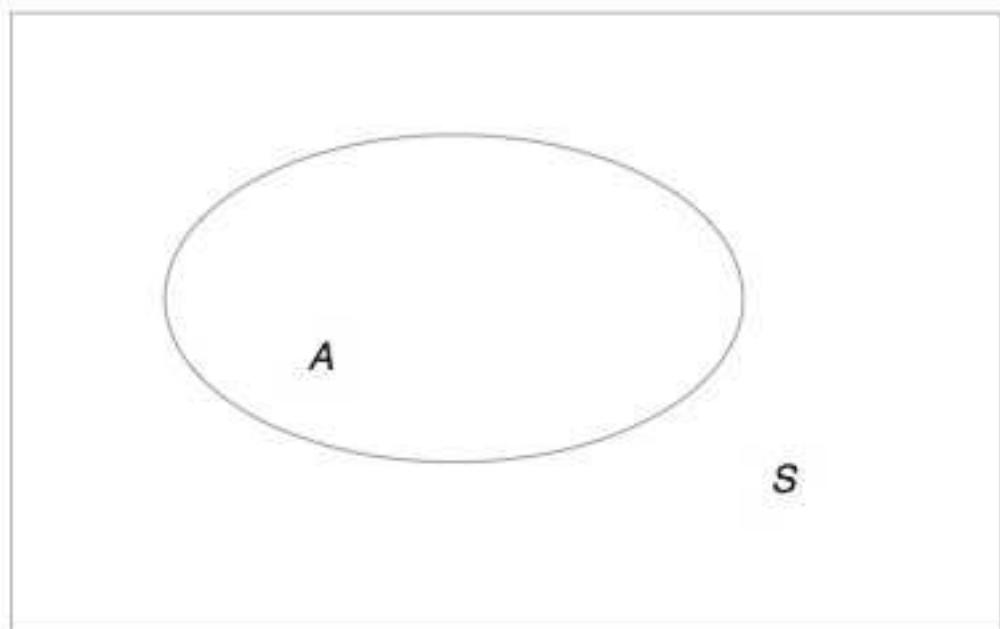
Todos esses números, evidentemente, estão sujeitos à discussão. Para computar a probabilidade de ganhar na Loteria Esportiva, essa avaliação teria de ser feita jogo a jogo.

1.3 Probabilidade do *e* e do *ou*

No início do capítulo, chamamos de espaço amostral o *conjunto* de todos os eventos possíveis. O uso do termo conjunto não foi por acaso. De fato, há uma associação muito grande entre a teoria dos conjuntos (e sua linguagem) e a de probabilidade.

Chamando de S o espaço amostral (que equivale a todos os eventos, portanto $P(S) = 1$) e sendo A um evento desse espaço amostral (isto é, A é um subconjunto de S), uma representação gráfica da probabilidade de A é mostrada na Figura 1.1.

Figura 1.1 – Probabilidade de A no espaço amostral S

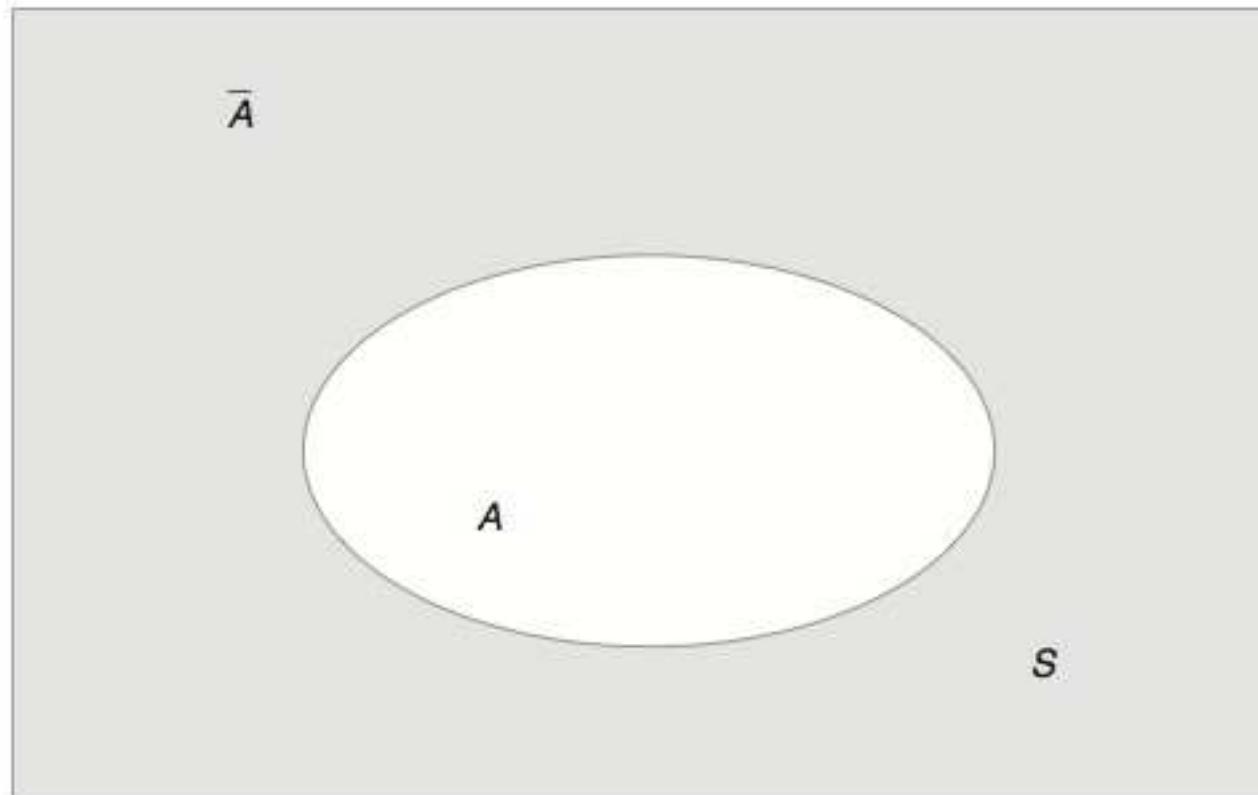


A região em que o conjunto A está representado denota sua probabilidade em relação ao espaço amostral S . Essa representação gráfica de probabilidade é conhecida como **diagrama de Venn**.

Um caso particular importante é um evento que não está em S (impossível de ocorrer), como o dado cair no número 7 ou a moeda não dar nem cara nem coroa, representado pelo conjunto vazio (\emptyset), em que, evidentemente,⁵ $P(\emptyset) = 0$.

Pelo diagrama de Venn, podemos verificar uma relação importante: a probabilidade de não A , ou seja, o complementar de A , representado⁶ por \bar{A} . O conjunto \bar{A} é representado por todos os pontos que pertencem a S , mas não pertencem a A , o que no diagrama de Venn da Figura 1.2 é representado pela região sombreada.

Figura 1.2 – Probabilidade de \bar{A} no espaço amostral S



A probabilidade de \bar{A} será dada, então, por:

$$P(\bar{A}) = P(S) - P(A)$$

Mas, como $P(S) = 1$, então:

$$P(\bar{A}) = 1 - P(A)$$

Ou:

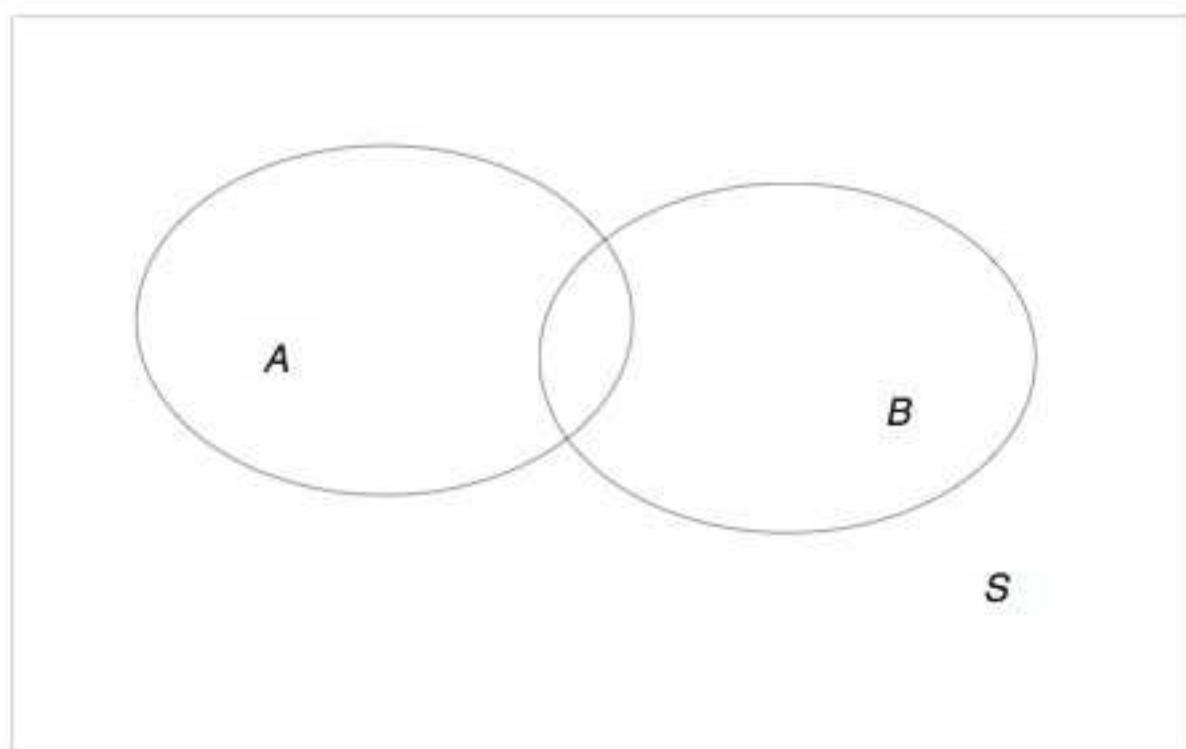
$$P(\bar{A}) + P(A) = 1$$

Assim, a soma da probabilidade de um evento com a de seu complementar é sempre igual a 1.

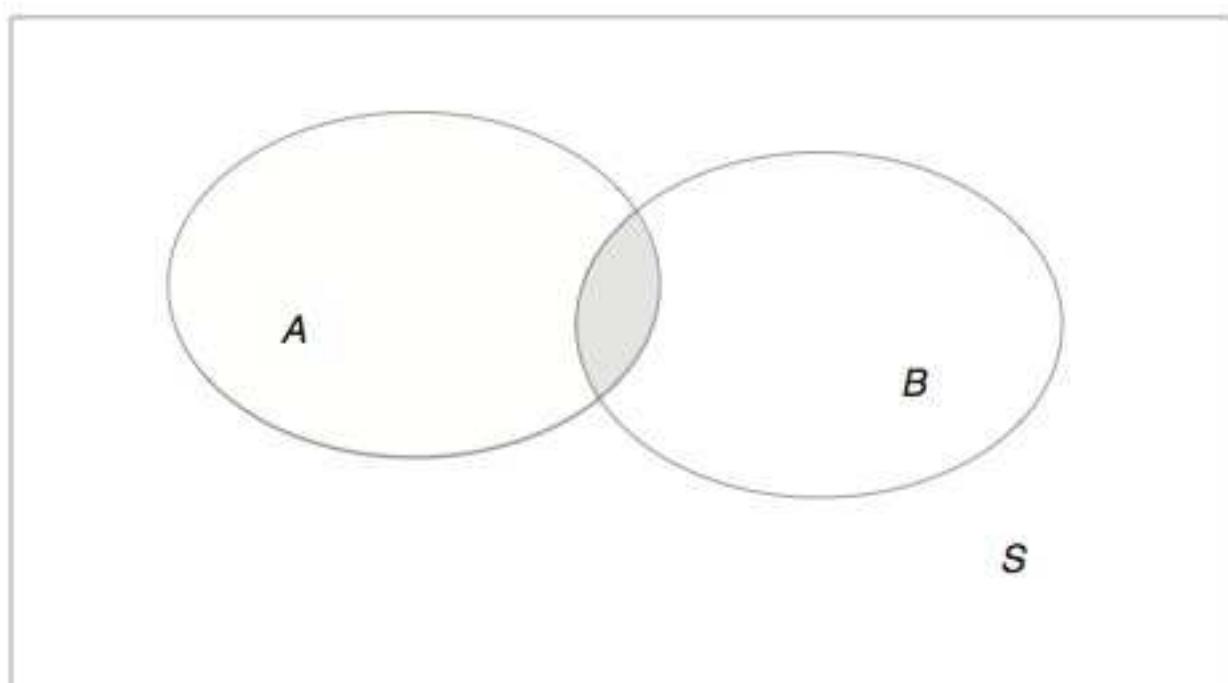
Suponhamos agora dois eventos quaisquer de S , A e B . A representação no diagrama de Venn é mostrada na Figura 1.3.

⁵ A recíproca não é verdadeira. Pelo Exemplo 1.1.2, vimos que $P(A)$ pode ser igual a zero mesmo que A não seja um conjunto vazio. No exemplo, $P(x = 1) = 0$, não porque x não pudesse ser igual a 1, mas por fazer parte de um conjunto contínuo.

⁶ Há quem prefira a notação A^c .

Figura 1.3 – Probabilidade de A e B no espaço amostral S 

Dados dois eventos, poderemos ter a probabilidade de ocorrer A e B , isto é, ocorrer A e *também* B . Por exemplo, jogar dois dados e dar 6 no primeiro e 1 no segundo; ser aprovado em Estatística e em Cálculo. Em linguagem de conjuntos, a ocorrência de um evento e também de outro é representada pela **intersecção** dos dois conjuntos ($A \cap B$). No diagrama de Venn, a intersecção aparece marcada pela área sombreada da Figura 1.4.

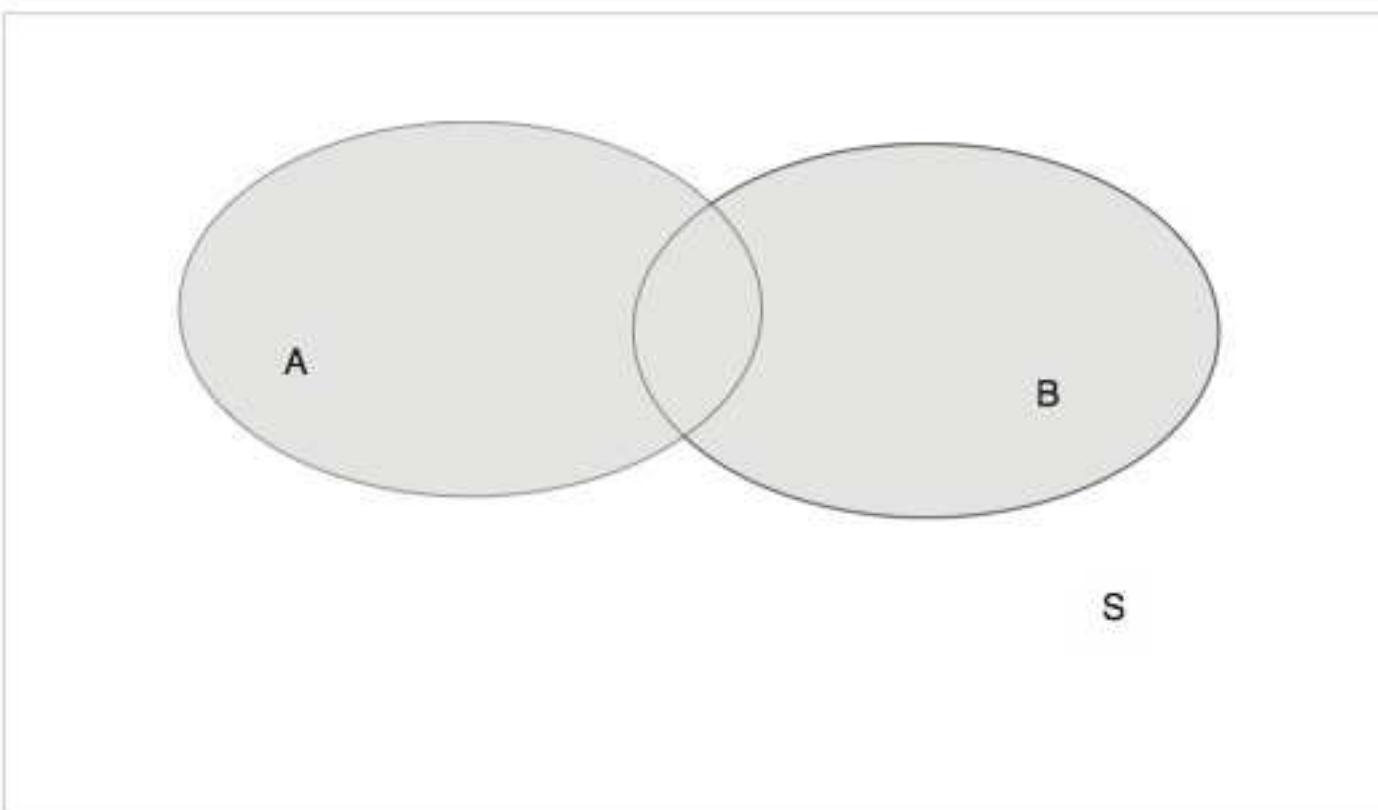
Figura 1.4 – Intersecção de A e B 

$$P(A \text{ e } B) = P(A \cap B)$$

Há ainda a probabilidade de ocorrência de A ou B . Isso equivale a ocorrer A , ou B , ou ambos.⁷ Em linguagem de conjuntos, equivale à **união** de A e B ($A \cup B$), representada na Figura 1.5 pela área sombreada.

⁷ Não confundir com o chamado **ou exclusivo**, em que ocorre um dos eventos: ou A , ou B , mas não ambos.

Figura 1.5 – União de A e B



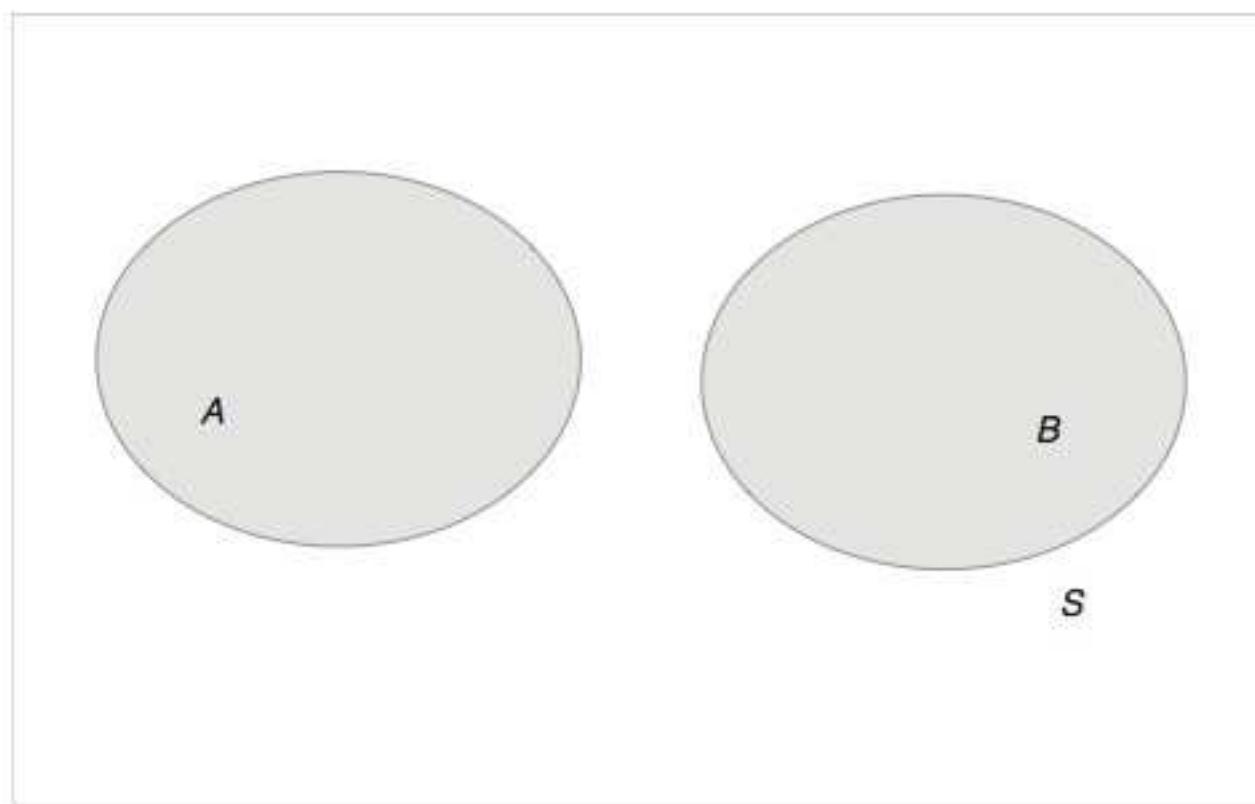
$$P(A \text{ ou } B) = P(A \cup B)$$

Podemos verificar que, se somarmos as probabilidades de A e B , a região comum a ambos (a intersecção) será somada duas vezes. Para retirar esse efeito, basta subtrairmos a intersecção (uma vez). Portanto:

$$P(A \text{ ou } B) = P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Um caso particular dessa regra é aquele em que A e B jamais ocorrem juntos: são eventos ditos **mutuamente exclusivos** (ocorrer um implica não ocorrer o outro). Os conjuntos não terão pontos em comum, portanto a intersecção é o conjunto vazio e A e B são ditos **disjuntos**, como mostrado na Figura 1.6.

Figura 1.6 – A e B mutuamente exclusivos



Nesse caso, não há dúvida:

$$P(A \text{ ou } B) = P(A \cup B) = P(A) + P(B)$$

Portanto, a chamada regra do ou pode ser resumida assim:

Se A e B são eventos quaisquer:

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

Se A e B são eventos mutuamente exclusivos (disjuntos):

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

Exemplo 1.3.1

Qual a probabilidade de, ao jogar um dado, obter um número maior que 4?

Os números maiores do que 4 no dado são o 5 e o 6, portanto:

$$P(\text{maior que } 4) = P(5 \text{ ou } 6)$$

Trata-se de eventos disjuntos, já que, se der 5, é impossível dar 6 e vice-versa.

$$P(5 \text{ ou } 6) = P(5) + P(6) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

Exemplo 1.3.2

Duas crianças gêmeas têm o seguinte comportamento: uma delas, a mais chorona, chora 65% do dia; a outra chora 45% do dia, e ambas choram, *ao mesmo tempo*, 30% do dia. Qual a probabilidade (qual o percentual do dia) de que pelo menos uma delas chore? E qual a probabilidade de que nenhuma delas chore?

A probabilidade de que pelo menos uma delas chore é a probabilidade de que a primeira chore *ou* a segunda chore. Chamando de C_1 o evento *a primeira criança chora* e C_2 , *a segunda criança chora*, temos:

$$P(C_1 \text{ ou } C_2) = P(C_1) + P(C_2) - P(C_1 \text{ e } C_2) = 0,65 + 0,45 - 0,3 = 0,8.$$

Portanto, pelo menos uma criança estará chorando 80% do tempo. *Nenhuma das crianças chora* é o evento complementar:

$$P(\text{nenhuma chora}) = 1 - P(C_1 \text{ ou } C_2) = 1 - 0,8 = 0,2.$$

Assim, os pais dessas crianças terão paz em apenas 20% do tempo.

1.4 Probabilidade condicional

Qual a probabilidade de que o Banco Central aumente a taxa de juros? E qual a probabilidade de que ele aumente a taxa sabendo-se que ocorreu uma crise que pode ter impacto sobre a inflação?

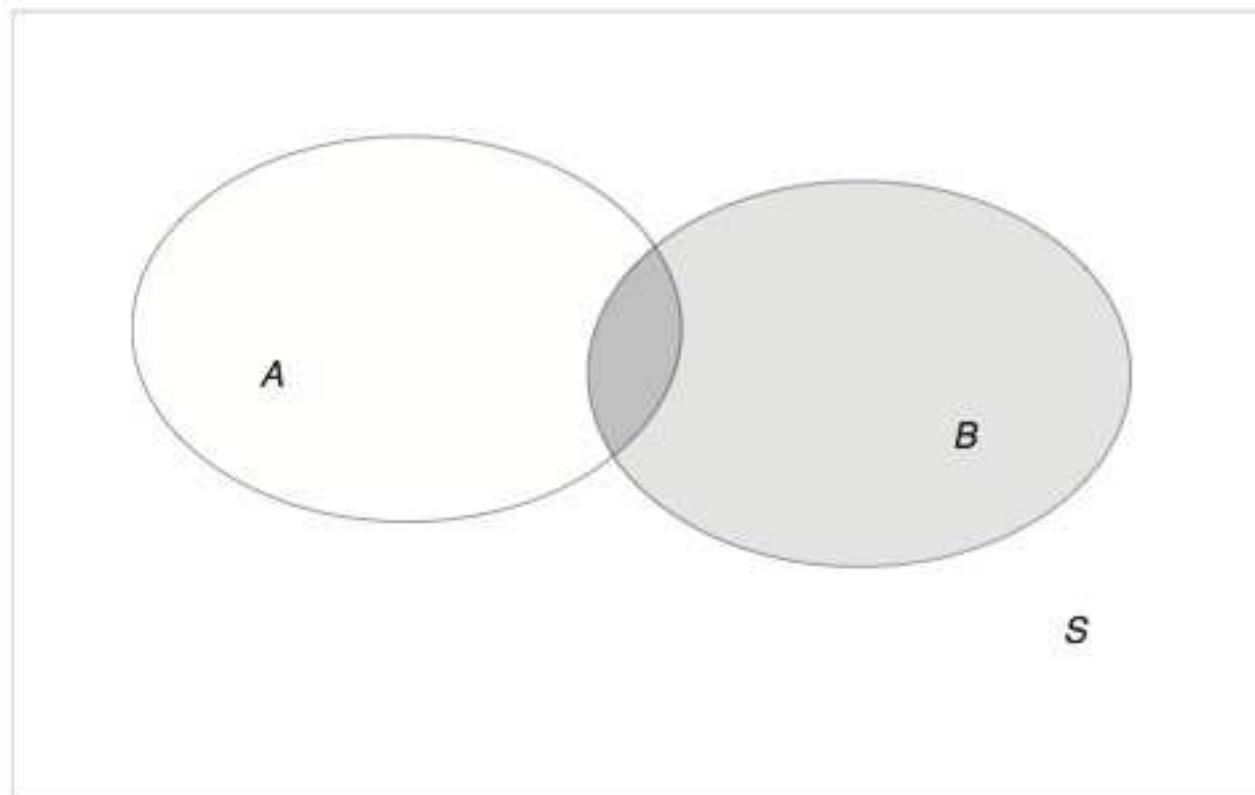
Qual a probabilidade de seu time ganhar o próximo jogo? E se já é sabido que o adversário jogará desfalcado de seu principal jogador?

Qual a probabilidade de, jogando dois dados em sequência, obter-se um total superior a 7? E se, na primeira jogada, já se tirou um 6?

Você acorda de manhã e o céu está azul e sem nuvens. Você pega o guarda-chuva ou não? É claro que, de posse dessa informação, a probabilidade estimada para o evento *chover* diminui. E assim vale para os três exemplos anteriores. O acontecimento de um evento afeta a probabilidade de ocorrência do outro.

Um casal que tem três filhos homens espera o quarto filho. Qual a probabilidade de ser, afinal, uma menina? Infelizmente, para o casal não é diferente daquela que seria caso fosse o primeiro. Não fazemos confusão: é claro que, para um casal que *vai ter* quatro filhos, a probabilidade de serem quatro meninas é pequena. Mas, se ele já teve três meninas, isso não afeta a probabilidade de o próximo filho ser menino ou menina (afinal, os pobres espermatozoides não têm a menor ideia do histórico familiar).

Figura 1.7 – Probabilidade de *A* dado *B*



A pergunta que se faz, seja em um caso ou em outro, é: qual a probabilidade de um evento acontecer sabendo-se que outro evento já ocorreu (ou vai ocorrer)? Qual a probabilidade de *A*, dado que *B* já é um fato da vida?

No diagrama de Venn da Figura 1.7, *B* já ocorreu. A probabilidade de *A* ocorrer, então, só pode ser naquele pedaço que *A* e *B* têm em comum (a interseção). Mas a probabilidade deve ser calculada não mais em relação a *S*, mas em relação a *B*, já que os pontos fora de *B* sabidamente não podem acontecer (uma vez que *B* ocorreu). Portanto, a probabilidade de *A* tendo em vista que *B* ocorreu (ou ocorrerá), representada por $P(A|B)$ (lê-se "probabilidade de *A*, dado *B*"), será dada por:

$$P(A|B) = \frac{P(A \text{ e } B)}{P(B)} \quad (1.4.1)$$

A *regra do e*, já apresentada na seção anterior, ganha uma nova forma:

$$P(A \text{ e } B) = P(A|B) \times P(B)$$

ou

$$P(A \text{ e } B) = P(B|A) \times P(A)$$

Se o evento B não tiver qualquer efeito sobre a probabilidade do evento A , teremos:

$$P(A|B) = P(A)$$

e

$$P(B|A) = P(B)$$

E A e B são ditos **eventos independentes** (a probabilidade condicional é igual à não condicional).

Serão **eventos dependentes** em caso contrário, isto é:

$$P(A|B) \neq P(A)$$

e

$$P(B|A) \neq P(B)$$

Então, se A e B forem eventos independentes, vale:

$$P(A \text{ e } B) = P(A) \times P(B)$$

Não confunda: o fato de dois eventos serem independentes não quer dizer que eles sejam mutuamente exclusivos. Pelo contrário: se dois eventos (não vazios) são mutuamente exclusivos (disjuntos), eles são, necessariamente, dependentes, já que a ocorrência de um implica a não ocorrência do outro.

Resumindo, para dois eventos independentes temos:

$$P(A \text{ e } B) = P(A) \times P(B)$$

$$P(A \text{ ou } B) = P(A) + P(B) - P(A) \times P(B)$$

Para dois eventos disjuntos (mutuamente exclusivos):

$$P(A \text{ e } B) = 0$$

$$P(A \text{ ou } B) = P(A) + P(B)$$

Para dois eventos quaisquer:

$$P(A \text{ e } B) = P(A) \times P(B|A) = P(B) \times P(A|B)$$

$$P(A \text{ ou } B) = P(A) + P(B) - P(A \text{ e } B)$$

Exemplo 1.4.1

Qual a probabilidade de que, jogando dois dados em sequência, obtenhamos *exatamente* 7? E se na primeira jogada já obtivemos um 6?

Para obter um total de 7, temos os seguintes resultados possíveis: 1 e 6, 2 e 5, 3 e 4, 4 e 3, 5 e 2, 6 e 1. O resultado de cada dado é independente do resultado do outro, de modo que:

$$P(1 \text{ e } 6) = P(2 \text{ e } 5) = P(3 \text{ e } 4) = P(4 \text{ e } 3) = P(5 \text{ e } 2) = P(6 \text{ e } 1) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$$

A probabilidade de que ocorra qualquer um desses resultados, tendo em vista que eles são mutuamente exclusivos, é:

$$P[(1 \text{ e } 6) \text{ ou } (2 \text{ e } 5) \text{ ou } (3 \text{ e } 4) \text{ ou } (4 \text{ e } 3) \text{ ou } (5 \text{ e } 2) \text{ ou } (6 \text{ e } 1)] = \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{1}{6}$$

Se já deu 6 no primeiro dado, o único resultado possível para somar 7 é que dê 1 no segundo dado. A probabilidade é $\frac{1}{6}$, portanto. De fato, usando a Definição 1.4.1:

$$P(\text{soma} = 7 | \text{primeiro dado} = 6) = \frac{P(\text{soma} = 7 \text{ e } \text{primeiro dado} = 6)}{P(\text{primeiro dado} = 6)} =$$

$$\frac{P(\text{segundo dado} = 1 \text{ e } \text{primeiro dado} = 6)}{P(\text{primeiro dado} = 6)} = \frac{\frac{1}{36}}{\frac{1}{6}} = \frac{1}{6}$$

Note que:

$$P(\text{soma} = 7 | \text{primeiro dado} = 6) = P(\text{soma} = 7)$$

Portanto, o evento *a soma dar exatamente 7* e o resultado⁸ do primeiro dado são independentes.

Exemplo 1.4.2

No Exemplo 1.3.2, os eventos são independentes? Caso não sejam, qual é a probabilidade de que a primeira criança chore, *dado* que a segunda chora? E qual a probabilidade de que a segunda criança chore, *dado* que a primeira chora?

Os eventos C_1 e C_2 *não são independentes* (são dependentes), dado que:

$$P(C_1) \times P(C_2) = 0,65 \times 0,45 = 0,2925 \text{ é diferente de } P(C_1 \text{ e } C_2) = 0,3$$

Para calcular as probabilidades condicionais, temos:

$$P(C_1 \text{ e } C_2) = P(C_1) \times P(C_2 | C_1)$$

$$0,3 = 0,65 \times P(C_2 | C_1)$$

$$P(C_2 | C_1) = \frac{0,3}{0,65} \cong 0,4615$$

$$P(C_1 \text{ e } C_2) = P(C_2) \times P(C_1 | C_2)$$

⁸ Note que a conclusão é válida para qualquer resultado no primeiro dado.

$$0,3 = 0,45 \times P(C_1|C_2)$$

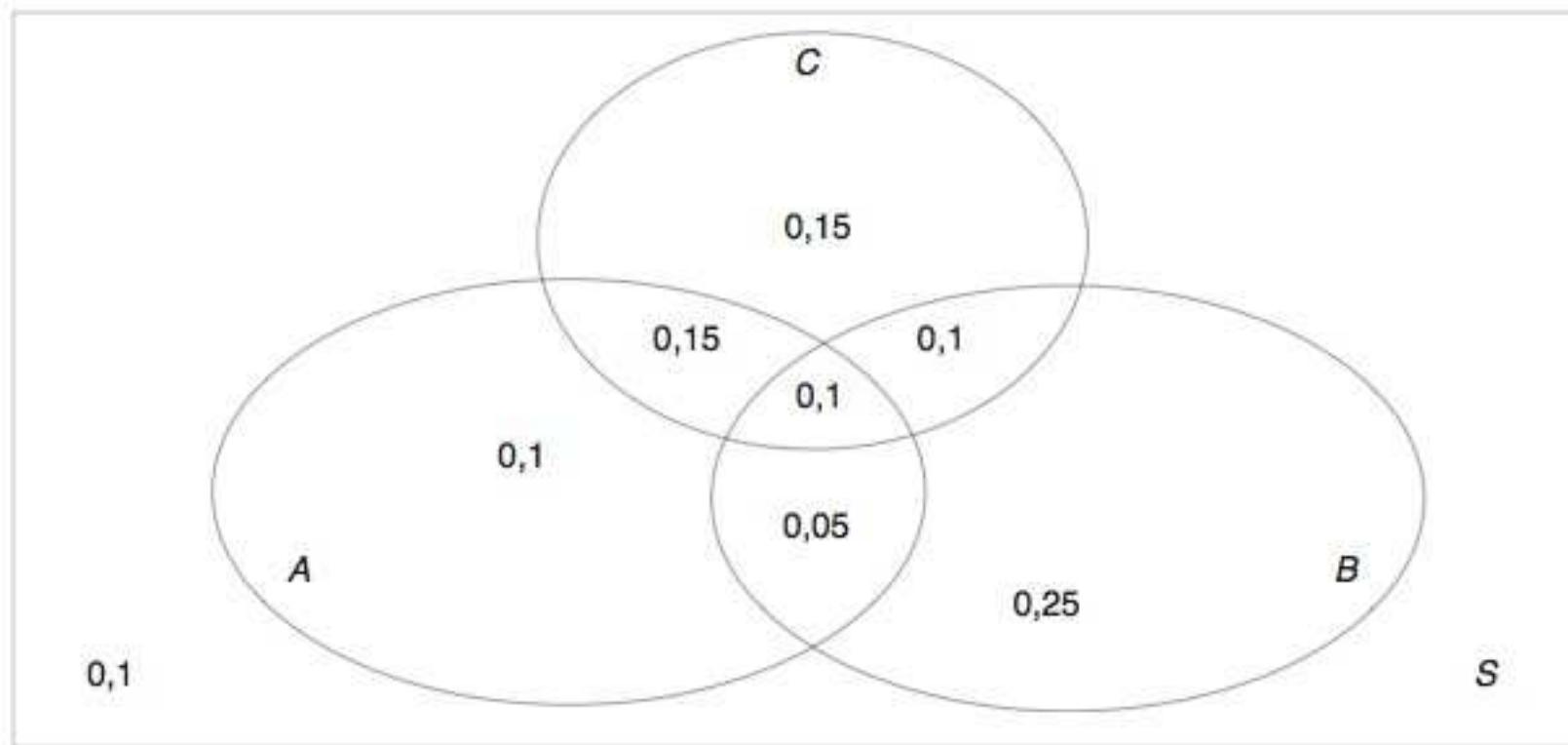
$$P(C_1|C_2) = \frac{0,3}{0,45} \cong 0,6667.$$

Portanto, se a primeira criança chorar, há probabilidade de 46,15% de que a segunda criança chore; se a segunda criança chorar, a probabilidade de que a primeira chore é de 66,67%. Como as probabilidades incondicionais eram de 45% e 65%, respectivamente, percebe-se que o fato de uma criança chorar aumenta a chance de a outra chorar também.

Exemplo 1.4.3

Por meio do diagrama de Venn da Figura 1.8, em que os valores marcados correspondem às probabilidades das áreas delimitadas, isto é, $P(A) = 0,1 + 0,15 + 0,1 + 0,05 = 0,4$, $P(B) = 0,25 + 0,05 + 0,1 + 0,1 = 0,5$ e $P(C) = 0,15 + 0,15 + 0,1 + 0,1 = 0,5$, verificaremos que, apesar de $P(A \cap B \cap C) = P(A) \times P(B) \times P(C)$, A , B e C não são eventos independentes.

Figura 1.8 – Intersecção de três eventos



De fato, $P(A \cap B \cap C) = P(A) \times P(B) \times P(C) = 0,1$, mas:

$$P(A \cap B) = 0,15 \neq P(A) \times P(B)$$

$$P(B \cap C) = 0,2 \neq P(B) \times P(C)$$

$$P(A \cap C) = 0,25 \neq P(A) \times P(C)$$

Portanto, A , B e C são dependentes.

Exemplo 1.4.4

Foi feita uma pesquisa com 100 pessoas sobre as preferências a respeito de programas na televisão. Os resultados obtidos são mostrados na Tabela 1.2.

Tabela 1.2 – Preferências por programas de televisão

	Homens	Mulheres	Total
Futebol	40	20	60
Novela	5	35	40
Total	45	55	100

Entre o grupo de entrevistados, qual a probabilidade de alguém preferir novela? E futebol?

$$P(\text{novela}) = \frac{40}{100} = 0,4 = 40\%.$$

$$P(\text{futebol}) = \frac{60}{100} = 0,6 = 60\%.$$

Qual a probabilidade de ser mulher e preferir futebol?

$$P(\text{mulher e futebol}) = \frac{20}{100} = 0,2 = 20\%.$$

Qual a probabilidade de, sendo homem, preferir futebol?

Podemos resolver diretamente, já que, pela tabela, dos 45 homens, 40 preferem futebol:

$$P(\text{futebol|homem}) = \frac{40}{45} = 0,888\dots \cong 88,8\%.$$

Ou, pela definição de probabilidade condicional:

$$P(\text{futebol|homem}) = \frac{P(\text{homem e futebol})}{P(\text{homem})} = \frac{\frac{40}{100}}{\frac{45}{100}} = \frac{40}{45} = 0,888\dots \cong 88,8\%$$

Qual a probabilidade de, se preferir novela, ser mulher?

De novo é possível resolver diretamente pela tabela, tendo em vista que, dos 40 que preferem novela, 35 são mulheres:

$$P(\text{mulher|novela}) = \frac{35}{40} = 0,875 = 87,5\%.$$

Ou, pela definição de probabilidade condicional:

$$P(\text{mulher|novela}) = \frac{P(\text{mulher e novela})}{P(\text{novela})} = \frac{\frac{35}{100}}{\frac{40}{100}} = \frac{35}{40} = 0,875 = 87,5\%$$

Note que a preferência por um tipo de programa ou outro e o sexo não são eventos independentes, já que:

$$\begin{aligned} P(\text{mulher}|\text{novela}) &\neq P(\text{mulher}) \\ P(\text{futebol}|\text{homem}) &\neq P(\text{futebol}) \end{aligned}$$

No exemplo anterior, repare que a soma de homens que gostam de novela e futebol é igual ao total de homens e o mesmo ocorre com as mulheres. Faz sentido, não? Mas só “deu certo” porque foi perguntado qual é “o” programa favorito, isto é, a cada entrevistado foi pedido que escolhesse um único programa. Como ficaria se a pergunta fosse, por exemplo, “a qual desses programas você assiste?”, de modo que seria possível que o entrevistado respondesse “os dois”?

Exemplo 1.4.5

Foi feita uma pesquisa com 100 pessoas sobre quais programas veem na televisão. Os resultados obtidos são mostrados na Tabela 1.3.

Tabela 1.3 – Programas de televisão vistos			
	Homens	Mulheres	Total
Futebol	40	30	70
Novela	20	50	70
Total	45	55	100

Nesse caso, os totais “não batem”. Por quê? Simples, porque há homens e mulheres que assistem tanto ao futebol como à novela.

Quantos homens assistem aos dois programas? A mesma regra da intersecção se aplica aqui. Se somarmos o número de homens que gostam de futebol com os que gostam de novela, teremos os que gostam dos dois (a intersecção desses conjuntos) somados duas vezes. Assim temos:

$$40 + 20 - \text{número dos que gostam de ambos} = 45$$

$$\text{número dos que gostam de ambos} = 60 - 45 = 15.$$

A mesma coisa para as mulheres:

$$30 + 50 - \text{número das que gostam de ambos} = 55$$

$$\text{número das que gostam de ambos} = 80 - 55 = 25.$$

Assim sendo, poderíamos refazer a Tabela 1.3 de uma maneira mais detalhada:

Tabela 1.4 – Programas de televisão vistos			
	Homens	Mulheres	Total
Só futebol	25	5	30
Só novela	5	25	30
Futebol e novela	15	25	40
Total	45	55	100

Agora os totais “batem”. Fica mais fácil responder a qualquer pergunta. Por exemplo, qual a probabilidade de ser mulher e assistir a futebol? $P(\text{mulher e futebol}) = \frac{30}{100} = 0,3 = 30\%$.

Ou ainda, qual a probabilidade de, sendo mulher, gostar **apenas** de futebol?

$$P(\text{só futebol} \mid \text{mulher}) = \frac{5}{55} = \frac{1}{11} \cong 0,0909 = 9,09\%.$$

1.5 Regra de Bayes

Qual a chance de você sair no fim de semana se chover? E se não chover? Suponha, entretanto, que a pergunta seja invertida: se você saiu no último fim de semana, qual é a probabilidade de ter chovido? Como responder a perguntas como essa é o que é mostrado no exemplo a seguir.

Exemplo 1.5.1

Suponha que, numa eleição para governador em um estado norte-americano, haja um candidato democrata e um republicano. Entre os eleitores brancos, 30% votam no democrata, proporção que sobe para 60% entre os eleitores negros e é de 50% entre os eleitores de outras etnias. Sabendo-se que há 70% de eleitores brancos, 20% de negros e 10% de outras etnias, se um voto democrata é retirado ao acaso, qual a probabilidade de que ele tenha sido dado por um eleitor negro?

Utilizaremos as seguintes abreviações:

B – branco D – democrata

N – negro R – republicano

O – outras etnias

Pelo enunciado, sabemos que:

$$P(B) = 0,7$$

$$P(N) = 0,2$$

$$P(O) = 0,1$$

$$P(D|N) = 0,6$$

$$P(D|B) = 0,3$$

$$P(D|O) = 0,5$$

E pede-se qual probabilidade de o voto ser de um eleitor negro, dado que o voto é para o candidato democrata, isto é:

$$P(N|D) = ?$$

$$P(N|D) = \frac{P(N \text{ e } D)}{P(D)}$$

A probabilidade de ser negro e democrata é dada por:

$$P(N \text{ e } D) = P(N) \times P(D|N) = 0,2 \times 0,6 = 0,12$$

E a probabilidade de ser democrata será dada pela soma dos votos brancos e democratas, negros e democratas e outras e democratas:

$$P(D) = P(D \text{ e } B) + P(D \text{ e } N) + P(D \text{ e } O) = 0,7 \times 0,3 + 0,2 \times 0,6 + 0,1 \times 0,5 = 0,38$$

Assim:

$$P(N|D) = \frac{0,12}{0,38} \cong 0,3158 = 31,58\%.$$

Portanto, 31,58% dos votos democratas são de eleitores negros.

O exemplo anterior partiu de probabilidades condicionais para calcular uma probabilidade com a *condição invertida*. A generalização do resultado obtido é conhecida como **regra de Bayes**, a seguir enunciada:

Se existem as probabilidades condicionais de um evento B , dados todos os eventos do tipo A_i ($i = 1, 2, \dots, n$), e queremos encontrar a probabilidade condicional de certo evento A_j , dado B , essa será dada por:⁹

$$P(A_j|B) = \frac{P(B|A_j) \times P(A_j)}{\sum_{i=1}^n P(B|A_i) \times P(A_i)}$$

Exercícios

- Em uma caixa há sete lâmpadas, sendo quatro boas e três queimadas. Retirando três lâmpadas ao acaso, sem reposição, calcule a probabilidade de que:
 - todas sejam boas;
 - todas estejam queimadas;
 - exatamente duas sejam boas;
 - pelo menos duas sejam boas.
- Calcule a probabilidade de que, no lançamento de um dado, o número que der seja:
 - ímpar;
 - primo;
 - no mínimo 4;
 - no máximo 5.
- Ao lançar dois dados em sequência, quer-se atingir um total de 11 pontos.
 - Qual a probabilidade de que isso ocorra?
 - Qual a probabilidade de que isso ocorra supondo que o primeiro dado deu 4?
 - Qual a probabilidade de que isso ocorra, supondo que o primeiro dado deu 6?
 - O evento *total de 11 pontos* é independente do resultado do primeiro dado? Justifique.
- Um jogador aposta no lançamento de um dado em um único número. Calcule probabilidade de:

⁹ Evidentemente, essa expressão não precisa ser memorizada se for repetido o raciocínio do Exemplo 1.5.1.

- a) em três jogadas, ganhar as três;
 b) em quatro jogadas, ganhar exatamente as duas primeiras;
 c) em quatro jogadas, ganhar exatamente duas (quaisquer);
 d) em quatro jogadas, ganhar pelo menos duas;
 e) em quatro jogadas, ganhar duas seguidas.
5. Na primeira loteria de números lançada no país, o apostador deveria acertar cinco dezenas em um total de 100 possíveis, apostando para isso em cinco, seis, sete, oito, nove ou dez dezenas.
 a) Qual a probabilidade de acertar as cinco dezenas em cada uma das situações?
 b) Se a aposta em cinco dezenas custasse R\$ 1,00, qual deveria ser o preço dos demais tipos de apostas, levando-se em consideração a probabilidade de acerto?
6. Considerando que, em jogos de futebol, a probabilidade de cada resultado (vitória de um time, do outro ou empate) é igual, qual a probabilidade de fazer os 13 pontos da Loteria nos seguintes casos:
 a) sem duplos ou triplos;
 b) com um único duplo;
 c) com um único triplo;
 d) com dois duplos e três triplos.
7. Represente no diagrama de Venn:
 a) $\bar{A} \cap B$;
 b) $\bar{A} \cap \bar{B}$;
 c) $\bar{A} \cup B$;
 d) $\bar{A} \cup \bar{B}$.
8. Verifique que a probabilidade do *ou exclusivo* é dada por:
- $$P(A \text{ ou exclusivo } B) = P[(\bar{A} \cap B) \cup (A \cap \bar{B})]$$
- Sugestão: utilize o diagrama de Venn.
9. Foram selecionados 200 prontuários de motoristas e o resultado foi o apresentado na Tabela 1.5.

Tabela 1.5 – Prontuários de motoristas

	Homens	Mulheres	Total
Com multa	65	50	115
Sem multa	45	40	85
Total	110	90	200

- a) Qual a probabilidade de que um motorista desse grupo tenha sido multado?
- b) Qual a probabilidade de que um motorista (homem) desse grupo tenha sido multado?
- c) Qual a probabilidade de que *uma* motorista desse grupo tenha sido multada?
- d) Qual a probabilidade de que, sendo o motorista homem, ele tenha sido multado?
- e) Qual a probabilidade de que, sendo mulher, a motorista tenha sido multada?
- f) Qual a probabilidade de que, tendo sido multado, o motorista seja homem?
- g) A probabilidade de ter sido multado é independente do sexo? Justifique.
10. Perguntou-se para 300 estudantes o que fariam após a faculdade: se procurariam emprego ou cursariam pós-graduação (ou ambos, já que uma pessoa pode optar por fazer as duas coisas). As respostas foram as da Tabela 1.6.

Tabela 1.6 – Planos para depois da faculdade		
	Homens	Mulheres
Emprego	110	90
Pós-graduação	90	80
Total	160	140

Calcule a probabilidade de um estudante, escolhido ao acaso:

- a) ser homem e procurar emprego;
- b) ser mulher e continuar estudando;
- c) ser homem e não continuar estudando;
- d) ser mulher ou não procurar emprego;
- e) sendo homem, querer continuar apenas estudando;
- f) se quer apenas trabalhar, ser mulher.
11. Um cubo de madeira é pintado e a seguir dividido em 64 cubinhos de mesmo tamanho. Qual a probabilidade de que, se pegarmos um desses cubinhos ao acaso, ele tenha apenas uma face pintada?
12. É dado um conjunto $X = \{x \in \text{IN} \mid 0 < x < 8\}$, em que IN representa o conjunto dos números naturais. Se escolhermos ao acaso um número desse intervalo, calcule as probabilidades pedidas:
- a) $P(x = 2)$;
- b) $P(x > 2)$;
- c) $P(x < 5)$;
- d) $P(x = 8)$.

13. É dado um conjunto $X = \{x \in IR \mid 0 < x < 8\}$, em que IR representa o conjunto dos números reais. Se escolhermos ao acaso um número desse intervalo, calcule as probabilidades pedidas:

- a) $P(x = 2)$;
- b) $P(x > 2)$;
- c) $P(x < 5)$;
- d) $P(0 \leq x \leq 8)$.

14. Em um colégio de ensino médio há 120 alunos no 1º ano, 100 no 2º ano e 80 no 3º ano. Se dois alunos são escolhidos ao acaso e o primeiro está mais adiantado do que o segundo, qual a probabilidade de que ele esteja no 3º ano?

15. Verifique se são verdadeiras ou falsas as afirmações a seguir e justifique.

- a) Sendo S o espaço amostral, então $P(S) = 1$.
- b) Se $P(A) = 1$, então $A = S$.
- c) Se $P(A) = 0$, então $A = \emptyset$.
- d) Se A e B são mutuamente exclusivos, então $P(A \cap B) = 0$.
- e) Se $P(A \cap B) = 0$, então A e B são disjuntos.
- f) Se A e B são independentes, então $P(A \cup B) = P(A) + P(B)$.
- g) Se $P(A \cap B) = 0$, então A e B são independentes.
- h) Se $P(A \cap B) = 1$, então $A = B = S$.
- i) Se $P(A \cap B) = 1$, então $A = S$ ou $B = S$.
- j) Se A , B e C são independentes, então $P(A \cap B \cap C) = P(A) \times P(B) \times P(C)$.
Se $P(A \cap B \cap C) = P(A) \times P(B) \times P(C)$, então A , B e C são independentes.
- k) Se $P(\bar{A}) = 1$, então $A = \emptyset$.
- l) Se A e B são independentes, então \bar{A} e \bar{B} são independentes.

16. Há 60% de probabilidade de que haja desvalorização cambial. Se a desvalorização ocorrer, há 70% de chance de o governo lançar um pacote emergencial de medidas. Se não ocorrer, as chances de esse pacote ser lançado caem para 40%. Se o pacote foi lançado, qual a probabilidade de que tenha ocorrido desvalorização cambial?

17. Num jogo de dominó, uma peça com dois valores iguais é tirada. Qual a probabilidade de que a peça seguinte se encaixe?

18. Num jogo de pôquer, cada jogador tem cinco cartas. Considerando que seja utilizado o baralho completo, calcule a probabilidade de o jogador obter:

- a) um par;
- b) uma trinca;

- c) dois pares;
- d) um par e uma trinca (*full house*);
- e) uma quadra;
- f) todas as cartas do mesmo naipe, mas não em sequência (*flush*);
- g) uma sequência (por exemplo: 7, 8, 9, 10 e J), mas não do mesmo naipe;
- h) uma sequência (exceto a maior) com o mesmo naipe (*straight flush*);
- i) a maior sequência (10, J, Q, K e A) com o mesmo naipe (*royal straight flush*).

19. Num dado viciado, a probabilidade de cair cada número é proporcional a esse número.

- a) Qual a probabilidade de cada número?
- b) Qual a probabilidade de, em uma jogada, o número ser no mínimo 4?
- c) Qual a probabilidade de, em duas jogadas, a soma ser no máximo 9?

20. Considere que a probabilidade de um recém-nascido ser menino é igual à de ser menina. Nesse caso, calcule a probabilidade de um casal com quatro filhos:

- a) ter exatamente duas meninas;
- b) ter, no máximo, dois meninos;
- c) ter pelo menos uma menina;
- d) o mais velho ser um menino.

21. Em 1 milhão de nascimentos, foram registrados 509.718 meninos e 490.282 meninas. Considerando essa proporção (aproximadamente) uma estimativa mais realista para a probabilidade de nascimento de meninas e meninos, refaça os cálculos do exercício anterior.

22. Entre as mulheres solteiras de uma cidade, 70% são morenas e 30% loiras. Entre as morenas, 60% têm olhos castanhos, 30% têm olhos verdes e 10% têm olhos azuis. Já entre as loiras, 40% têm olhos castanhos, 30% verdes e 30% azuis. Para um homem que vai a um *encontro às escuras*, calcule a probabilidade de que a pessoa que ele vai encontrar:

- a) tenha olhos azuis;
- b) seja loira de olhos verdes;
- c) seja morena de olhos castanhos;
- d) caso tenha olhos castanhos, seja loira;
- e) caso tenha olhos verdes, seja morena.

23. Dado um espaço amostral definido num plano cartesiano:

$$S = \{(x,y) \in \text{IR}^2 \mid -1 \leq x \leq 3; 2 \leq y \leq 4\}$$

E dado o conjunto A:

$$A = \{(x,y) \in \text{IR}^2 \mid 1 \leq x < 2; 3 < y < 4\}$$

Calcule P(A). Sugestão: encontre graficamente S e A.

24. Dados os conjuntos A , B e C não vazios cujas probabilidades são dadas por $P(A)$, $P(B)$ e $P(C)$, determine $P(A \cup B \cup C)$. Sugestão: use um diagrama semelhante ao do Exemplo 1.4.3.
25. Segundo as pesquisas eleitorais, o candidato A tem 30% das preferências dos eleitores. Admitindo que esse valor esteja correto, se tomarmos cinco eleitores ao acaso, calcule a probabilidade de:
- exatamente três deles votarem no candidato A;
 - no máximo dois deles votarem no candidato A;
 - pelo menos um deles votar no candidato A.
26. Em uma urna há seis bolas que podem ser brancas ou pretas. Se três bolas retiradas ao acaso, com reposição, são brancas, qual a probabilidade de não haver bolas pretas?
27. A probabilidade de que um jogador de basquete acerte um arremesso é p . Determine o valor de p para que a probabilidade de fazer pelo menos uma cesta a cada dois arremessos seja de 80%.
28. Mostre que, se é válida a expressão $P(A|B) = P(A|\bar{B})$, A e B são independentes.
29. Num país, 80% da população pertence à etnia majoritária e há 10% de pobres, sendo que, dos pobres, 60% pertencem à etnia minoritária. Qual a probabilidade de, uma vez pertencendo à etnia minoritária, ser pobre?
30. Se uma família tem exatamente n crianças ($n \geq 2$) e supondo que a probabilidade de ser menino é igual à de ser menina, determine:
- a probabilidade de a família ter só meninas;
 - a probabilidade de ter, no mínimo, uma menina;
 - a probabilidade de ter, no mínimo, um menino, dado que tem pelo menos uma menina.
31. Num torneio de tênis entre três jogadores, A, B e C, cada um dos competidores enfrenta todos os demais uma única vez. As probabilidades de cada jogo são $P(A \text{ vença } B) = 0,7$, $P(A \text{ vença } C) = 0,8$ e $P(B \text{ vença } C) = 0,6$. Determine:
- a probabilidade de que cada jogador vença o torneio;
 - a probabilidade de que o torneio termine empatado;
 - a probabilidade de que cada jogador vença pelo menos tantas vezes quanto vencerem seus oponentes.

32. Uma fábrica produz bolas coloridas na seguinte proporção: 60% são vermelhas, 30% azuis e 10% amarelas.

a) Qual a probabilidade de, escolhidas 5 bolas ao acaso, 2 serem vermelhas, 2 azuis e 1 amarela?

b) Qual a probabilidade de, escolhidas 6 bolas ao acaso, 2 serem vermelhas, 2 azuis e 2 amarelas?

c) Qual a probabilidade de, escolhidas 4 bolas ao acaso, nenhuma ser vermelha?

33. Uma fábrica produz bolas coloridas na seguinte proporção: 40% são vermelhas, 30% azuis, 20% verdes e 10% amarelas.

a) Qual a probabilidade de, escolhidas 6 bolas ao acaso, 2 serem vermelhas, 2 azuis, 1 verde e 1 amarela?

b) Qual a probabilidade de, escolhidas 8 bolas ao acaso, 3 serem vermelhas, 2 azuis, 2 verdes e 1 amarela?

c) Qual a probabilidade de, escolhidas 4 bolas ao acaso, nenhuma ser vermelha?

Apêndice 1.A – Revisão de análise combinatória

1.A.1 Fatorial

Define-se como fatorial de um número n ($n!$), sendo esse número um inteiro maior do que 1:

$$n! = n \times (n - 1) \times \dots \times 1$$

Assim:

$$2! = 2 \times 1 = 2$$

$$3! = 3 \times 2 \times 1 = 6$$

$$4! = 4 \times 3 \times 2 \times 1 = 24$$

$$5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$$

$$6! = 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720$$

E assim sucessivamente.

Note que:

$$3! = 3 \times 2!$$

$$4! = 4 \times 3!$$

$$5! = 5 \times 4!$$

$$6! = 6 \times 5!$$

Ou, generalizando:

$$n! = n \times (n - 1)!, \quad n > 2$$

Se estendermos essa propriedade para $n = 2$:

$$2! = 2 \times 1!$$

$$1! = \frac{2!}{2} = 1$$

Então, convenientemente, definimos:

$$1! = 1$$

Se continuarmos para $n = 1$:

$$1! = 1 \times 0!$$

$$0! = \frac{1!}{1} = 1$$

Portanto, temos:

$$n! = n \times (n - 1) \times \dots \times 1, n > 1$$

$$1! = 1$$

$$0! = 1$$

1.A.2 Permutações

Quantos anagramas são possíveis a partir da palavra *amor*?

AMOR MAOR OAMR RAMO
AMRO MARO OARM RAOM
ARMO MORA OMRA RMOA
AROM MOAR OMAR RMAO
AOMR MRAO ORAM ROAM
AORM MROA ORMA ROMA

Portanto, são possíveis 24 anagramas. Os anagramas são as permutações (*trocas de lugar*) das letras da palavra. Temos, então, no caso P_4 (lê-se “permutações de quatro elementos”) anagramas.

Se a palavra fosse *castelo*, o exercício anterior seria muito mais trabalhoso. Como fazer, então? Na palavra *amor*, são quatro *espaços* onde é possível colocar as quatro letras. No primeiro espaço, pode-se colocar qualquer uma das quatro letras. Para cada letra colocada no primeiro espaço, sobram três letras para preencher o segundo espaço; uma vez preenchido esse espaço, sobram apenas duas para o terceiro; finalmente, sobrará uma última letra no quarto espaço. Assim:

$$P_4 = 4 \times 3 \times 2 \times 1 = 4! = 24$$

Generalizando:

$$P_n = n!$$

Portanto, o total de anagramas da palavra *castelo* é:

$$P_7 = 7! = 5.040$$

1.A.3 Arranjos

Utiliza-se um **arranjo** quando se quer formar grupos a partir de um conjunto maior em que a ordem é importante. Por exemplo, de um grupo de cinco pessoas, deseja-se montar uma chapa eleitoral composta por um presidente, um vice e um tesoureiro.

Há três vagas. Para a vaga de presidente, são cinco opções; escolhido o presidente, há quatro opções para vice, sobrando três opções para tesoureiro. Então, o número total de chapas será dado por $A_{5,3}$ (lê-se “arranjos de cinco elementos, três a três”), calculado assim:

$$A_{5,3} = 5 \times 4 \times 3 = 60$$

Seriam 60 chapas possíveis, portanto. Faltaria, para completar o $5!$, multiplicar por 2 e por 1. Multiplicando e dividindo, temos:

$$A_{5,3} = \frac{5 \times 4 \times 3 \times 2 \times 1}{2 \times 1} = \frac{5!}{2!}$$

Generalizando:

$$A_{n,k} = \frac{n!}{(n-k)!}$$

1.A.4 Combinações

Quando se fala em **combinações**, como em arranjos, queremos formar grupos a partir de um conjunto de elementos, a diferença é que a ordem *não importa*.

Suponhamos que, no exemplo anterior, a chapa não tenha cargos (é uma chapa para um conselho, por exemplo), então, não importa quem é escolhido primeiro. O total de chapas possíveis será dado pelo número de arranjos, mas é preciso descontar o número de vezes em que a chapa é repetida (há simplesmente troca de posições); uma vez escolhida a chapa, trocando-se suas posições (isto é, fazendo permutações), obtém-se uma chapa idêntica. Portanto, o número de chapas será dado por $C_{5,3}$ (lê-se “combinações de cinco elementos, três a três”), calculado por:

$$C_{5,3} = \frac{A_{5,3}}{P_3} = \frac{5!}{2! \times 3!} = 10$$

Generalizando:

$$C_{n,k} = \frac{n!}{k!(n-k)!}$$

1.A.5 Triângulo de Pascal

Uma maneira simples de calcular combinações é por meio do triângulo de Pascal:

0	1
1	1 1
2	1 2 1
3	1 3 3 1
4	1 4 6 4 1
5	1 5 10 10 5 1
6	1 6 15 20 15 6 1
7	1 7 21 35 35 21 7 1

A construção do triângulo é simples. Cada linha começa e termina com 1. Os outros números são obtidos por meio da soma do número acima com o número à sua esquerda. Por exemplo, o terceiro número da linha correspondente ao número 5 (que é 10) foi obtido pela soma do segundo e terceiro números da linha acima ($4 + 6$). E assim pode ser feito com qualquer número apresentado no triângulo, inclusive para linhas que não foram mostradas (8, 9, 10 etc.).

As combinações podem ser obtidas imediatamente. Por exemplo, se quisermos combinações de seis elementos, devemos utilizar os números da linha correspondente, que são 1, 6, 15, 21, 15, 6 e 1. Temos que (verifique!):

$$C_{6,0} = 1$$

$$C_{6,1} = 6$$

$$C_{6,2} = 15$$

$$C_{6,3} = 21$$

$$C_{6,4} = 15$$

$$C_{6,5} = 6$$

$$C_{6,6} = 1$$

E assim é possível obter quaisquer combinações diretamente do triângulo.

Finalmente, outra propriedade, entre muitas, do triângulo de Pascal é que a soma dos números de uma linha é exatamente 2 elevado à potência do número correspondente. Por exemplo, observe a mesma linha correspondente ao número 6:

$$1 + 6 + 15 + 21 + 15 + 6 + 1 = 64 = 2^6$$

Apêndice 1.B – Definição axiomática de probabilidade

A ideia de definir probabilidade por meio de axiomas vem do desejo de tratar o assunto de maneira mais rigorosa.

Propor axiomas significa estabelecer um conjunto de regras, que devem ser no menor número possível. O conjunto de axiomas, entretanto, deve ser completo, na medida em que qualquer afirmação envolvendo probabilidades possa ser demonstrada utilizando apenas esses axiomas.

Seguem algumas definições.

O conjunto S de todos os resultados possíveis de um experimento aleatório é chamado de **espaço amostral**.

Seja σ um conjunto de subconjuntos de S , para o qual a probabilidade será definida. A esse conjunto denominamos **espaço de eventos**.

A definição de que subconjuntos de S farão parte do espaço de eventos é simples se S for discreto, pois, nesse caso, basta que definamos σ como o conjunto de todos os subconjuntos possíveis de S (incluindo o próprio S e o vazio). No caso de um conjunto S contínuo, ou mesmo no caso de um S muito grande, devemos nos contentar com uma definição mais restrita para σ .

O espaço de eventos σ deverá ter as seguintes propriedades:¹⁰

- $S \in \sigma$;
- se $A \in \sigma$, então $\bar{A} \in \sigma$;
- se A e $B \in \sigma$, então $A \cup B \in \sigma$;
- se $A_1, A_2, \dots \in \sigma$, então $\bigcup_{i=1}^{\infty} A_i \in \sigma$.

A probabilidade é, então, uma função que associa um elemento de σ a um número real, isto é:

$$P: \sigma \rightarrow \mathbb{R}$$

Obedecendo aos seguintes axiomas:

Axioma 1

Para qualquer $A \in \sigma$, $P(A) \geq 0$.

Axioma 2

$$P(S) = 1$$

Axioma 3

Dados $A_1, A_2, \dots, A_n \in \sigma$, disjuntos dois a dois, temos:

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i)$$

Isto é, a probabilidade da união dos eventos, sendo disjuntos, é a soma das probabilidades de cada um deles.

O **espaço de probabilidade** será a terna (S, σ, P) , em que S é o conjunto universo (espaço amostral); σ , um conjunto de subconjuntos de S ; e P , uma função que associa as probabilidades aos elementos de σ .

Todas as propriedades de probabilidade podem ser estabelecidas a partir dos três axiomas estabelecidos anteriormente.¹¹ Vejamos algumas delas:

Teorema 1.B.1

Se $A \in \sigma$, então $P(A) = 1 - P(\bar{A})$

Demonstração:

Pela própria definição de complementar, temos:

$$A \cup \bar{A} = S$$

¹⁰ Se σ segue essas propriedades, é dito um σ -field (*sigma-field*).

¹¹ Esses axiomas foram estabelecidos por Andrei Kolmogorov, matemático russo considerado o fundador da moderna teoria de probabilidades, em 1933. Antes de Kolmogorov, o Axioma 3 era limitado ao caso de dois conjuntos, isto é: se A e B são disjuntos, então $P(A \cup B) = P(A) + P(B)$.

Pelo Axioma 2:

$$P(S) = P(A \cup \bar{A}) = 1$$

E como A e \bar{A} são disjuntos, temos, pelo Axioma 3:

$$P(A \cup \bar{A}) = P(A) + P(\bar{A}) = 1$$

Portanto:

$$P(A) = 1 - P(\bar{A})$$

Teorema 1.B.2

$$P(\emptyset) = 0.$$

Demonstração:

Se $A = \emptyset$, então $\bar{A} = S$. Lembrando que $P(S) = 1$ pelo Axioma 2 e utilizando o Teorema 1.B.1:

$$P(\emptyset) = 1 - P(S) = 1 - 1 = 0$$

Teorema 1.B.3

Se $A, B \in \sigma$, então $P(A) = P(A \cap B) + P(A \cap \bar{B})$.

Demonstração:

$$A \cap S = A$$

Pela definição de complementar:

$$A \cap (B \cup \bar{B}) = A$$

Como a interseção tem propriedade distributiva:

$$(A \cap B) \cup (A \cap \bar{B}) = A$$

E, sendo os conjuntos $A \cap B$ e $A \cap \bar{B}$ disjuntos, temos, pelo Axioma 3:

$$P(A) = P[(A \cap B) \cup (A \cap \bar{B})] = P(A \cap B) + P(A \cap \bar{B})$$

Teorema 1.B.4

Se $A, B \in \sigma$, então $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Demonstração:

Temos que:

$$(A \cup B) \cap S = A \cup B$$

Pela definição de complementar:

$$(A \cup B) \cap (B \cup \bar{B}) = A \cup B$$

Como a união também tem propriedade distributiva, colocando B em evidência:

$$B \cup (A \cap \bar{B}) = A \cup B$$

Como os eventos B e $A \cap \bar{B}$ são disjuntos, pelo Axioma 3, temos:

$$P[B \cup (A \cap \bar{B})] = P(B) + P(A \cap \bar{B})$$

E, pelo Teorema 1.B.3, temos:

$$P(A) = P(A \cap B) + P(A \cap \bar{B})$$

$$P(A \cap \bar{B}) = P(A) - P(A \cap B)$$

Logo:

$$P(A \cup B) = P[B \cup (A \cap \bar{B})] = P(B) + P(A) - P(A \cap B)$$



Capítulo 2

MEDIDAS DE POSIÇÃO E DISPERSÃO

2.1 Variável aleatória

Variável aleatória (v.a.) é uma variável que está associada a uma *distribuição de probabilidade*.¹ Portanto, é uma variável que não tem valor fixo, podendo assumir vários valores.

O valor que cai ao se jogar um dado, por exemplo, pode ser 1, 2, 3, 4, 5 ou 6, com probabilidade igual a 1/6 para cada um dos valores (se o dado não estiver viciado); é, portanto, uma variável aleatória.

Assim como são variáveis aleatórias: o valor de uma ação ao final do dia de amanhã; o número de pontos de um time num campeonato que está começando esta semana; a quantidade de chuva que vai cair no mês que vem; a altura de uma criança em fase de crescimento daqui a seis meses; a taxa de inflação no mês que vem. Todas essas variáveis podem assumir diferentes valores, que, por sua vez, estão associados a probabilidades.

E *não* são variáveis aleatórias: o valor de uma ação no final do pregão de *ontem*; o número de pontos de um time num campeonato que já acabou; a altura de uma pessoa na faixa dos 30 anos de idade daqui a seis meses; a área útil de um apartamento; a velocidade de processamento de um computador. Todas essas variáveis têm valores fixos.

2.2 Medidas de posição central

2.2.1 Média

Há diferentes tipos de média: a **média aritmética**, a mais comum, é a soma dos elementos de um conjunto dividido pelo número de elementos desse conjunto. Assim, um grupo de cinco pessoas, com idades de 21, 23, 25, 28 e 31, terá média (aritmética) de idade dada por:

$$\bar{X} = \frac{21 + 23 + 25 + 28 + 31}{5} = 25,6 \text{ anos}$$

¹ Voltaremos ao conceito de distribuição de probabilidade no próximo capítulo.

De modo geral, a média aritmética será dada por:

$$\bar{X} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Ou, escrevendo de maneira mais resumida:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

A média aritmética também pode ser ponderada, o que não a torna um tipo diferente de média. Ponderar significa *atribuir pesos*; logo, se um valor tem peso maior, significa simplesmente que ele entrará *mais vezes* na média. Digamos, por exemplo, que em três provas um aluno tenha tirado 4, 6 e 8. Se a média não for ponderada, é óbvio que será 6.

Se, no entanto, a média for ponderada da seguinte forma: a primeira prova com peso 1, a segunda com 2 e a terceira com 3, a média será calculada como se as provas com maior peso tivessem ocorrido mais vezes, ou seja:

$$\bar{X} = \frac{4+6+6+8+8+8}{6}$$

Ou, simplesmente:

$$\bar{X} = \frac{4 \times 1 + 6 \times 2 + 8 \times 3}{6} \approx 6,7$$

Os pesos podem ser o número de vezes que um valor aparece. Suponhamos que numa classe de 20 alunos haja oito com idade de 22 anos, sete de 23, três de 25, um de 28 e um de 30. A quantidade em que cada número aparece no conjunto é chamada de **frequência** (frequência absoluta, nesse caso, pois se trata da quantidade de alunos com determinada idade). A média de idade, então, será dada por:

$$\bar{X} = \frac{22 \times 8 + 23 \times 7 + 25 \times 3 + 28 \times 1 + 30 \times 1}{20} = 23,5 \text{ anos}$$

A frequência também pode ser expressa em proporções, sendo chamada nesse caso de **frequência relativa**. No exemplo anterior, há oito alunos com 22 anos de idade em um total de 20, portanto, nessa classe há $8/20 = 0,4 = 40\%$ dos alunos com essa idade. Da mesma forma, temos 35% com 23, 15% com 25 e 5% com 28 e 30, respectivamente. A média de idade, então, pode ser calculada da seguinte forma:

$$\bar{X} = 22 \times 0,4 + 23 \times 0,35 + 25 \times 0,15 + 28 \times 0,05 + 30 \times 0,05 = 23,5$$

Repare que o segundo *modo* de calcular (usando a frequência relativa) nada mais é do que o primeiro (usando a frequência absoluta) com a fração simplificada (dividiu-se o valor dos pesos pelo número total).

Outro tipo de média é a **média geométrica**. A média geométrica para o aluno que tirou notas 4, 6 e 8 será:

$$G = \sqrt[3]{4 \times 6 \times 8} \cong 5,8$$

Ou, genericamente:

$$G = \sqrt[n]{X_1 \times X_2 \times \dots \times X_n}$$

Ou, ainda, de maneira mais resumida:

$$G = \left(\prod_{i=1}^n X_i \right)^{\frac{1}{n}}$$

Repare que a média geométrica *zera* se um dos elementos for zero.

A média geométrica também pode ser ponderada: se os pesos das provas forem 1, 2 e 3, ela será dada por:

$$G = \sqrt[6]{4^1 \times 6^2 \times 8^3} \cong 6,5$$

Há ainda um terceiro tipo de média, a **média harmônica**. No exemplo das notas, ela será dada por:

$$H = \frac{1}{\frac{1}{4} + \frac{1}{6} + \frac{1}{8}} = \frac{3}{\frac{1}{4} + \frac{1}{6} + \frac{1}{8}} \cong 5,5$$

De modo geral:

$$H = \frac{n}{\frac{1}{X_1} + \frac{1}{X_2} + \dots + \frac{1}{X_n}}$$

Ou ainda:

$$H = \frac{n}{\sum_{i=1}^n \frac{1}{X_i}}$$

Também é possível que a média harmônica seja ponderada. Repetindo o exemplo anterior:

$$H = \frac{6}{\frac{1}{4} \times 1 + \frac{1}{6} \times 2 + \frac{1}{8} \times 3} \cong 6,3$$

É possível notar, tanto para as médias simples (sem pesos) como para as ponderadas, que, em geral, a média aritmética é maior do que a média geométrica e essa, por sua vez, é maior do que a harmônica. Isso é verdade, exceto, obviamente, quando os valores são todos iguais. Temos, então, que:

$$\bar{X} \geq G \geq H$$

Exemplo 2.2.1.1

Um aluno tira as seguintes notas bimestrais: 3, 4, 5, 7 e 8,5. Determinemos qual seria sua média final se esta fosse calculada dos três modos (aritmética, geométrica e harmônica), em cada um dos seguintes casos.

- a) As notas dos bimestres têm os mesmos pesos.

Nesse caso, a média aritmética final seria:

$$\bar{X} = \frac{3 + 4,5 + 7 + 8,5}{4} = \frac{23}{4}$$

$$\boxed{\bar{X} = 5,75}$$

A média geométrica seria:

$$G = \sqrt[4]{3 \times 4,5 \times 7 \times 8,5} = \sqrt[4]{803,25}$$

$$\boxed{G \approx 5,32}$$

E a harmônica seria:

$$H = \frac{4}{\frac{1}{3} + \frac{1}{4,5} + \frac{1}{7} + \frac{1}{8,5}}$$

$$\boxed{H \approx 4,90}$$

- b) Supondo que os pesos para as notas bimestrais sejam 1, 2, 3 e 4.

Agora os pesos dos quatro bimestres totalizam 10, portanto, a média aritmética final será:

$$\bar{X} = \frac{1 \times 3 + 2 \times 4,5 + 3 \times 7 + 4 \times 8,5}{10} = \frac{67}{10}$$

$$\boxed{\bar{X} = 6,7}$$

A geométrica será:

$$G = \sqrt[10]{3^1 \times 4,5^2 \times 7^3 \times 8,5^4}$$

$$\boxed{G \approx 6,36}$$

E a harmônica:

$$H = \frac{10}{\frac{1}{3} + \frac{2}{2,45} + \frac{3}{7} + \frac{4}{8,5}}$$

$$H \cong 5,96$$

c) Supondo que os pesos sejam, respectivamente, 30%, 25%, 25% e 20%.

Agora os pesos são dados em termos relativos (percentuais) e somam, portanto, 1.

O cálculo da média aritmética será, então:

$$\bar{X} = 0,3 \times 3 + 0,25 \times 4,5 + 0,25 \times 7 + 0,2 \times 8,5$$

$$\bar{X} = 5,475$$

O da média geométrica será:

$$G = 3^{0,3} \times 4,5^{0,25} \times 7^{0,25} \times 8,5^{0,2}$$

$$G \cong 5,05$$

E a harmônica:

$$H = \frac{1}{\frac{1}{3} \times 0,3 + \frac{1}{4,5} \times 0,25 + \frac{1}{7} \times 0,25 + \frac{1}{8,5} \times 0,2}$$

$$H \cong 4,66$$

Exemplo 2.2.1.2

Foi medida a altura de 30 pessoas. Os valores estão na Tabela 2.1.

Tabela 2.1 – Altura de 30 pessoas (em cm)

159	168	172	175	181
161	168	173	176	183
162	169	173	177	185
164	170	174	178	190
166	171	174	179	194
167	171	174	180	201

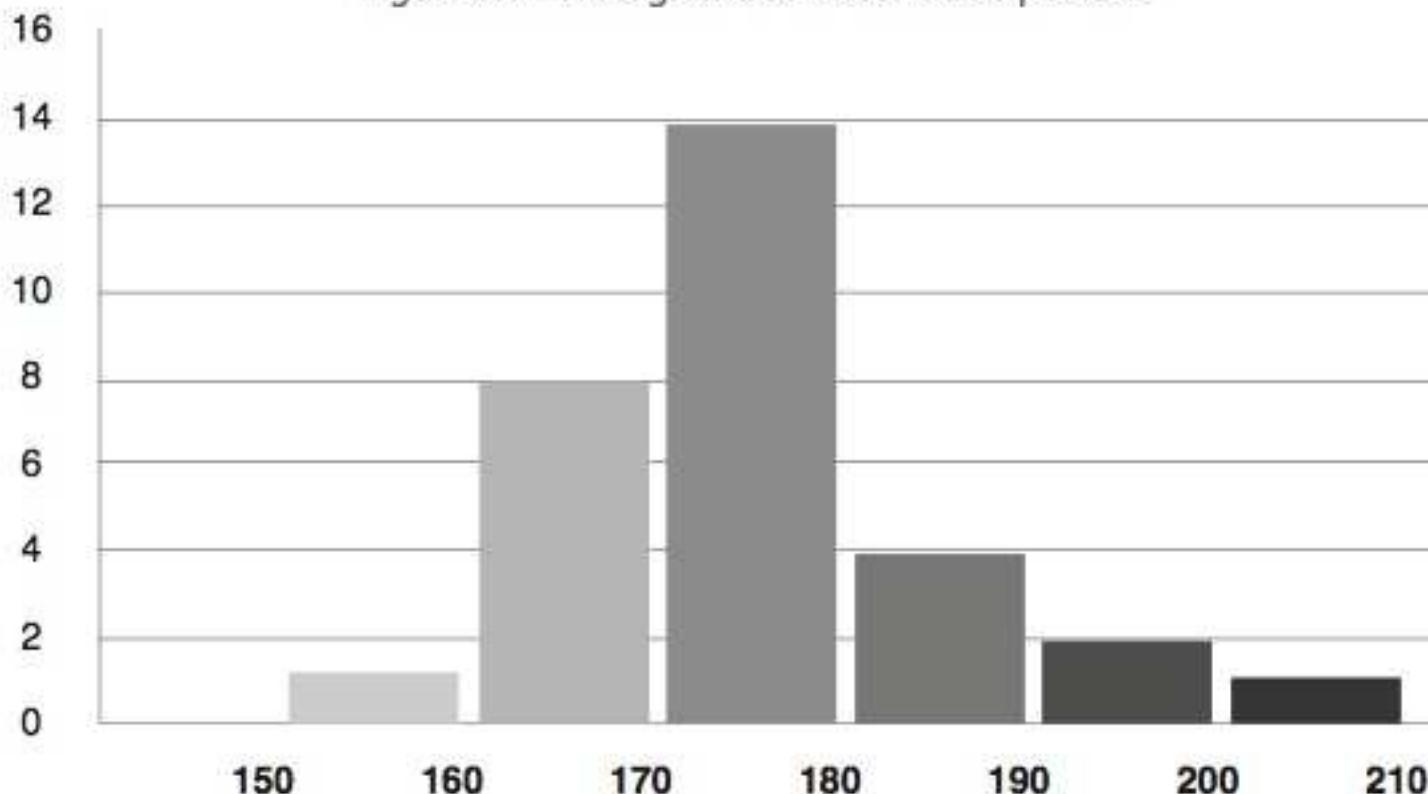
Agrupemos essas pessoas em **classes** de 10 cm e façamos o **histograma** correspondente. Para agrupar em classes de 10 cm, o mais lógico (mas não obrigatório) seria agrupar em: de 150 a 160; de 160 a 170, e assim sucessivamente. O problema é: onde incluir aqueles que têm, por exemplo, exatamente 170 cm? Na classe de 160 a 170 ou na de 170 a 180? Há de se escolher uma, mas essa escolha é completamente arbitrária. Vamos optar por incluir

sempre o limite inferior, por exemplo, a classe de 170 a 180 inclui todas as pessoas com 170 cm (inclusive) até 180 cm (exclusive),² para o que utilizaremos a notação [170; 180[. Então, para os valores da Tabela 2.1, teremos:

Tabela 2.2 – Classes de 10 cm	
[150; 160[1
[160; 170[8
[170; 180[14
[180; 190[4
[190; 200[2
[200; 210[1

Um histograma é uma maneira gráfica de representar esse agrupamento, utilizando-se retângulos, cuja altura é proporcional ao número de elementos em cada classe. O histograma para o agrupamento realizado é mostrado na Figura 2.1.

Figura 2.1 – Histograma da altura de 30 pessoas



Exemplo 2.2.1.3

A partir dos dados agrupados do exemplo anterior, calculemos a média.³

Utilizaremos como dados os agrupamentos, é como se – e com frequência isso acontece – não tivéssemos conhecimento dos dados que originaram esse agrupamento.

Uma vez que nossa única informação é o agrupamento (seja pela tabela, seja pelo histograma), não é possível saber *como* os dados se distribuem pelo agrupamento, então, o melhor a fazer, na falta de outra opção, é supor que os dados se distribuem igualmente em cada agrupamento, de modo que, por exemplo, no agrupamento que vai de 170 a 180 é como se houvesse 14 pessoas com altura de 175 cm.

² Em linguagem de conjuntos, equivaleria a dizer que o conjunto é fechado em 170 e aberto em 180.

³ Quando se fala *média*, sem especificar, supõe-se que seja média aritmética.

Em outras palavras, tomaremos a média de cada classe para o cálculo da média total. Obviamente, a não ser por uma grande coincidência, esse não será o valor correto da média, mas é uma aproximação e, de novo, é o melhor que se pode fazer, dada a limitação da informação. Então:

$$\bar{X} = \frac{155 \times 1 + 165 \times 8 + 175 \times 14 + 185 \times 4 + 195 \times 2 + 205 \times 1}{30}$$

$$\bar{X} = 175,33 \text{ cm}$$

Repare que o valor correto da média, tomando-se os 30 dados originais, é 174,5 cm.

2.2.2 Moda

Moda é o elemento de maior frequência, ou seja, que aparece o maior número de vezes.⁴ No exemplo das idades na classe com 20 alunos, a moda é 22 anos, que é a idade mais frequente nesse conjunto.

Pode haver, entretanto, mais de uma moda em um conjunto de valores. Se houver apenas uma moda, a distribuição é chamada de **unimodal**; se houver duas, de **bimodal**; caso haja muitas modas, **multimodal** e, se não houver moda, o que é possível, **amodal**.

2.2.3 Mediana

Mediana é o valor que divide um conjunto ao meio. Por exemplo, num grupo de cinco pessoas com alturas de 1,60 m, 1,65 m, 1,68 m, 1,70 m e 1,73 m, a mediana é 1,68 m, pois há o mesmo número de pessoas mais altas e mais baixas (duas).

A mediana apresenta uma vantagem em relação à média, como veremos a seguir. No grupo, a média é 1,672 m; então, nesse caso, tanto a média como a mediana nos dão uma ideia razoável do grupo de pessoas que estamos considerando. Se, no entanto, retirarmos a pessoa de 1,73 m, substituindo-a por outra de 2,10 m, a média passará a ser 1,746 m. Nesse caso, a média não seria muito representativa de um grupo que, afinal de contas, tem apenas uma pessoa acima de 1,70 m. A mediana, entretanto, fica inalterada.

A mediana, ao contrário da média, *não é sensível a valores extremos*.

Seguindo a mesma lógica, os **quartis** são os elementos que dividem o conjunto em quatro partes iguais. Assim, o primeiro quartil é aquele elemento que é maior do que 1/4 dos elementos e, portanto, menor do que 3/4 deles; o segundo quartil (que coincide com a mediana) é aquele que divide 2/4 para cima e 2/4 para baixo; finalmente, o terceiro quartil é aquele elemento que tem 3/4 abaixo de si e 1/4 acima.

Da mesma forma, se dividirmos o conjunto em oito pedaços iguais, teremos os **octis**; se dividirmos em 10 e, mais genericamente, os **percentis**: o percentil de ordem 20 é aquele que tem abaixo de si 20% dos elementos e acima, 80%.

⁴ Na linguagem cotidiana, também dizemos que algo está na moda quando é adotado pela maioria das pessoas.

Exemplo 2.2.3.1

A partir da Tabela 2.1, determinemos a moda, a mediana e os quartis do conjunto.

a) A moda:

O elemento que aparece mais vezes (três) é 174 cm, portanto:

$$Mo = 174 \text{ cm}$$

E só há uma moda, o que não é necessário que ocorra. No caso deste exemplo, bastaria que houvesse mais uma pessoa com 168 cm de altura para que essa distribuição se tornasse bimodal.

b) A mediana:

Há 30 dados. Do menor para o maior, o 15º dado é, pela ordem, 173 cm, enquanto o 16º é 174 cm. Como a mediana deve ter 15 elementos abaixo e 15 acima, tomaremos o ponto médio entre o 15º e o 16º dado:

$$Md = \frac{173 + 174}{2}$$

$$Md = 173,5 \text{ cm}$$

c) O primeiro e o segundo quartis:

Devemos dividir o total de elementos por quatro, o que dá 7,5. Como o sétimo e o oitavo elementos, indo do menor para o maior, são iguais:

$$\text{Primeiro quartil} = 168 \text{ cm}$$

O segundo quartil coincide com a mediana:

$$\text{Segundo quartil} = Md = 173,5 \text{ cm}$$

2.3 Medidas de dispersão

É muito comum se ouvir: “em estatística, quando uma pessoa come dois frangos enquanto outra passa fome, na média ambas comem um frango e estão, portanto, bem alimentadas” ou “se uma pessoa está com os pés no forno e a cabeça no freezer, na média experimenta uma temperatura agradável”. É claro que essas distorções têm de ser percebidas – e são! – pela estatística. É para isso que servem as **medidas de dispersão**, isto é, medidas de como os dados estão *agrupados*: mais ou menos próximos entre si (menos ou mais dispersos).

2.3.1 Variância

Uma das medidas mais comuns de dispersão é a **variância**. Tomemos o exemplo de três frangos para três indivíduos, ilustrado na Tabela 2.3. Na situação 1, há uma divisão equitativa, enquanto na situação 2, um indivíduo come demais e o outro passa fome.

Tabela 2.3 – Frangos consumidos

	Situação 1	Situação 2
Indivíduo 1	1	2
Indivíduo 2	1	1
Indivíduo 3	1	0

É claro que, em ambas as situações, a média é um frango por indivíduo. Para encontrar uma maneira de distinguir numericamente os dois casos, uma tentativa poderia ser subtrair a média de cada valor, como mostrado na Tabela 2.4.

Tabela 2.4 – Frangos consumidos (subtração de médias)

	Situação 1	Situação 2
Indivíduo 1	$1 - 1 = 0$	$2 - 1 = 1$
Indivíduo 2	$1 - 1 = 0$	$1 - 1 = 0$
Indivíduo 3	$1 - 1 = 0$	$0 - 1 = -1$
Média	0	0

O que não resolveu muito, pois a média dos desvios em relação à média⁵ (valor menos a média) continua igual. Mais precisamente, ambas são zero. Isso ocorre porque, na situação 2, os valores abaixo da média, que ficam negativos, compensam os que ficam acima da média, os positivos.

Para eliminar esse inconveniente dos sinais, pode-se elevar todos os valores encontrados ao quadrado, como realizado na Tabela 2.5.

Tabela 2.5 – Frangos consumidos (valores elevados ao quadrado)

	Situação 1	Situação 2
Indivíduo 1	$(1 - 1)^2 = 0$	$(2 - 1)^2 = 1$
Indivíduo 2	$(1 - 1)^2 = 0$	$(1 - 1)^2 = 0$
Indivíduo 3	$(1 - 1)^2 = 0$	$(0 - 1)^2 = 1$
Média	0	2/3

Dessa forma, conseguimos encontrar uma medida que distingue a dispersão entre as duas situações.

Na situação 1, não há dispersão, pois todos os dados são iguais, e a variância é zero.

Na situação 2, a dispersão é, obviamente, maior – encontramos uma variância de $2/3 \approx 0,67$.

Basicamente, para se encontrar a variância: (1) subtraem-se todos os elementos do conjunto pela média, (2) elevam-se os resultados ao quadrado; (3) tira-se a média dos

⁵ Aliás, vale a pena lembrar que a soma dos desvios em relação à média é zero, sempre.

valores encontrados. Portanto, a variância de um conjunto de valores X , que chamaremos de $\text{var}(X)$ ou σ_x^2 , será dada por:

$$\text{var}(X) = \sigma_x^2 = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2}{n}$$

Ou ainda:

$$\text{var}(X) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

Variância é, portanto, uma *medida de dispersão que lembra quadrados*. Esse último aspecto, aliás, pode ser um problema na utilização da variância.

Na situação 2 do exemplo dos frangos, encontramos uma variância de 0,67... *frangos ao quadrado*? Sim, porque elevamos, por exemplo, um frango ao quadrado. Da mesma forma que, na geometria, um quadrado de lado 2 m tem área $(2\text{m})^2 = 4\text{m}^2$, temos que $(\text{um frango})^2 = \text{um frango}^2$! E assim também valeria para outras variáveis: renda medida em reais ou dólares teria variância medida em reais ao quadrado ou dólares ao quadrado.

Além de causar estranheza, isso dificulta, por exemplo, uma comparação com a média. Para eliminar esse efeito, utiliza-se outra medida de dispersão que é, na verdade, uma pequena alteração da variância.

Exemplo 2.3.1.1

Utilizando o agrupamento da Tabela 2.2, determinemos a variância.

A variância é calculada com o mesmo princípio utilizado para a média, ou seja, tomando-se o valor médio de cada classe como seu representante. Assim:

$$\begin{aligned} \text{var}(X) &= \frac{1}{30} [(155 - 175,33)^2 \times 1 + (165 - 175,33)^2 \times 8 + (175 - 175,33)^2 \times 14 \\ &\quad + (185 - 175,33)^2 \times 4 + (195 - 175,33)^2 \times 2 + (205 - 175,33)^2 \times 1] \\ \text{var}(X) &\approx 108,89 \end{aligned}$$

Mais uma vez, é uma aproximação. Observe que o valor correto da variância (utilizando os dados da Tabela 2.1) é 86,92.

2.3.2 Desvio padrão

Para eliminar o efeito dos quadrados existente na variância, basta extrair a raiz quadrada. Chamaremos essa nova medida de **desvio padrão** da variável X ($dp(X)$ ou σ_X):

$$dp(X) \equiv \sigma_X = \sqrt{\text{var}(x)}$$

Portanto, o desvio padrão na situação 2 do exemplo dos frangos é dado por:

$$dp(X) = \sqrt{0,67} \equiv 0,8 \text{ frangos}$$

Estando na mesma unidade os dados (e a média) – no caso específico, frangos –, é possível comparar o desvio padrão com a média: nesse caso, o desvio padrão equivale a 80%⁶ da média.

Note que, se o objetivo é a comparação entre dois conjuntos de dados, tanto faz usar a variância ou o desvio padrão. Se a variância é maior, o desvio padrão também será maior (e vice-versa), necessariamente.

2.3.3 Outra maneira de calcular a variância

Se, a partir da definição de variância, ela for desenvolvida algebraicamente, obtém-se:

$$\text{var}(X) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

$$\text{var}(X) = \frac{1}{n} \sum_{i=1}^n (X_i^2 - 2X_i\bar{X} + \bar{X}^2)$$

$$\text{var}(X) = \frac{1}{n} \sum_{i=1}^n X_i^2 - \frac{1}{n} \sum_{i=1}^n 2X_i\bar{X} + \frac{1}{n} \sum_{i=1}^n \bar{X}^2$$

$$\text{var}(X) = \frac{1}{n} \sum_{i=1}^n X_i^2 - 2\bar{X} \frac{1}{n} \sum_{i=1}^n X_i + \frac{1}{n} n\bar{X}^2$$

$$\text{var}(X) = \frac{1}{n} \sum_{i=1}^n X_i^2 - 2\bar{X}^2 + \bar{X}^2$$

$$\text{var}(X) = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2$$

Ou, em outras palavras:

$$\text{var}(X) = \text{média dos quadrados} - \text{quadrado da média}$$

Utilizando esse método para calcular a variância da situação 2 do exemplo dos frangos, chega-se à Tabela 2.6.

Tabela 2.6 – Frangos consumidos (variância)

	Situação 2	Valor ao quadrado
Indivíduo 1	2	4
Indivíduo 2	1	1
Indivíduo 3	0	0
Média	1	5/3

$$\text{var}(X) = \text{média dos quadrados} - \text{quadrado da média} = 5/3 - 1^2 = 2/3$$

⁶ Essa proporção, obtida por meio da divisão do desvio padrão pela média, é também chamada de coeficiente de variação.

Nesse caso, esse valor é igual ao que foi calculado usando a definição inicial de variância. Tome, contudo, o exemplo de um aluno que tem as seguintes notas em três disciplinas, como apresentado na Tabela 2.7.

Tabela 2.7 – Notas do aluno A		
Aluno A	Notas	Notas ao quadrado
Economia	3	9
Contabilidade	2	4
Administração	4	16
Matemática	1	1
Média	2,5	7,5

Para esse aluno:

$$\bar{X} = 2,5$$

$$\text{var}(X) = 7,5 - 2,5^2 = 1,25$$

$$\text{dp}(X) = 1,12$$

Suponha agora um aluno B com notas que são exatamente o dobro, conforme a Tabela 2.8.

Tabela 2.8 – Notas do aluno B		
Aluno B	Notas	Notas ao quadrado
Economia	6	36
Contabilidade	4	16
Administração	8	64
Matemática	2	4
Média	5	30

Para o aluno B, a média é:

$$\bar{X} = 5$$

Ou seja, se os valores dobraram, a média dobra. Quanto à variância:

$$\text{var}(X) = 30 - 5^2 = 5 = 4 \times 1,25$$

Ou seja, se os valores dobraram, a variância quadruplica. Isso porque variância lembra quadrados. Em outras palavras, vale a relação:⁷

$$\text{var}(aX) = a^2 \text{var}(X) \tag{2.3.3.1}$$

$$\text{dp}(X) = 2,24$$

⁷ Veja demonstrações nos Apêndices 2.B.1 e 2.B.2.

Isto é, o desvio padrão dobra, assim como a média. Vale, portanto, a relação:

$$dp(aX) = |a| \cdot dp(X) \quad (2.3.3.2)$$

Note que o a está em módulo, pois se fosse -2 , por exemplo, ao quadrado seria 4 e a raiz quadrada 2 do mesmo jeito, ou seja, o desvio padrão dobraria do mesmo jeito, não ficaria negativo. Aliás, não existe desvio padrão negativo.

Agora tomemos um aluno C, que tira cinco pontos a mais do que o aluno A em todas as matérias, como na Tabela 2.9.

Tabela 2.9 – Notas do aluno C

Aluno C	Notas	Notas ao quadrado
Economia	8	64
Contabilidade	7	49
Administração	9	81
Matemática	6	36
Média	7,5	57,5

Para esse aluno:

$$\bar{X} = 7,5$$

Se o aluno tira cinco pontos a mais em cada disciplina, a média também será de cinco pontos a mais:

$$\text{var}(X) = 57,5 - 7,5^2 = 1,25$$

$$dp(X) = 1,12$$

A variância e o desvio padrão são os mesmos do aluno A. Isso porque são medidas de dispersão – se se somar o mesmo valor a todas as notas de A, elas continuarão dispersas, espalhadas da mesma forma, apenas mudarão de posição. Valem, portanto, as relações:⁸

$$\text{var}(X + a) = \text{var}(X) \quad (2.3.3.3)$$

$$dp(X + a) = dp(X) \quad (2.3.3.4)$$

2.3.4 Relações entre variáveis – covariância

A **covariância** pode ser entendida como uma “variância conjunta” entre duas variáveis. Enquanto a variância sai de quadrados (da variável menos a média), a covariância é definida por meio de produtos:

$$\text{cov}(X, Y) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$$

⁸ Cuja demonstração podem ser vistas nos Apêndices 2.B.3 e 2.B.4.

Assim como a variância, a covariância pode ser calculada de outra forma, cuja demonstração está no Apêndice 2.B.5:

$$\text{cov}(X, Y) = \text{média dos produtos} - \text{produto das médias} \quad (2.3.4.1)$$

A Tabela 2.10 traz um exemplo do consumo e da taxa de juros de um país.

Tabela 2.10 – Consumo e taxa de juros

Ano	Consumo (X)	Taxa de juros (Y)	Produto (XY)
1	800	10	8.000
2	700	11	7.700
3	600	13	7.800
4	500	14	7.000
Média	650	12	7.625

$$\text{cov}(X, Y) = 7.625 - 650 \times 12 = -175$$

E agora entre o consumo e a renda, como mostra a Tabela 2.11.

Tabela 2.11 – Consumo e renda

Ano	Consumo (X)	Renda (Y)	Produto (XY)
1	600	1.000	600.000
2	700	1.100	770.000
3	800	1.300	1.040.000
4	900	1.400	1.260.000
Média	750	1.200	917.500

$$\text{cov}(X, Y) = 917.500 - 750 \times 1.200 = 17.500$$

A primeira diferença que se nota entre os dados das Tabelas 2.10 e 2.11 é o sinal da covariância em cada um deles. A covariância é negativa entre o consumo e a taxa de juros, e positiva entre o consumo e a renda. Isso porque consumo e renda caminham na mesma direção (quando aumenta um, aumenta o outro), quando isso ocorre, o sinal da covariância é positivo.

Já o consumo e a taxa de juros se movem em sentidos opostos (quando aumenta um, cai o outro), assim, o sinal da covariância é negativo.

A covariância entre duas variáveis é influenciada pela *importância* que uma variável tem sobre a outra, de tal modo que duas variáveis independentes têm covariância zero, mas a recíproca não é verdadeira.

Entretanto, não é possível concluir, pelos valores obtidos, que a renda é mais importante do que a taxa de juros para a determinação do consumo só porque o valor da covariância entre o consumo e a renda é bem maior do que o valor entre o consumo e a taxa de juros. Isso porque a covariância também é afetada pelos valores das variáveis. A covariância entre consumo e renda é maior porque os valores da renda são bem maiores que os da taxa de juros.

2.3.5 Coeficiente de correlação

O **coeficiente de correlação** é obtido retirando-se o efeito dos valores de cada uma das variáveis sobre a covariância. Isso é feito dividindo-se esta última pelos desvios padrão das variáveis.

O coeficiente de correlação é dado, então, por:

$$\text{corr}(X,Y) \equiv \rho_{XY} = \frac{\text{cov}(X,Y)}{\text{dp}(X).\text{dp}(Y)}$$

No exemplo do consumo e da renda, os desvios padrão são, respectivamente, 111,8 e 158,1 (verifique!). O coeficiente de correlação é dado, então, por:

$$\rho_{XY} = \frac{17.500}{111,8 \times 158,1} = 0,99$$

O sinal do coeficiente de correlação é o mesmo da covariância (e deve ser interpretado da mesma forma). Seus valores variam apenas no intervalo de -1 a 1 e podem ser interpretados como um percentual.⁹ Portanto, um valor de 0,99 (quase 1) indica que a renda é muito importante para a determinação do consumo.

O valor de 1 (ou -1) para o coeficiente de correlação só é encontrado para duas variáveis que tenham uma relação exata e dada por uma função do primeiro grau. Por exemplo, o número de cadeiras e de assentos em uma sala de aula; o número de pessoas e dedos das mãos (supondo que não haja indivíduos polidáctilos, acidentados ou com defeitos congênitos); a área útil e a área total em apartamentos de um mesmo edifício.

Valores muito pequenos (em módulo) indicam que as variáveis têm pouca influência uma sobre a outra ou que essa influência é pouco explicada por uma função de 1º grau. Vejamos os exemplos a seguir.

Exemplo 2.3.5.1

Considere a Tabela 2.12 com valores de X e Y. Determine o coeficiente de correlação entre essas duas variáveis.

Tabela 2.12 – Valores de X e Y

X	Y
2	3
6	11
11	21
13	25
18	35

* Com ressalvas, pois ele é calculado sem considerar a influência de outras variáveis.

Calculando as médias, quadrados e produtos:

Tabela 2.13 – Valores de X e Y com cálculos

X	Y	X^2	Y^2	XY
2	3	4	9	6
6	11	36	121	66
11	21	121	441	231
13	25	169	625	325
18	35	324	1.225	630

Da Tabela 2.13, obtém-se as seguintes informações: $\bar{X} = 10$, $\bar{Y} = 19$, média dos quadrados de $X = 130,8$, média dos quadrados de $Y = 484,2$, média dos produtos de X e $Y = 251,6$. Portanto:

$$\text{var}(X) = 130,8 - 10^2 = 30,8.$$

$$\text{var}(Y) = 484,2 - 19^2 = 123,2.$$

$$\text{cov}(X,Y) = 251,6 - 10 \times 19 = 61,6.$$

$$\text{corr}(X,Y) = \frac{61,6}{\sqrt{30,8 \times 123,2}} = 1.$$

Exatamente 1! Por quê? Repare que $Y = 2X - 1$, exatamente. A relação entre essas duas variáveis é descrita por uma reta. Nesse caso, o coeficiente será exatamente 1, ou -1, se for uma reta decrescente. Verifique inventando uma variável Z qualquer como sendo $Z = aX + b$. Quando você escolher valores positivos para a , o coeficiente de correlação será 1; nos casos em que a for negativo, será -1.

Exemplo 2.3.5.2

Considere a Tabela 2.14 com valores de X e Y . Determine o coeficiente de correlação entre essas duas variáveis.

Tabela 2.14 – Valores de X e Y

X	Y
-2	4
-1	1
0	0
1	1
2	4

Novamente calculando médias, quadrados e produtos:

Tabela 2.15 – Valores de X e Y com cálculos				
X	Y	X ²	Y ²	XY
-2	4	4	16	-8
-1	1	1	1	-1
0	0	0	0	0
1	1	1	1	1
2	4	4	16	8

Da Tabela 2.15, obtém-se as seguintes informações: $\bar{X} = 0$, $\bar{Y} = 2$, média dos quadrados de X = 2, média dos quadrados de Y = 6,8, média dos produtos de X e Y = 0. Portanto:

$$\text{var}(X) = 2 - 0^2 = 2.$$

$$\text{var}(Y) = 6,8 - 2^2 = 2,8.$$

$$\text{cov}(X,Y) = 0 - 0 \times 2 = 0.$$

$$\text{corr}(X,Y) = \frac{0}{\sqrt{2 \times 2,8}} = 0.$$

Zero! Significa que não há qualquer relação entre X e Y? Na verdade não, pois $Y = X^2$. Não só há relação, como essa relação também é exata. Apesar disso, o coeficiente de correlação foi zero, porque não há nenhuma reta possível que “conte a história” de X e Y, nesse caso. Note que, se tomarmos apenas as três primeiras linhas da Tabela 2.14 (ou mesmo as três últimas), o resultado será bem diferente (verifique!), porque, nesse caso, uma reta, ainda que não perfeitamente, pode ser usada para descrever essa relação.

2.3.6 Outras propriedades

No exemplo do consumo e da taxa de juros, multipliquemos o consumo por três e a taxa de juros por dois, como na Tabela 2.16.

Tabela 2.16 – Consumo e taxa de juros (multiplicados por três e por dois, respectivamente)			
Ano	3X	2Y	Produto
1	2.400	20	48.000
2	2.100	22	46.200
3	1.800	26	46.800
4	1.500	28	42.000
Média	1.950	24	45.750

A nova covariância é dada por:

$$\text{cov}(3X, 2Y) = 45.750 - 1.950 \times 24 = -1.050 = 6 \times (-175)$$

Ou seja, o sétuplo da covariância entre as variáveis originais. A propriedade apresentada aqui pode ser assim resumida:

$$\text{cov}(aX, bY) = a.b.\text{cov}(X, Y) \quad (2.3.6.1)^{10}$$

Tomemos agora, como mostra a Tabela 2.17, duas variáveis X e Y .

	X	Y	X^2	Y^2	XY
	10	1	100	1	10
	12	3	144	9	36
	18	2	324	4	36
	20	2	400	4	40
Média	15	2	242	4,5	30,5

Podemos calcular:

$$\text{var}(X) = 242 - 15^2 = 17$$

$$\text{var}(Y) = 4,5 - 2^2 = 0,5$$

$$\text{cov}(X, Y) = 30,5 - 15 \times 2 = 0,5$$

Vamos inventar duas variáveis: $X + Y$ e $X - Y$, como na Tabela 2.18.

	$X + Y$	$X - Y$	$(X + Y)^2$	$(X - Y)^2$
	11	9	121	81
	15	9	225	81
	20	16	400	256
	22	18	484	324
Média	17	13	307,5	185,5

Então, temos:

$$\text{var}(X + Y) = 307,5 - 17^2 = 18,5$$

$$\text{var}(X - Y) = 185,5 - 13^2 = 16,5$$

Note que poderíamos obter esses resultados a partir dos valores anteriores, da seguinte forma:

$$\text{var}(X + Y) = 17 + 0,5 + 2 \times 0,5 = 18,5$$

$$\text{var}(X - Y) = 17 + 0,5 - 2 \times 0,5 = 16,5$$

Generalizando, vem:¹¹

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y) \quad (2.3.6.2)^{12}$$

$$\text{var}(X - Y) = \text{var}(X) + \text{var}(Y) - 2\text{cov}(X, Y) \quad (2.3.6.3)^{13}$$

¹⁰ A demonstração desta expressão está no Apêndice 2.B.6.

¹¹ Note que é muito semelhante à forma do produto notável $(a+b)^2 = a^2 + b^2 + 2ab$, fazendo a variância análoga ao quadrado e a covariância análoga ao produto.

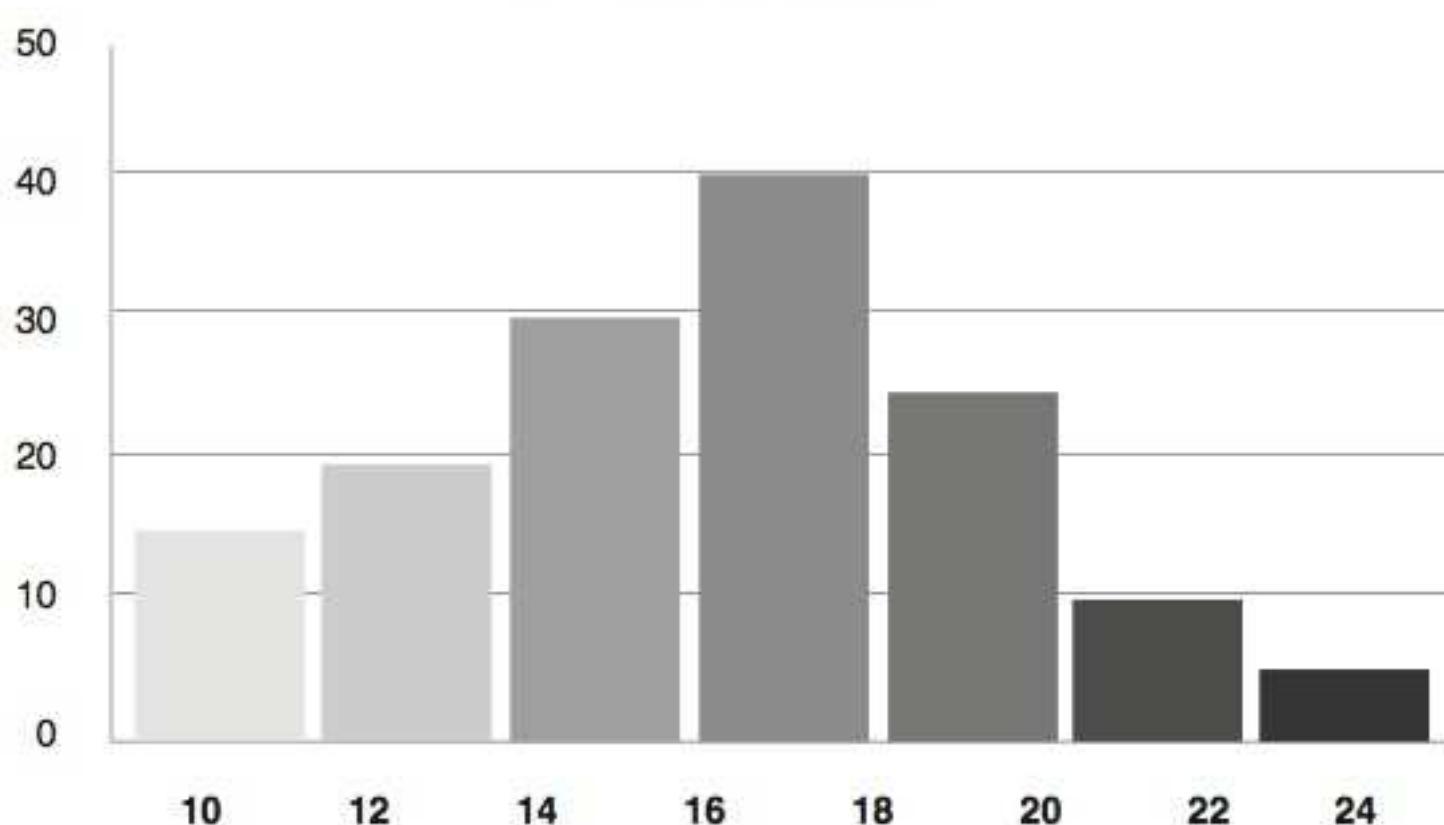
¹² A demonstração desta expressão está no Apêndice 2.B.7.

¹³ Veja o Apêndice 2.B.8 para a demonstração desta expressão.

Exercícios

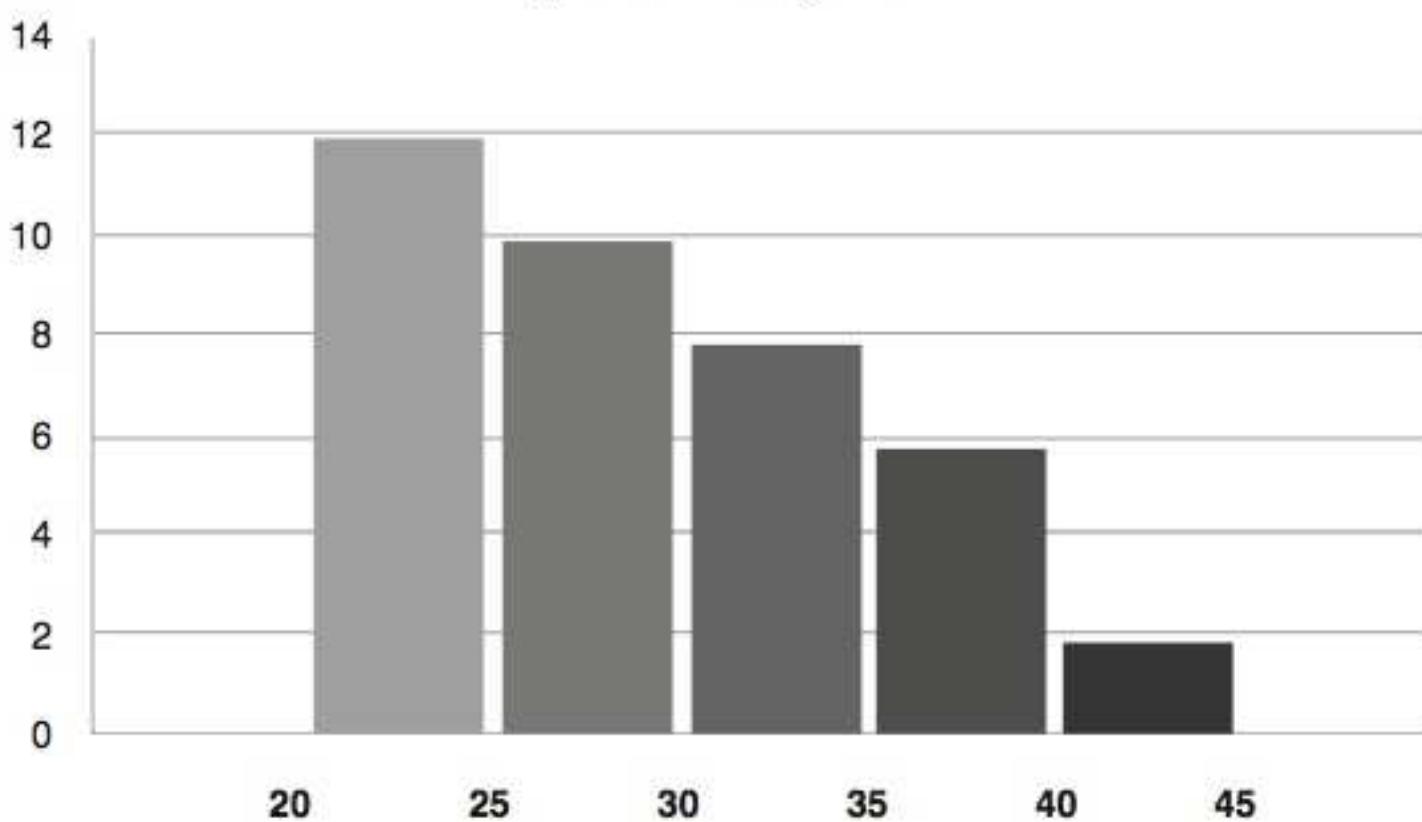
1. Num sistema de avaliação, há duas provas (com notas variando de zero a dez) e, para ser aprovado, o aluno deve ter média final cinco. Qual é a nota mínima que é preciso tirar na primeira prova para ter chance de ser aprovado, supondo:
 - a) média aritmética ponderada, com a primeira prova tendo peso dois e a segunda, um;
 - b) média geométrica (simples);
 - c) média harmônica (simples).
2. Dado o conjunto $\{2; 3; 5; 8; 12\}$, calcule as médias aritmética, geométrica e harmônica, supondo:
 - a) pesos iguais;
 - b) pesos nove, sete, cinco, três e um;
 - c) pesos 10%, 20%, 30%, 25% e 15%.
3. A partir dos dados da Tabela 2.1:
 - a) agrupe os dados em classes de 5 cm;
 - b) calcule a média e a variância;
 - c) comente os resultados obtidos no item anterior;
 - d) trace o histograma correspondente.
4. Com base nos histogramas das Figuras 2.2 e 2.3, calcule a média, a variância e o desvio padrão.
 - a)

Figura 2.2 – Histograma 1



b)

Figura 2.3 – Histograma 2



5. Calcule o coeficiente de correlação entre o consumo e a taxa de juros da Tabela 2.10.
6. Para os dados das Tabelas 2.19 e 2.20, calcule:
 - a) a variância e o desvio padrão de X ;
 - b) a variância e o desvio padrão de Y ;
 - c) a covariância entre X e Y ;
 - d) o coeficiente de correlação entre X e Y .

Tabela 2.19 – X e Y

X	Y
20	12
30	13
40	14
45	13
36	15
27	11

Tabela 2.20 – X e Y

X	Y
114	55
112	61
109	77
123	66
111	81
99	95
121	75
113	77
98	90
103	87

7. Considere duas variáveis aleatórias independentes, X e Y , cujas médias são 10 e 12, respectivamente, e cujas variâncias são 25 e 16. Usando as abreviações a seguir:

$m(X)$ = média aritmética de X ;
 $\text{var}(X)$ = variância de X ;
 $\text{dp}(X)$ = desvio padrão de X ;

Determine:

- a) $m(X + 5);$
- b) $m(5Y);$
- c) $m(3X - 4Y + 7);$
- d) $\text{var}(2X);$
- e) $\text{var}(Y + 6);$
- f) $\text{var}(4X) - \text{var}(2Y + 12);$
- g) $\text{dp}(5X) + \text{dp}(6Y);$
- h) $\text{dp}(3X - 5) - \text{dp}(4Y - 8).$

8. Dadas as variáveis aleatórias X, Y e Z , sendo:

$$\text{var}(X) = 4 \quad \text{cov}(Y, Z) = -3$$

$$\text{var}(Y) = 9 \quad X \text{ e } Y \text{ independentes}$$

$$\text{var}(Z) = 1 \quad X \text{ e } Z \text{ independentes}$$

Calcule:

- a) $\text{var}(X + Y);$
- b) $\text{var}(X - Y);$
- c) $\text{var}(2X + 3Y);$
- d) $\text{var}(Y + Z);$
- e) $\text{var}(2Y - 3Z + 5);$
- f) $\text{var}(4X - 2);$
- g) $\text{corr}(Z, Y);$
- h) $\text{cov}(4Z, 5Y);$
- i) $\text{cov}(2Z, -2Y);$
- j) $\text{corr}(1, 5Z; 2Y).$

9. O coeficiente de correlação entre X e Y é 0,6. Se $W = 3 + 4X$ e $Z = 2 - 2Y$, determine o coeficiente de correlação entre W e Z .

10. O coeficiente de correlação entre X e Y é ρ . Se $W = a + bX$ e $Z = c + dY$, determine o coeficiente de correlação entre W e Z .

11. Uma carteira de investimentos é composta do ativo A, que tem valor médio igual a \$100 e variância igual a 100; um ativo B, que tem valor médio igual a \$90 e variância igual a 64. O coeficiente de correlação entre os dois ativos é -0,5.

- a) Calcule a média e a variância de uma carteira de investimentos com 50% do ativo A e 50% do ativo B.

- b) Calcule a média e a variância de uma carteira de investimentos com 70% do ativo A e 30% do ativo B.
- c) Calcule a média e a variância de uma carteira de investimentos com 20% do ativo A e 80% do ativo B.
- d) Qual a composição da carteira que minimiza o risco do investidor (variância)? Qual o seu retorno (média)?

Apêndice 2.B – Demonstrações

2.B.1 Demonstração da Expressão 2.3.3.1

$$\text{var}(aX) = a^2\text{var}(X)$$

$$\text{var}(aX) = \frac{1}{n} \sum_{i=1}^n (aX_i - a\bar{X})^2$$

$$\text{var}(aX) = \frac{1}{n} \sum_{i=1}^n a(X_i - \bar{X})^2$$

$$\text{var}(aX) = a^2 \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

$$\text{var}(aX) = a^2\text{var}(X) \text{ (c.q.d.)}$$

2.B.2 Demonstração da Expressão 2.3.3.2

$$\text{dp}(aX) = a.\text{dp}(X)$$

$$\text{dp}(aX) = \sqrt{\text{var}(aX)}$$

$$\text{dp}(aX) = \sqrt{a^2 \text{var}(X)}$$

$$\text{dp}(aX) = |a| \sqrt{\text{var}(X)}$$

$$\text{dp}(aX) = |a|.\text{dp}(X) \text{ (c.q.d.)}$$

2.B.3 Demonstração da Expressão 2.3.3.3

$$\text{var}(X + a) = \text{var}(X)$$

$$\text{var}(X + a) = \frac{1}{n} \sum_{i=1}^n [X_i + a - (\bar{X} + a)]^2$$

$$\text{var}(X + a) = \frac{1}{n} \sum_{i=1}^n [X_i + a - \bar{X} - a]^2$$

$$\text{var}(X + a) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

$$\text{var}(X + a) = \text{var}(X) \text{ (c.q.d.)}$$

2.B.4 Demonstração da Expressão 2.3.3.4

$$\text{dp}(X + a) = \text{dp}(X)$$

$$\text{dp}(X + a) = \sqrt{\text{var}(X + a)}$$

$$\text{dp}(X + a) = \sqrt{\text{var}(X)}$$

$$\text{dp}(X + a) = \text{dp}(X) \text{ (c.q.d.)}$$

2.B.5 Demonstração da Expressão 2.3.4.1

$$\text{cov}(X, Y) = \text{média dos produtos} - \text{produto das médias}$$

$$\text{cov}(X, Y) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$$

$$\text{cov}(X, Y) = \frac{1}{n} \sum_{i=1}^n (X_i Y_i - X_i \bar{Y} - \bar{X} Y_i + \bar{X} \bar{Y})$$

$$\text{cov}(X, Y) = \frac{1}{n} \sum_{i=1}^n X_i Y_i - \frac{1}{n} \sum_{i=1}^n X_i \bar{Y} - \frac{1}{n} \sum_{i=1}^n \bar{X} Y_i + \frac{1}{n} \sum_{i=1}^n \bar{X} \bar{Y}$$

$$\text{cov}(X, Y) = \frac{1}{n} \sum_{i=1}^n X_i Y_i - \bar{Y} \frac{1}{n} \sum_{i=1}^n X_i - X \frac{1}{n} \sum_{i=1}^n Y_i + \frac{1}{n} n X \bar{Y}$$

$$\text{cov}(X, Y) = \frac{1}{n} \sum_{i=1}^n X_i Y_i - \bar{X} \bar{Y} - \bar{X} \bar{Y} + \bar{X} \bar{Y}$$

$$\text{cov}(X, Y) = \frac{1}{n} \sum_{i=1}^n X_i Y_i - \bar{X} \bar{Y}$$

$\text{cov}(X, Y)$ = média dos produtos – produto das médias (c.q.d.)

2.B.6 Demonstração da Expressão 2.3.6.1

$$\text{cov}(aX, bY) = a.b.\text{cov}(X, Y)$$

$$\text{cov}(aX, bY) = \frac{1}{n} \sum_{i=1}^n (aX_i - a\bar{X})(bY_i - b\bar{Y})$$

$$\text{cov}(aX, bY) = \frac{1}{n} \sum_{i=1}^n a(X_i - \bar{X})b(Y_i - \bar{Y})$$

$$\text{cov}(aX, bY) = a.b. \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$$

$$\text{cov}(aX, bY) = a.b.\text{cov}(X, Y)$$

2.B.7 Demonstração da Expressão 2.3.6.2

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y)$$

$$\text{var}(X + Y) = \frac{1}{n} \sum_{i=1}^n (X_i + Y_i)^2 - (\bar{X} + \bar{Y})^2$$

$$\text{var}(X + Y) = \frac{1}{n} \sum_{i=1}^n (X_i^2 + Y_i^2 + 2X_i Y_i) - (\bar{X}^2 + \bar{Y}^2 + 2\bar{X}\bar{Y})$$

$$\text{var}(X + Y) = \left(\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 \right) + \left(\frac{1}{n} \sum_{i=1}^n Y_i^2 - \bar{Y}^2 \right) + 2 \left(\frac{1}{n} \sum_{i=1}^n X_i Y_i - \bar{X} \bar{Y} \right)$$

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y) \text{ (c.q.d.)}$$

2.B.8 Demonstração da Expressão 2.3.6.3

$$\text{var}(X - Y) = \text{var}(X) + \text{var}(Y) - 2\text{cov}(X, Y)$$

$$\text{var}(X - Y) = \text{var}[X + (-Y)]$$

$$\text{var}(X - Y) = \text{var}(X) + \text{var}(-Y) + 2\text{cov}(X, -Y)$$

$$\text{var}(X - Y) = \text{var}(X) + \text{var}(Y) - 2\text{cov}(X, Y) \text{ (c.q.d.)}$$



Capítulo 3

DISTRIBUIÇÃO DE PROBABILIDADE

Suponha que você compre uma ação de uma companhia ao preço de R\$ 20,00 e, após um mês, pretenda vendê-la. Suponha ainda que, por algum motivo qualquer, ao final desse período, essa ação só pode: (1) estar valendo os mesmos R\$ 20,00, com probabilidade de 50%; (2) ter caído para R\$ 15,00, com probabilidade de 30%; (3) ou, ainda, ter subido para R\$ 25,00, com probabilidade de 20%. Só esses três valores são possíveis, tendo em vista que as respectivas probabilidades somam exatamente 100%.

Isso corresponde a uma **distribuição de probabilidade** associada ao preço da ação, isto é, cada um dos valores possíveis dessa ação (só três, nesse caso) tem uma probabilidade correspondente. Como definimos no Capítulo 2, isso caracteriza o preço da ação como uma variável aleatória.

Como o conjunto de valores do preço da ação é um conjunto discreto, essa é uma distribuição de probabilidade discreta ou, em outras palavras, é uma distribuição de probabilidade de uma variável aleatória discreta. Poderíamos ter uma distribuição contínua (o que, aliás, provavelmente seria mais adequado, considerando que se trata do preço de uma ação), mas isso será abordado no Capítulo 4. Por enquanto, trataremos de distribuições discretas.

3.1 Esperança matemática

Quando você compra uma ação, *pode* sair ganhando, *pode* perder ou até ficar na mesma, dependendo do que aconteça com o preço da ação. Então, na média, dá na mesma, certo?

Errado! A probabilidade de que a ação caia é maior que a probabilidade de que ela suba. O valor médio do preço da ação é:

$$15 \times 0,3 + 20 \times 0,5 + 25 \times 0,2 = \text{R\$ } 19,50$$

O valor médio é 50 centavos abaixo do preço inicial da ação, o que significa que, em média, quem comprar essa ação sairá perdendo.

Esse é um valor médio *esperado*. É uma média do que *pode* acontecer com a variável, baseada em sua distribuição de probabilidade. É o que chamamos de **esperança matemática** ou, simplesmente, **esperança**.

A esperança de uma variável aleatória discreta X , $E(X)$, pode ser definida, então, como:

$$E(X) = X_1 P(X_1) + X_2 P(X_2) + \dots + X_n P(X_n) = \sum_{i=1}^n X_i P(X_i)$$

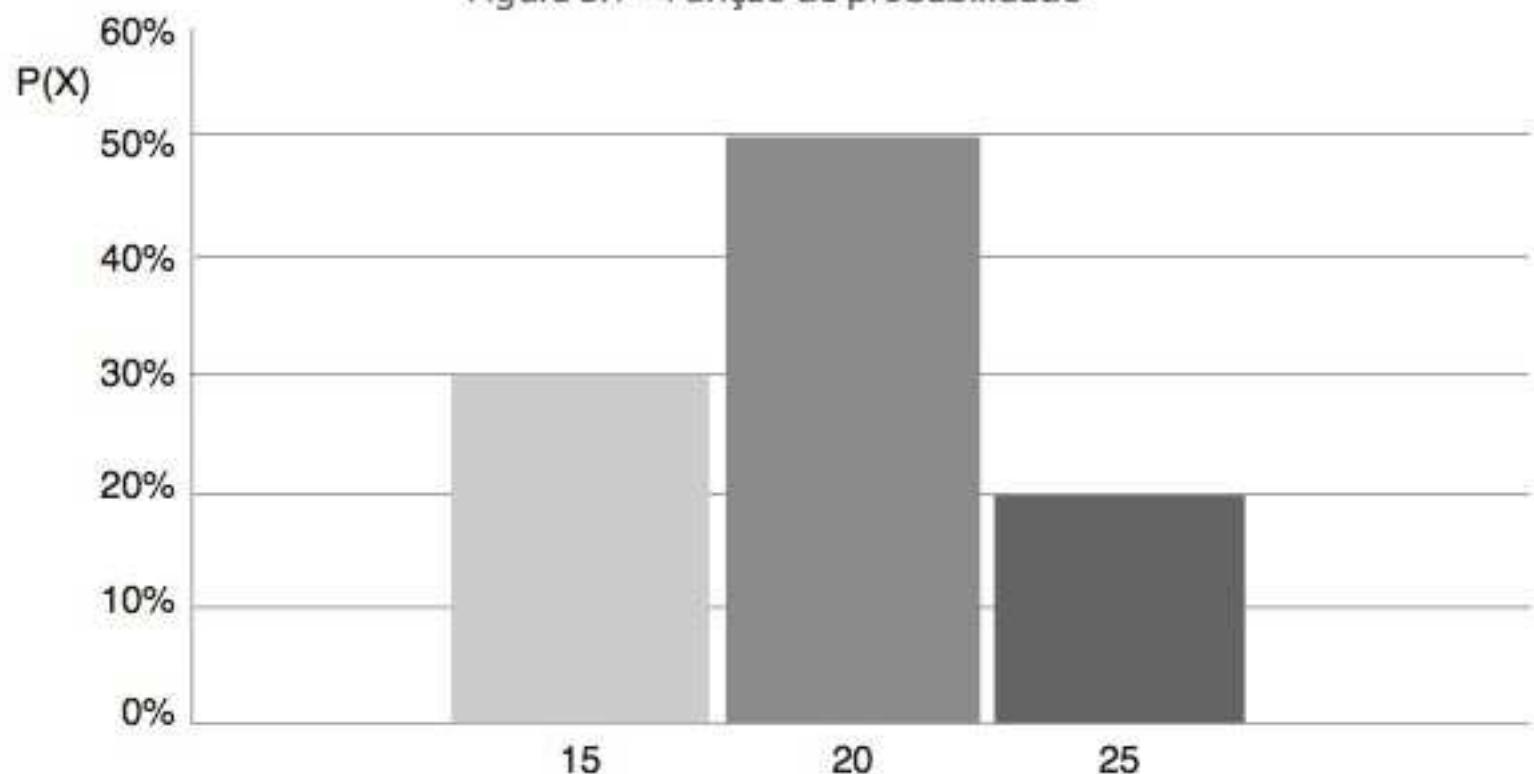
A probabilidade aqui tem o mesmo papel da frequência relativa, apresentada no Capítulo 2. A diferença é que, quando se fala em frequência relativa, usualmente nos referimos a uma quantidade obtida, já a ideia de probabilidade remete a proporções em que a variável pode assumir determinado valor.¹

Aliás, pode-se pensar em $P(X)$ como uma função que associa o valor de X à sua probabilidade, que é chamada de **função de probabilidade**.

Outra função importante que pode ser associada a probabilidades é a função que, dado o valor de X , fornece a probabilidade acumulada. É a **função distribuição acumulada** ou, simplesmente, **função distribuição**, representada por $F(X)$.

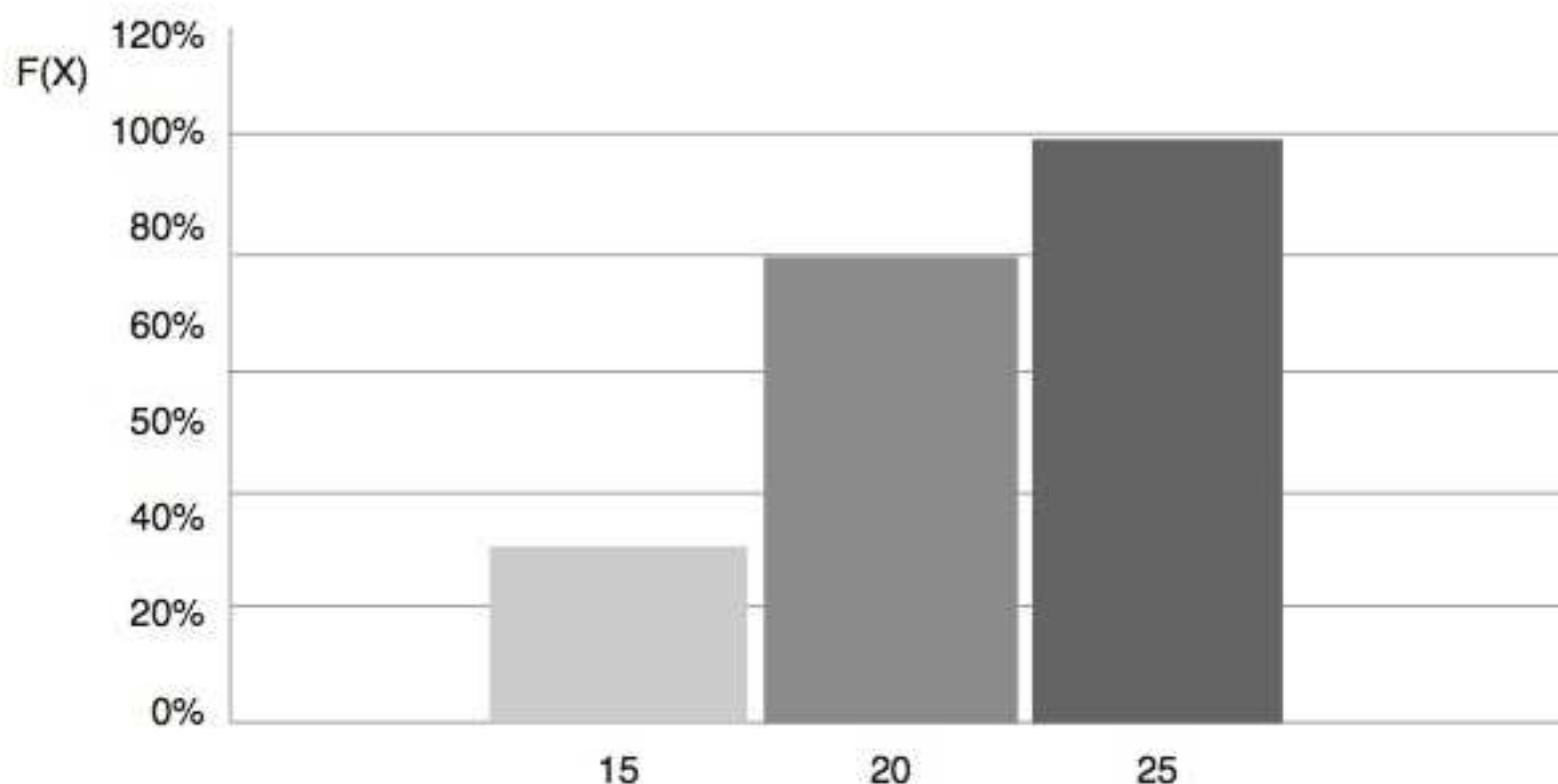
Se X for o preço de uma ação, por exemplo, a do início deste capítulo, então X só pode assumir três valores, isto é, 15, 20 e 25. $F(15)$ seria a probabilidade de o preço da ação ser, no máximo, 15, o que é exatamente 30%. $F(20)$ é a probabilidade de ser até 20, que, nesse caso, equivale à probabilidade de ser 15 ou 20 e é de 80%. Finalmente, $F(25)$ é a probabilidade de o preço ser, no máximo, 25, isto é, de ser 15, 20, ou 25 e é de 100%. Essa é uma característica das funções distribuição, o *último* valor da função (ou o limite para quando X tende ao infinito) é 1 (ou 100%).

Figura 3.1 – Função de probabilidade



¹ A diferença ficará mais clara no Capítulo 7, quando falarmos em valores amostrais e populacionais. Você pode imaginar a frequência relativa como sendo o valor amostral, enquanto a probabilidade seria o valor populacional. Ou ainda, lembrando o Capítulo 1, pela abordagem frequentista, a probabilidade é o limite da frequência relativa quando há um número muito grande de experimentos.

Figura 3.2 – Função distribuição acumulada



Nas Figuras 3.1 e 3.2, o formato de histograma foi utilizado para proporcionar melhor visualização, não sendo obrigatório, embora seja adequado para uma variável aleatória discreta.

Exemplo 3.1.1

Num sorteio de números inteiros de 1 a 5, a probabilidade de um número ser sorteado é proporcional a esse número (isto é, a probabilidade de o número 5 ser sorteado é cinco vezes a probabilidade de o número 1 ser sorteado). Qual a probabilidade de cada número ser sorteado?

Se chamarmos a probabilidade de o número 1 ser sorteado de $P(1)$, de uma constante desconhecida A , temos que:

$$P(2) = 2A$$

$$P(3) = 3A$$

$$P(4) = 4A$$

$$P(5) = 5A$$

Ora, sabemos que a soma de todas as probabilidades, sendo os eventos mutuamente exclusivos, tem de ser igual a 1:

$$P(1) + P(2) + P(3) + P(4) + P(5) = 1$$

$$A + 2A + 3A + 4A + 5A = 1$$

$$15A = 1$$

$$A = \frac{1}{15}$$

Portanto:

$$\begin{aligned}P(1) &= 1/15 \\P(2) &= 2/15 \\P(3) &= 3/15 = 1/5 \\P(4) &= 4/15 \\P(5) &= 5/15 = 1/3\end{aligned}$$

Voltando à esperança, trata-se de uma média ponderada pelas probabilidades. Valem, portanto, para ela, as mesmas propriedades da média:

$$\begin{aligned}E(aX + b) &= aE(X) + b \\E(X + Y) &= E(X) + E(Y)\end{aligned}$$

Você pode, inclusive, escrever a variância em termos da esperança. Como a variância é definida como a média dos quadrados dos desvios em relação à média, vale:

$$\text{var}(X) = E[X - E(X)]^2$$

Ou, ainda, você pode calcular a variância como sendo a média dos quadrados menos o quadrado da média, portanto:

$$\text{var}(X) = E(X^2) - [E(X)]^2$$

Da mesma forma, a covariância entre duas variáveis pode ser escrita utilizando a esperança:

$$\text{cov}(X, Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y)$$

Exemplo 3.1.2

Uma ação comprada por R\$ 10,00 pode assumir, após 30 dias, os seguintes valores: R\$ 5,00, com 20% de probabilidade; R\$ 10,00, com 30% de probabilidade; R\$ 16,00, com 25% de probabilidade; ou R\$ 20,00 com 25% de probabilidade. Determine o valor esperado da ação e sua variância.

O valor esperado (esperança) da ação é dado por:

$$\begin{aligned}E(X) &= 5 \times 0,2 + 10 \times 0,3 + 16 \times 0,25 + 20 \times 0,25 \\E(X) &= 1 + 3 + 4 + 5 = 13.\end{aligned}$$

Como o preço da ação foi de R\$ 10,00, o lucro médio (esperado) dessa ação é de R\$ 3,00.

Quanto à variância:

$$\begin{aligned}E(X^2) &= 5^2 \times 0,2 + 10^2 \times 0,3 + 16^2 \times 0,25 + 20^2 \times 0,25 \\E(X^2) &= 25 \times 0,2 + 100 \times 0,3 + 256 \times 0,25 + 400 \times 0,25 \\E(X^2) &= 5 + 30 + 64 + 100 = 199 \\ \text{var}(X) &= E(X^2) - [E(X)]^2 \\ \text{var}(X) &= 199 - 13^2 \\ \text{var}(X) &= 30.\end{aligned}$$

Repare que a variância, ao medir a dispersão dos possíveis valores da ação, é uma medida do risco da ação.

3.2 Algumas distribuições discretas especiais

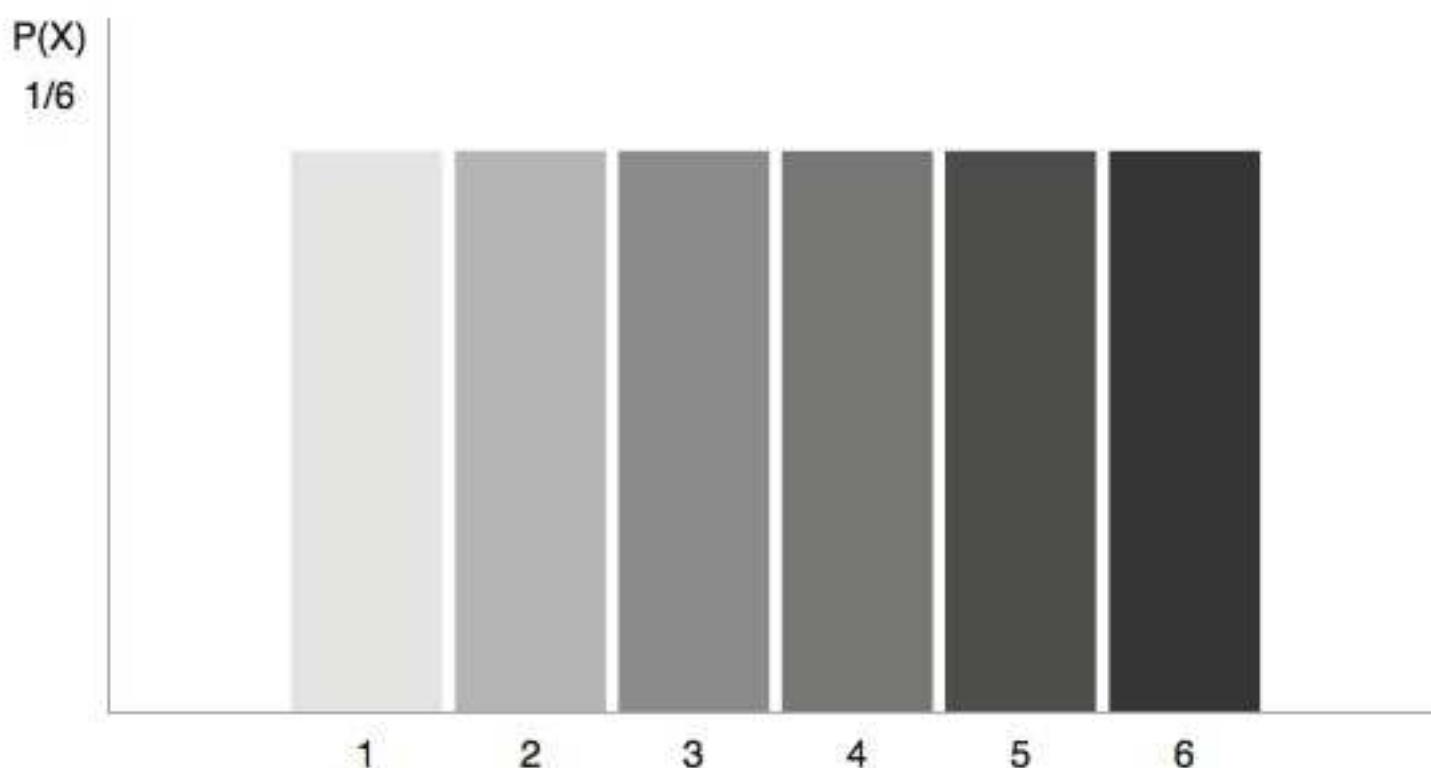
Há distribuições que, por sua importância, até recebem um *nome*. Trataremos de algumas delas agora.

3.2.1 Distribuição uniforme discreta

Na **distribuição uniforme** todos os elementos têm a mesma probabilidade de ocorrer. Imagine, por exemplo, o marcador das horas em um relógio digital. Qual a probabilidade de que, ao olhar para ele num momento qualquer do dia, esteja sendo mostrado um número particular? É $1/12$ para qualquer número, considerando um mostrador de 12 horas, ou $1/24$, se o mostrador é de 24 horas.

Já em um dado não viciado, a probabilidade de ocorrência de um número qualquer é $1/6$. Logo, também nesse caso, se trata de uma distribuição uniforme. O gráfico da função de probabilidade para o lançamento de um dado é mostrado na Figura 3.3 (de novo, em forma de histograma).

Figura 3.3 – Função de probabilidade para um dado não viciado



Exemplo 3.2.1.1

Joga-se um dado uma única vez. Qual o valor esperado do número obtido? E sua variância? O valor esperado (esperança) é dado por:

$$E(X) = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} = \frac{21}{6} = 3,5$$

Repare que, não por coincidência:

$$E(X) = 3,5 = \frac{1+6}{2}$$

Ou seja, no caso de uma distribuição uniforme discreta, a média é a própria média aritmética dos valores extremos (desde que esses valores cresçam num intervalo constante). E a variância é:

$$E(X^2) = 1^2 \times \frac{1}{6} + 2^2 \times \frac{1}{6} + 3^2 \times \frac{1}{6} + 4^2 \times \frac{1}{6} + 5^2 \times \frac{1}{6} + 6^2 \times \frac{1}{6}$$

$$E(X^2) = 1 \times \frac{1}{6} + 4 \times \frac{1}{6} + 9 \times \frac{1}{6} + 16 \times \frac{1}{6} + 25 \times \frac{1}{6} + 36 \times \frac{1}{6} = \frac{91}{6}$$

$$\text{var}(X) = E(X^2) - [E(X)]^2$$

$$\text{var}(X) = \frac{91}{6} - \left(\frac{21}{6}\right)^2 = \frac{105}{36} \cong 2,92$$

3.2.2 Distribuição de Bernoulli

A **distribuição de Bernoulli** se caracteriza pela existência de apenas dois eventos, mutuamente exclusivos, denominados *sucesso* e *fracasso*, num experimento que é realizado uma única vez. Se a probabilidade de sucesso é p , a probabilidade de fracasso é $1 - p$, uma vez que só existem esses dois eventos e eles são mutuamente exclusivos.

Uma distribuição desse tipo é o lançamento de uma moeda uma única vez. Se você aposta na cara, sendo esta, então, o sucesso, a probabilidade de sucesso é $p = 1/2$, e a probabilidade de fracasso (coroa) é $1 - p = 1/2$.

Da mesma forma, num lançamento de um dado, se você aposta num número, por exemplo, o 3, este será o sucesso, sendo qualquer um dos outros cinco números fracasso. Nesse caso, a probabilidade de sucesso é $p = 1/6$, e a probabilidade de fracasso é $1 - p = 5/6$.

Há outros exemplos: digamos que a intenção de voto para um candidato é 30%. Se, ao escolher um eleitor ao acaso, for definido como sucesso esse eleitor pretender votar no referido candidato, a probabilidade de sucesso é $p = 0,3$, e a probabilidade de fracasso é $1 - p = 0,7$. Da mesma forma, se há 5% de peças defeituosas em um lote, definindo como sucesso escolher, ao acaso, uma peça que não seja defeituosa, a probabilidade é $p = 0,95$, enquanto a probabilidade de fracasso será $1 - p = 0,05$.

Exemplo 3.2.2.1

No caso da cara ou coroa, atribuindo o valor 1 para o sucesso e 0 para o fracasso, determine a média e a variância do resultado após uma jogada.

A média é dada por:

$$E(X) = 1 \times \frac{1}{2} + 0 \times \frac{1}{2} = \frac{1}{2} = 0,5$$

E a variância:

$$E(X^2) = 1^2 \times \frac{1}{2} + 0^2 \times \frac{1}{2} = \frac{1}{2} = 0,5$$

$$\text{var}(X) = E(X^2) - [E(X)]^2 = 0,5 - 0,5^2 = 0,25$$

Exemplo 3.2.2.2

No caso do dado, em que se aposta em um único número, atribuindo o valor 1 para o sucesso e 0 para o fracasso, determine a média e a variância do resultado após uma jogada. A média é dada por:

$$E(X) = 1 \times \frac{1}{6} + 0 \times \frac{5}{6} = \frac{1}{6}$$

E a variância por:

$$E(X^2) = 1^2 \times \frac{1}{6} + 0^2 \times \frac{5}{6} = \frac{1}{6}$$

$$\text{var}(X) = E(X^2) - [E(X)]^2$$

Pelos dois exemplos anteriores, você pode verificar que numa distribuição de Bernoulli:²

$$E(X) = p$$

$$\text{var}(X) = p(1-p)$$

Assim, você pode utilizar o resultado para o caso do candidato que tem 30% das intenções de voto. Verifique que:

$$E(X) = p = 0,3$$

$$\text{var}(X) = p(1-p) = 0,3 \times 0,7 = 0,21$$

E o mesmo vale para o caso das peças defeituosas ou para qualquer situação que se enquadre em uma distribuição de Bernoulli.

Especificamente no caso do candidato, é possível, como você verá no Capítulo 7, montar as chamadas *margens de erro* das pesquisas eleitorais por meio da variância.

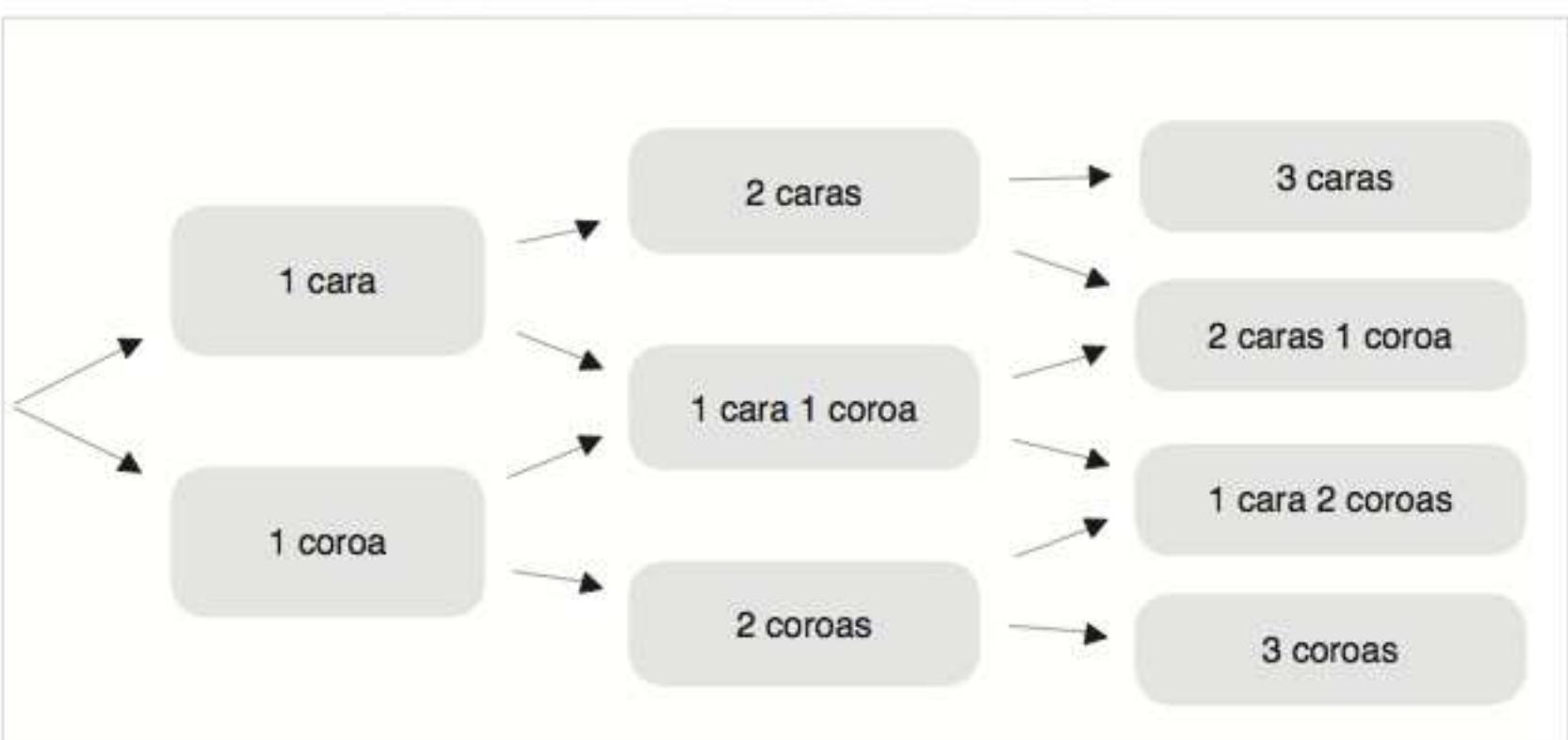
3.2.3 Distribuição binomial

A **distribuição binomial** nada mais é do que a generalização da distribuição de Bernoulli. Há um sucesso, com probabilidade p , e um fracasso, com probabilidade $1 - p$, mas o número de experimentos (de *jogadas*) pode ser qualquer um.

Tome um exemplo simples, o da cara ou coroa, com três jogadas, que está representado na árvore da Figura 3.4.

² A demonstração é dada no Apêndice 3.B.

Figura 3.4 – Árvore para três jogadas de cara ou coroa



Você já conhece o resultado da primeira jogada:

$$P(1 \text{ cara}) = p = \frac{1}{2}$$

$$P(1 \text{ coroa}) = 1 - p = \frac{1}{2}$$

Para a segunda jogada, observando a árvore, verifique que, da origem, há quatro caminhos possíveis e, nesse caso, todos com a mesma probabilidade. Desses quatro, em um deles chega-se a duas caras ou duas coroas. Entretanto, para uma cara e uma coroa há dois caminhos possíveis. Portanto, para duas jogadas:

$$P(2 \text{ caras}) = \frac{1}{4}$$

$$P(1 \text{ cara e } 1 \text{ coroa}) = \frac{2}{4}$$

$$P(2 \text{ coroas}) = \frac{1}{4}$$

Repare que:

$$P(2 \text{ caras}) = p \times p$$

$$P(1 \text{ cara e } 1 \text{ coroa}) = 2 \times p \times (1 - p)$$

$$P(2 \text{ coroas}) = (1 - p) \times (1 - p)$$

O número 2 que aparece para uma cara e uma coroa se deve ao fato de que esse resultado é possível de ocorrer de duas maneiras, isto é, dando cara na primeira jogada ou dando coroa logo na primeira.

Para três jogadas, verifique que há oito caminhos possíveis. Desses oito, em apenas um ocorrem só caras ou só coroas. Em três deles, ocorrem duas caras e uma coroa e, em outros três, duas coroas e uma cara.

$$P(3 \text{ caras}) = \frac{1}{8}$$

$$P(2 \text{ caras e } 1 \text{ coroa}) = \frac{3}{8}$$

$$P(1 \text{ cara e } 2 \text{ coroas}) = \frac{3}{8}$$

$$P(3 \text{ coroas}) = \frac{1}{8}$$

Então:

$$P(3 \text{ caras}) = p \times p \times p$$

$$P(2 \text{ caras e } 1 \text{ coroa}) = 3 \times p \times p \times (1-p)$$

$$P(1 \text{ cara e } 2 \text{ coroas}) = 3 \times p \times (1-p) \times (1-p)$$

$$P(3 \text{ coroas}) = (1-p) \times (1-p) \times (1-p)$$

E agora aparece o número 3 para duas caras e uma coroa (ou uma cara e duas coroas). De onde? Há realmente três possibilidades: primeira cara, segunda cara e terceira coroa; ou, primeira cara, segunda coroa e terceira cara; ou ainda, primeira coroa, segunda cara, terceira cara. As posições das duas caras podem ser combinadas de três maneiras diferentes. O número 3, na verdade, é a quantidade de combinações de três elementos em grupos de dois (veja o Apêndice 1.A).

Portanto:

$$P(3 \text{ caras}) = C_{3,3} \times p \times p \times p$$

$$P(2 \text{ caras e } 1 \text{ coroa}) = C_{3,2} \times p \times p \times (1-p)$$

$$P(1 \text{ cara e } 2 \text{ coroas}) = C_{3,1} \times p \times (1-p) \times (1-p)$$

$$P(3 \text{ coroas}) = C_{3,0} \times (1-p) \times (1-p) \times (1-p)$$

Nota: as combinações de n elementos em grupos de k também podem ser escritas como:

$$C_{n,k} = \binom{n}{k}$$

Essa expressão é lida como binomial de n,k . Portanto, as probabilidades para três jogadas podem ser escritas assim:

$$P(3 \text{ caras}) = \binom{3}{3} \times p \times p \times p$$

$$P(2 \text{ caras e } 1 \text{ coroa}) = \binom{3}{2} \times p \times p \times (1-p)$$

$$P(1 \text{ cara e } 2 \text{ coroas}) = \binom{3}{1} \times p \times (1-p) \times (1-p)$$

$$P(3 \text{ coroas}) = \binom{3}{0} \times (1-p) \times (1-p) \times (1-p)$$

É possível generalizar: para um experimento qualquer, em que a probabilidade de sucesso é p e a probabilidade de fracasso é $1 - p$, a probabilidade de que, em n jogadas, ocorram k sucessos é:

$$P(x=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Exemplo 3.2.3.1

Suponha um jogo de dados em que se aposta em um único número. Determine as seguintes probabilidades.

a) Em três jogadas, ganhar duas.

É uma distribuição binomial em que $p = 1/6$: três jogadas e o sucesso ocorre em duas delas:

$$P(x=2) = \binom{3}{2} \times \left(\frac{1}{6}\right)^2 \times \left(\frac{5}{6}\right)^1$$

$$P(x=2) = 3 \times \frac{1}{36} \times \frac{5}{6}$$

$$P(x=2) = \frac{15}{216}$$

b) Em quatro jogadas, ganhar duas.

$$P(x=2) = \binom{4}{2} \times \left(\frac{1}{6}\right)^2 \times \left(\frac{5}{6}\right)^2$$

$$P(x=2) = 6 \times \frac{1}{36} \times \frac{25}{36}$$

$$P(x=2) = \frac{150}{1.296}$$

c) Em cinco jogadas, ganhar três.

$$P(x=3) = \binom{5}{3} \times \left(\frac{1}{6}\right)^3 \times \left(\frac{5}{6}\right)^2$$

$$P(x=3) = 10 \times \frac{1}{216} \times \frac{25}{36}$$

$$P(x=3) = \frac{250}{7.776}$$

Exemplo 3.2.3.2

Calcule a média e a variância no jogo de cara ou coroa, atribuindo valor 1 para cara e 0 para coroa, considerando uma, duas e três jogadas.

Para uma jogada, fica-se reduzido ao caso particular da distribuição de Bernoulli, cujo resultado você já conhece:

$$E(x) = p = \frac{1}{2}$$

$$\text{var}(x) = p(1-p) = \frac{1}{4}$$

Faça, então, o cálculo para duas e três jogadas. Para duas jogadas:

$$E(x) = 2 \times \frac{1}{4} + 1 \times \frac{2}{4} + 0 \times \frac{1}{4} = \frac{4}{4} = 1$$

$$E(x^2) = 2^2 \times \frac{1}{4} + 1^2 \times \frac{2}{4} + 0^2 \times \frac{1}{4} = \frac{6}{4} = 1,5$$

$$\text{var}(x) = 1,5 - 1^2 = 0,5$$

Para três jogadas:

$$E(x) = 3 \times \frac{1}{8} + 2 \times \frac{3}{8} + 1 \times \frac{3}{8} + 0 \times \frac{1}{8} = \frac{12}{8} = 1,5$$

$$E(x^2) = 3^2 \times \frac{1}{8} + 2^2 \times \frac{3}{8} + 1^2 \times \frac{3}{8} + 0^2 \times \frac{1}{8} = \frac{24}{8} = 3$$

$$\text{var}(x) = 3 - 1,5^2 = 0,75$$

Note que é válido que (veja o Apêndice 3.B):

$$E(x) = np$$

$$\text{var}(x) = np(1-p)$$

3.2.4 Distribuição geométrica

A **distribuição geométrica** também se refere a sucessos e fracassos, mas, diferentemente da binomial, é a probabilidade de que o sucesso ocorra (exatamente) na k -ésima jogada. Por exemplo, na cara ou coroa, qual a probabilidade de que a cara só ocorra na terceira jogada? Ou qual a probabilidade de que o dado só dê o número desejado na quarta jogada?

Assim, a forma geral da distribuição geométrica é dada por:

$$P(x = k) = (1 - p)^{k-1}p$$

Ou seja, uma sequência de fracassos nas $k - 1$ primeiras jogadas, culminando com sucesso apenas na k -ésima jogada.

Exemplo 3.2.4.1

Um time de basquete não está muito bem nesta temporada, de tal forma que a probabilidade de que ganhe um jogo qualquer é de 20%.

- a) Qual a probabilidade de que a primeira vitória ocorra na primeira partida?

Nesse caso, o sucesso é imediato:

$$P(x = 1) = 0,2 = 20\%$$

- b) E na segunda partida?

$$P(x = 2) = 0,8 \times 0,2 = 0,16 = 16\%$$

- c) E na quinta partida?

$$P(x = 5) = 0,8^4 \times 0,2 = 0,08192 \cong 8,2\%$$

Exemplo 3.2.4.2

E qual é a partida esperada em que ocorrerá a primeira vitória?

O valor esperado da k -ésima partida em que ocorrerá a tão sonhada vitória é:

$$E(x) = 1 \times 0,2 + 2 \times 0,8 \times 0,2 + 3 \times 0,8^2 \times 0,2 + 4 \times 0,8^3 \times 0,2 + \dots$$

$$E(x) = 0,2 \times [1 + 2 \times 0,8 + 3 \times 0,8^2 + 4 \times 0,8^3 + \dots]$$

A expressão entre colchetes é quase uma progressão geométrica, exceto pelos números 1, 2, 3, 4 etc. Na verdade, é uma soma de progressões geométricas:

$$1 + 0,8 + 0,8^2 + 0,8^3 + \dots$$

$$0,8 + 0,8^2 + 0,8^3 + \dots$$

$$0,8^2 + 0,8^3 + \dots$$

$$0,8^3 + \dots$$

$$1 + 2 \times 0,8 + 3 \times 0,8^2 + 4 \times 0,8^3 + \dots$$

Relembrando que a soma de uma progressão geométrica infinita cujo primeiro termo é a , e cuja razão (q) é menor do que 1, em módulo, é dada por:³

$$S = \frac{a}{1 - q}$$

³ O que é mostrado no Apêndice 3.A.

Então:

$$E(x) = 0,2 \times \left(\frac{1}{1-0,8} + \frac{0,8^2}{1-0,8} + \frac{0,8^3}{1-0,8} + \dots \right)$$

$$E(x) = \frac{0,2}{1-0,8} \times (1 + 0,8 + 0,8^2 + 0,8^3 + \dots)$$

O termo entre parênteses é também uma progressão geométrica, enquanto o termo multiplicando é exatamente 1:

$$E(x) = \frac{1}{1-0,8} = \frac{1}{0,2} = 5$$

Portanto, o esperado é que a vitória ocorra na quinta partida.

Repare que o resultado obtido pode ser generalizado para $E(x) = \frac{1}{p}$, que é a média de uma distribuição geométrica:

$$\text{Média de uma distribuição geométrica} = E(x) = \frac{1}{p}$$

3.2.5 Distribuição hipergeométrica

A **distribuição hipergeométrica** refere-se à probabilidade de, ao retirarmos, *sem reposição*, n elementos de um conjunto de N , saiam k elementos com o atributo sucesso, considerando-se que, do total de N elementos, s possuem esse atributo e, portanto, $N - s$ possuem o atributo fracasso. Fica claro que, da maneira como foi definido p :

$$p = \frac{s}{N}$$

A pergunta aqui, então, é: qual a probabilidade de que, retirando-se n elementos, k possua o atributo sucesso e $n - k$, o atributo fracasso?

Do total de N elementos, é possível tirar $\binom{N}{n}$ grupos de n elementos. Dos s que possuem o atributo sucesso, há $\binom{s}{k}$ grupos de k elementos que poderiam sair nessa extração. Finalmente, dos $N - s$ elementos que possuem o atributo fracasso, há $\binom{N-s}{n-k}$ grupos de $n - k$ elementos. Então, a probabilidade de você encontrar k elementos com o atributo sucesso é:

$$P(x=k) = \frac{\binom{s}{k} \binom{N-s}{n-k}}{\binom{N}{n}}$$

Exemplo 3.2.5.1

Sabe-se que há 10% de peças defeituosas em um lote de 50. Ao retirar oito peças desse lote, *sem reposição*, qual a probabilidade de que duas delas sejam defeituosas?

Como são 10% de peças defeituosas em um total de 50, há cinco peças defeituosas. Pede-se a probabilidade de retirar duas (do total de cinco) peças defeituosas e seis (de um total de 45) peças em bom estado.

Essa probabilidade é calculada como se segue:

$$P(x=2) = \frac{\binom{5}{2} \binom{45}{6}}{\binom{50}{8}} \cong 0,1517 = 15,17\%$$

3.2.6 Distribuição de Poisson

Você é capaz de dizer quantas vezes, em média, toca o telefone por dia em sua casa ou em seu escritório? Provavelmente, sim. Mas quantas vezes *não* toca o telefone? Essa pergunta é muito difícil de responder. Quando uma variável aleatória tem um comportamento parecido com esse, dizemos que ela segue uma **distribuição de Poisson**.

Se você considerar que sucesso é tocar o telefone, é muito difícil calcular o p , a probabilidade de isso ocorrer, já que não há como calcular a não ocorrência do evento.

A solução é imaginar que o p é muito pequeno, já que o toque do telefone dura apenas alguns segundos em um dia de 24 horas. Portanto, o número de vezes em que esse experimento é realizado (telefone toca ou não toca), que é o n da distribuição binomial, é realizado muitas vezes.

Assim é possível modelar esse tipo de distribuição: partindo de uma distribuição binomial, considere que p é muito pequeno (tende a zero) e n é muito grande (tende a infinito).

$$p \rightarrow 0$$

$$n \rightarrow \infty$$

Mas de tal modo que o produto np é um número finito diferente de zero.

$$np = \lambda$$

E o que significa esse novo parâmetro λ ? Como partimos de uma distribuição binomial:

$$E(x) = np = 1$$

Portanto, λ é exatamente o número médio de vezes em que o evento ocorre. No exemplo do telefone, é o número de vezes em que o telefone toca por dia.

Ainda é possível calcular a variância partindo de uma distribuição binomial:

$$\text{var}(x) = np(1-p)$$

Mas, como p tende a zero, $1-p$ tende a 1. Portanto:

$$\text{var}(x) = np = \lambda$$

A distribuição de Poisson se caracteriza, dessa forma, por ter média igual à variância. Para calcular a probabilidade de uma variável como essa, parte-se da distribuição binomial e fazemos $p \rightarrow 0$ e $n \rightarrow \infty$. Fazendo isso (veja demonstração no Apêndice 3.B), chega-se a:

$$P(x=k) = \frac{e^{-\lambda} \lambda^k}{k!}$$

Exemplo 3.2.6.1

Suponha que, em média, o telefone toque quatro vezes ao dia em uma casa. Qual a probabilidade de que, em certo dia, ele toque, no máximo, duas vezes?

É uma distribuição de Poisson cujo parâmetro é $\lambda = 4$. A probabilidade de tocar no máximo duas vezes é equivalente à probabilidade de tocar zero, uma ou duas vezes.

$$P(x=0) = \frac{e^{-4} 4^0}{0!} = e^{-4}$$

$$P(x=1) = \frac{e^{-4} 4^1}{1!} = 4e^{-4}$$

$$P(x=2) = \frac{e^{-4} 4^2}{2!} = 8e^{-4}$$

Portanto:

$$P(x \leq 2) = 13e^{-4} \cong 0,2381 = 23,81\%$$

A distribuição de Poisson também pode ser útil como uma aproximação da binomial quando, embora não seja impossível, o valor de p seja tão pequeno que os cálculos se tornem um tanto quanto trabalhosos, como no exemplo a seguir.

Exemplo 3.2.6.2

Um candidato tem apenas 2% das intenções de voto. Qual a probabilidade de que, em 100 eleitores escolhidos ao acaso, encontrem-se cinco que desejem votar nesse candidato?

Usando a binomial pura e simplesmente:

$$P(x=5) = \binom{100}{5} 0,02^5 \times 0,98^{95} \cong 0,0353 = 3,53\%$$

Você pode, entretanto, usar a distribuição de Poisson como aproximação, tendo como parâmetro $\lambda = np = 100 \times 0,02 = 2$.

$$P(x=5) = \frac{e^{-2} 2^5}{5!} \cong 0,0361 = 3,61\%$$

Note que 3,61% é um valor bem próximo do encontrado por meio da binomial.

Exercícios

1. Calcule a média, a variância e o desvio padrão das seguintes variáveis aleatórias discretas:
 - a) valor de uma ação:
 - i) \$ 50 com probabilidade de 35%;
 - ii) \$ 40 com probabilidade de 30%;
 - iii) \$ 30 com probabilidade de 20%;
 - iv) \$ 20 com probabilidade de 15%;
 - b) pontos de um time ao final do campeonato:
 - i) 40 com probabilidade de 5%;
 - ii) 36 com probabilidade de 10%;
 - iii) 32 com probabilidade de 25%;
 - iv) 28 com probabilidade de 25%;
 - v) 24 com probabilidade de 20%;
 - vi) 20 com probabilidade de 15%;
 - c) o valor em uma jogada de um dado não viciado;
 - d) o valor em uma jogada de um dado viciado em que a probabilidade é inversamente proporcional a cada número (isto é, a probabilidade de dar 1 é seis vezes maior que a de dar 6);
 - e) ganhos em um jogo de cara ou coroa (com uma moeda não viciada), no qual, após quatro jogadas:
 - i) para quatro vitórias, o prêmio é de \$ 60;
 - ii) para três vitórias seguidas, o prêmio é de \$ 30;
 - iii) para três vitórias alternadas, o prêmio é de \$ 20;
 - iv) para duas vitórias seguidas, o prêmio é de \$ 10;
 - v) para duas vitórias alternadas, o prêmio é de \$ 0;
 - vi) para uma vitória, a penalidade é de \$ 20;
 - vii) para nenhuma vitória, a penalidade é de \$ 50;

- f) ganhos em um jogo de dados tetraédricos (apostando em um único número), no qual, após três jogadas:
- para três vitórias, o prêmio é de \$ 20;
 - para duas vitórias seguidas, o prêmio é de \$ 10;
 - para duas vitórias alternadas, o prêmio é de \$ 0;
 - para uma vitória, a penalidade é de \$ 10;
 - para nenhuma vitória, a penalidade é de \$ 20;
- g) $Z = 1, 2, 3, 4$

$$P(Z = k) = \frac{0,48}{k}.$$

- Dada uma **variável aleatória** (v.a.) X , tal que X é um número inteiro positivo cuja probabilidade é $P(X = k) = A(0,8)^k$, determine o valor de A .
- A probabilidade de que um aluno atrasse a mensalidade é de 10%. Qual a probabilidade de que, num grupo de cinco alunos, no máximo dois atrasem a mensalidade?
- Um candidato tem 20% das intenções de voto. Qual a probabilidade de que, entre eleitores escolhidos ao acaso, sete tenham a intenção de votar nesse candidato?
- Num grupo de 20 pessoas, 12 são casadas. Qual a probabilidade de que, num subgrupo de pessoas escolhidas ao acaso, duas sejam solteiras?
- Uma pessoa está interessada em vender um imóvel e foi informada de que a probabilidade de encontrar um comprador disposto a pagar o preço pedido em qualquer dia é de 30%. Qual a probabilidade de que ela consiga vender o imóvel em até três dias?
- Numa grande cidade brasileira, ocorrem, em média, cinco enchentes por ano. Qual a probabilidade de que em determinado ano ocorram no máximo três enchentes?
- Uma aluna, quando assiste a aulas em salas com ar-condicionado, espirra, em média, três vezes por hora.
 - Qual a probabilidade de que, em três horas, ela espirre dez vezes?
 - Qual a probabilidade de que, em duas horas, ela espirre no mínimo uma vez?
 - Qual a probabilidade de que, em uma hora, ela espirre no máximo três vezes?
- Calcule a probabilidade pedida usando a binomial e a respectiva aproximação pela Poisson.
 - Em um lote de 1.000 peças, 1% é defeituosa; qual a probabilidade de que um lote de 20 peças não apresente nenhuma defeituosa?
 - Um candidato tem 30% das intenções de voto; qual a probabilidade de que, entre-vistados 100 eleitores, 35 afirmem que vão votar nesse candidato?

10. Um jogador de futebol converte em gol 25% das suas cobranças de falta.

- a) Quantas faltas, em média, ele deve cobrar para fazer um gol?
- b) Qual a probabilidade de ele marcar um gol após duas cobranças?

11. Assinale verdadeiro ou falso:

- a) Se retirarmos um grupo de n elementos de um total de N elementos, **sem reposição**, a distribuição de probabilidade é a binomial.
- b) Se retirarmos um grupo de n elementos de um total de N elementos, **sem reposição**, a distribuição de probabilidade é aproximadamente a binomial, desde que n seja muito pequeno em relação a N .
- c) A distribuição de Bernoulli é um caso particular da distribuição binomial.
- d) Na distribuição de Poisson, a média é igual à variância.
- e) A distribuição geométrica é usada quando não é possível contar as “não ocorrências” do evento.
- f) A distribuição binomial se caracteriza por n repetições de um experimento, sendo que as repetições são independentes e cada experimento apresenta apenas dois resultados possíveis, que são mutuamente exclusivos.

12. Suponha que o número de vezes durante um ano em que um indivíduo pega uma doença seja uma variável aleatória com distribuição de Poisson com média 3. Uma nova droga reduz essa média para 1, para 70% da população, mas não tem nenhum efeito para os 30% restantes.

- a) Qual a probabilidade de que um indivíduo que se beneficia da nova droga pegue a doença duas vezes em determinado ano?
- b) Qual a probabilidade de que um indivíduo que não se beneficia da nova droga pegue a doença duas vezes em determinado ano?
- c) Qual a probabilidade de que um indivíduo escolhido aleatoriamente pegue a doença duas vezes em determinado ano?
- d) Um indivíduo escolhido aleatoriamente pegou duas vezes a doença em determinado ano. Qual a probabilidade de ele fazer parte da parcela da população que se beneficia da droga?

13. Em uma classe de 30 estudantes, sendo 18 mulheres e 12 homens, 4 serão escolhidos aleatoriamente para fazer parte de uma comissão.

- a) Qual a probabilidade de que todos os escolhidos sejam homens?
- b) Qual a probabilidade de que sejam escolhidos no mínimo 2 homens?
- c) Qual a probabilidade de que sejam escolhidos no máximo 3 homens?

14. A quantidade demandada de ingressos para uma peça de teatro é uma distribuição de Poisson com parâmetro igual a $100 - 2d$, em que d é o número de dias em que a peça está em cartaz.

- Quantas pessoas são esperadas para o 20º dia, em média? Qual é o desvio padrão?
- Qual a probabilidade de que venham mais do que 80 pessoas no 18º dia?
- Qual a probabilidade de que venha pelo menos uma pessoa no 49º dia?

Apêndice 3.A – Progressão geométrica

3.A.1 Progressão geométrica

Progressão geométrica (ou, simplesmente, PG) é uma sequência de números em que, dado um número da série, o número seguinte será encontrado multiplicando-se o número dado por um valor fixo.

Por exemplo, a sequência de números abaixo:

$$\{2, 6, 18, 54, 162\}$$

É uma PG, pois, partindo do 2, multiplicando-o por 3, temos $2 \times 3 = 6$, que é o número seguinte; para acharmos o próximo, fazemos $6 \times 3 = 18$, e assim sucessivamente para encontrarmos os seguintes.

Essa é uma PG de cinco termos; o número 3, que é o que se multiplica para encontrar o próximo número da sequência, é chamado de **razão** da PG.

Nosso principal interesse é a soma dos termos de uma PG. No caso específico, porém, ela pode ser facilmente encontrada, pois são poucos termos:

$$S = 2 + 6 + 18 + 54 + 162 \quad (3.A.1)$$

$$S = 242$$

Há de se encontrar, no entanto, uma fórmula geral que possa ser aplicada a qualquer PG, não importa seu tamanho. Para isso, multipliquemos a Expressão 3.A.1 por 3, que é a razão da PG.

$$3S = 6 + 18 + 54 + 162 + 486 \quad (3.A.2)$$

Note que todos os termos se repetiram, exceto o primeiro. Subtraiamos a Expressão 3.A.1 da Expressão 3.A.2:

$$3S = 6 + 18 + 54 + 162 + 486$$

$$- (S = 2 + 6 + 18 + 54 + 162)$$

$$2S = 486 - 2$$

$$2S = 484$$

$$S = \frac{484}{2} = 242$$

Dessa forma, podemos repetir o procedimento para uma PG qualquer de n termos, com o primeiro termo denominado a , e a razão, q . A soma dessa PG é dada por:

$$S = a + aq + aq^2 + aq^3 + \dots + aq^{n-1} \quad (3.A.3)$$

Multiplicando a Expressão 3.A.3 por q , obtém-se:

$$qS = aq + aq^2 + aq^3 + \dots + aq^{n-1} + aq^n \quad (3.A.4)$$

Subtraindo 3.A.3 de 3.A.4, temos:

$$qS = aq + aq^2 + aq^3 + \dots + aq^{n-1} + aq^n$$

$$- (S = a + aq + aq^2 + aq^3 + \dots + aq^{n-1})$$

$$qS - S = aq^n - a$$

$$S(q - 1) = a(q^n - 1)$$

$$S = \frac{a(q^n - 1)}{q - 1}$$

Assim, conseguimos encontrar um termo geral para calcular a soma de uma PG. Para isso, devemos identificar o primeiro termo da série (o a da fórmula), a razão (q) e o número de termos (n).

E se a PG for infinita? É possível que a soma seja finita? A resposta é sim. Tomemos, por exemplo, uma pessoa que come um chocolate seguindo uma regra: em cada mordida, ela come exatamente metade do que falta. Quantos chocolates ela comerá ao final de infinitas mordidas? Um chocolate. Mas isso só acontece porque em cada mordida ela come sempre uma fração do que falta. Isto é, é necessário que a razão seja (em módulo) menor do que 1.

A soma que representa as mordidas do chocolate é dada por:

$$S = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \dots = 1$$

Que é uma PG com infinitos termos, cujo primeiro é $1/2$ e cuja razão também é $1/2$ e que, sabemos, é igual a 1 (porque há um, e somente um, chocolate).

Nesse caso, temos uma PG infinita, portanto:

$$S = a + aq + aq^2 + aq^3 + \dots \quad (3.A.5)$$

E, se multiplicarmos essa soma por q e subtrairmos, teremos:

$$S = a + aq + aq^2 + aq^3 + \dots$$

$$- (qS = aq + aq^2 + aq^3 + \dots)$$

$$S - qS = a$$

$$(1 - q)S = a$$

$$S = \frac{a}{1 - q}$$

Apêndice 3.B – Tópicos adicionais em distribuições de probabilidade

3.B.1 Média e variância de uma distribuição de Bernoulli

$$E(X) = 1 \times p + 0 \times (1 - p)$$

$$E(X) = p$$

$$E(X^2) = 1^2 \times p + 0^2 \times (1 - p)$$

$$E(X^2) = p$$

$$\text{var}(X) = E(X^2) - [E(X)]^2$$

$$\text{var}(X) = p - p^2$$

$$\text{var}(X) = p(1 - p)$$

3.B.2 Média e variância de uma distribuição binomial

Usando o mesmo método da distribuição de Bernoulli, façamos $n = 2$:

$$E(X) = 2 \times p^2 + 1 \times 2 p (1 - p) + 0 \times (1 - p)^2 = 2 p^2 + 2 p - 2 p^2 = 2p.$$

Para $n = 3$, teríamos:

$$\begin{aligned} E(X) &= 3 \times p^3 + 2 \times 3 p^2 (1 - p) + 1 \times 3 p (1 - p)^2 + 0 \times (1 - p)^3 = \\ &= 3 p^3 + 6p^2 - 6p^3 + 3p - 6p^2 + 3p^3 = 3p. \end{aligned}$$

Generalizando, temos:

$$E(X) = np$$

Para os quadrados, quando $n = 2$:

$$E(X) = 2^2 \times p^2 + 1^2 \times 2 p (1 - p) + 0^2 \times (1 - p)^2 = 4 p^2 + 2p - 2 p^2 = 2p^2 + 2p.$$

$$\text{var}(X) = 2p^2 + 2p - (2p)^2 = 2p^2 + 2p - 4p^2 = 2p - 4p^2 = 2p(1 - p).$$

Se repetirmos o processo para $n = 2, 3, \dots$, verificaremos que:

$$\text{var}(X) = np(1 - p)$$

3.B.3 Da binomial à Poisson

A probabilidade em uma distribuição binomial é dada por:

$$P(x = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Pela definição de binomial (combinações):

$$P(x = k) = \frac{n!}{(n - k)! k!} p^k (1 - p)^{n-k}$$

$$P(x=k) = \frac{n(n-1)(n-2)\dots(n-k+1)(n-k)!}{(n-k)!k!} p^k (1-p)^{n-k}$$

$$P(x=k) = \frac{n(n-1)(n-2)\dots(n-k+1)}{k!} p^k (1-p)^{n-k}$$

No numerador da fração anterior há k fatores. Colocando n em evidência em cada um deles:

$$P(x=k) = \frac{1}{k!} n^k [(1 - \frac{1}{n})(1 - \frac{2}{n}) \dots (1 - \frac{k-1}{n})] p^k (1-p)^{n-k}$$

Como n tende ao infinito, $1/n, 2/n$ etc. tendem a zero.

$$P(x=k) = \frac{1}{k!} n^k p^k (1-p)^{n-k}$$

Como, por definição, $\lambda = np$, então $p = \frac{\lambda}{n}$.

$$P(x=k) = \frac{1}{k!} n^k \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

Do cálculo diferencial, você sabe que:

$$\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{n-k} = e^{-\lambda}$$

E assim chega-se a:

$$P(x=k) = \frac{e^{-\lambda} \lambda^k}{k!}$$

3.B.4 Média e variância das principais distribuições discretas

Quadro 3.1 – Resumo das principais distribuições discretas

Distribuição	Forma geral $P(x=k)$	Média	Variância
Binomial	$\binom{n}{k} p^k (1-p)^{n-k}$	np	$np(1-p)$
Geométrica	$(1-p)^{k-1} p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Hipergeométrica	$\frac{\binom{s}{k} \binom{N-s}{n-k}}{\binom{N}{n}}$	$np = n \frac{s}{N}$	$n \frac{s}{N} \times \frac{N-s}{N} \times \frac{N-n}{N-1}$
Poisson	$\frac{e^{-\lambda} \lambda^k}{k!}$	$np = \lambda$	λ



Capítulo 4

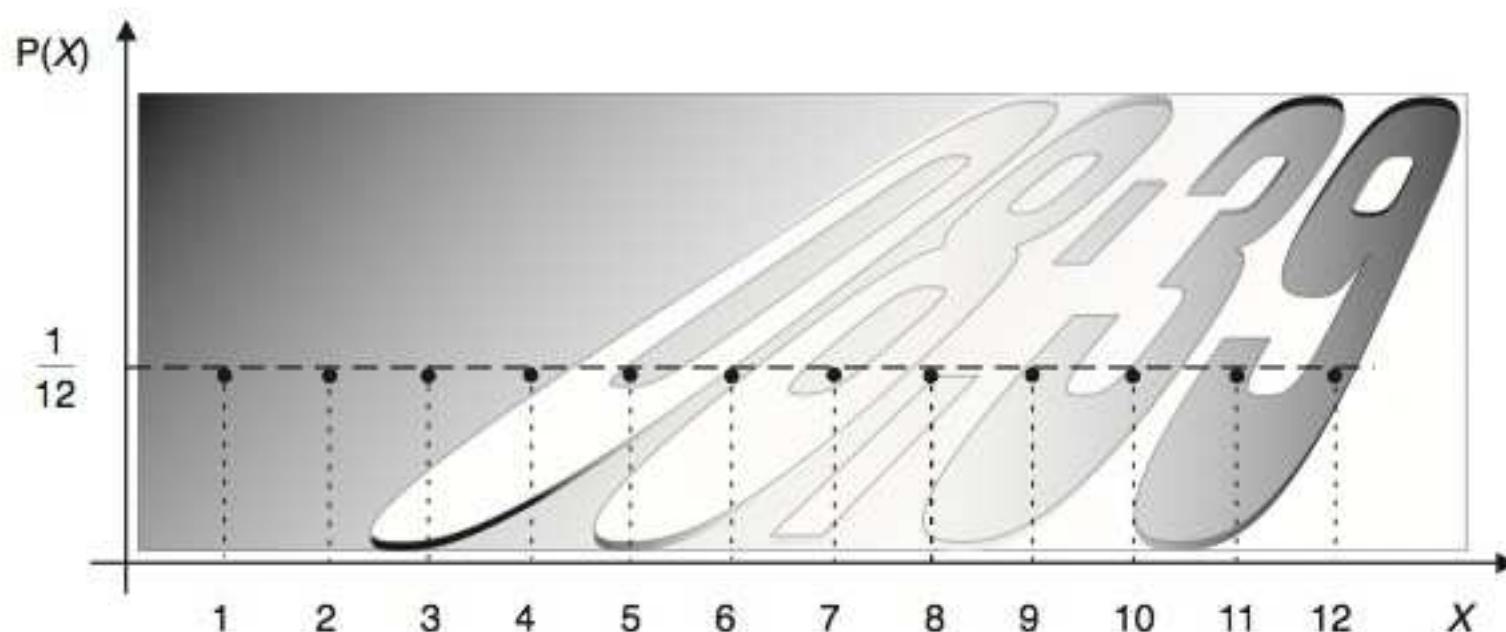
DISTRIBUIÇÕES CONTÍNUAS E TEOREMA DE TCHEBICHEV

4.1 Distribuições contínuas

Imagine o marcador das horas de um relógio digital. Agora, pense no ponteiro das horas de um relógio analógico. Há uma diferença significativa, além da tecnologia empregada. Enquanto o ponteiro passa por qualquer posição do marcador, se se atribui essa sua posição a um valor, esse será exatamente 2 quando for pontualmente 2h, valerá 2,5 quando forem 2h30, 3,25 às 3h25 e assim sucessivamente. O que se quer dizer aqui é que o valor atribuído à posição do ponteiro das horas pode ser qualquer um entre 0 (exclusive) e 12 (inclusive). Já no relógio digital, o mostrador só assume valores inteiros.

Essa diferença pode ser vista graficamente. Primeiro, como mostra a Figura 4.1, num gráfico para o relógio digital.

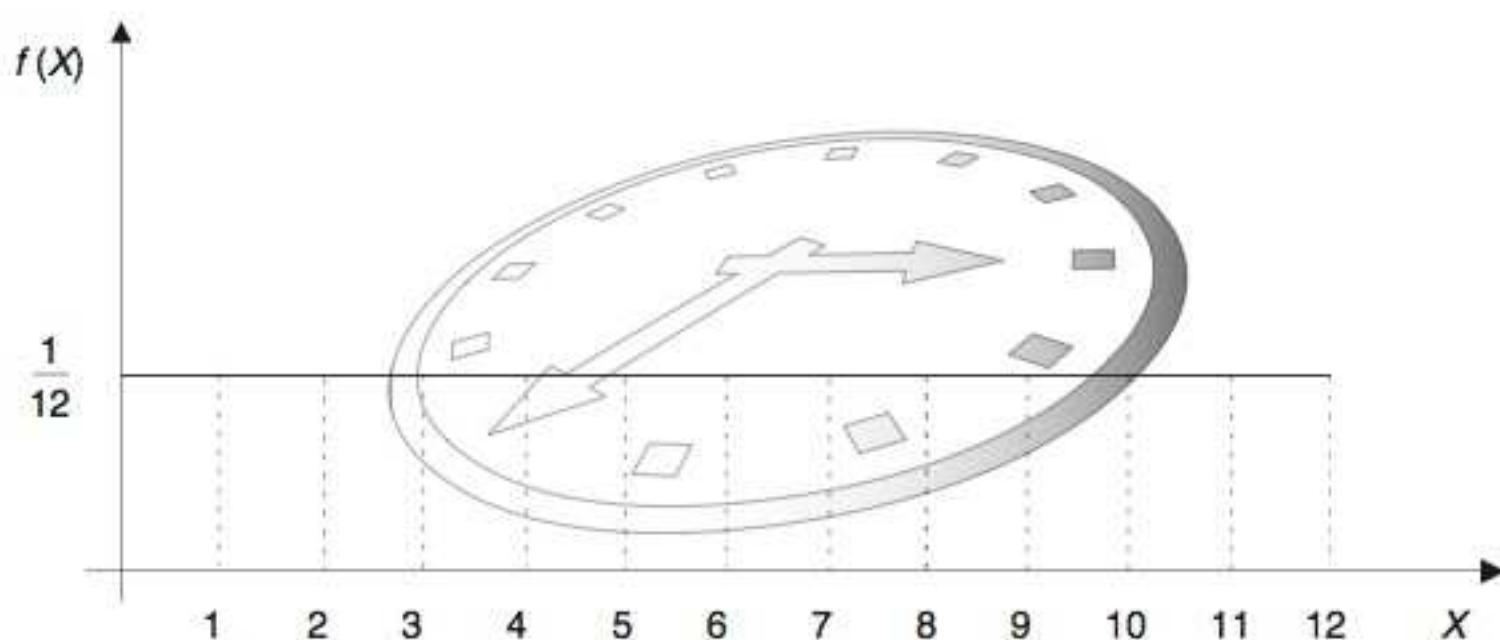
Figura 4.1 – Probabilidade de cada hora em um relógio digital



A variável x é o valor assumido pelo marcador das horas do relógio digital. Se você olhar para ele numa hora qualquer do dia, a probabilidade de que ela tenha um dos 12 valores inteiros da abscissa da Figura 4.1 é exatamente $\frac{1}{12}$. Não há a possibilidade de que a hora assuma outros valores.

A diferença no gráfico para o relógio analógico é que ele assume, em princípio, qualquer valor e, portanto, como apresenta a Figura 4.2, devemos *preencher* a linha que une os 12 pontos.

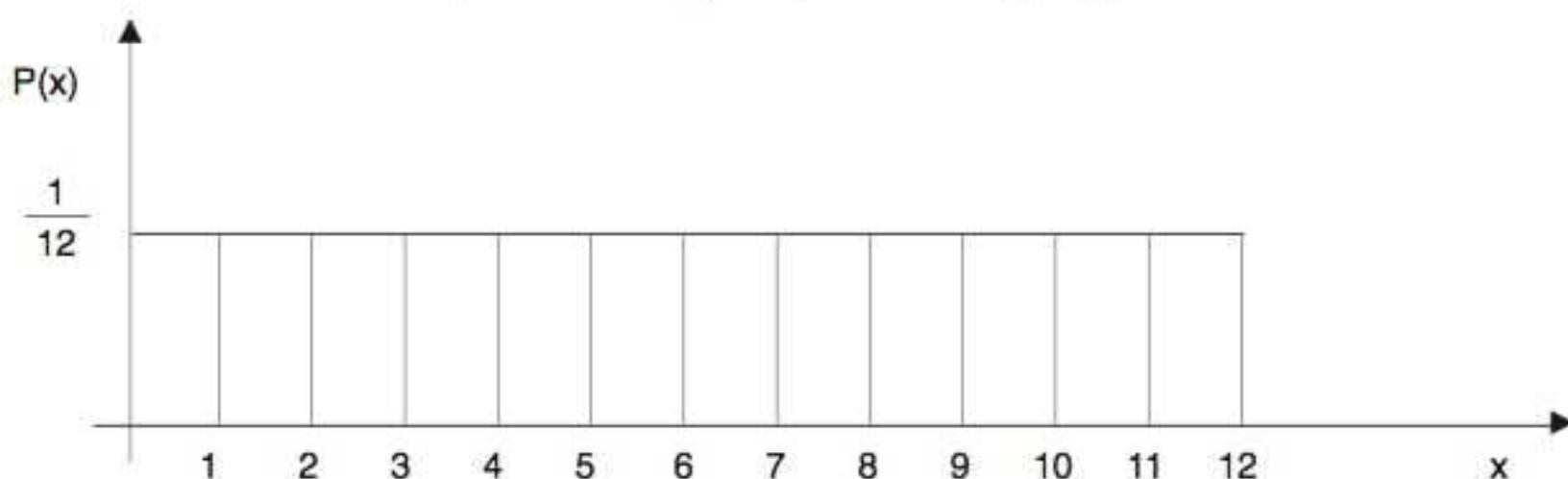
Figura 4.2 – Probabilidade de cada hora em um relógio analógico



A variável x pode assumir, portanto, infinitos valores. Como vimos no Capítulo 1, embora o ponteiro das horas passe pelo 2, a probabilidade de que x seja *exatamente* igual a 2 é zero, já que é um valor entre infinitos possíveis. Como calcular, então, a probabilidade de que x assuma um valor entre, digamos, 2 e 3? Do Capítulo 1, já sabemos a resposta, que é o mesmo $\frac{1}{12}$, já que o intervalo de 2 a 3 é $\frac{1}{12}$ do intervalo total (e todos os intervalos do mesmo tamanho têm a mesma probabilidade de ocorrer).

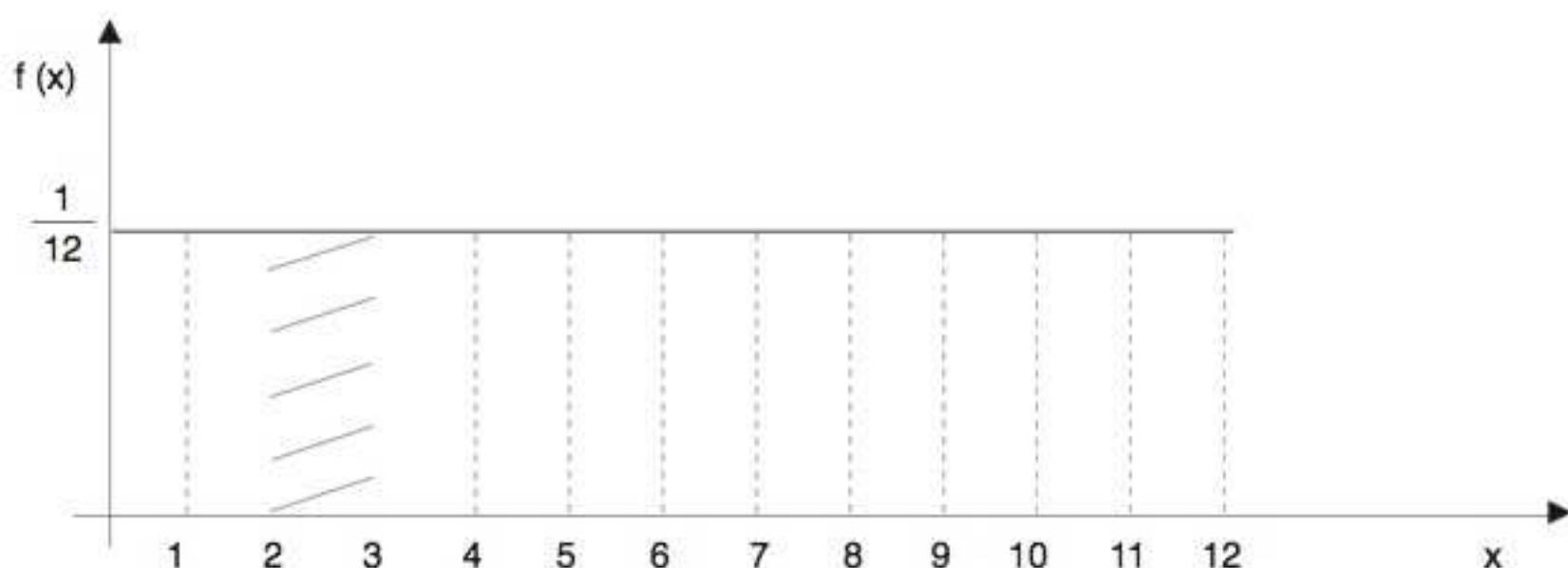
Outra maneira de chegar a esse cálculo é retomar o gráfico da Figura 4.1, mas dessa vez em forma de histograma, na Figura 4.3.

Figura 4.3 – Histograma para um relógio digital



Uma maneira de se interpretar a probabilidade de o mostrador estar indicando duas horas, isto é, $P(x = 2)$, é a área do retângulo correspondente a $x = 2$. A base desse retângulo é 1 e a altura, $1/12$. A área corresponde, portanto, a $1 \times \frac{1}{12} = \frac{1}{12}$.

Para uma distribuição contínua, você pode usar um raciocínio análogo, isto é, para determinar a probabilidade de x estar entre 2 e 3, calcule, na Figura 4.4, a área definida pela função nesse intervalo.

Figura 4.4 – Probabilidade de x estar entre 2 e 3

A área é, de novo, de um retângulo, cuja base é 1 e a altura $1/12$. Portanto:

$$P(2 < x < 3) = 1 \times \frac{1}{12} = \frac{1}{12}$$

Repare que, como a probabilidade de um ponto é igual a zero, tanto faz, nesse caso, se forem utilizados os símbolos de *menor* ou *menor ou igual*, pois a probabilidade será a mesma:

$$P(2 < x < 3) = P(2 \leq x < 3) = P(2 < x \leq 3) = P(2 \leq x \leq 3) = \frac{1}{12}$$

Uma distribuição como essa do relógio analógico é uniforme (contínua).

Note uma coisa importante: a função $f(x)$ não fornece diretamente a probabilidade de x , até porque essa é zero, já que se trata de uma distribuição contínua. Ela é chamada de **função densidade de probabilidade** (f.d.p.), e as probabilidades são obtidas por meio das *áreas* definidas por essa função.

As propriedades de probabilidade, entretanto, devem ser mantidas para que $f(x)$ seja uma função densidade de probabilidade. A soma das probabilidades tem de ser igual a 1, o que vale dizer que a *área total deve ser igual* a 1, embora $f(x)$ possa ser maior do que 1.

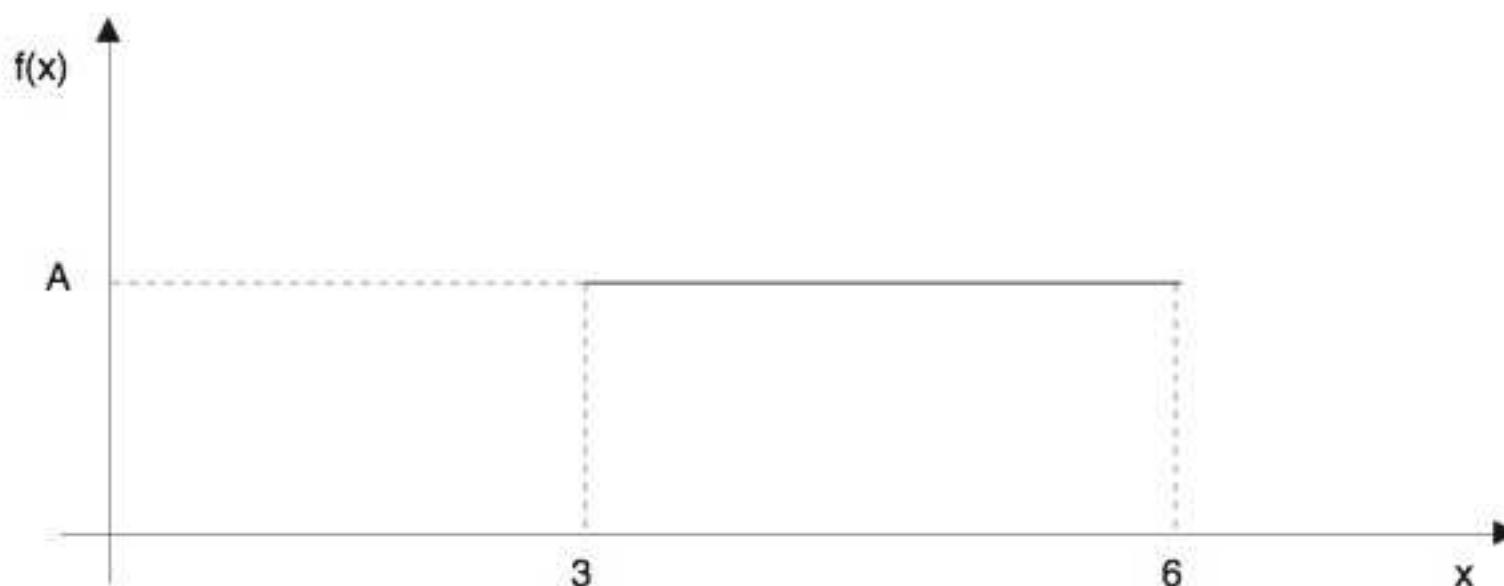
De fato, a área total definida por $f(x)$ é $12 \times \frac{1}{12} = 1$.

Além disso, a probabilidade não pode ser negativa. Portanto, $f(x)$ tem de ser não negativo, isto é, *maior ou igual a zero*.

Exemplo 4.1.1

Uma variável aleatória (v.a.) contínua, com distribuição uniforme, pode assumir qualquer valor real entre 3 e 6. Determine a função densidade de probabilidade dessa função.

O gráfico dessa função é mostrado da Figura 4.5.

Figura 4.5 – Distribuição uniforme (x entre 3 e 6)

E A é um valor que ainda é preciso determinar. Como $f(x)$ é sempre positiva ou zero, pode-se aplicar a condição de que a área total delimitada pelo gráfico tem de ser igual a 1. A base do retângulo é 3 ($= 6 - 3$) e a altura é igual a A . Portanto:

$$A \times 3 = 1$$

$$A = \frac{1}{3}$$

Ou seja, $f(x) = 1/3$ quando x está entre 3 e 6 e é igual a zero para todos os demais valores de x , o que pode ser representado como se segue:

$$f(x) = \begin{cases} 0, & x < 3 \text{ ou } x > 6 \\ \frac{1}{3}, & 3 \leq x \leq 6 \end{cases}$$

Exemplo 4.1.2

Partindo da função densidade de probabilidade do exemplo anterior, determine as probabilidades a seguir.

a) $x = 4$.

Embora seja possível, como se trata de distribuição contínua, a probabilidade de x ser exatamente igual a um valor é igual a zero. Portanto:

$$P(x = 4) = 0$$

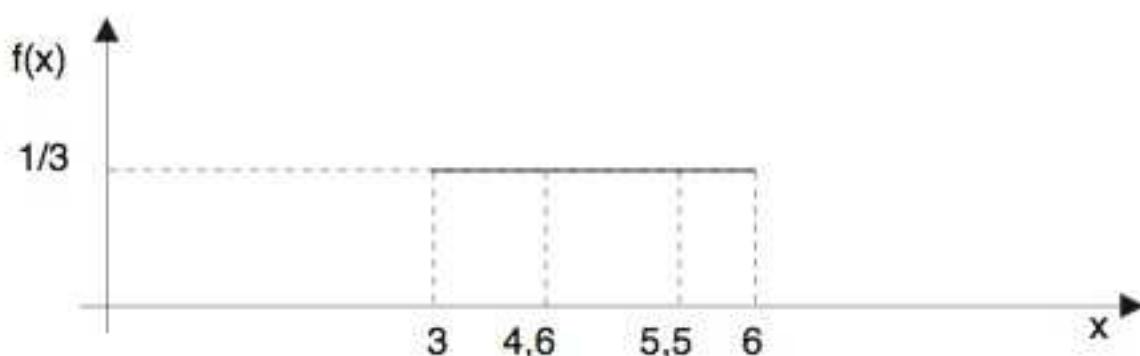
b) x entre 4,6 e 5,5.

A função é dada por:

$$f(x) = \begin{cases} 0, & x < 3 \text{ ou } x > 6 \\ \frac{1}{3}, & 3 \leq x \leq 6 \end{cases}$$

Seu gráfico é mostrado na Figura 4.6.

Figura 4.6 – Função de x entre 4,6 e 5,5



A probabilidade é dada pela área delimitada no gráfico da Figura 4.6, que corresponde a um triângulo de base 0,9 e altura $1/3$.

$$P(4,6 \leq x \leq 5,5) = 0,9 \times \frac{1}{3} = 0,3$$

c) x entre 2 e 4.

Como x só assume valores entre 3 e 6, a área relevante a ser calculada corresponde aos pontos entre 3 e 4, já que, para qualquer intervalo antes de 3, a probabilidade é igual a zero.

$$P(2 \leq x \leq 4) = P(2 \leq x \leq 3) + P(3 \leq x \leq 4)$$

$$P(2 \leq x \leq 4) = 0 + 1 \times \frac{1}{3}$$

$$P(2 \leq x \leq 4) \approx 0,33$$

Exemplo 4.1.3

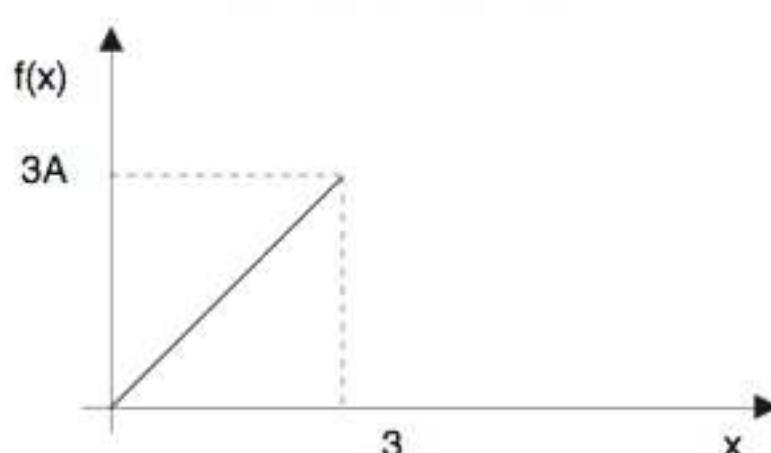
Dada a função densidade de probabilidade da variável aleatória contínua mostrada a seguir:

$$f(x) = \begin{cases} Ax, & 0 \leq x \leq 3 \\ 0, & x < 0 \text{ ou } x > 3 \end{cases}$$

a) Determine o valor de A.

O gráfico dessa função é dado na Figura 4.7.

Figura 4.7 – $f(x) = Ax$



Como $f(x) = Ax$, $f(3) = 3A$ e $f(0) = 0$. A forma definida pelo gráfico da Figura 4.7 é um triângulo de base 3 e altura $3A$. Sabemos que $f(x)$ é sempre não negativo, portanto, basta aplicarmos a propriedade de que a área total seja igual a 1:

$$\frac{3A \times 3}{2} = 1$$

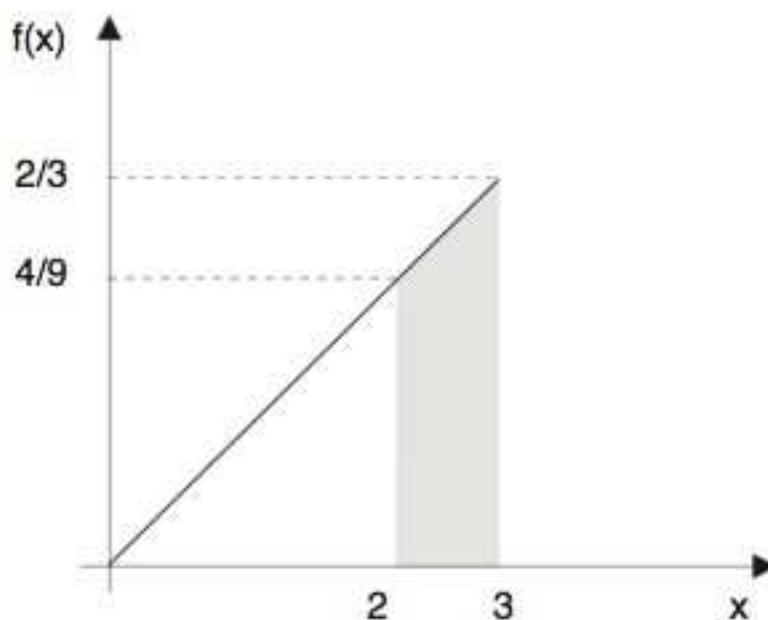
$$\frac{9A}{2} = 1$$

$$A = 2/9$$

b) Determine a probabilidade de que x esteja entre 2 e 3.

Agora $f(2) = 2 \times \frac{2}{9} = \frac{4}{9}$ e $f(3) = 3 \times \frac{2}{9} = \frac{6}{9} = \frac{2}{3}$. A área correspondente a essa probabilidade está assinalada no gráfico da Figura 4.8.

Figura 4.8 – Área sobre a função $f(x) = Ax$



Esse gráfico determina um trapézio. Pode-se calcular diretamente a área do trapézio ou, então, a diferença entre a área dos dois triângulos (o maior, cuja base vai de 0 a 3, e o menor, cuja base vai de 0 a 2):

$$P(2 \leq x \leq 3) = 3 \times \frac{2}{3} \times \frac{1}{2} - 2 \times \frac{4}{9} \times \frac{1}{2}$$

$$P(2 \leq x \leq 3) = 1 - \frac{4}{9} = \frac{5}{9}$$

Exemplo 4.1.4

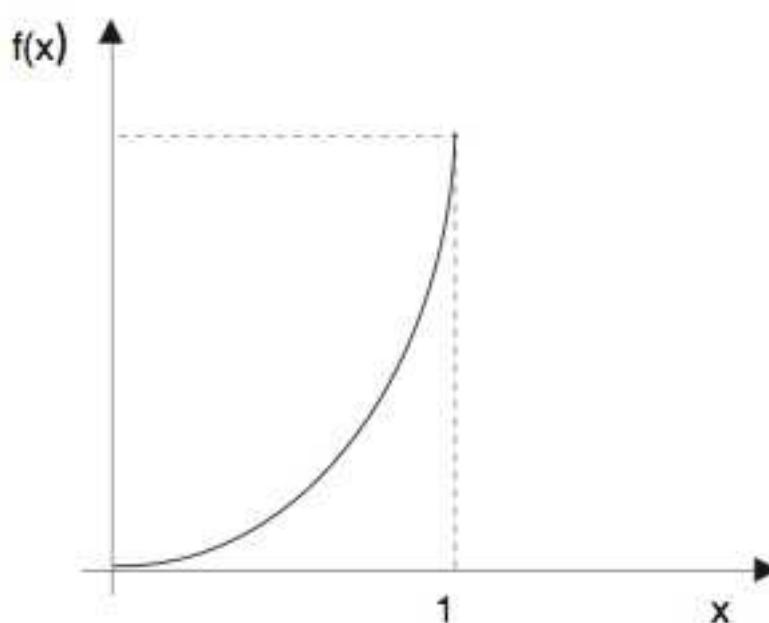
Dada a f.d.p. de uma variável aleatória contínua a seguir:

$$f(x) = \begin{cases} Ax^2, & 0 \leq x \leq 1 \\ 0, & x < 0 \text{ ou } x > 1 \end{cases}$$

a) Determine o valor da constante A .

O gráfico dessa função é dado na Figura 4.9.

Figura 4.9 – $f(x) = Ax^2$



Uma vez que não se trata mais de uma função cujo gráfico é retilíneo, como as funções anteriores, é preciso recorrer ao cálculo integral. Sabe-se que a área sobre uma curva é dada pela integral da função correspondente (veja Apêndice 4.A). Portanto, a condição de que a área total tem de ser igual a 1 pode ser escrita como:

$$\int_{-\infty}^{+\infty} f(x) dx = 1$$

Nessa situação específica, a função vale zero para valores de x abaixo de 0 ou acima de 1. Portanto, os limites de integração relevantes são, nesse caso, 0 e 1:

$$\int_0^1 f(x) dx = 1$$

$$\int_0^1 Ax^2 dx = 1$$

$$A \int_0^1 x^2 dx = 1$$

$$A \left[\frac{x^3}{3} \right]_0^1 = 1$$

$$A \left[\frac{1}{3} - \frac{0}{3} \right] = 1$$

$$A \times \frac{1}{3} = 1$$

$$A = 3.$$

b) Determine a probabilidade de que x esteja entre 0,5 e 1.

De novo, para calcular a área entre $x = 0,5$ e $x = 1$, determinando, assim, a probabilidade, basta encontrar a integral com esses limites de integração:

$$P(0,5 \leq x \leq 1) = \int_{0,5}^1 3x^2 dx$$

$$P(0,5 \leq x \leq 1) = \left[x^3 \right]_{0,5}^1$$

$$P(0,5 \leq x \leq 1) = 1^3 - 0,5^3$$

$$P(0,5 \leq x \leq 1) = 1 - 0,125$$

$$P(0,5 \leq x \leq 1) = 0,875 = 87,5\%$$

É possível usar o cálculo integral para os exemplos anteriores também. Assim, pode-se resumir as condições para que uma função qualquer seja uma função densidade de probabilidade:

$$\int_{-\infty}^{+\infty} f(x)dx = 1, \text{ e}$$

$f(x) \geq 0$ para todos os valores de x .

Exemplo 4.1.5

[Distribuição exponencial]

Dada a seguinte f.d.p. da variável aleatória contínua x :

$$f(x) = \begin{cases} Ae^{-\alpha x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Determine o valor de A .

Essa particular distribuição é conhecida como distribuição exponencial.

Sabe-se que:

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

Como essa função é nula para valores de x negativos:

$$\int_0^{+\infty} Ae^{-\alpha x} dx = 1$$

$$A \int_0^{+\infty} e^{-\alpha x} dx = 1$$

$$A \left[\frac{-e^{-\alpha x}}{\alpha} \right]_0^{+\infty} = 1$$

$$A \left[0 - \left(-\frac{1}{\alpha} \right) \right] = 1$$

$$A \times \frac{1}{\alpha} = 1$$

$$A = \alpha$$

4.2 Função de distribuição de variáveis contínuas

A **função de distribuição acumulada**, ou simplesmente função de distribuição, no caso de variáveis contínuas, segue a mesma lógica do caso discreto.

No caso discreto, a função de distribuição $F(x)$ é a soma das probabilidades de todos os valores possíveis que a variável x pode assumir até o valor de x propriamente dito. Assim, se x é um número inteiro não negativo, a função de distribuição é dada por:

$$F(0) = P(0)$$

$$F(1) = P(0) + P(1)$$

$$F(2) = P(0) + P(1) + P(2)$$

$$F(3) = P(0) + P(1) + P(3)$$

E assim sucessivamente. Para o caso de uma variável contínua, porém, todos os valores possíveis devem ser somados, o que é feito pela integral. Dessa forma:

$$F(x) = \int_{-\infty}^x f(t)dt$$

Portanto, do ponto de vista matemático, $f(x)$ é a derivada da função $F(x)$:

$$f(x) = \frac{dF(x)}{dx}$$

Exemplo 4.2.1

Dada a seguinte f.d.p. da distribuição exponencial, determine a função de distribuição correspondente:

$$f(x) = \begin{cases} e^{-x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Como a função só é definida para $x \geq 0$, o limite de integração inferior é zero.

$$F(x) = \int_0^x f(t)dt$$

$$F(x) = \int_0^x e^{-t} dt$$

$$F(x) = [-e^{-t}]_0^x$$

$$F(x) = -e^{-x} + e^0$$

$$F(x) = 1 - e^{-x}$$

A função distribuição é dada, então, por:

$$F(x) = \begin{cases} 1 - e^{-x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Exemplo 4.2.2

Dada a seguinte função distribuição, determine a função densidade de probabilidade correspondente.

$$F(x) = \begin{cases} 0,5(x^3 + 1), & -1 \leq x \leq 1 \\ 0, & x < -1 \\ 1, & x > 1 \end{cases}$$

A função densidade de probabilidade é dada por:

$$f(x) = \frac{dF(x)}{dx}$$

$$f(x) = \frac{d(0,5x^3 + 1)}{dx}$$

$$f(x) = 3 \times 0,5x^2 + 0$$

$$f(x) = 1,5x^2$$

Portanto, a função densidade de probabilidade é:

$$f(x) = \begin{cases} 1,5x^2, & -1 \leq x \leq 1 \\ 0, & x < -1 \text{ ou } x > 1 \end{cases}$$

A função distribuição $F(x)$, assim como a função densidade, deve preencher alguns *requisitos*: o primeiro é que, em se tratando de uma soma de probabilidades, ela jamais pode ser negativa.

Como a soma das probabilidades tem de ser 1, $F(x)$ não pode ser nunca maior do que 1 e, além disso, seu valor *final* tem de ser, necessariamente, 1. Portanto:

$$0 \leq F(x) \leq 1$$

$$\lim_{x \rightarrow \infty} F(x) = 1$$

Você pode verificar que, tanto no Exemplo 4.2.1 como no 4.2.2, as funções $F(x)$ apresentadas atendem a essas condições.

4.3 Esperança e variância de variáveis aleatórias contínuas

Para uma variável aleatória discreta, a esperança é dada por:

$$E(X) = X_1 P(X_1) + X_2 P(X_2) + \dots + X_n P(X_n) = \sum_{i=1}^n X_i P(X_i)$$

Para uma variável aleatória contínua, seria preciso de somar continuamente todos os valores de x com suas respectivas probabilidades. Uma soma contínua é a integral e, por sua vez, a probabilidade é encontrada pela função densidade de probabilidade. Então, no caso contínuo:

$$E(x) = \int_{-\infty}^{+\infty} xf(x)dx$$

A variância, por sua vez, é:

$$\text{var}(X) = E[X - E(X)]^2$$

Chamando, por simplicidade, $E(X)$ (que é a média de X) de μ :

$$\text{var}(X) = E(X - \mu)^2$$

Para o caso contínuo, bastaria substituir $(x - \mu)^2$ na expressão da esperança anterior e:

$$\text{var}(x) = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x)dx$$

É possível utilizar a seguinte expressão da variância:

$$\text{var}(x) = E(x^2) - [E(x)]^2$$

Em que:

$$E(x) = \int_{-\infty}^{+\infty} xf(x)dx \text{ e}$$

$$E(x^2) = \int_{-\infty}^{+\infty} x^2 f(x)dx$$

Exemplo 4.3.1

Da f.d.p. do Exemplo 4.1.4, determine o valor médio de x .

Trata-se aqui de calcular a esperança de x :

$$E(x) = \int_{-\infty}^{+\infty} xf(x)dx$$

O que, para essa variável, equivale a:

$$E(x) = \int_0^1 x^3 x^2 dx$$

$$E(x) = 3 \int_0^1 x^3 dx$$

$$E(x) = 3 \left[\frac{x^4}{4} \right]_0^1$$

$$E(x) = 3 \times \frac{1}{4}$$

$$E(x) = \frac{3}{4} = 0,75$$

Determine, agora, a variância de x. A média dos quadrados de x é dada por:

$$E(x^2) = \int_{-\infty}^{+\infty} x^2 f(x) dx$$

$$E(x^2) = \int_0^1 x^2 3x^2 dx$$

$$E(x^2) = 3 \int_0^1 x^4 dx$$

$$E(x^2) = 3 \left[\frac{x^5}{5} \right]_0^1$$

$$E(x^2) = 3 \times \frac{1}{5}$$

$$E(x^2) = \frac{3}{5} = 0,6$$

Assim, pode-se calcular a variância:

$$\text{var}(x) = E(x^2) - [E(x)]^2$$

$$\text{var}(x) = 0,6 - 0,75^2$$

$$\text{var}(x) = 0,6 - 0,5625$$

$$\text{var}(x) = 0,0375$$

Determine, finalmente, o desvio padrão de x.

$$dp(x) = \sqrt{0,0375}$$

$$dp(x) \cong 0,194$$

Exemplo 4.3.2

Dada a seguinte distribuição exponencial:

$$f(x) = \begin{cases} e^{-x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Determine a média de x.

$$E(x) = \int_0^{+\infty} x f(x) dx$$

$$E(x) = \int_0^{+\infty} x e^{-x} dx$$

$$E(x) = \left[-xe^{-x} - e^{-x} \right]_0^{+\infty}$$

$$E(x) = 1$$

Calcule também a mediana de x , levando em conta que a mediana de uma variável é o valor que divide a distribuição em duas. Se a mediana é m , vale dizer que, para uma variável aleatória contínua:

$$P(x > m) = \int_m^{+\infty} f(x)dx = 0,5$$

$$P(x < m) = \int_{-\infty}^m f(x)dx = 0,5$$

Utilizando a primeira delas (poderia ser qualquer uma) da f.d.p. em questão:

$$\int_m^{+\infty} e^{-x}dx = 0,5$$

$$[-e^{-x}]_m^{+\infty} = 0,5$$

$$e^{-m} = 0,5$$

Aplicando logaritmo natural em ambos os lados:

$$\ln(e^{-m}) = \ln 0,5$$

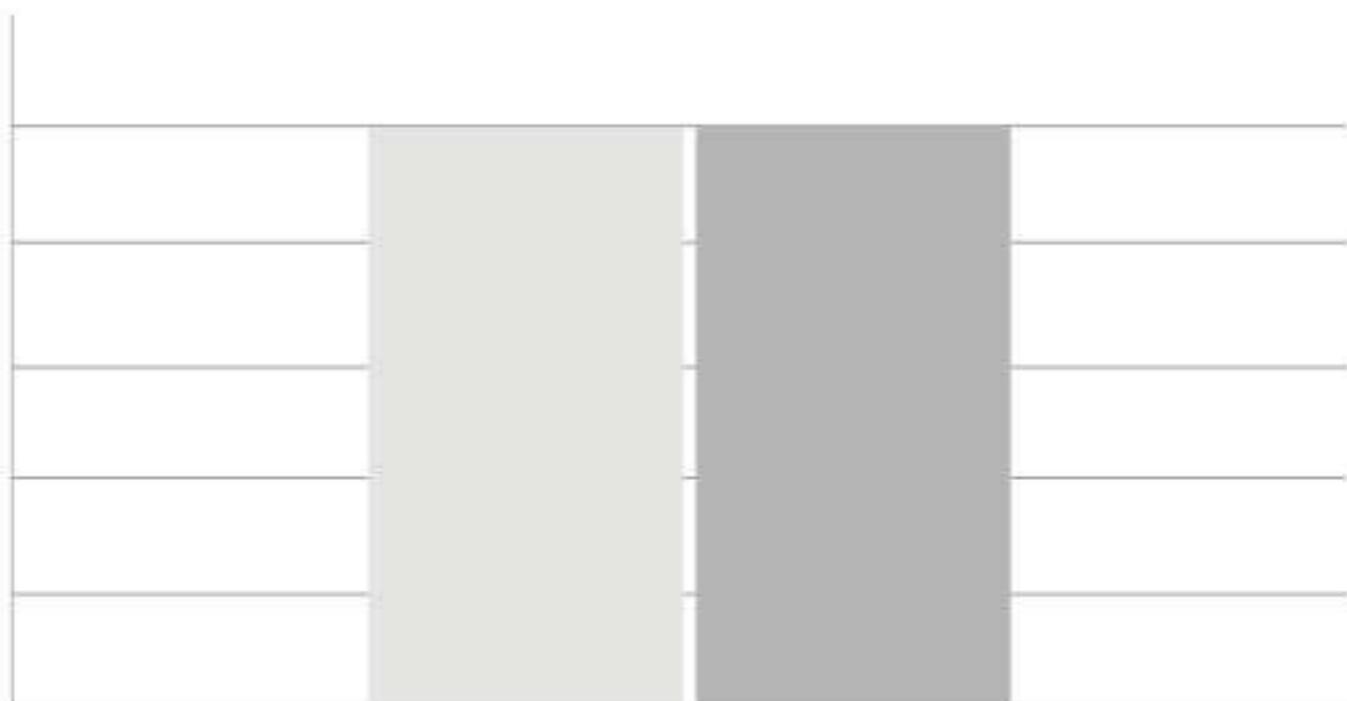
$$-m \cong -0,693$$

$$m \cong 0,693$$

4.4 A distribuição normal

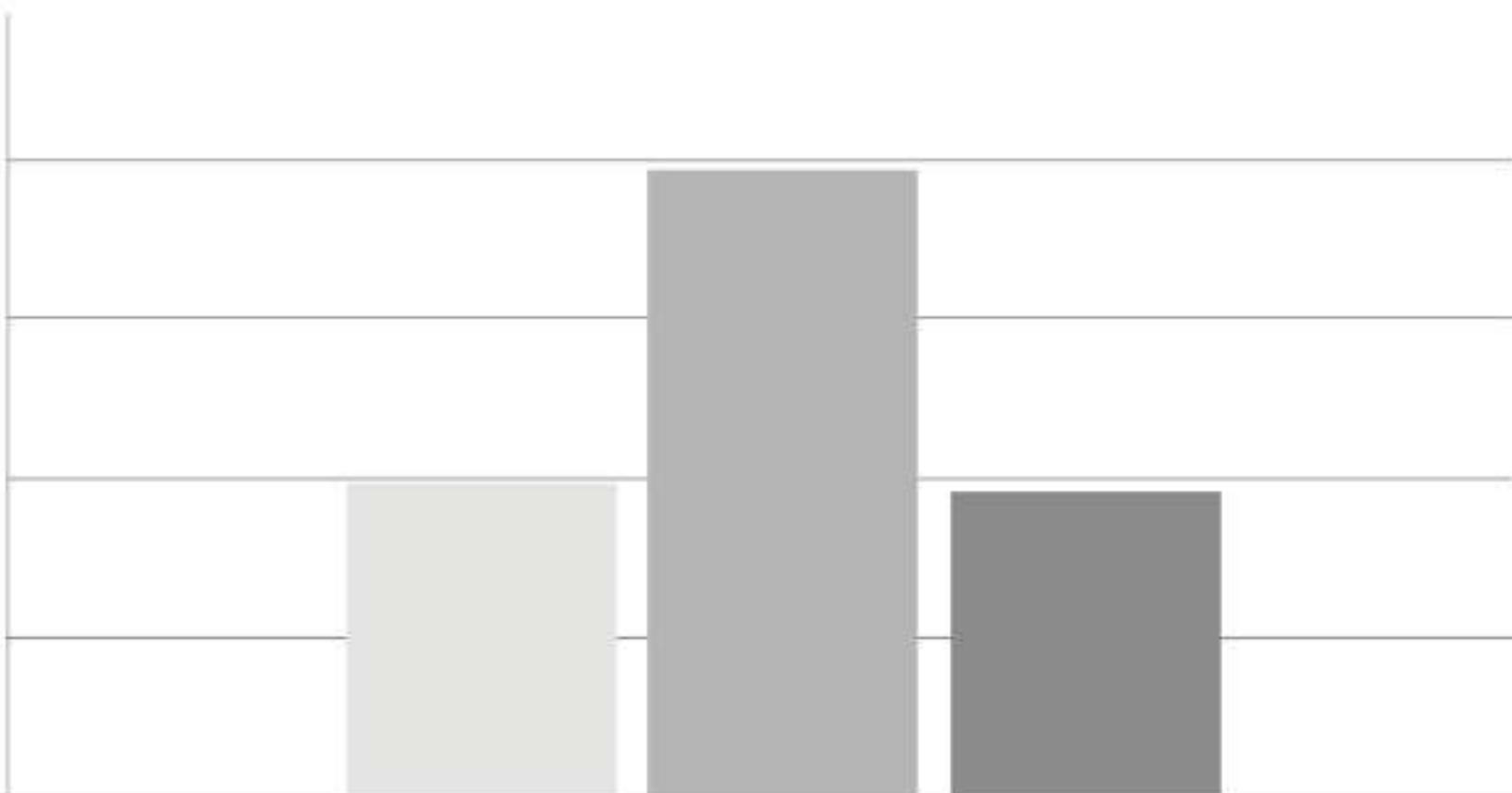
Voltemos à distribuição binomial. Se $n = 1$, ela recai na distribuição de Bernoulli. Supondo que $p = 0,5$, o gráfico em forma de histograma dessa distribuição é dado na Figura 4.10.

Figura 4.10 – Distribuição binomial com $n = 1$



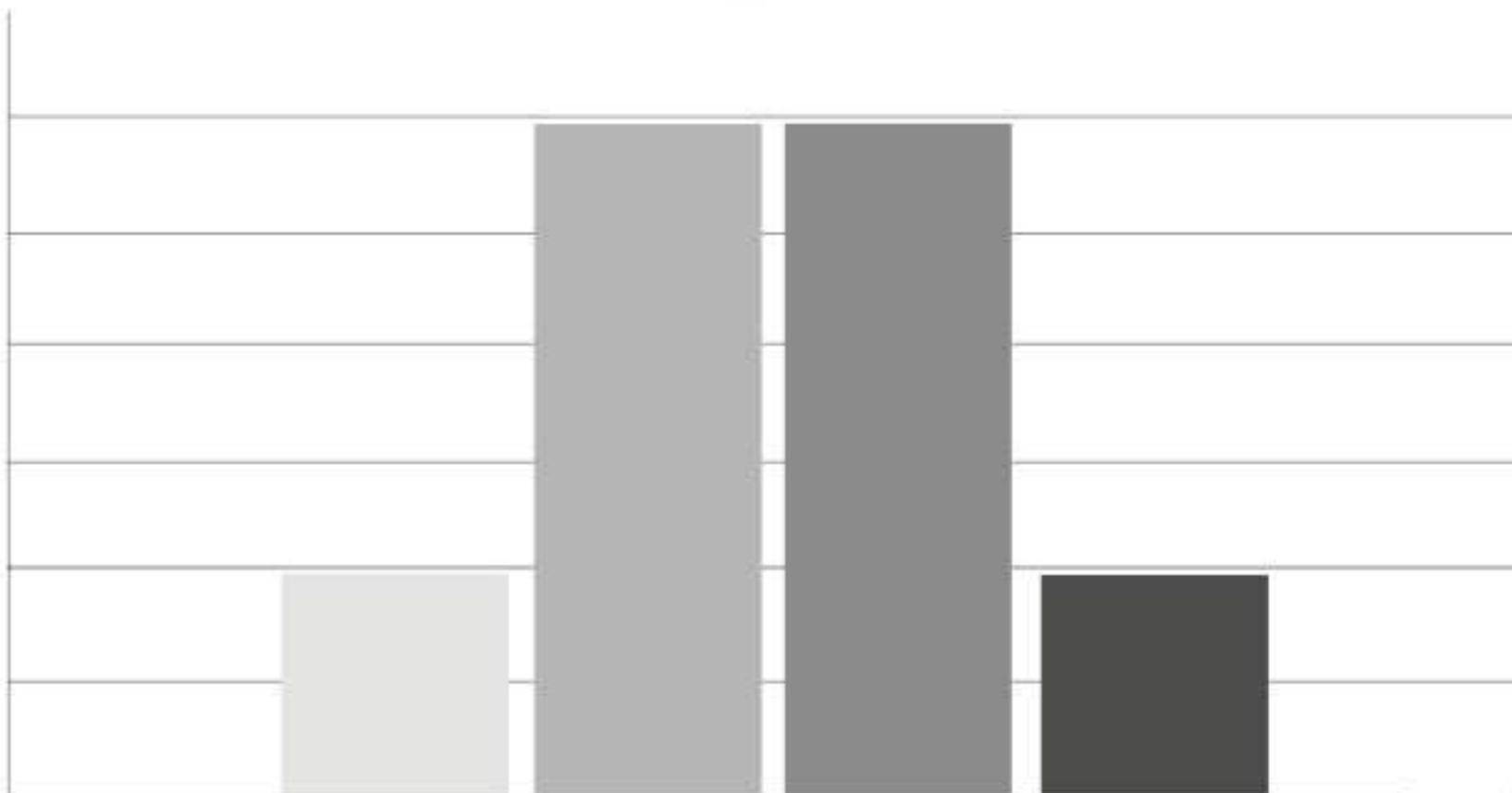
Para $n = 2$, veja o gráfico da Figura 4.11.

Figura 4.11 – Distribuição binomial com $n = 2$



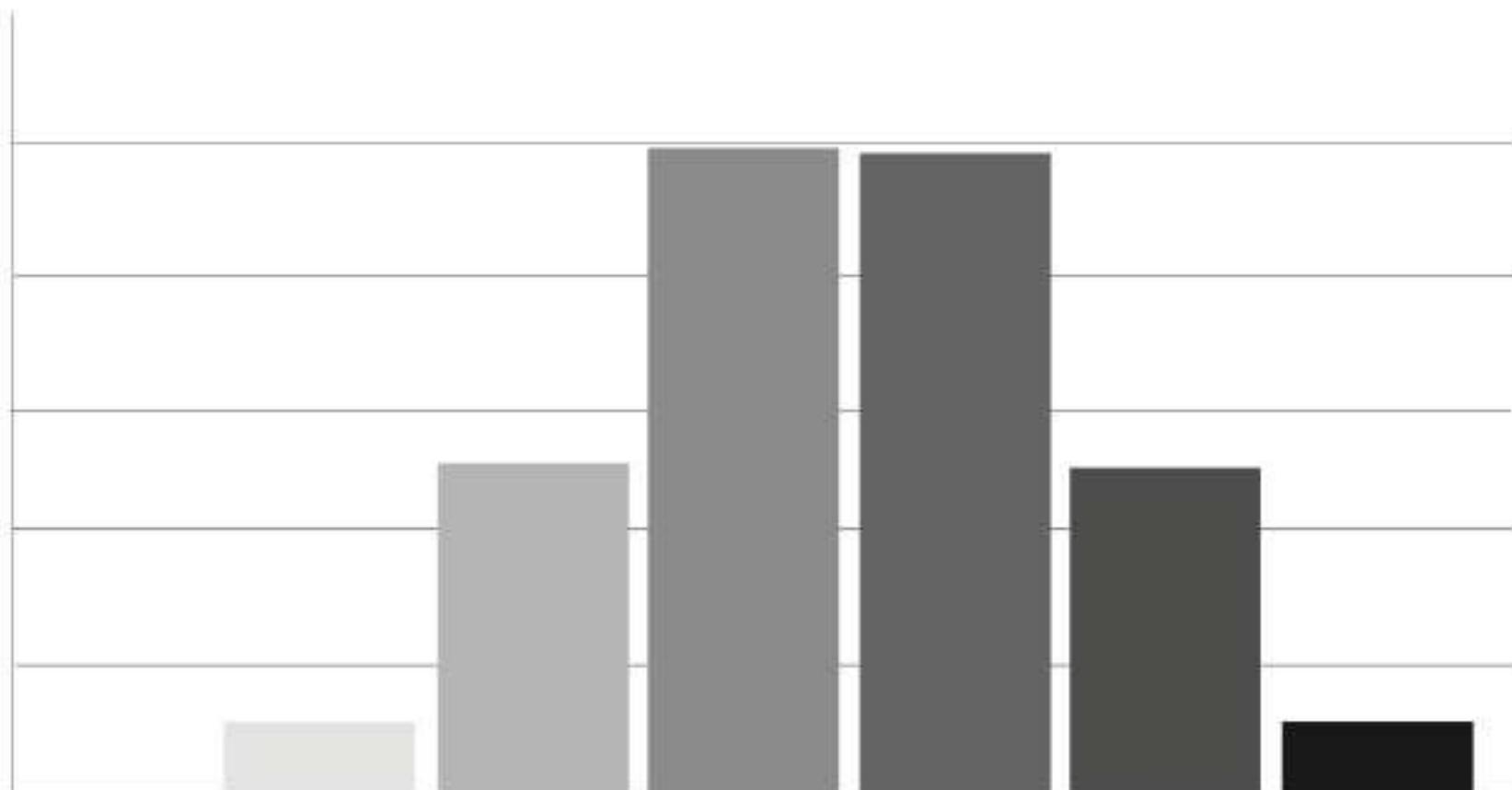
Assim, para $n = 3$, o histograma está na Figura 4.12.

Figura 4.12 – Distribuição binomial com $n = 3$



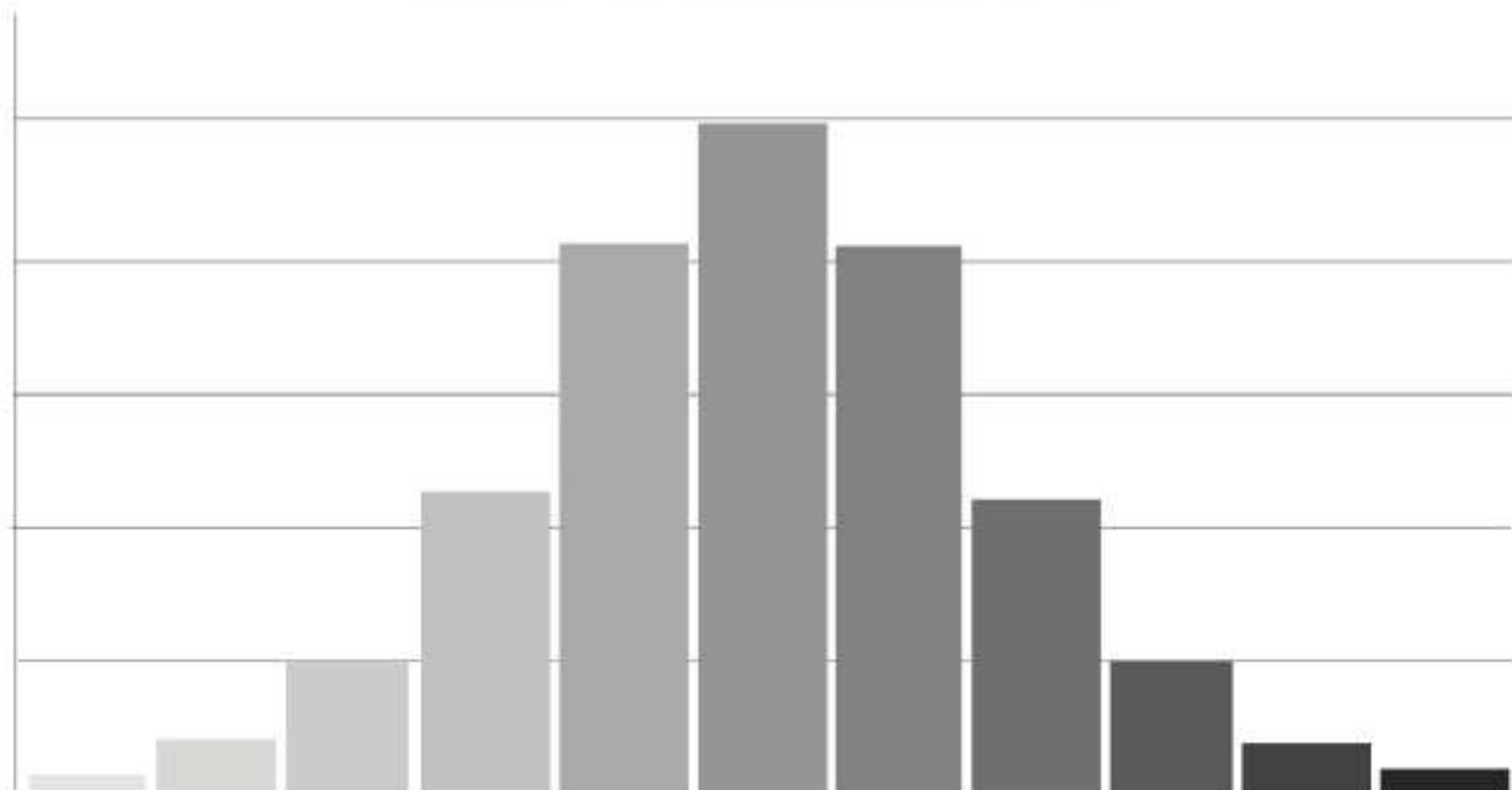
Para $n = 5$, o gráfico é retratado na Figura 4.13.

Figura 4.13 – Distribuição binomial com $n = 5$



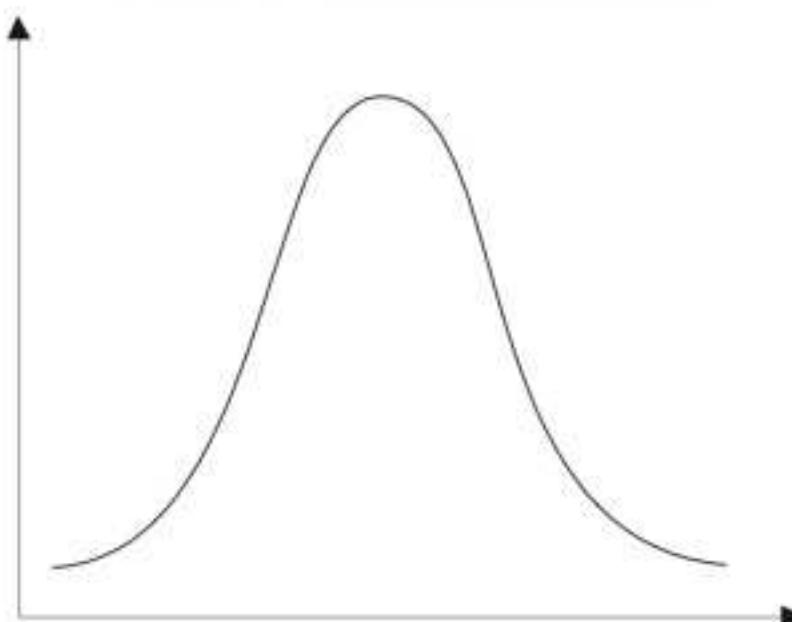
Ou mesmo para $n = 10$, veja a Figura 4.14.

Figura 4.14 – Distribuição binomial com $n = 10$



Suponha que se aumente n indefinidamente, de tal forma que os retângulos do histograma se tornem cada vez mais *espremidos* ou os pontos de um gráfico comum colidam, tornando-se uma função contínua. Essa função teria a aparência da Figura 4.15.

Figura 4.15 – A distribuição normal



Essa distribuição de probabilidade é conhecida como **normal** ou **gaussiana**¹, e sua função densidade de probabilidade é dada por:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Em que μ é a média e σ é o desvio padrão. Se a variável x tem distribuição normal (isto é, é normalmente distribuída), costuma-se simbolizar essa característica por:

$$x \sim N(\mu, \sigma)$$

Essa expressão deve ser lida da seguinte forma: x segue uma distribuição normal com média μ e desvio padrão σ .

Note que se define completamente uma distribuição normal com a média e o desvio padrão (ou a variância), já que não há nenhum outro parâmetro a ser especificado na função anterior. A média determina a posição da curva em relação à origem, enquanto o desvio padrão determina se a curva será mais *gorda* (mais dispersa, com maior desvio padrão) ou mais *magra* (mais concentrada, com menor desvio padrão).

O cálculo das probabilidades sob uma distribuição normal pode se tornar um tanto quanto trabalhoso, já que não há uma função cuja derivada é e^{-x^2} . Esse cálculo deve ser feito por métodos numéricos.

Uma particular distribuição normal, conhecida por **normal padronizada**, que tem média 0 e desvio padrão igual a 1, tem seus resultados das integrais tabelados. Você encontra essa tabela² no fim deste livro.

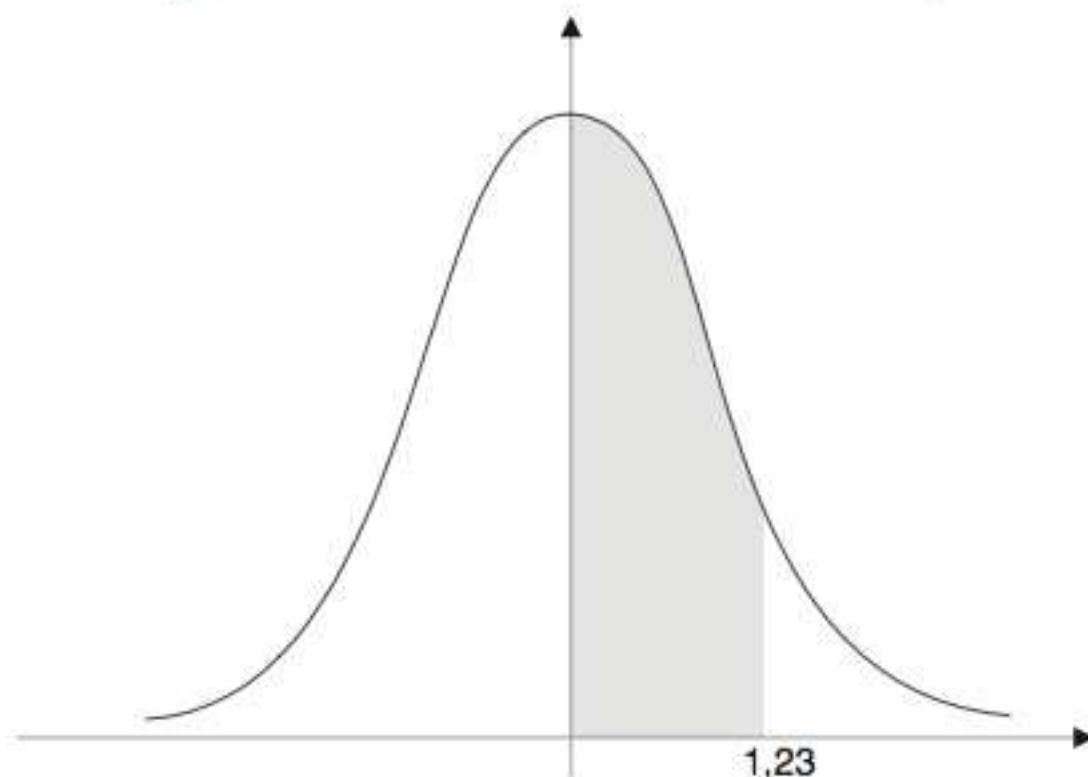
Chamando de z a variável normal padronizada, você encontra na tabela a probabilidade de z estar entre 0 e o valor especificado³. Por exemplo, se quiser encontrar a

¹ Em homenagem ao matemático alemão Carl Friedrich Gauss (1777-1855).

² A utilidade dessa tabela é limitada hoje em dia, tendo em vista que há vários softwares de computador que se valem desses métodos numéricos e calculam rapidamente as integrais sob a curva normal (a própria tabela no final do livro foi calculada assim). Atualmente, a tabela serve para fins didáticos e utilização em exames.

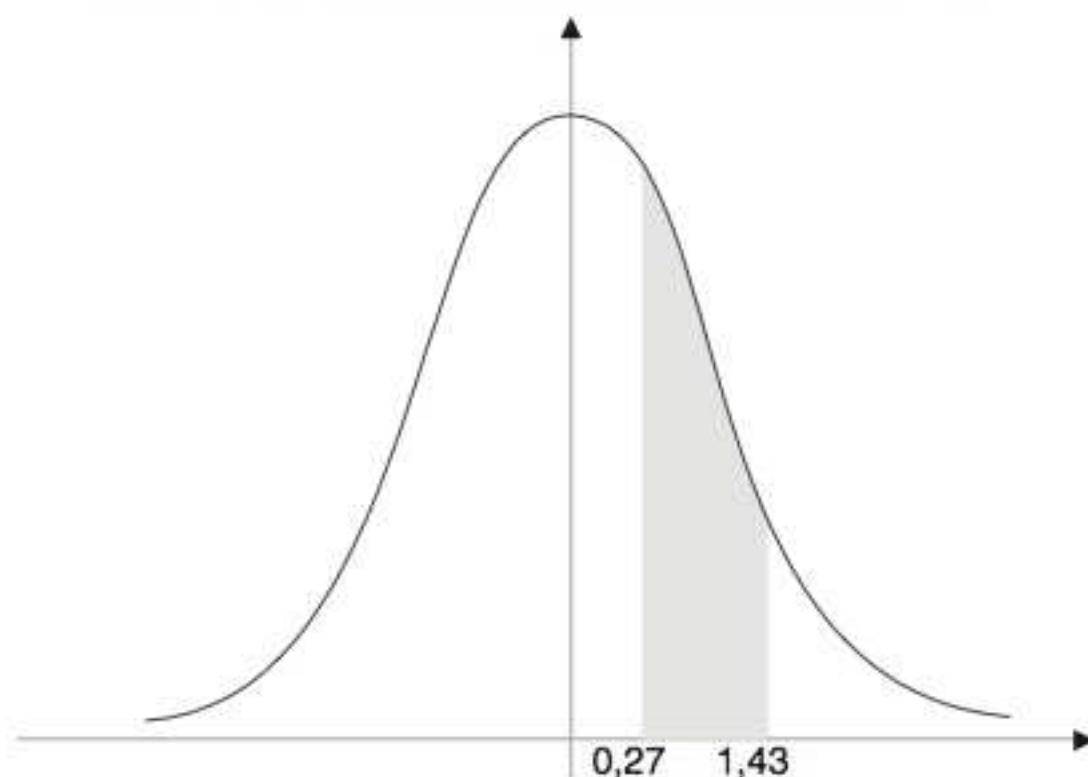
³ Nas linhas da tabela, você encontra o valor de z até a primeira casa decimal, enquanto os valores da segunda casa decimal se encontram

probabilidade de z estar entre 0 e 1,23, você encontra diretamente a probabilidade na tabela, como mostra a Figura 4.16.

Figura 4.16 – Probabilidade de z estar entre 0 e 1,23

$$P(0 < z < 1,23) \cong 0,3907 = 39,07\%$$

Para um valor de z que esteja entre 0,27 e 1,43, veja a Figura 4.17.

Figura 4.17 – Probabilidade de z estar entre 0,27 e 1,43

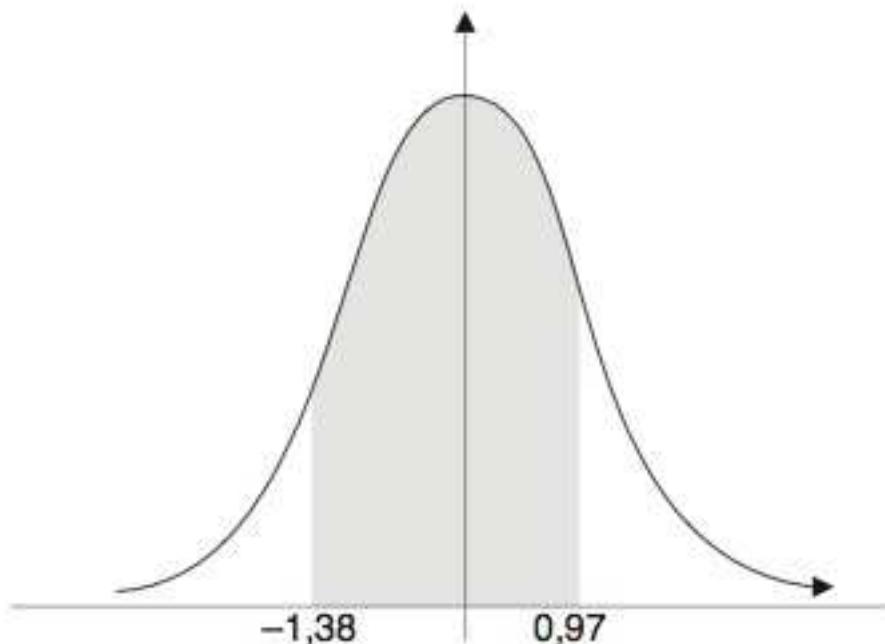
Os valores encontrados na tabela para $z = 0,27$ e $z = 1,43$ são as integrais de 0 até cada um deles. A área que vai de 0,27 a 1,43 é a diferença entre esses dois valores:

$$P(0,27 < z < 1,43) = P(0 < z < 1,43) - P(0 < z < 0,27)$$

$$P(0,27 < z < 1,43) \cong 0,4236 - 0,1064 = 0,3172 = 31,72\%$$

nas colunas.

Para valores negativos (como a média é zero, vale dizer para valores abaixo da média), há de se notar que a normal é simétrica, portanto, o que vale para os valores de z positivos vale também para os negativos. Suponha, então, que você queira calcular a probabilidade de z estar entre $-1,38$ e $0,97$, como mostra a Figura 4.18.

Figura 4.18 – Probabilidade de z estar entre $-1,38$ e $0,97$ 

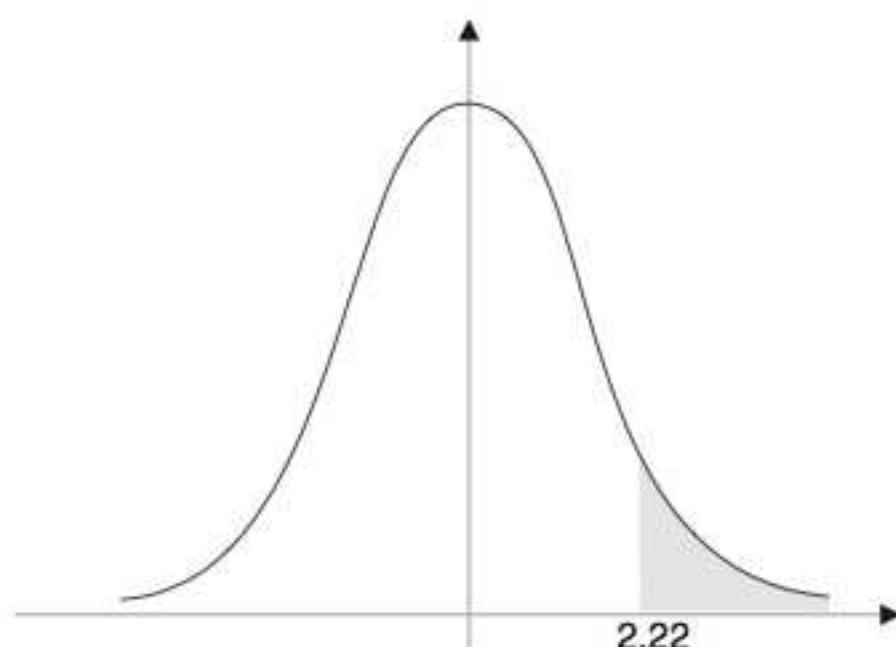
Nesse caso, somam-se as duas áreas hachuradas da Figura 4.18.

$$P(-1,38 < z < 0,97) = P(-1,38 < z < 0) + P(0 < z < 0,97)$$

$$P(-1,38 < z < 0,97) = P(0 < z < 1,38) + P(0 < z < 0,97)$$

$$P(-1,38 < z < 0,97) \cong 0,4162 + 0,3340 = 0,7502 = 75,02\%$$

A Figura 4.19 ajuda a calcular a probabilidade de z ser maior do que $2,22$.

Figura 4.19 – Probabilidade de z ser maior do que $2,22$ 

Aí vale lembrar que, como a distribuição é simétrica, em cada metade há uma probabilidade total de 0,5. Pela tabela, sabemos a probabilidade de z estar entre 0 e 2,22; para saber de 2,22 em diante, basta subtrair de 0,5.

$$P(z > 2,22) = 0,5 - P(0 < z < 2,22)$$

$$P(z > 2,22) \cong 0,5 - 0,4868 = 0,0132 = 1,32\%$$

O problema é que nem todas as variáveis normalmente distribuídas têm média 0 e desvio padrão 1.

A primeira questão é fácil de resolver: basta subtrair a média da variável. Essa nova variável tem média zero.

Quanto ao desvio padrão, basta lembrar que:

$$dp(ax) = adp(x)$$

Portanto, se o desvio padrão de uma variável aleatória x é σ , o desvio padrão da variável $\frac{x}{\sigma}$ é:

$$dp\left(\frac{x}{\sigma}\right) = \frac{1}{\sigma} dp(x) = \frac{1}{\sigma} \times \sigma = 1$$

Portanto, para que a variável tenha desvio padrão igual a 1, é preciso dividi-la pelo seu desvio padrão.

O processo de transformar uma variável qualquer em uma cuja média seja 0 e cujo desvio padrão seja 1, chamado de **padronização**, consiste em subtrair a média e dividir pelo desvio padrão. Portanto, se uma variável aleatória x possui média μ e desvio padrão σ , a variável z é assim definida:

$$z = \frac{x - \mu}{\sigma}$$

Ela tem média igual a 0 e desvio padrão igual a 1. Se for normalmente distribuída, é possível utilizar os valores da tabela com valores de integrais do final deste livro para calcular as probabilidades.

Exemplo 4.4.1

O faturamento mensal de uma loja segue uma distribuição normal com média R\$ 20.000,00 e desvio padrão R\$ 4.000,00. Calcule a probabilidade de que, em determinado mês, o faturamento esteja entre R\$ 19.000,00 e R\$ 25.000,00.

A variável é normal, mas não padronizada. Você deve, portanto, padronizar os valores antes de utilizar a tabela:

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{19.000 - 20.000}{4.000} = -0,25$$

$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{25.000 - 20.000}{4.000} = 1,25$$

Portanto:

$$P(19.000 < x < 25.000) = P(-0,25 < z < 1,25)$$

Mais um caso em que há um valor acima e outro abaixo de zero.

$$P(19.000 < x < 25.000) = P(-0,25 < z < 0) + P(0 < z < 1,25)$$

$$P(19.000 < x < 25.000) = P(0 < z < 0,25) + P(0 < z < 1,25)$$

$$P(19.000 < x < 25.000) \cong 0,0987 + 0,3944 = 0,4931 = 49,31\%$$

4.5 Transformações de variáveis

Suponha que uma variável aleatória x cuja função densidade seja dada por $f(x)$. Se y é função de x , de modo que $y = u(x)$, qual é a função densidade de probabilidade de y ? Para responder a essa pergunta, parta de um caso simples (em que $u(x)$ é uma função afim), como o mostrado no Exemplo 4.5.1

Exemplo 4.5.1

Para certa variável aleatória x , contínua, é dada a função densidade $f(x)$. Se $y = ax + b$, com a e b positivos, determine a função densidade de probabilidade de y .

Se $f(x)$ é a função densidade de probabilidade de x , então:

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

Como $y = ax + b$, tem-se que:

$$x = \frac{y - b}{a} \quad (4.5.1)$$

Então:

$$\int_{-\infty}^{+\infty} f\left(\frac{y - b}{a}\right) dx = 1$$

Mas a função densidade de y , por exemplo, $g(y)$, deve ser tal que:

$$\int_{-\infty}^{+\infty} g(y)dy = 1$$

Isto é, a função *integrada em relação a y* (e não a x) deve ser igual a 1. Mas, diferenciando a Expressão 4.5.1, chega-se a:

$$dx = \frac{1}{a} dy$$

Substituindo:

$$\int_{-\infty}^{+\infty} f\left(\frac{y - b}{a}\right) \frac{1}{a} dy = 1$$

Portanto, a função

$$g(y) = \frac{1}{a} f\left(\frac{y - b}{a}\right)$$

tem as características de uma função densidade de probabilidade e é, portanto, a f.d.p. da variável y .

Esse resultado é um caso particular de um teorema mais geral, enunciado a seguir.

Teorema 4.5.1

Dada uma variável aleatória x com função densidade de probabilidade dada por $f(x)$, e sendo $y = u(x)$, existindo uma função inversa $x = v(y)$ e $v'(y)$ sua derivada, a função densidade de probabilidade de y é dada por:

$$g(y) = |v'(y)|f(v(y))$$

Nos pontos em que $v(y)$ existir e $u'(x) \neq 0$, e 0 em caso contrário.

A presença do módulo é necessária para garantir a não negatividade da função densidade de probabilidade de y .

A aplicação direta do teorema no exemplo anterior nos levaria a:

$$u(x) = ax + b$$

$$v(y) = \frac{y - b}{a}$$

$$v'(y) = \frac{1}{a}$$

$$g(y) = |v'(y)|f(v(y))$$

$$g(y) = \left| \frac{1}{a} \right| f\left(\frac{y - b}{a} \right)$$

Como a é positivo:

$$g(y) = \frac{1}{a} f\left(\frac{y - b}{a} \right)$$

Exemplo 4.5.2

Dada a variável aleatória x , cuja função densidade de probabilidade é:

$$f(x) = \begin{cases} e^{-x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Supondo $y = x^2$, determine a f.d.p. de y .

Admitindo que $u(x) = x^2$, portanto, $v(y) = \sqrt{y}$, desde que y seja positivo e:

$$v'(y) = \frac{1}{2\sqrt{y}}$$

Aplicando o Teorema 4.5.1, vem:

$$g(y) = \left| \frac{1}{2\sqrt{y}} \right| e^{-\sqrt{y}}$$

Como y tem de ser positivo, assim como \sqrt{y} , a f.d.p. de y será dada por:

$$g(y) = \begin{cases} \frac{1}{2\sqrt{y}} e^{-\sqrt{y}}, & y \geq 0 \\ 0, & y < 0 \end{cases}$$

4.6 Teorema de Tchebichev

Se a função densidade de uma variável é conhecida, pode-se conhecer sua média e variância. A recíproca não é verdadeira, mas é possível estabelecer um limite para uma distribuição de probabilidade qualquer (seja discreta ou contínua), limite esse que é dado pelo teorema de Tchebichev⁴.

Teorema 4.6.1 [teorema de Tchebichev]

Admita uma variável aleatória x com média μ e desvio padrão σ . A probabilidade de essa variável estar acima ou abaixo da média, no *máximo*, k desvios padrão (k é uma constante positiva) é, no mínimo, igual a $1 - \frac{1}{k^2}$. Ou:

$$P(|x - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}$$

Consequentemente, a probabilidade de ultrapassar esse valor é, no máximo, $\frac{1}{k^2}$, isto é:

$$P(|x - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

O que vale dizer que a probabilidade de uma variável aleatória qualquer estar entre dois desvios padrão acima ou abaixo é de, no mínimo⁵, $1 - \frac{1}{4} = \frac{3}{4} = 75\%$.

Exemplo 4.6.1

Uma variável aleatória contínua x tem média 50 e desvio padrão 10. Calcule a probabilidade mínima de que x esteja entre 35 e 65.

Pede-se, portanto:

$$P(35 < x < 50) = ?$$

O que é a probabilidade de x estar 1,5 desvio padrão acima ou abaixo da média, ou seja:

$$P(35 < x < 50) = P(|x - \mu| < 1,5\sigma)$$

Pelo teorema de Tchebichev:

⁴ Em homenagem ao matemático russo Pafnuti Lvocitch Tchebichev (1821-1894).

⁵ Note que, para a distribuição normal, essa probabilidade é de cerca de 95%.

$$P(35 < x < 50) \geq 1 - \frac{1}{1,5^2}$$

$$P(35 < x < 50) \geq 0,5556 = 55,56\%.$$

4.7. Momentos de uma distribuição

O **momento de uma distribuição** (de uma variável aleatória x) de ordem k , em relação à média (M_k), é:

$$M_k = E(x - \mu)^k$$

É imediato que o primeiro momento em relação à média é sempre zero:

$$M_1 = E(x - \mu) = E(x) - \mu = \mu - \mu = 0$$

E o segundo momento é a variância:

$$M_2 = E(x - \mu)^2 = \sigma^2$$

O terceiro momento é:

$$M_3 = E(x - \mu)^3$$

Esse momento tem relação com o grau de simetria da distribuição. Uma distribuição simétrica (como a normal) tem o terceiro momento em relação à média igual a zero. Define-se, inclusive, um coeficiente de assimetria por:

$$a_3 = \frac{M_3}{\sigma^3}$$

Que é tão maior (em módulo) quanto mais assimétrica for a distribuição.

O quarto momento:

$$M_4 = E(x - \mu)^4$$

Tem que ver com a **curtose**, que é o grau de *achatamento* de uma distribuição. Se uma distribuição é muito achatada, ela é dita **platicúrtica**; se é mais para pontiaguda, é chamada **leptocúrtica**. A referência para essa definição é a distribuição normal, dita **mesocúrtica**.

Define-se o coeficiente de curtose como:

$$\alpha_4 = \frac{M_4}{\sigma^4}$$

Cujo valor, para a normal, é três. Se for maior do que três, a distribuição é leptocúrtica; caso contrário, platicúrtica.

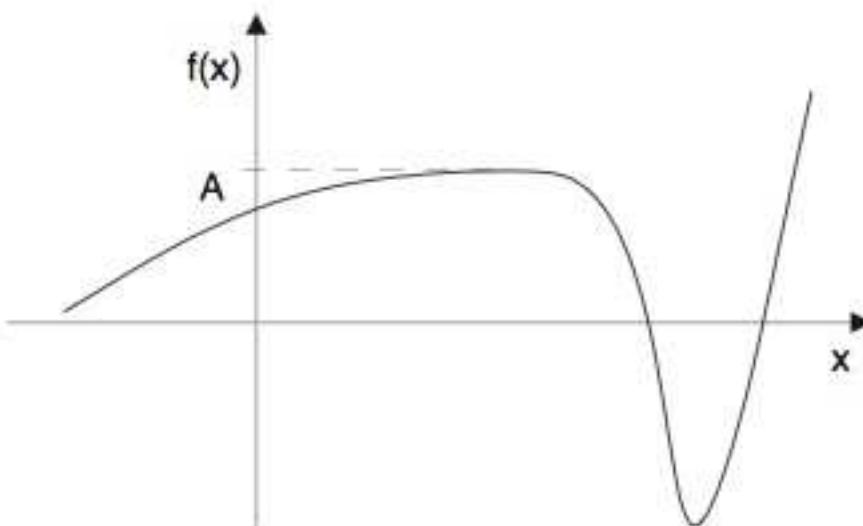
Os momentos podem ser definidos também em relação à origem. O momento de uma distribuição de ordem k em torno da origem é dado por:

$$M'_k = E(x)^k$$

Assim, o momento em torno da origem de ordem 1 é a própria média. O de ordem 2, a média dos quadrados. E assim sucessivamente.

Exercícios

1. É possível encontrar um valor de A para que a função $f(x)$ representada na Figura 4.20 seja uma função densidade de probabilidade? Justifique.

Figura 4.20 - Função $f(x)$ 

2. Determine os valores de A para que as funções a seguir sejam funções densidade de probabilidade:

$$a) f(x) = \begin{cases} 0, & x < 2 \text{ ou } x > 8 \\ A, & 2 \leq x \leq 8 \end{cases}$$

$$b) f(x) = \begin{cases} 0, & x < 0 \text{ ou } x > 4 \\ Ax, & 0 \leq x \leq 4 \end{cases}$$

$$c) f(x) = \begin{cases} 0, & x < 1 \text{ ou } x > 3 \\ Ax, & 1 \leq x \leq 3 \end{cases}$$

$$d) f(x) = \begin{cases} 0, & x < -1 \text{ ou } x > 3 \\ A(x+1), & -1 \leq x \leq 3 \end{cases}$$

$$e) f(x) = \begin{cases} 0, & x < 0 \\ Ae^{-3x}, & x \geq 0 \end{cases}$$

$$f) f(x) = \begin{cases} 0, & x < -2 \text{ ou } x > 2 \\ Ax^2, & -2 \leq x \leq 2 \end{cases}$$

$$g) f(x) = \begin{cases} 0, & x < -2 \text{ ou } x > 0 \\ Ax^3, & -2 \leq x \leq 0 \end{cases}$$

$$h) f(x) = \begin{cases} 0, & x < -1 \text{ ou } x > 1 \\ |Ax|, & -1 \leq x \leq 1 \end{cases}$$

3. Para as variáveis apresentadas no Exercício 2 (itens a e b), determine a média, a variância, o desvio padrão e a mediana.
4. Determine a função densidade de probabilidade de uma variável x que pode assumir qualquer valor no intervalo $[a, b]$ e tem distribuição uniforme.
5. Dada a função densidade de probabilidade a seguir:

$$f(x) = \begin{cases} 0, & x < 1 \text{ ou } x > 9 \\ 1/8, & 1 \leq x \leq 9 \end{cases}$$

Determine as probabilidades de:

- a) $x > 5$;
 - b) $x \leq 6$;
 - c) $x = 4$;
 - d) $0 < x < 7$;
 - e) $2 \leq x < 4$;
 - f) $4 < x \leq 8$.
6. Dada a seguinte f.d.p.:

$$f(x) = \begin{cases} 0, & x < 0 \text{ ou } x > 1 \\ 4x^3, & 0 \leq x \leq 1 \end{cases}$$

Determine as probabilidades de:

- a) $x > 0,5$;
 - b) $x \leq 0,7$;
 - c) $0,2 < x < 0,6$;
 - d) $0,1 \leq x < 0,3$;
 - e) $0,4 < x \leq 1,2$.
7. Dada a função densidade de probabilidade:

$$f(x) = \begin{cases} 0, & x < 0 \\ 2e^{-2x}, & x \geq 0 \end{cases}$$

Determine as probabilidades de:

- a) $x > 1$;
- b) $x \leq -1$;
- c) $2 < x < 5$;
- d) $x < 3$;
- e) $4 < x \leq 10$.

8. Numa normal padronizada, determine a probabilidade de z estar entre:
- 1 desvio padrão acima ou abaixo da média;
 - 2 desvios padrão acima ou abaixo da média;
 - 3 desvios padrão acima ou abaixo da média.
9. Os lucros anuais de uma firma seguem uma distribuição normal com média R\$ 700 mil e desvio padrão R\$ 150 mil. Calcule a probabilidade de, em dado ano, os lucros:
- serem maiores do que R\$ 800 mil;
 - serem maiores do que R\$ 600 mil;
 - serem menores do que R\$ 900 mil;
 - serem menores do que R\$ 650 mil;
 - estarem entre R\$ 550 mil e R\$ 770 mil;
 - estarem entre R\$ 350 mil e R\$ 500 mil;
 - estarem entre R\$ 720 mil e R\$ 850 mil.
10. As notas bimestrais de um aluno seguem uma distribuição normal com média 5 e variância 4,84. Calcule a probabilidade de, num dado bimestre, sua nota:
- ser maior do que 8;
 - ser maior do que 4,5;
 - ser menor do que 9;
 - ser menor do que 4;
 - estar entre 3,5 e 6,5;
 - estar entre 2,5 e 4,5;
 - estar entre 6 e 8,5.
11. As notas bimestrais de um aluno são, em média, quatro e têm variância 2,56, mas a distribuição não é conhecida. Determine um limite para a probabilidade de, em dado bimestre, sua nota:
- estar entre 1,5 e 6,5;
 - estar entre 2 e 6;
 - ser menor do que 1 ou maior do que 7.

12. Uma variável aleatória x , $x \geq 0$, tem f.d.p. dada por $f(x)$. Se $y = \sqrt{y}$, determine a f.d.p. de y .

13. Se $y = \frac{1}{x}$ e x é uma variável aleatória contínua cuja f.d.p. é dada por:

$$f(x) = \begin{cases} 3x^2, & 0 < x \leq 1 \\ 0, & x \leq 0 \text{ ou } x > 1 \end{cases}$$

Determine a f.d.p. de y .

14. Determine a média e a variância de uma variável aleatória x cuja f.d.p. é dada por:

$$f(x) = \begin{cases} ae^{-ax}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

15. Dada uma variável aleatória contínua x cuja média é 20 e a variância é 25, determine limites para as seguintes probabilidades:

- a) $P(10 < x < 30)$;
- b) $P(14 < x < 26)$;
- c) $P(x < 12,5 \text{ ou } x > 27,5)$.

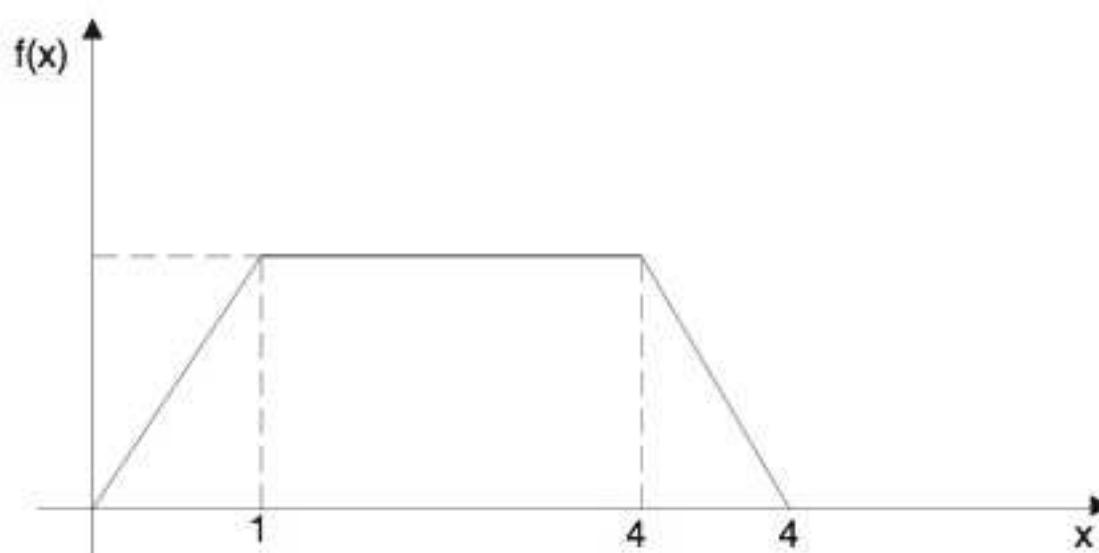
16. Mostre que, para uma variável aleatória com média μ e variância σ^2 , é válida a expressão:

$$P(|x - \mu| < k) \geq 1 - \frac{\sigma^2}{k^2}$$

17. Assinale verdadeiro ou falso:

- a) Uma f.d.p. nunca pode assumir valores maiores do que 1.
- b) Dada a f.d.p. de uma variável aleatória, é possível encontrar a média e o desvio padrão dessa variável.
- c) Dados a média e o desvio padrão de uma variável aleatória, é possível calcular exatamente a probabilidade dessa variável.
- d) Dados a média e o desvio padrão de uma variável aleatória, é possível estabelecer limites para a probabilidade dessa variável.
- e) Se x é uma variável aleatória com f.d.p. contínua $f(x)$, a probabilidade de x ser igual a um certo valor a é igual a $f(a)$.
- f) A distribuição normal é perfeitamente definida pela média e pela variância.
- g) A distribuição normal é perfeitamente definida pela média e pelo desvio padrão.

18. Determine a função densidade, a função distribuição, a média e a variância da variável aleatória representada no gráfico a seguir:



Apêndice 4.A – Cálculo diferencial e integral

4.A.1 Cálculo diferencial e integral

4.A.1.1 Derivadas

Derivada é a variação instantânea. Se você percorre, com seu carro, 100km em 1h, sua velocidade média é de 100km/h. É pouco provável, entretanto, que durante todo o percurso a velocidade tenha sido constante. A velocidade que marca o velocímetro (ou o radar) é a velocidade do carro em determinado instante, ou seja, sua variação instantânea.

A definição formal é a seguinte:

$$\frac{dy}{dx} = \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x}$$

Em que $\Delta y/\Delta x$ é a taxa de variação média (a velocidade média, por exemplo). Se uma variação de x for muito pequena, a taxa de variação média tende a coincidir com a taxa de variação instantânea (a derivada).

Os termos dy e dx (diferenciais de y e x) indicam que se trata de uma variação (diferença) infinitamente pequena dessas variáveis, em contraste com os símbolos Δy e Δx , que representam a diferença (variação) finita.

Se se usa a notação $y = f(x)$, a derivada também pode ser escrita como $f'(x)$.

4.A.1.2 Regras de derivação

A partir da definição formal, é possível calcular a derivada de qualquer função, se ela existir. Entretanto, normalmente se usam algumas regras gerais, mostradas no Quadro 4.1.

Quadro 4.1 – Regras de derivação

$f(x)$	$f'(x)$
a (constante)	0
x	1
x^2	$2x$
x^n	nx^{n-1}
e^x	e^x
$\ln x$	$1/x$
$\operatorname{sen} x$	$\cos x$
$\cos x$	$-\operatorname{sen} x$
$ag(x)$	$ag'(x)$
$g(x) + h(x)$	$g'(x) + h'(x)$
$g(x).h(x)$	$g'(x).h(x) + g(x).h'(x)$
$g(x)/h(x)$	$[g'(x).h(x) - g(x).h'(x)]/[h(x)]^2$
$g(h(x))$	$h'(x).g'(h(x))$

4.A.1.3 Integral

A **integral** de uma função é o limite de uma soma.

$$\int_a^b f(x)dx = \lim_{n \rightarrow \infty} \sum_{i=1}^n f(x_i)\Delta x_i$$

Daí sua utilidade em cálculos de áreas, por exemplo. É como se aproximássemos a curva em questão por meio de um conjunto de retângulos e calculássemos a área desses retângulos. Quanto maior o número de retângulos, e, portanto, menor sua área, a soma dessas áreas estaria mais próxima da área da figura.

Demonstra-se essa afirmação por meio do teorema do valor médio:

$$\int_a^b f(x)dx = F(b) - F(a)$$

Em que $F(x)$ é chamada de **primitiva** de $f(x)$, isto é, é a função cuja derivada é $f(x)$, ou seja: $F'(x) = f(x)$

No Quadro 4.2, você encontra algumas primitivas:

Quadro 4.2 – Primitivas

$f(x)$	$F(x)$
a	ax
x	$x^2/2$
x^n	$x^{n+1}/(n+1)$, com $(n \neq -1)$
$1/x$	$\ln x$
e^x	e^x
e^{-x}	$-e^{-x}$
x^{-x}	$-xe^{-x} - e^{-x}$
x^2e^{-x}	$-e^{-x}(x^2 + 2x + 2)$

4.A.1.4 Máximos e mínimos

É possível encontrar os **máximos e mínimos** de uma função resolvendo a seguinte equação:

$$f'(x) = 0$$

Isto é, derivando e igualando a zero.

Para saber se é ponto de máximo, você substitui o(s) valor(es) encontrado(s) em $f'(x) = 0$, por exemplo, x_0 , na derivada segunda (condição de segunda ordem), em que valem as seguintes regras:

$$f''(x_0) > 0 \Rightarrow \text{ponto de mínimo}$$

$$f''(x_0) < 0 \Rightarrow \text{ponto de máximo}$$

Apêndice 4.B – Demonstração dos teoremas e momentos de uma distribuição

4.B.1 Demonstração do Teorema 4.5.1

Considere os seguintes dois casos: (1) $u(x)$ é uma função crescente (assim, sua derivada é positiva); (2) $u(x)$ é uma função decrescente (com derivada negativa).

Relembrando que $y = u(x)$, cuja função inversa é dada por $x = v(y)$.

Para o caso de $u(x)$ crescente, tomando duas constantes a e b quaisquer, tem-se:

$$\begin{aligned} P(a < y < b) &= P[v(a) < x < v(b)] \\ P(a < y < b) &= \int_{v(a)}^{v(b)} f(x) dx \end{aligned}$$

Como $f(x) = f(v(y))$ e $dx = v'(y)dy$ e:

se $x = v(a)$, então $y = a$

se $x = v(b)$, então $y = b$

Substituindo, chega-se a:

$$P(a < y < b) = \int_a^b f(v(y))v'(y)dy$$

Portanto, a f.d.p. de y , nesse caso, é:

$$g(y) = v'(y)f(v(y))$$

Para $u(x)$ decrescente, há de se fazer uma inversão:

$$\begin{aligned} P(a < y < b) &= P[v(b) < x < v(a)] \\ P(a < y < b) &= \int_{v(a)}^{v(b)} f(x) dx \end{aligned}$$

De novo, substituindo, tem-se:

$$P(a < y < b) = \int_b^a f(v(y))v'(y)dy$$

O que é equivalente a:

$$P(a < y < b) = - \int_a^b f(v(y))v'(y)dy$$

Assim, agora a f.d.p. de y é:

$$g(y) = -v'(y)f(v(y))$$

Ou seja, $v'(y)$, quando é negativo, fica com o sinal de menos à frente, de modo a torná-lo positivo, o que equivale a calcular seu módulo.

Então, vale a regra geral:

$$g(y) = |v'(y)|f(v(y))$$

4.B.2 Demonstração do Teorema de Tchebichev

Vamos nos limitar aqui ao caso de distribuições contínuas.

Sabe-se que:

$$\sigma^2 = \text{var}(x) = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx$$

Dividindo essa integral em três partes, chega-se a:

$$\sigma^2 = \int_{-\infty}^{\mu-k\sigma} (x - \mu)^2 f(x) dx + \int_{\mu-k\sigma}^{\mu+k\sigma} (x - \mu)^2 f(x) dx + \int_{\mu+k\sigma}^{+\infty} (x - \mu)^2 f(x) dx$$

Como todos os três termos são não negativos, já que $f(x)$ é não negativa, e $(x - \mu)$ está elevada ao quadrado, se a integral for retirada do meio:

$$\sigma^2 \geq \int_{-\infty}^{\mu-k\sigma} (x - \mu)^2 f(x) dx + \int_{\mu+k\sigma}^{+\infty} (x - \mu)^2 f(x) dx$$

Agora há x em dois intervalos: um, em que $x \leq \mu - k\sigma$; e o outro, em que $x \geq \mu + k\sigma$. Em ambos os casos, chega-se a $(x - \mu)^2 \geq k^2 \sigma^2$. Portanto, é válido que:

$$\sigma^2 \geq \int_{-\infty}^{\mu-k\sigma} k^2 \sigma^2 f(x) dx + \int_{\mu+k\sigma}^{+\infty} k^2 \sigma^2 f(x) dx$$

Dividindo por $k^2 \sigma^2$ ambos os lados:

$$\frac{1}{k^2} \geq \int_{-\infty}^{\mu-k\sigma} f(x) dx + \int_{\mu+k\sigma}^{+\infty} f(x) dx$$

Sabe-se que:

$$\int_{-\infty}^{\mu-k\sigma} f(x) dx = P(x \leq \mu - k\sigma) = P(x - \mu \leq -k\sigma)$$

$$\int_{\mu+k\sigma}^{+\infty} f(x) dx = P(x \geq \mu + k\sigma) = P(x - \mu \geq k\sigma)$$

Substituindo:

$$\frac{1}{k^2} \geq P(x - \mu \leq -k\sigma) + P(x - \mu \geq k\sigma)$$

O que equivale a:

$$P(|x - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

Cujo complementar é:

$$P(|x - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}$$

4.B.3 Distribuição log-normal

Se x é uma variável cuja distribuição é normal com média μ e desvio padrão σ , e seja y definida como $y = e^x$ (ou seja, $x = \ln y$), y segue uma distribuição conhecida como log-normal.

Aplicando o Teorema 4.5.1:

$$u(x) = e^x$$

$$v(y) = \ln y$$

$$v'(y) = \frac{1}{y}$$

A f.d.p. de uma variável normal é:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

A f.d.p. da variável log-normal (y) é, então:

$$g(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln y - \mu)^2}{2\sigma^2}}$$

Cuja média é $e^{\mu + \frac{\sigma^2}{2}}$ e cuja variância é $e^{2\mu}(e^{2\sigma^2} - e^{\sigma^2})$.

4.B.4 Momentos da distribuição normal

A média e a variância da distribuição normal são parâmetros da própria distribuição.

Dada sua f.d.p., $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, é imediato que os dois primeiros momentos em torno da média são dados por:

$$M_1 = E(x - \mu) = E(x) - \mu = \mu - \mu = 0$$

$$M_2 = E(x - \mu)^2 = \sigma^2.$$

No caso do primeiro momento, aliás, como em qualquer outra distribuição. O primeiro momento em torno da origem é, evidentemente, a média μ .

O terceiro momento em torno da origem é:

$$M_3 = E(x - \mu)^3 = \int_{-\infty}^{+\infty} (x - \mu)^3 f(x) dx$$

$$M_3 = \int_{-\infty}^{+\infty} (x - \mu)^3 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

$$M_3 = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} (x - \mu)^3 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

$$M_3 = \frac{1}{\sqrt{2\pi\sigma^2}} \left[\int_{-\infty}^{\mu} (x - \mu)^3 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx + \int_{\mu}^{+\infty} (x - \mu)^3 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \right]$$

Note que a função cúbica tem a seguinte propriedade: $f(-x) = -f(x)$. Como, por outro lado, a função e^{-x^2} é sempre positiva, a primeira integral da lista anterior tem o mesmo valor da segunda, apenas com o sinal trocado. Então:

$$M_3 = \frac{1}{\sqrt{2\pi\sigma^2}} \left[- \int_{\mu}^{+\infty} (x - \mu)^3 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx + \int_{\mu}^{+\infty} (x - \mu)^3 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \right]$$

$$M_3 = 0.$$

O terceiro momento em torno da média da distribuição normal é zero, uma vez que se trata de uma distribuição simétrica.

E o quarto momento é dado por:

$$M_4 = E(x - \mu)^4 = \int_{-\infty}^{+\infty} (x - \mu)^4 f(x) dx$$

$$M_4 = \int_{-\infty}^{+\infty} (x - \mu)^4 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

$$M_4 = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} (x - \mu)^4 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

A derivada de $e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ é dada por $-2 \frac{(x-\mu)}{2\sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = -\frac{(x-\mu)}{\sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. Multipli- cando e dividindo a expressão anterior por $-\sigma^5$, chega-se a:

$$M_4 = -\frac{\sigma^5}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} \frac{(x-\mu)^3}{\sigma^3} \left[-\frac{(x-\mu)}{\sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \right] dx$$

Assim, fazendo $f = \frac{(x-\mu)^3}{\sigma^3}$ e $g' = -\frac{(x-\mu)}{\sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ (portanto, $g = e^{-\frac{(x-\mu)^2}{2\sigma^2}}$), pode-se resolver essa integral por partes:

$$M_4 = -\frac{\sigma^5}{\sqrt{2\pi\sigma^2}} \left[\frac{(x-\mu)^3}{\sigma^3} e^{-\frac{(x-\mu)^2}{2\sigma^2}} - \int_{-\infty}^{+\infty} 3 \frac{(x-\mu)^2}{\sigma^3} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \right]$$

$$M_4 = -\frac{\sigma^5}{\sigma^3 \sqrt{2\pi\sigma^2}} \left[(x-\mu)^3 e^{-\frac{(x-\mu)^2}{2\sigma^2}} - 3 \int_{-\infty}^{+\infty} (x-\mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \right]$$

$$M_4 = -\frac{\sigma^2}{\sqrt{2\pi\sigma^2}} \left[(x-\mu)^3 e^{-\frac{(x-\mu)^2}{2\sigma^2}} - 3 \int_{-\infty}^{+\infty} (x-\mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \right]$$

$$M_4 = -\frac{\sigma^2}{\sqrt{2\pi\sigma^2}} \left\{ \left[(x-\mu)^3 e^{-\frac{(x-\mu)^2}{2\sigma^2}} \right]_{-\infty}^{+\infty} - 3 \int_{-\infty}^{+\infty} (x-\mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \right\}$$

É fácil ver que a expressão em colchetes se anula, pois o limite de $e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ se anula tanto para x tendendo a mais infinito como a menos infinito. Arranjando a integral restante:

$$M_4 = 3\sigma^2 \int_{-\infty}^{+\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

A expressão calculada pela integral é a própria variância da distribuição normal, que é igual a σ^2 . Então:

$$M_4 = 3\sigma^2 \times \sigma^2 = 3\sigma^4.$$

Conclui-se que o coeficiente de curtose de uma distribuição normal é 3.



Capítulo 5

DISTRIBUIÇÃO DE PROBABILIDADE CONJUNTA

Conjunta é a probabilidade que se refere a duas (ou mais) variáveis aleatórias simultaneamente ou é a distribuição de probabilidade de um **vetor aleatório** (X, Y) ¹ com duas variáveis, isto é, para o caso bidimensional.

Essas variáveis podem ser discretas ou contínuas.

5.1 Distribuição conjunta de variáveis discretas

Imagine um time de vôlei que vai disputar um campeonato muito equilibrado, em que a probabilidade de ganhar ou perder uma partida seja 0,5. O técnico pede ao estatístico da equipe que faça uma análise das probabilidades das três primeiras partidas, consideradas vitais para o restante das competições. Em particular, a vitória na primeira partida é considerada decisiva pela comissão técnica.

O estatístico, então, define duas variáveis, X e Y , dessa forma: X é o número de vitórias obtidas nos três primeiros jogos, e Y é igual a 1, caso ocorra vitória no primeiro jogo, ou a 0, caso ocorra o contrário. Por ora, pense se X e Y são variáveis independentes. Mais adiante, neste capítulo, você verá a resposta.

Há oito possíveis resultados nas três primeiras partidas ($2 \times 2 \times 2$, duas em cada partida), todos com a mesma probabilidade (já que a probabilidade de vitória em cada jogo é 0,5). Os possíveis resultados – e os correspondentes valores de X e Y – são mostrados na Tabela 5.1.

Tabela 5.1 – Resultados possíveis em três partidas

Resultados possíveis	X	Y
VVV	3	1
VVD	2	1
VDV	2	1
VDD	1	1

¹ Quando X e Y forem variáveis aleatórias, o vetor (X, Y) é chamado de vetor aleatório.

Resultados possíveis	X	Y
DVV	2	0
DDV	1	0
DVD	1	0
DDD	0	0

Nesse esquema, V representa vitória e D, derrota. O resultado VDV, por exemplo, representa vitória no primeiro jogo, derrota no segundo e vitória no terceiro.

A seguir, o estatístico constrói uma tabela que apresenta as probabilidades conjuntas de X e Y . O preenchimento dessa tabela é feito por meio da Tabela 5.1. Assim, na posição da tabela que corresponde a $X = 2$ e $Y = 1$, deve-se colocar a probabilidade de isso ocorrer, isto é, $P(X = 2 \text{ e } Y = 1)$. Pela Tabela 5.1, verifica-se que, em oito resultados possíveis, existem dois em que há duas vitórias ($X = 2$) e há vitória no primeiro jogo ($Y = 1$). Portanto, $P(X = 2 \text{ e } Y = 1) = \frac{2}{8}$. E assim procedendo, constrói-se a Tabela 5.2.

		Tabela 5.2 – Probabilidades conjuntas de X e Y			
		X			
		0	1	2	3
Y	0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	0
	1	0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$

Com a Tabela 5.2 pronta, torna-se desnecessário utilizar a Tabela 5.1 para obter as probabilidades conjuntas. Assim, diretamente pela Tabela 5.2, é possível saber, por exemplo:

$$P(X = 1 \text{ e } Y = 1) = \frac{2}{8}$$

$$P(X = 2 \text{ e } Y = 0) = \frac{1}{8}$$

$$P(X = 3 \text{ e } Y = 0) = 0$$

Da Tabela 5.2 pode-se obter também as distribuições de probabilidade só de X e só de Y . Como? A probabilidade de X ser igual a 1, independentemente do valor de Y , é a probabilidade de $X = 1$ e $Y = 0$ ou $X = 1$ e $Y = 1$. Lembre-se de que $Y = 0$ e $Y = 1$ são eventos mutuamente exclusivos, portanto, vale a regra $P(A \text{ ou } B) = P(A) + P(B)$. Portanto:

$$P(X = 1) = P[(X = 1 \text{ e } Y = 0) \text{ ou } (X = 1 \text{ e } Y = 1)] =$$

$$\frac{2}{8} + \frac{1}{8} = \frac{3}{8}$$

Como se nota, a probabilidade de X (só de X , sem considerar o que ocorre com Y) é dada pela soma das probabilidades ao longo da coluna, ou seja, somando-se as probabilidades de todos os valores possíveis de Y .

Então, a Tabela 5.3, além da distribuição conjunta de X e Y , mostra também a **distribuição marginal²** de X , a distribuição só de X , representada por $P(X)$:

		X			
		0	1	2	3
Y					
0		$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	0
1		0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$
P(X)		$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

A distribuição de probabilidade só de Y é obtida da mesma forma, ou seja, somando-se as probabilidades ao longo da linha, isto é, somam-se todos os valores possíveis de X . Por exemplo, a probabilidade de Y ser igual a 0 é dada por:

$$P(Y=0) = P(Y=0 \text{ e } X=0) + P(Y=0 \text{ e } X=1) + P(Y=0 \text{ e } X=2) + P(Y=0 \text{ e } X=3)$$

$$P(Y=0) = \frac{1}{8} + \frac{2}{8} + \frac{1}{8} + 0 = \frac{4}{8} = \frac{1}{2}$$

Fazendo o mesmo para Y igual a 1, obtém-se a distribuição marginal de Y , representada por $P(Y)$ na Tabela 5.4.

		X				
		0	1	2	3	P(Y)
Y						
0		$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	0	$\frac{1}{2}$
1		0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	$\frac{1}{2}$
P(X)		$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	1

O número 1 colocado no canto inferior direito da tabela representa a soma das probabilidades marginais (e da conjunta também), que tem de ser igual a 1.

² Tal distribuição chama-se marginal – à margem – porque foi obtida de uma distribuição conjunta.

Repare que as probabilidades marginais de X e Y obtidas pela soma das probabilidades conjuntas são as mesmas (nem poderia ser diferente) que seriam obtidas diretamente da Tabela 5.1. Por exemplo, dos oito resultados possíveis, há três em que X é igual a 1, portanto, $P(X = 1) = \frac{3}{8}$; e há quatro em que Y é igual a 0, portanto, $P(Y = 0) = \frac{4}{8} = \frac{1}{2}$.

É possível utilizar a Tabela 5.4 para calcular as probabilidades condicionais, embora elas não possam ser obtidas diretamente dessa fonte. Suponha que você queira saber qual a probabilidade de X ser igual a 1, dado que Y é 1 (isto é, se acontecer uma vitória no primeiro jogo, qual a probabilidade de que só aconteça uma vitória nos três jogos?).

Pela definição de probabilidade condicional:

$$P(X = 1|Y = 1) = \frac{P(X = 1 \text{ e } Y = 1)}{P(Y = 1)}$$

Da Tabela 5.4:

$$P(X = 1|Y = 1) = \frac{\frac{1}{8}}{\frac{1}{2}} = \frac{1}{4}$$

Esse resultado também é compatível com as informações da Tabela 5.1, pois, se Y já é 1, só há quatro resultados possíveis, dos quais em apenas um deles X é igual a 1.

Da mesma forma, pode-se calcular a probabilidade de Y ser igual a 0, dado que X é igual a 2 (isto é, se duas vitórias ocorreram, a probabilidade de que o primeiro jogo tenha sido derrota).

$$P(Y = 0|X = 2) = \frac{P(Y = 0 \text{ e } X = 2)}{P(X = 2)} = \frac{\frac{1}{8}}{\frac{3}{8}} = \frac{1}{3}$$

Ou seja, se ocorreram duas vitórias, os resultados possíveis se reduzem a três. Desses, em apenas um no primeiro jogo ocorre uma derrota.

Voltando à pergunta formulada no início do capítulo: X e Y são independentes? Quando se sabe o que representam X e Y , a resposta é simples: se no primeiro jogo o time for derrotado, é impossível que haja vitória em três jogos (portanto, se Y é igual a 0, é impossível que X seja 3); da mesma forma, se Y é igual a 1, é impossível que X seja 0. Portanto, X e Y não são independentes. Isso, no entanto, poderia ser verificado mesmo que não houvesse outra informação além da Tabela 5.4, uma vez que, por exemplo:

$$P(X = 1|Y = 1) = \frac{1}{4} \text{ e } P(X = 1) = \frac{3}{8}$$

Portanto:

$$P(X = 1|Y = 1) \neq P(X = 1)$$

Pela definição de dependência dada no Capítulo 1, X e Y são dependentes, já que não vale a igualdade entre a probabilidade condicional e a incondicional.

Para mostrar que as variáveis não são independentes, basta encontrar uma situação em que a igualdade não vale. Para o contrário, no entanto, é necessário que a igualdade valha para todos os valores de X e Y , pois é possível que, para um par de valores particulares de X e Y , valha, por coincidência, a igualdade, ainda que X e Y não sejam independentes.

Exemplo 5.1.1

Calcule o valor esperado e a variância das variáveis aleatórias X e Y definidas no texto, bem como a covariância e o coeficiente de correlação entre elas.

As distribuições conjunta e marginal de X e Y foram apresentadas na Tabela 5.4.

Para calcular $E(X)$ e $\text{var}(X)$, usam-se as probabilidades dadas pela distribuição marginal de X , que pode assumir os valores 0, 1, 2 e 3:

$$E(X) = 0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 2 \times \frac{3}{8} + 3 \times \frac{1}{8} = \frac{12}{8} = 1,5$$

$$E(X^2) = 0^2 \times \frac{1}{8} + 1^2 \times \frac{3}{8} + 2^2 \times \frac{3}{8} + 3^2 \times \frac{1}{8} =$$

$$0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 4 \times \frac{3}{8} + 9 \times \frac{1}{8} = \frac{24}{8} = 3$$

$$\text{var}(X) = E(X^2) - [E(X)]^2 = 3 - 1,5^2 = 3 - 2,25 = 0,75$$

Para Y , vale o mesmo raciocínio:

$$E(Y) = 0 \times \frac{1}{2} + 1 \times \frac{1}{2} = 0,5$$

$$E(Y^2) = 0^2 \times \frac{1}{2} + 1^2 \times \frac{1}{2} = 0 \times \frac{1}{2} + 1 \times \frac{1}{2} = 0,5$$

$$\text{var}(Y) = E(Y^2) - [E(Y)]^2 = 0,5 - 0,5^2 = 0,5 - 0,25 = 0,25$$

Para calcular a covariância de X e Y , você pode utilizar a expressão:

$$\text{cov}(X, Y) = E(XY) - E(X)E(Y)$$

Como você já conhece as esperanças de X e Y , calcule a esperança dos produtos. Os produtos são mostrados na Tabela 5.5:

Tabela 5.5 – Produto de X e Y

X	Y	XY
3	1	3
2	1	2
2	1	2
1	1	1

X	Y	XY
2	0	0
1	0	0
1	0	0
0	0	0

Por essa tabela:

$$P(XY = 0) = \frac{4}{8}$$

$$P(XY = 1) = \frac{1}{8}$$

$$P(XY = 2) = \frac{2}{8}$$

$$P(XY = 3) = \frac{1}{8}$$

Portanto, a esperança dos produtos é dada por:

$$E(XY) = 0 \times \frac{1}{8} + 1 \times \frac{1}{8} + 2 \times \frac{2}{8} + 3 \times \frac{1}{8} = \frac{8}{8} = 1$$

E a covariância:

$$\text{cov}(X, Y) = E(XY) - E(X)E(Y) = 1 - 1,5 \times 0,5 = 1 - 0,75 = 0,25$$

E o coeficiente de correlação:

$$\rho_{XY} = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\text{var}(Y)}} = \frac{0,25}{\sqrt{0,75 \times 0,25}} \cong 0,5774$$

Exemplo 5.1.2

Dadas as variáveis aleatórias X e Y definidas no texto, determine $E(X|Y = 0)$.

Para calcular a esperança condicionada, você precisa das probabilidades condicionais para todos os valores de X :

$$P(X = 0|Y = 0) = \frac{1}{4}$$

$$P(X = 1|Y = 0) = \frac{1}{2}$$

$$P(X = 2|Y = 0) = \frac{1}{4}$$

$$P(X = 3|Y = 0) = 0$$

Portanto:

$$E(X|Y = 0) = 0 \times \frac{1}{4} + 1 \times \frac{1}{2} + 2 \times \frac{1}{4} + 3 \times 0 = 1$$

Exemplo 5.1.3

Com as variáveis aleatórias X e Y definidas no texto, determine $E[E(X|Y)]$.

Você já calculou $E(X|Y = 0)$. De maneira análoga, calcule agora $E(X|Y = 1)$:

$$E(X|Y = 1) = 0 \times 0 + 1 \times \frac{1}{4} + 2 \times \frac{1}{2} + 3 \times \frac{1}{4} = 2$$

O valor de $E[E(X|Y)]$ é dado pela média da esperança de X , dado Y pelos diversos valores de Y (0 e 1 neste caso):

$$E[E(X|Y)] = E(X|Y = 0) \times P(Y = 0) + E(X|Y = 1) \times P(Y = 1)$$

$$E[E(X|Y)] = 1 \times \frac{1}{2} + 2 \times \frac{1}{2} = 1,5$$

Note que:

$$E[E(X|Y)] = E(X).$$

Essa fórmula é conhecida como Lei das Expectativas Iteradas.

Exemplo 5.1.4

Dadas as variáveis aleatórias X e Y definidas no texto, determine $\text{var}(Y|X = 1)$.

De novo, são necessárias as probabilidades condicionais:

$$P(Y = 0|X = 1) = \frac{2}{3}$$

$$P(Y = 1|X = 1) = \frac{1}{3}$$

Então:

$$E(Y|X = 1) = 0 \times \frac{2}{3} + 1 \times \frac{1}{3} = \frac{1}{3}$$

$$E(Y^2|X = 1) = 0^2 \times \frac{2}{3} + 1^2 \times \frac{1}{3} = 0 \times \frac{2}{3} + 1 \times \frac{1}{3} = \frac{1}{3}$$

$$\text{var}(Y|X = 1) = E(Y^2|X = 1) - [E(Y|X = 1)]^2 = \frac{1}{3} - \left(\frac{1}{3}\right)^2 =$$

$$\frac{1}{3} - \frac{1}{9} = \frac{2}{9} = 0,222\dots$$

Exemplo 5.1.5

Dadas as variáveis aleatórias X e Y definidas no texto, determine $\text{var}(X|Y = 0)$ e $\text{var}(X|Y = 1)$.

É preciso todo o conjunto de probabilidades condicionais, cujos valores são (verifique!):

$$P(X = 0|Y = 0) = \frac{1}{4}; P(X = 1|Y = 0) = \frac{1}{2}; P(X = 2|Y = 0) = \frac{1}{4}; P(X = 3|Y = 0) = 0.$$

$$P(X=0|Y=1)=0; P(X=1|Y=1)=\frac{1}{4}; P(X=2|Y=1)=\frac{1}{2}; P(X=3|Y=1)=\frac{1}{4}.$$

Você já viu em exemplos anteriores que $E(X|Y=0) = 1$ e $E(X|Y=1) = 2$. As médias dos quadrados são:

$$E(X^2|Y=0) = 0^2 \times \frac{1}{4} + 1^2 \times \frac{1}{2} + 2^2 \times \frac{1}{4} + 3^2 \times 0 = \frac{6}{4} = 1,5.$$

$$E(X^2|Y=1) = 0^2 \times 0 + 1^2 \times \frac{1}{4} + 2^2 \times \frac{1}{2} + 3^2 \times \frac{1}{4} = \frac{18}{4} = 4,5$$

As variâncias são, portanto:

$$\text{var}(X|Y=0) = E(X^2|Y=0) - [E(X|Y=0)]^2 = 1,5 - 1^2 = 0,5.$$

$$\text{var}(X|Y=1) = E(X^2|Y=1) - [E(X|Y=1)]^2 = 4,5 - 2^2 = 0,5.$$

Repare que a variância condicional é sempre (ou quase sempre) menor do que a variância condicional, isto é, $\text{var}(X|Y) \leq \text{var}(X)$ ou $\text{var}(Y|X) \leq \text{var}(Y)$, enquanto a esperança condicional pode ser maior ou menor (ou igual) à esperança incondicional. Por que isso acontece?

Tome o exemplo do início deste capítulo. Se é garantido que o primeiro jogo é vitória (ou seja, é dado que $Y=1$), o número de vitórias esperado sobe. Do mesmo modo, se o primeiro jogo é, com certeza, derrota, o número de vitórias esperado cai.

Já a variância – condicional ou não – é uma medida de dispersão. O fato de o resultado do primeiro jogo ser dado, seja ele qual for, diminui o número de resultados possíveis, daí a variância condicional ser (quase) sempre menor.

Quase sempre porque pode ser igual. Se as variáveis X e Y forem independentes, isso significa que $P(X|Y) = P(X)$ e, portanto $E(X|Y) = E(X)$ e $\text{var}(X|Y) = \text{var}(X)$.

Exemplo 5.1.6

Usando resultados dos exemplos anteriores, calcule $E[\text{var}(X|Y)]$ e $\text{var}[E(X|Y)]$.

$$E[\text{var}(X|Y)] = \text{var}(X|Y=0) \times P(Y=0) + \text{var}(X|Y=1) \times P(Y=1)$$

$$E[\text{var}(X|Y)] = 0,5 \times 0,5 + 0,5 \times 0,5 = 0,5.$$

$$E[E(X|Y)]^2 = [E(X|Y=0)]^2 \times P(Y=0) + [E(X|Y=1)]^2 \times P(Y=1)$$

$$E[E(X^2|Y)] = 1^2 \times 0,5 + 2^2 \times 0,5 = 2,5.$$

$$\text{var}[E(X|Y)] = E[E(X^2|Y)] - [E[E(X|Y)]]^2 = 2,5 - 1,5^2 = 0,25.$$

Note que $\text{var}[E(X|Y)] + E[\text{var}(X|Y)] = \text{var}(X)$, isso vale sempre e é conhecido como **decomposição da variância**.

Exemplo 5.1.7

Para casais de dois filhos, definem-se duas variáveis, W e Z . W é o sexo do primeiro filho, sendo 0 para masculino e 1 para feminino. Z é igual a 1 se as duas crianças são

do mesmo sexo, 0 se formam um casal. Construa uma tabela com as distribuições conjunta e marginal de W e Z e determine se são variáveis independentes. Para um casal com dois filhos, há quatro possibilidades. Representando os meninos por H e as meninas por M:

Tabela 5.6 – Possibilidades de meninos ou meninas em casais de dois filhos

Possibilidades	W	Z
HH	0	1
HM	0	0
MM	1	1
MH	1	0

As probabilidades para cada possibilidade são mostradas na Tabela 5.7:

Tabela 5.7 – Probabilidades de meninos ou meninas em casais de dois filhos

	Z		
W	0	1	P(W)
0	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{2}$
1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
P(Z)	$\frac{1}{2}$	$\frac{1}{2}$	1

Note que, para quaisquer valores de Z ou W :

$$P(Z = Z_0 | W = W_0) = P(Z = Z_0) \text{ e}$$

$$P(W = W_0 | Z = Z_0) = P(W = W_0)$$

Por exemplo:

$$P(Z = 1 | W = 1) = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{2}{4} = \frac{1}{2} \text{ e}$$

$$P(Z = 1) = \frac{1}{2}$$

Portanto, Z e W são independentes, o que é lógico, pois o fato de os dois filhos serem ou não do mesmo sexo independe do sexo do primeiro filho.

Exemplo 5.1.8

A Tabela 5.8 mostra a distribuição conjunta das variáveis aleatórias discretas U e V . Encontre suas distribuições marginais, verifique se U e V são independentes e calcule a covariância das duas variáveis.

		U		
		0	1	2
V		0		
		$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$
-1	0	$\frac{1}{8}$	0	$\frac{1}{8}$
1	1	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

As distribuições marginais de U e V são dadas pela soma ao longo das linhas (a de V) e ao longo das colunas (a de U). A Tabela 5.9 mostra também as distribuições marginais.

		U			P(V)
		0	1	2	
V		0			P(V)
		$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	
-1	0	$\frac{1}{8}$	0	$\frac{1}{8}$	$\frac{2}{8}$
1	1	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{8}$
P(U)		$\frac{3}{8}$	$\frac{2}{8}$	$\frac{3}{8}$	1

Observe que:

$$P(U = 1|V = 0) = 0 \quad \text{e}$$

$$P(U = 1) = \frac{2}{8}$$

Portanto:

$$P(U = 1|V = 0) \neq P(U = 1)$$

Então, U e V não são independentes.

Os valores esperados de U e V são:

$$E(U) = \frac{3}{8} \times 0 + \frac{2}{8} \times 1 + \frac{3}{8} \times 2 = \frac{8}{8} = 1$$

$$E(V) = \frac{3}{8} \times (-1) + \frac{2}{8} \times 0 + \frac{3}{8} \times 1 = 0$$

Para calcular a covariância de U e V , são necessárias as probabilidades do produto UV :

$$E(UV) = \frac{1}{8} \times (-2) + \frac{1}{8} \times (-1) + \frac{4}{8} \times 0 + \frac{1}{8} \times 1 + \frac{1}{8} \times 2 = 0$$

Então:

$$\text{cov}(U,V) = E(UV) - E(U)E(V) = 0 - 1 \times 0 = 0$$

Isto é, apesar de a covariância ser zero, as variáveis U e V são dependentes. Lembre-se de que, se as variáveis são independentes, a covariância é zero, mas a recíproca não é verdadeira, isto é, covariância zero não implica independência, como pode ser visto neste exemplo.

5.2 Distribuição conjunta de variáveis contínuas

Se as variáveis aleatórias forem contínuas, o procedimento é similar àquele para uma variável. Define-se uma função densidade de probabilidade (f.d.p.) conjunta $f(x,y)$, de tal modo que a probabilidade de x estar entre os valores a e b e y entre c e d é dada por:

$$P(a < x < b \text{ e } c < y < d) = \int_c^d \int_a^b f(x,y) dx dy$$

Ou seja, a função densidade de probabilidade conjunta, assim como a distribuição de probabilidade conjunta discreta, dá a probabilidade do e. Em se tratando de variáveis contínuas (seja uma ou mais de uma), a probabilidade só pode ser calculada para um intervalo, isto é:

$$P(x = x_0 \text{ e } y = y_0) = 0$$

Mesmo que $x = x_0$ e $y = y_0$ sejam eventos possíveis.

A função densidade de probabilidade conjunta deve seguir as mesmas propriedades da f.d.p. para uma variável, isto é, não pode ser negativa:

$$f(x,y) \geq 0$$

E a soma de todas as probabilidades tem de ser igual a 1:

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y) dx dy = 1$$

Exemplo 5.2.1

Dada a função:

$$f(x,y) = \begin{cases} Axy, \text{ para } 0 < x < 1 \text{ e } 0 < y < 1 \\ 0, \text{ demais valores} \end{cases}$$

Determine o valor de A para que $f(x,y)$ seja uma função densidade de probabilidade.

Para ser uma f.d.p., deve obedecer a:

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y) dx dy = 1$$

Ou, no caso específico, uma vez que tanto x como y variam entre 0 e 1:

$$\int_0^1 \int_0^1 f(x,y) dx dy = 1$$

$$\int_0^1 \int_0^1 Axy dx dy = 1$$

$$\int_0^1 Ay \int_0^1 x dx dy = 1$$

$$\int_0^1 Ay \left[\frac{x^2}{2} \right]_0^1 dy = 1$$

$$\int_0^1 Ay \frac{1}{2} dy = 1$$

$$\frac{A}{2} \int_0^1 y dy = 1$$

$$\frac{A}{2} \left[\frac{y^2}{2} \right]_0^1 = 1$$

$$\frac{A}{2} \times \frac{1}{2} = 1$$

$$\frac{A}{4} = 1$$

$$A = 4$$

Exemplo 5.2.2

Dada a função densidade de probabilidade do Exemplo 5.2.1, determine a probabilidade de x estar entre 0,2 e 0,4 e y estar entre 0,6 e 0,8.

A f.d.p. é dada por:

$$f(x,y) = \begin{cases} 4xy, & \text{para } 0 < x < 1 \text{ e } 0 < y < 1 \\ 0, & \text{demais valores} \end{cases}$$

A probabilidade desejada é dada diretamente pela integral da função densidade de probabilidade:

$$P(0,2 < x < 0,4 \text{ e } 0,6 < y < 0,8) = \int_{0,6}^{0,8} \int_{0,2}^{0,4} f(x,y) dx dy$$

$$P(0,2 < x < 0,4 \text{ e } 0,6 < y < 0,8) = \int_{0,6}^{0,8} \int_{0,2}^{0,4} 4xy dx dy$$

$$\begin{aligned}
 P(0,2 < x < 0,4 \text{ e } 0,6 < y < 0,8) &= \int_{0,6}^{0,8} \int_{0,2}^{0,4} 4y \, x \, dx \, dy \\
 P(0,2 < x < 0,4 \text{ e } 0,6 < y < 0,8) &= \int_{0,6}^{0,8} 4y \left[\frac{x^2}{2} \right]_{0,2}^{0,4} \, dy \\
 P(0,2 < x < 0,4 \text{ e } 0,6 < y < 0,8) &= \int_{0,6}^{0,8} 4y \left[\frac{0,4^2}{2} - \frac{0,2^2}{2} \right]_{0,2}^{0,4} \, dy \\
 P(0,2 < x < 0,4 \text{ e } 0,6 < y < 0,8) &= \int_{0,6}^{0,8} 0,24y \, dy \\
 P(0,2 < x < 0,4 \text{ e } 0,6 < y < 0,8) &= 0,24 \int_{0,6}^{0,8} y \, dy \\
 P(0,2 < x < 0,4 \text{ e } 0,6 < y < 0,8) &= 0,24 \left[\frac{y^2}{2} \right]_{0,6}^{0,8} \\
 P(0,2 < x < 0,4 \text{ e } 0,6 < y < 0,8) &= 0,24 \left[\frac{0,8^2}{2} - \frac{0,6^2}{2} \right] \\
 P(0,2 < x < 0,4 \text{ e } 0,6 < y < 0,8) &= 0,0336
 \end{aligned}$$

Exemplo 5.2.3

Dada a função densidade de probabilidade do Exemplo 5.2.1, determine as f.d.p. marginais de x e y .

No caso de variáveis aleatórias discretas, a distribuição marginal de x era encontrada somando-se as probabilidades para todos os y , e vice-versa. Com variáveis contínuas, a função densidade de probabilidade marginal de x (chamada aqui de $g(x)$) é encontrada de forma análoga, isto é, integrando (somando) em y .

De modo geral, a f.d.p. marginal de x pode ser encontrada assim:

$$g(x) = \int_{-\infty}^{+\infty} f(x,y) \, dy$$

No caso específico:

$$g(x) = \int_0^1 4xy \, dy$$

$$g(x) = 4x \int_0^1 y \, dy$$

$$g(x) = 4x \left[\frac{y^2}{2} \right]_0^1$$

$$g(x) = 4x \times \frac{1}{2}$$

$$g(x) = 2x$$

De forma análoga, a função densidade de probabilidade marginal de y , chamada aqui de $h(y)$, é dada por:

$$h(y) = \int_0^1 4xy dx$$

$$h(y) = 2y$$

Exemplo 5.2.4

Dada a função densidade de probabilidade conjunta do Exemplo 5.2.1, determine a probabilidade de x estar entre 0,3 e 0,7.

Como só se quer a probabilidade de x , utilize a função densidade de probabilidade marginal de x :

$$P(0,3 < x < 0,7) = \int_{0,3}^{0,7} 2x dx = \left[x^2 \right]_{0,3}^{0,7} = 0,7^2 - 0,3^2 = 0,49 - 0,09 = 0,4$$

Exemplo 5.2.5

Dada a função densidade de probabilidade conjunta do Exemplo 5.2.1, determine as f.d.p. condicionais de x e y .

A probabilidade condicional para dois eventos A e B quaisquer é dada por:

$$P(A|B) = \frac{P(A \text{ e } B)}{P(B)}$$

A probabilidade da interseção (do e) é a própria probabilidade conjunta, isto é, a probabilidade de x e y é obtida pela função densidade de probabilidade conjunta. Portanto, a f.d.p. condicional de x (dado y), representada por $f_x|y$, é dada por:

$$f_x|y = \frac{f(x,y)}{h(y)}$$

No caso da função densidade de probabilidade conjunta do Exemplo 5.2.1:

$$f_x|y = \frac{4xy}{2y}$$

$$f_x|y = 2x$$

Da mesma forma, para a função densidade de probabilidade condicional de y (dado x), denominada $f_y|x$, vale:

$$f_y|x = \frac{f(x,y)}{g(x)}$$

$$f_y|x = \frac{4xy}{2x}$$

$$f_y|x = 2y$$

Note que:

$$f_x|y = g(x) \text{ e } f_y|x = h(y)$$

Ou seja, as probabilidades condicionais são iguais às não condicionais. Portanto, x e y são variáveis independentes.

Repare que, para essa função, é válida a igualdade:

$$f(x,y) = g(x)h(y) \quad (5.2.1)$$

Já que:

$$4xy = 2x \times 2y$$

A igualdade (5.2.1) é válida sempre que as variáveis forem independentes (veja demonstração no Apêndice 5.B).

Assim, uma maneira de verificar se as variáveis em uma função densidade de probabilidade conjunta são independentes é analisar se essa função pode ser fatorada em uma função só de x e outra só de y , ou seja, se for possível separar x e y .

Exemplo 5.2.6

Dada a função densidade de probabilidade do Exemplo 5.2.1, determine $E(x)$. Pode-se calcular o valor esperado de x diretamente da f.d.p. conjunta. De modo geral, de maneira análoga às f.d.p. com uma única variável, vale:

$$E(x) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xf(x,y)dx dy$$

Para o caso particular da função densidade de probabilidade apresentada no Exemplo 5.2.1:

$$E(x) = \int_0^1 \int_0^1 xf(x,y)dx dy$$

$$E(x) = \int_0^1 \int_0^1 x4xydx dy$$

$$E(x) = 4 \int_0^1 \int_0^1 x^2 4 dx dy$$

$$E(x) = 4 \int_0^1 y \left[\frac{x^3}{3} \right]_0^1 dy$$

$$E(x) = \frac{4}{3} \int_0^1 y dy$$

$$E(x) = \frac{4}{3} \left[\frac{y^2}{2} \right]_0^1$$

$$E(x) = \frac{4}{3} \times \frac{1}{2}$$

$$E(x) = \frac{2}{3}$$

Ou você pode utilizar simplesmente a função densidade de probabilidade marginal de x , cálculo cuja forma geral é:

$$E(x) = \int_{-\infty}^{+\infty} xg(x)dx$$

Para o caso específico deste exemplo:

$$E(x) = \int_0^1 x^2 dx$$

$$E(x) = 2 \int_0^1 x^2 dx$$

$$E(x) = 2 \left[\frac{x^3}{3} \right]_0^1$$

$$E(x) = 2 \times \frac{1}{3}$$

$$E(x) = \frac{2}{3}$$

Exemplo 5.2.7

Dada a função densidade de probabilidade do Exemplo 5.2.1, determine a variância de x . De novo, pode-se calcular a variância diretamente da f.d.p. conjunta, que, de forma análoga às f.d.p. de uma única variável, é dada por:

$$\text{var}(x) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} [x - E(x)]^2 f(x,y) dx dy = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^2 f(x,y) dx dy - \left[\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x f(x,y) dx dy \right]^2$$

Sendo o último termo nada mais do que uma nova forma para uma já conhecida expressão (média dos quadrados menos o quadrado das médias).

Ou você pode utilizar diretamente a função marginal, como fez para a esperança de x :

$$\text{var}(x) = \int_{-\infty}^{+\infty} [x - E(x)]^2 g(x) dx = \int_{-\infty}^{+\infty} x^2 g(x) dx - \left[\int_{-\infty}^{+\infty} x g(x) dx \right]^2$$

Como você já calculou a média no exemplo anterior, use a última expressão:

$$\text{var}(x) = \int_{-\infty}^{+\infty} x^2 g(x) dx - \left[\int_{-\infty}^{+\infty} x g(x) dx \right]^2$$

Que, neste exemplo, é:

$$\text{var}(x) = \int_0^1 x^2 g(x) dx - \left[\frac{2}{3} \right]^2$$

$$\text{var}(x) = \int_0^1 x^2 dx - \frac{4}{9}$$

$$\text{var}(x) = 2 \int_0^1 x^3 dx - \frac{4}{9}$$

$$\text{var}(x) = 2 \left[\frac{x^4}{4} \right]_0^1 - \frac{4}{9}$$

$$\text{var}(x) = \frac{2}{4} - \frac{4}{9}$$

$$\boxed{\text{var}(x) = \frac{1}{18}}$$

Exemplo 5.2.8

Dada a função densidade de probabilidade do Exemplo 5.2.1, determine $\text{cov}(x,y)$:

Lembre que:

$$\text{cov}(x,y) = E[(x - E(x)) \times (y - E(y))] = E(xy) - E(x)E(y)$$

O que, para uma função densidade de probabilidade conjunta, pode ser escrito como:

$$\text{cov}(x,y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - E(x)) (y - E(y)) dx dy = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy f(x, y) dx dy - \\ \int_{-\infty}^{+\infty} x g(x) dx \int_{-\infty}^{+\infty} y h(y) dy$$

Como você já calculou a média de x (e viu que ela é igual à média de y), fique com a segunda expressão, que, para este exemplo, é dada por:

$$\text{cov}(x,y) = \int_0^1 \int_0^1 xy 4xy dx dy - \frac{2}{3} \times \frac{2}{3}$$

$$\text{cov}(x,y) = 4 \int_0^1 y^2 \int_0^1 x^2 dx dy - \frac{4}{9}$$

$$\text{cov}(x,y) = 4 \int_0^1 y^2 \left[\frac{x^3}{3} \right]_0^1 dy - \frac{4}{9}$$

$$\text{cov}(x,y) = \frac{4}{3} \int_0^1 y^2 dy - \frac{4}{9}$$

$$\text{cov}(x,y) = \frac{4}{3} \left[\frac{y^3}{3} \right]_0^1 - \frac{4}{9}$$

$$\text{cov}(x,y) = \frac{4}{3} \times \frac{1}{3} - \frac{4}{9}$$

$$\text{cov}(x,y) = \frac{4}{9} - \frac{4}{9}$$

$$\boxed{\text{cov}(x,y) = 0}$$

O que, diga-se de passagem, já era um resultado esperado, tendo em vista que as duas variáveis são independentes, como já visto.

Exemplo 5.2.9

Dada a função:

$$f(x,y) = \begin{cases} B(x^2 + y^2), & \text{para } 0 < x < 1 \text{ e } 0 < y < 1 \\ 0, & \text{demais valores} \end{cases}$$

- a) determine o valor da constante B , de modo que a função dada seja uma f.d.p.;
- b) determine as f.d.p. marginais de x e y ;
- c) determine as f.d.p. condicionais de x e y ;
- d) verifique se x e y são variáveis aleatórias independentes;
- e) calcule $P(x < 0,5 | y = 0,5)$.

Seguem as soluções.

- a) Para ser uma f.d.p., a função deve obedecer à condição:

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = 1$$

Como no Exemplo 5.2.1, tanto x como y variam entre 0 e 1:

$$\int_0^1 \int_0^1 f(x, y) dx dy = 1$$

$$\int_0^1 \int_0^1 B(x^2 + y^2) dx dy = 1$$

$$B \int_0^1 \int_0^1 (x^2 + y^2) dx dy = 1$$

$$B \int_0^1 \left(\frac{x^3}{3} + y^2 x \right)_0^1 dy = 1$$

$$B \int_0^1 \left(\frac{1}{3} + y^2 \right) dy = 1$$

$$B \left(\frac{1}{3}y + \frac{y^3}{3} \right)_0^1 = 1$$

$$B \left(\frac{1}{3} + \frac{1}{3} \right) = 1$$

$$B \times \frac{2}{3} = 1$$

$$B = \frac{3}{2}$$

b) Para encontrar a f.d.p. marginal de x , integra-se (soma-se) em y :

$$g(x) = \frac{2}{3} \int_0^1 (x^2 + y^2) dy = \frac{3}{2} \left[x^2 y + \frac{y^3}{3} \right]_0^1 = \frac{3}{2} \left(x^2 + \frac{1}{3} \right)$$

Da mesma forma para a f.d.p. marginal de y :

$$h(y) = \frac{3}{2} \int_0^1 (x^2 + y^2) dx = \frac{3}{2} \left[\frac{x^3}{3} + y^2 x \right]_0^1 = \frac{3}{2} \left(\frac{1}{3} + y^2 \right)$$

c) As f.d.p. marginais de x e y são dadas por:

$$f_{x|y} = \frac{f(x,y)}{h(y)} = \frac{\frac{2}{3}(x^2 + y^2)}{\frac{3}{2}\left(\frac{1}{3} + y^2\right)} = \frac{x^2 + y^2}{\frac{1}{3} + y^2}$$

$$f_{y|x} = \frac{f(x,y)}{g(x)} = \frac{\frac{3}{2}(x^2 + y^2)}{\frac{3}{2}\left(\frac{1}{3} + x^2\right)} = \frac{x^2 + y^2}{\frac{1}{3} + x^2}$$

d) As variáveis x e y são dependentes, já que, pelos resultados obtidos nos itens anteriores:

$$f_{x|y} \neq g(x) \text{ e } f_{y|x} \neq h(y)$$

Mas essa conclusão já poderia ser tirada antes mesmo da resolução dos itens (b) e (c), uma vez que é impossível fatorar a função $x^2 + y^2$ em uma função só de x e outra só de y .

e) Para calcular a probabilidade pedida, usa-se a f.d.p. condicional de x (dado que $y = 0,5$).

$$f_{x|y} = \frac{x^2 + y^2}{\frac{1}{3}y^2} = \frac{x^2 + \left(\frac{1}{2}\right)^2}{\frac{1}{3} + \left(\frac{1}{2}\right)^2} = \frac{x^2 + \frac{1}{4}}{\frac{1}{3} + \frac{1}{4}} = \frac{x^2 + \frac{1}{4}}{\frac{7}{12}} = \frac{12}{7} \left(x^2 + \frac{1}{4} \right)$$

Nesse caso, a probabilidade de x ser menor do que 0,5 (dado que y é igual a 0,5) é dada por:

$$\begin{aligned} P(x < 0,5 | y = 0,5) &= \frac{12}{7} \int_0^{0,5} \left(x^2 + \frac{1}{4} \right) dx = \frac{12}{7} \left[\frac{x^3}{3} + \frac{1}{4}x \right]_0^{0,5} = \\ &= \frac{12}{7} \left(\frac{1}{3} \times \frac{1}{8} \times \frac{1}{4} \times \frac{1}{2} \right) \frac{2}{7} \cong 0,2857 \end{aligned}$$

Exemplo 5.2.10

Com a f.d.p. do Exemplo 5.2.9, determine $E(x|y = 0,5)$.

Do exemplo anterior, sabe-se que:

$$f_{x|y=0,5} = \frac{12}{7} \left(x^2 + \frac{1}{4} \right)$$

A esperança condicional de x é dada por:

$$E(x|y=0,5) = \int_{-\infty}^{+\infty} xf_{x|y=0,5} dx$$

O que, nesse exemplo, é calculado como se segue:

$$E(x|y=0,5) = \frac{12}{7} \int_0^1 x \left(x^2 + \frac{1}{4} \right) dx$$

$$E(x|y=0,5) = \frac{12}{7} \int_0^1 \left(x^3 + \frac{1}{4}x \right) dx$$

$$E(x|y=0,5) = \frac{12}{7} \left(\frac{x^4}{4} + \frac{x^2}{8} \right)_0^1$$

$$E(x|y=0,5) = \frac{12}{7} \left(\frac{1}{4} + \frac{1}{8} \right)$$

$$E(x|y=0,5) = \frac{12}{7} \times \frac{3}{8}$$

$$E(x|y=0,5) = \frac{9}{14}$$

Exemplo 5.2.11

Dada a função:

$$f(x,y) = \begin{cases} C, & \text{para } 0 < x < y < 1 \\ 0, & \text{demais valores} \end{cases}$$

Determine o valor da constante C para que essa função seja uma f.d.p.

Aqui é preciso tomar o cuidado, porque os limites de integração são diferentes, pois, embora x e y variem de 0 a 1, note que x na verdade vai de 0 a y (se y é igual a 1, então x vai de 0 a 1 mesmo; mas, se y for, por exemplo, 0,34, x vai de 0 a 0,34).

Portanto, os limites de integração quando se integra em relação a x devem ser 0 e y . Uma vez eliminado x , os limites de integração para y são mesmo 0 e 1.

Assim, aplicando a condição de que a soma de todas as probabilidades deve ser igual a 1:

$$\int_0^1 \int_0^y C dx dy = 1$$

$$\int_0^1 [Cx]_0^y dy = 1$$

$$\int_0^1 Cy dy = 1$$

$$\left[C \frac{y^2}{2} \right]_0^1 = 1$$

$$C \times \frac{1}{2} = 1$$

$$C = 2$$

Repare que a ordem em que as variáveis são integradas, mesmo nesse caso, não é importante. Se quiser integrar primeiro em relação a y , deve-se notar que y vai de x a 1 e, uma vez eliminado y , x varia de 0 a 1.

$$\int_0^1 \int_x^1 Cd y dx = 1$$

$$\int_0^1 [Cy]_x^1 dx = 1$$

$$\int_0^1 (C - Cx) dx = 1$$

$$\left[Cx - \frac{Cx^2}{2} \right]_0^1 = 1$$

$$C - \frac{C}{2} = 1$$

$$\frac{C}{2} = 1$$

$$C = 2$$

Exemplo 5.2.12

Suponha que x e y são duas variáveis aleatórias independentes, com distribuição normal, identicamente distribuídas (mesma média e mesmo desvio padrão)³. Determine a f.d.p. conjunta para essas duas variáveis.

Em se tratando de variáveis cuja distribuição é normal, a f.d.p. de cada uma delas é dada por:

$$g(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$h(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}$$

³ Já que a média e o desvio padrão definem uma distribuição normal.

Como são variáveis independentes, vale:

$$f(x,y) = g(x)h(y)$$

$$f(x,y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \times \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}$$

$$f(x,y) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^2 e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 - \frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}$$

$$f(x,y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}[(x-\mu)^2 + (y-\mu)^2]}$$

Essa é uma f.d.p. de uma distribuição normal **bivariada** (em que as variáveis são independentes).

Exercícios

1. Dadas as distribuições de probabilidade mostradas nas Tabelas 5.10 e 5.11, determine:
 - a) as distribuições marginais de X e Y ;
 - b) as seguintes probabilidades:
 - i) $P(X = 1)$;
 - ii) $P(Y = 1)$;
 - iii) $P(X = 2)$;
 - iv) $P(X = 2 \text{ e } Y = -1)$;
 - v) $P(X = 3 \text{ e } Y = 1)$;
 - vi) $P(X = 1 | Y = -1)$;
 - vii) $P(X = 2 | Y = 1)$;
 - viii) $P(Y = 1 | X = 2)$;
 - c) se X e Y são variáveis independentes (justifique);
 - d) $E(X)$, $E(Y)$, $\text{var}(X)$, $\text{var}(Y)$, $\text{covar}(X,Y)$ e ρ_{xy} ;
 - e) $E(X|Y = -1)$; $E(Y|X = 1)$;
 - f) $\text{var}(X|Y = 1)$.

Tabela 5.10 – Distribuição conjunta de X e Y

		X			
		0	1	2	3
Y	-1	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$
	1	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	0

Tabela 5.11 – Distribuição conjunta de X e Y

		X			
		0	1	2	3
Y	-1	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	0
	1	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

Enunciado para os Exercícios 2 a 4: suponha que o estatístico mencionado no início deste capítulo trabalhasse para um time de futebol, em vez de um time de vôlei. Ele define, então, três variáveis para os *dois* primeiros jogos: X é o número de pontos do time (três pontos para vitória, um para empate); Y é o número de vitórias; Z é igual a 1 se o resultado de um jogo é o mesmo do anterior (por exemplo, para duas vitórias seguidas, $Z = 1$; para uma vitória e um empate $Z = 0$). A probabilidade em cada jogo é a mesma para vitória, empate ou derrota.

2. Numa tabela, mostre a distribuição conjunta e as marginais de X e Y. Calcule a covariância de X e Y e determine se são variáveis independentes.
3. Numa tabela, mostre a distribuição conjunta e as marginais de Y e Z. Calcule a covariância de Y e Z e determine se são variáveis independentes.
4. Numa tabela, mostre a distribuição conjunta e as marginais de X e Z. Calcule a covariância de X e Z e determine se são variáveis independentes.
5. Uma urna contém quatro bolas, duas vermelhas e duas brancas, numeradas, respectivamente, de 1 a 2 e de 3 a 4. Para três bolas sorteadas, sem reposição, defina X como o número de bolas vermelhas e Y como sendo 1 para número ímpar e 0 para número par na primeira bola sorteada.
 - a) Determine a distribuição conjunta de X e Y.
 - b) Determine as distribuições marginais de X e Y.

- c) X e Y são independentes?
- d) Calcule $E(X)$, $E(Y)$.
- e) Calcule $\text{var}(X)$, $\text{var}(Y)$.
- f) Calcule a covariância e o coeficiente de correlação entre X e Y .
6. Dada a seguinte distribuição de probabilidade conjunta:

		Tabela 5.12 – Probabilidade conjunta de K e L		
		L		
K		0	1	2
-1		0,1	0,1	0,15
0		0,15	0,1	0,1
1		0,05	0,15	0,1

- a) determine as distribuições marginais de K e L ;
- b) determine o valor esperado de K e L ;
- c) determine a covariância de K e L ;
- d) K e L são variáveis aleatórias independentes?
- e) determine $E(K|L = 1)$ e $E(L|K = 0)$.
7. Dadas as seguintes distribuições de probabilidade, preencha o espaço vazio com o valor apropriado e determine as distribuições marginais.
- a)

		Tabela 5.13 – Probabilidade conjunta de W e Z			
		Z			
W		0	1	2	3
1		$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$
2		$\frac{1}{9}$	$\frac{1}{3}$	$\frac{1}{9}$	

b)

		Tabela 5.14 – Probabilidade conjunta de F e G		
		G		
F		2	4	6
1		0,1	0,1	0,1
3		0,15		0,05
5		0,05	0,2	0,05

8. Dada a f.d.p. conjunta do Exemplo 5.2.1, determine as seguintes probabilidades:
- $P(0,2 < x < 0,7);$
 - $P(0,1 < y < 0,4);$
 - $P(x > 0,5);$
 - $P(y < 0,8);$
 - $P(x < 0,7 \text{ e } y > 0,2);$
 - $P(0,1 < x < 0,3 \text{ e } 0,4 < y < 0,8);$
 - $P(x < 0,9 | y = 0,2);$
 - $P(y > 0,6 | x = 0,45).$
9. Determine o valor da constante A em cada uma das seguintes funções, de tal modo que elas sejam f.d.p.

$$\text{a) } f(x,y) = \begin{cases} Ax^2y, \text{ para } -1 < x < 1 \text{ e } 0 < y < 2 \\ 0, \text{ demais valores} \end{cases}$$

$$\text{b) } f(x,y) = \begin{cases} A(x+y^2), \text{ para } 0 < x < 2 \text{ e } -1 < y < 0 \\ 0, \text{ demais valores} \end{cases}$$

$$\text{c) } f(x,y) = \begin{cases} Ae^{-(x+y)}, \text{ para } x > 0 \text{ e } y > 0 \\ 0, \text{ demais valores} \end{cases}$$

$$\text{d) } f(x,y) = \begin{cases} A, \text{ para } 3 < x < 7 \text{ e } -2 < y < 1 \\ 0, \text{ demais valores} \end{cases}$$

10. Dada a f.d.p. conjunta abaixo:

$$f(x,y) = \begin{cases} 6x^2y, \text{ para } 0 < x < 1 \text{ e } 0 < y < 1 \\ 0, \text{ demais valores} \end{cases}$$

Determine:

- as f.d.p. marginais de x e y ;
- as f.d.p. condicionais de x e y ;
- se x e y são independentes;
- $P(x > 0,4);$

- e) $P(y < 0,8)$;
f) $P(x < 0,2 \text{ e } y > 0,3)$.

11. Dada a função abaixo:

$$f(x,y) = \begin{cases} Ce^{-ay}, \text{ para } 0 < x < y \\ 0, \text{ demais valores} \end{cases}$$

- a) determine o valor de C para que $f(x,y)$ seja uma f.d.p.;
b) determine as f.d.p. marginais e condicionais de x e y .

12. Dada a função abaixo:

$$f(x,y) = \begin{cases} B, \text{ para } 0 < x < y < 2 \\ 0, \text{ demais valores} \end{cases}$$

- a) determine o valor de B para que $f(x,y)$ seja uma f.d.p.;
b) determine as f.d.p. marginais e condicionais de x e y ;
c) calcule $E(x)$ e $E(y)$;
d) calcule $E(y|x=1)$.

13. Se as variáveis X e Y forem definidas como se segue:

$X = 1$ se o evento A ocorre, e 0 em caso contrário;

$Y = 1$ se o evento B ocorre, e 0 em caso contrário.

Se $P(A)$ e $P(B)$ são não nulas, mostre que, nesse caso, se o coeficiente de correlação entre X e Y for igual a zero, então X e Y são independentes.

14. Suponha x e y duas variáveis aleatórias independentes com distribuição normal e média e desvio padrão dados, respectivamente, por 0 e 2 (para x) e -1 e 1 (para y). Determine a f.d.p. conjunta de x e y .

15. Suponha w e z duas variáveis aleatórias independentes com distribuição exponencial e média dadas, respectivamente, por 0,5 e 0,75. Determine a f.d.p. conjunta de w e z .

16. Com os dados das Tabelas 5.1 a 5.4, determine $E[\text{var}(Y|X)]$ e $\text{var}[E(Y|X)]$. A decomposição das variâncias se aplica?

17. Considere duas variáveis aleatórias x e y . Suponha que x seja distribuída de acordo com a seguinte função de densidade:

$$g(x) = \begin{cases} 1, \text{ se } 0 < x < 1 \\ 0, \text{ caso contrário.} \end{cases}$$

Suponha ainda que

$$f_{y|x} = \begin{cases} \frac{1}{x}, & \text{se } 0 < y < x \\ 0, & \text{caso contrário.} \end{cases}$$

Calcule a média, a variância e a covariância das duas variáveis.

18. No começo do dia uma máquina de refrigerantes tem uma quantidade de litros aleatória y . No decorrer do mesmo dia, uma parte disso (x) é aleatoriamente descartada pela máquina. Como a máquina não é carregada, $x \leq y$. A função densidade conjunta de X e Y é:

$$f(x,y) = \begin{cases} A, & \text{para } 0 \leq x \leq y \leq 40 \\ 0, & \text{demais valores} \end{cases}$$

Determine o valor da constante A .

Calcule a probabilidade de que menos de 10 litros sejam descarregados no decorrer de um dia, dado que a máquina contém 20 litros no começo do dia.

19. Suponha que a função densidade de probabilidade conjunta das variáveis x e y seja dada por:

$$f(x,y) = \begin{cases} kx(x-y), & \text{para } 0 \leq y \leq x \leq 2 \\ 0, & \text{demais valores} \end{cases}$$

Calcule a média, a variância e a covariância das duas variáveis.

Apêndice 5.B – Tópicos adicionais em distribuição conjunta

5.B.1 Probabilidade condicional

Algum leitor mais desconfiado pode ter suspeitado da validade, por exemplo, da seguinte expressão para o caso de distribuições contínuas:

$$P(x > 0,5 | y = 0,5) = ?$$

A suspeita é válida, uma vez que $P(y = y_0) = 0$ para qualquer valor de y_0 quando se trata de uma distribuição contínua.

Uma probabilidade condicional, nesse caso, só poderia ser definida quando a condição fosse também um intervalo (e não um ponto). Em outras palavras, seria alguma coisa como:

$$P(a < x < b | c < y < d) = ?$$

E o resultado seria dado por:

$$P(a < x < b | c < y < d) = \frac{P[(a < x < b) \text{ e } (c < y < d)]}{P(c < y < d)}$$

O numerador dessa fração sairia automaticamente de uma (dada) f.d.p. conjunta:

$$P[(a < x < b) \text{ e } (c < y < d)] = \int_c^d \int_a^b f(x,y) dx dy$$

Já o denominador é obtido pela f.d.p. marginal de y , que, por sua vez, é dada por:

$$h(y) = \int_{-\infty}^{+\infty} f(x,y) dx$$

Portanto, a expressão no denominador é:

$$P(c < y < d) = \int_c^d \int_{-\infty}^{+\infty} f(x,y) dx dy = \int_c^d h(y) dy$$

Fazendo: $c = y_0$ e $d = y_0 + \Delta y$.

A desigualdade $c < y < d$ colapsa em $y = y_0$ quando d se aproxima de c , isto é, quando Δy se aproxima de (tende a) zero.

Portanto, pode-se interpretar a probabilidade condicional com uma igualdade na condição como um caso-limite do caso geral:

$$\lim_{d \rightarrow c} P(a < x < b | c < y < d) = \lim_{\Delta y \rightarrow 0} P(a < x < b | c < y < d) = \\ = P(a < x < b | y = y_0)$$

Mas, do cálculo diferencial, sabe-se que tomar o limite para $\Delta y \rightarrow 0$ equivale à derivada em relação a y no ponto em questão, no caso y_0 .

O denominador, então, é dado por:

$$\lim_{\Delta y \rightarrow 0} P(c < y < d) = \lim_{\Delta y \rightarrow 0} \int_{y_0}^{y_0 + \Delta y} h(y) dy$$

O que equivale a:

$$\lim_{\Delta y \rightarrow 0} = \int_{y_0}^{y_0 + \Delta y} h(y) dy = \frac{\partial}{\partial y} \int_0^y h(t) dt$$

Que é uma derivada de uma função definida por uma integral, a qual é o próprio valor da função a ser integrada, calculado no ponto y_0 , isto é:

$$\lim_{\Delta y \rightarrow 0} P(c < y < d) = \frac{\partial}{\partial y} \int_0^y h(t) dt = h(y_0)$$

Da mesma forma, para a expressão no numerador:

$$\lim_{\Delta y \rightarrow 0} P[(a < x < b) \text{ e } (c < y < d)] = \lim_{\Delta y \rightarrow 0} \int_{y_0}^{y_0 + \Delta y} \int_a^b f(x,y) dx dy$$

$$\lim_{\Delta y \rightarrow 0} P[(a < x < b) \text{ e } (c < y < d)] = \frac{\partial}{\partial y} \int_0^y \int_a^b f(x,t) dx dt$$

$$\lim_{\Delta y \rightarrow 0} P[(a < x < b) \text{ e } (c < y < d)] = \int_a^b f(x,y_0) dx$$

Portanto, a probabilidade condicional (com a condição equivalendo a um ponto) é dada por:

$$P(a < x < b | y = y_0) = \int_a^b \frac{f(x,y_0)}{h(y_0)} dx$$

Como $h(y_0)$ é uma constante em relação a x , pode-se escrever:

$$P(a < x < b | y = y_0) = \int_a^b \frac{f(x,y_0)}{h(y_0)} dx$$

Finalmente, definindo:

$$f_{x|y}(x,y_0) = \frac{f(x,y_0)}{h(y_0)}$$

Chega-se ao cálculo da probabilidade condicional como foi feito no texto:

$$P(a < x < b | y = y_0) = \int_a^b f_{x|y}(x,y_0) dx$$

Portanto, como um caso-limite da hipótese geral em que a condição é um intervalo.

5.B.2 Independência em uma distribuição conjunta

Nesta seção vamos demonstrar, no caso contínuo, que a Expressão 5.2.1 é válida se, e somente se, as variáveis x e y são independentes.

$$f(x,y) = g(x)h(y)$$

Se as variáveis são independentes, então é válido que:

$$f_{x|y} = g(x) \tag{5.B.2.1}$$

$$f_{y|x} = h(y) \tag{5.B.2.2}$$

Mas, pela definição de condicional:

$$f_{x|y} = \frac{f(x,y)}{h(y)}$$

Logo:

$$f(x,y) = f_{x|y}h(y)$$

Substituindo pela Expressão 5.B.2.1:

$$f(x,y) = g(x)h(y)$$

Como queríamos demonstrar.

5.B.3 Valor esperado de uma esperança condicional

O título desta seção foi propositalmente elaborado para evitar a redundância esperança da esperança condicional.

Problemas semânticos à parte, faz sentido falar nisso se se levar em conta que a seguinte esperança condicional é função do valor de x :

$$E(Y|X = x)$$

O valor esperado dessa esperança condicional é a média, considerando todos os possíveis valores de x :

$$\begin{aligned} E[E(Y|X)] &= E(Y|X = x_1) \times P(X = x_1) + E(Y|X = x_2) \times P(X = x_2) + \\ &+ \dots + E(Y|X = x_n) \times P(X = x_n) \end{aligned}$$

Ou, no caso contínuo:

$$E[E(Y|X)] = \int_{-\infty}^{+\infty} E(Y|X)g(x)dx$$

E como:

$$E(Y|X) = \int_{-\infty}^{+\infty} yfY|X dx$$

Vale:

$$E[E(Y|X)] = \int_{-\infty}^{+\infty} yfY|X g(x)dx dy$$

Mas, pela própria definição de f.d.p. condicional:

$$fY|X g(x) = f(x,y)$$

E chega-se a:

$$E[E(Y|X)] = \int_{-\infty}^{+\infty} yf(x,y)dxdy = E(Y)$$

Portanto, o valor esperado da esperança condicional de Y é o próprio valor esperado de Y . Esta demonstração foi feita para o caso contínuo, mas o resultado também é válido para o caso discreto.

5.B.4 Distribuição de probabilidade com três variáveis

Uma f.d.p. conjunta para três variáveis será uma função $f: \mathbb{R}^3 \rightarrow \mathbb{R}$ com as seguintes propriedades:

$$f(x,y,z) \geq 0 \text{ para todo } x,y,z \in \mathbb{R} \text{ e}$$

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y,z) dx dy dz = 1$$

Com ela, pode-se calcular a seguinte probabilidade:

$$P(a < x < b \text{ e } c < y < d \text{ e } e < z < f) = \int_e^f \int_c^d \int_a^b f(x,y,z) dx dy dz$$

As f.d.p. marginais são dadas por:

$$g(x) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y,z) dy dz$$

$$h(y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y,z) dx dz$$

$$k(z) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y,z) dx dy$$

As f.d.p. condicionais são dadas por:

$$f_{x|y} = \frac{\int_{-\infty}^{+\infty} f(x,y,z) dx}{h(y)}$$

De maneira análoga para y e z.

Note que é possível definir uma f.d.p. conjunta apenas para duas variáveis, por exemplo:

$$G(x,y) = \int_{-\infty}^{+\infty} f(x,y,z) dz$$

E mesmo uma f.d.p. condicional em que a condição seja dada por duas variáveis:

$$f_{x|y \text{ e } z} = \frac{f(x,y,z)}{\int_{-\infty}^{+\infty} f(x,y,z) dx}$$

Note que, de maneira análoga, é possível trabalhar com distribuições com um número qualquer de variáveis.

5.B.5 Esperança condicional em uma normal bivariada

Suponha que $E(X) = \mu_X$, $E(Y) = \mu_Y$, $\text{var}(X) = \sigma_X^2$, $\text{var}(Y) = \sigma_Y^2$ e que a correlação entre X e Y seja σ_{XY} .

A distribuição normal bivariada tem f.d.p. dada por:

$$f_{xy}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{Q}{2(1-\rho^2)}}$$

$$\text{onde } Q = \left(\frac{x-\mu_x}{\sigma_x}\right)^2 - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \left(\frac{y-\mu_y}{\sigma_y}\right)^2$$

cujas distribuições marginais são as normais univariadas, isto é:

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}}$$

$$h(y) = \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{(y-\mu_y)^2}{2\sigma_y^2}}$$

A f.d.p. condicional de y é dada por:

$$f_{y|x} = \frac{f(x,y)}{g(x)}$$

$$f_{y|x} = \frac{\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{Q}{2(1-\rho^2)}}}{\frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}}}$$

$$f_{y|x} = \frac{1}{\sqrt{2\pi(1-\rho^2)}\sigma_y} e^{-\frac{Q'}{2}}$$

$$\text{onde } Q' = \frac{Q}{(1-\rho^2)} - \frac{(x-\mu_x)^2}{\sigma_x^2}$$

$$Q' = \frac{1}{1-\rho^2} \left(\frac{x-\mu_x}{\sigma_x} \right)^2 - \frac{2\rho}{1-\rho^2} \frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{1}{1-\rho^2} \left(\frac{y-\mu_y}{\sigma_y} \right)^2 - \left(\frac{x-\mu_x}{\sigma_x} \right)^2$$

$$Q' = \left(\frac{1}{1-\rho^2} - 1 \right) \left(\frac{x-\mu_x}{\sigma_x} \right)^2 - \frac{2\rho}{1-\rho^2} \frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{1}{1-\rho^2} \left(\frac{y-\mu_y}{\sigma_y} \right)^2$$

$$Q' = \frac{\rho^2}{1-\rho^2} \left(\frac{x-\mu_x}{\sigma_x} \right)^2 - \frac{2\rho}{1-\rho^2} \frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{1}{1-\rho^2} \left(\frac{y-\mu_y}{\sigma_y} \right)^2$$

$$Q' = \frac{1}{(1-\rho^2)\sigma_y^2} \left[\rho^2 \frac{\sigma_y^2}{\sigma_x^2} (x-\mu_x)^2 - 2\rho \frac{\sigma_y}{\sigma_x} (x-\mu_x)(y-\mu_y) + (y-\mu_y)^2 \right]$$

$$Q' = \frac{1}{(1-\rho^2)\sigma_y^2} \left[y - \mu_y - \rho \frac{\sigma_y}{\sigma_x} (x - \mu_x) \right]^2$$

Dessa forma, Q' é justamente o expoente de uma f.d.p. normal univariada cuja média é dada por:

$$E(Y|X) = \mu_y + \rho \frac{\sigma_y}{\sigma_x} (x - \mu_x)$$

E a variância é dada por:

$$\text{var}(Y|X) = \frac{1}{(1-\rho^2)\sigma_y^2}$$



Capítulo 6

ESTIMAÇÃO

6.1 O que é inferência estatística?

Inferência é algo que todo mundo (ou, pelo menos, muita gente) já fez na vida. Ao cozinhar, por exemplo: para ver se um molho está bom, já no ponto para ser servido, não é necessário prová-lo por inteiro, basta uma *colheradinha*. Ao fazer um exame de sangue, não é necessário (ainda bem!) tirar todo o sangue do corpo.

Tanto no caso do molho como no do sangue, a informação sobre o todo é extraída de um pedaço. Nem sempre é tão simples assim, já que, às vezes, o todo sobre o qual queremos uma informação é mais complicado, mais heterogêneo do que o molho, por exemplo.

Numa pesquisa sobre as intenções de voto para prefeito, não basta o pesquisador tomar as opiniões somente dos moradores dos Jardins (se for em São Paulo), de São Conrado (se for no Rio) ou na Boa Viagem (se for em Recife). O resultado da eleição nesses bairros, tendo em vista serem regiões de renda elevada, pode ser (e muito provavelmente será) diferente do resultado em bairros mais pobres. A pesquisa só serviria para se ter uma ideia da intenção de voto naquele bairro, e não na cidade como um todo.

Quando o problema é, então, um pouco mais complicado do que o do molho, usam-se ferramentas estatísticas. É a isso que chamamos de **inferência estatística** ou **estatística inferencial**, isto é, a parte da estatística em que se fazem inferências, diferentemente da estatística descritiva (vista nos capítulos anteriores).

Na inferência estatística, o todo é denominado **população**; o pedaço é chamado de **amostra**. Portanto, a inferência estatística trata de, a partir da amostra, obter informações sobre a população.

6.2 Estimadores

Suponha que você quer conhecer alguma característica de determinada população – por exemplo, a média de idade, a variância da renda, o percentual de intenções de voto para um candidato –, e essa população é composta de milhares (às vezes, milhões) de

elementos (nesse caso, pessoas, mas poderia ser de qualquer tipo), de tal modo que seria muito difícil pesquisar o valor correto, pois seria inviável pesquisar todos os elementos. Nesse caso, é preciso recorrer aos valores encontrados em uma amostra.

Em São Paulo, há 11 milhões de habitantes e cerca de 8 milhões de eleitores. Para uma pesquisa eleitoral, são ouvidas mil, 2 mil, 3 mil pessoas. O número de elementos na amostra geralmente é muito pequeno quando comparado com o da população. Nesse tipo de situação, dizemos que a população é **infinita**. Porque, em termos práticos, não faz diferença se a população é de 5 milhões, 10 milhões, um bilhão ou... infinita! Quando a amostra representa uma fração importante da população, alguns aspectos devem ser considerados, o que você verá mais adiante neste capítulo.

Note que o que é às vezes muito difícil, por uma questão de número, pode ser impossível. Imagine uma pessoa que vai prestar um exame vestibular para uma faculdade. Ela pode estar nervosa no dia, e isso vai prejudicar seu desempenho, ou a prova pode abanger, em sua maioria, tópicos que ela tenha estudado melhor, o que, então, fará com que seu desempenho seja acima do esperado. Qual deveria ser o desempenho *verdadeiro* ou, se preferir, o desempenho médio? É uma pergunta para a qual não há resposta, pois, para responder a ela, precisaríamos de infinitas (ou, pelo menos, de um número muito grande) de repetições desse experimento, que, por definição, não vai repetir-se nunca. Não adianta utilizar nessa *amostra* o desempenho dessa pessoa no vestibular do ano seguinte, pois é outra situação (um ano a mais de estudo, por exemplo).

Há situações em que, mesmo não caindo na armadilha do exemplo dado no parágrafo anterior (em que só é possível obter uma amostra com um elemento), ainda assim é impossível obter uma amostra que se iguale à população. Imagine que você gostaria de obter o preço médio dos imóveis em determinado bairro. Para cada venda, é possível que o vendedor seja habilidoso e consiga um valor superior ao que normalmente seria obtido ou mesmo que o comprador pechinche e consiga um preço mais vantajoso. Para obter o valor correto (populacional) seria preciso que se calculasse a média de todas as *transações possíveis* de ocorrer, o que, evidentemente, não está disponível, ainda que você tenha as informações de todas as transações que foram efetivamente realizadas.

Seja qual for o caso (muito difícil ou impossível de pesquisar a população inteira), o fato é que, em muitas situações, é preciso obter as informações de uma amostra. O valor da população, chamado de **parâmetro populacional**, é desconhecido. O que é possível obter é um valor da amostra, que supostamente nos dá uma ideia do valor correto (populacional) do parâmetro. Esse valor amostral é chamado de **estimador** do parâmetro populacional.

Por exemplo, queremos saber a média de idade dos estudantes universitários na cidade de São Paulo. Como há muitos estudantes, recorre-se a uma amostra de, por exemplo,

100 elementos. A média da amostra encontrada foi de 22 anos, então, essa é a estimativa¹ para a média de idade de todos os estudantes universitários.

Porém a média de idade dos universitários é realmente 22 anos? Não dá para saber, a não ser que todos os estudantes universitários fossem pesquisados. Portanto, são coisas diferentes o parâmetro populacional e o estimador e devem ser representados de maneira diferente, por exemplo:

$$\mu = \text{média populacional} (\text{parâmetro populacional})$$

$$\bar{X} = \text{média amostral} (\text{estimador})$$

Essa não é só uma diferença de valores. Enquanto o parâmetro populacional é, em geral, um valor fixo, o estimador depende da amostra, portanto, está associado a uma distribuição de probabilidade e, assim, é uma variável aleatória.

Apenas como regra geral para a nomenclatura, adota-se a seguinte convenção: se o parâmetro populacional for $\hat{\theta}$, o estimador será $\hat{\theta}$.² A média, por ser um parâmetro especial, receberá tratamento diferente e será chamada como já foi definido.

Você já sabe que o estimador não é igual ao parâmetro populacional. É preciso (ou, pelo menos, é desejável), no entanto, que ele atenda a algumas propriedades.

6.3 Estimadores não viesados

A primeira propriedade (desejável) de um estimador é que, *na média*, ele acerte o valor correto. Ou seja, se fosse possível repetir a experiência (por exemplo, a da média de idade dos universitários) um número de vezes muito grande (infinito), o valor médio das estimativas encontradas em cada experimento seria o valor correto do parâmetro populacional. Resumindo:

$$E(\hat{\theta}) = \theta$$

A esperança do estimador deve ser o parâmetro populacional, o primeiro acerta, em média, o valor do último. Se isso ocorre, o estimador é chamado de **não viesado** ou **não tendencioso**.

Se, entretanto, o estimador erra, em média, fala-se que ele é **viesado**, e a diferença entre sua média e o valor verdadeiro do parâmetro é chamado de **viés**:

$$\hat{\theta} \text{ é viesado} \Leftrightarrow E(\hat{\theta}) = \theta + \text{viés}$$

Fica uma pergunta: a média amostral é um estimador não viesado da média amostral?

Para responder, veja o exemplo a seguir.

¹ Não confundir: estimador é a variável; estimativa é o valor encontrado para essa variável, isto é, o valor encontrado para o estimador *nessa amostra*.

² Há de se fazer uma distinção, pois se trata de coisas diferentes, mas não necessariamente precisa ser essa. Há autores que chamam o parâmetro populacional por uma letra grega (por exemplo, Θ) e o estimador por uma letra latina correspondente (nesse caso, T).

Exemplo 6.3.1

Tome uma população cuja distribuição é muito simples: uma cidade em que metade da população tem 1,80m (os *altos*) e a outra metade tem 1,60m (os *baixos*). Sem saber disso, um pesquisador quer conhecer a média de altura da população da cidade e utiliza, para isso, uma amostra de cinco elementos.

Se conhecesse a distribuição da população, ficaria fácil para ele (pois a média, 1,70m, pode ser facilmente calculada). Como o pobre coitado não sabe, ele pode, numa amostra de cinco pessoas, encontrar 32 possibilidades diferentes, listadas na Tabela 6.1 (em que A representa altos e B representa baixos):

Tabela 6.1 – Possibilidades de pessoas altas e baixas na cidade

Amostra encontrada	Média amostral	Amostra encontrada	Média amostral
BBBBB	1,60m	BBAAA	1,72m
BBBBA	1,64m	BABAA	1,72m
BBBAB	1,64m	BAABA	1,72m
BBABB	1,64m	BAAAB	1,72m
BABBB	1,64m	ABBA	1,72m
ABBBB	1,64m	ABAAB	1,72m
BBBAA	1,68m	ABABA	1,72m
BBAAB	1,68m	AABAB	1,72m
BAABB	1,68m	AAABB	1,72m
AABBB	1,68m	AABBA	1,72m
BBABA	1,68m	BAAAA	1,76m
BABBA	1,68m	ABAAA	1,76m
ABBBA	1,68m	AABAA	1,76m
BABAB	1,68m	AAABA	1,76m
ABBAB	1,68m	AAAAB	1,76m
ABABB	1,68m	AAAAA	1,80m

Repare que, em nenhuma das amostras, o valor populacional (1,70m) foi obtido. Mas a questão é: em média, chega-se ao valor correto? Listadas as possibilidades,³ verifica-se que em uma delas a média é de 1,60m; em cinco, a média é de 1,64m; em dez, 1,68m; para 1,72m há também dez possibilidades; cinco possibilidades para 1,76m; e, em uma delas, a média encontrada será 1,80m. Portanto, a média das médias é dada por:

$$E(\bar{X}) = \frac{1 \times 1,60 + 5 \times 1,64 + 10 \times 1,68 + 10 \times 1,72 + 5 \times 1,76 + 1 \times 1,80}{32} = 1,70m$$

Portanto, pelo menos nesse caso, a média amostral é um estimador não viesado da média populacional. Isso é válido sempre? Sim!

³ Seria absolutamente necessária a montagem da Tabela 6.1 para que encontrássemos esses valores?

Uma média amostral (qualquer) é dada por:

$$\bar{X} = \sum_{i=1}^n \frac{X_1 + X_2 + \dots + X_n}{n}$$

Para saber se esse estimador é ou não viesado, deve-se calcular sua esperança:

$$E(\bar{X}) = E\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right)$$

Pelas propriedades da esperança matemática, sabe-se que:

$$E(\bar{X}) = \frac{1}{n} E(X_1 + X_2 + \dots + X_n)$$

$$E(\bar{X}) = \frac{1}{n} [E(X_1) + E(X_2) + \dots + E(X_n)]$$

Mas qual é a esperança de X_1 (ou de X_2, X_3 , etc.)? Antes de se *sortearem* os elementos da amostra, o valor esperado de seu valor, já que não se sabe qual elemento será escolhido, é a própria média populacional.⁴ Assim:

$$E(\bar{X}) = \frac{1}{n} [\mu + \mu + \dots + \mu]$$

$$E(\bar{X}) = \frac{1}{n} [n\mu]$$

$$E(\bar{X}) = \mu$$

Portanto, a esperança da média amostral é (sempre) igual à média populacional, o que equivale a dizer que a média amostral é um estimador não viesado da média populacional.

Exemplo 6.3.2 [Média ponderada]

Dado o estimador para a média M_1 , definido a seguir, determine se ele é um estimador viesado e, caso seja, calcule o viés.

$$M_1 = \frac{2X_1 + 3X_2}{5}$$

Trata-se de uma média ponderada (com pesos dois e três) para uma amostra de dois elementos. Isso significa que o primeiro elemento a ser sorteado na amostra tem peso menor do que o segundo. Apesar disso, o estimador M_1 também é não viesado, como mostrado a seguir:

⁴ Por exemplo, no caso da cidade de pessoas altas e baixas, como metade da população é de cada tipo, há igual probabilidade de, ao sortear os elementos de uma amostra qualquer, encontrar um alto ou um baixo. Assim, a altura esperada para o elemento da amostra é $(1,60 + 1,80)/2 = 1,70$ m, que é a própria média populacional.

$$E(M_1) = E\left(\frac{2X_1 + 2X_2}{5}\right)$$

$$E(M_1) = \frac{1}{5} [E(2X_1) + E(3X_2)]$$

$$E(M_1) = \frac{1}{5} [2E(X_1) + 3E(X_2)]$$

$$E(M_1) = \frac{1}{5} [2\mu + 3\mu]$$

$$E(M_1) = \frac{1}{5} [5\mu]$$

$$E(M_1) = \mu$$

Portanto, M_1 é um estimador não viesado da média populacional (apesar da ponderação).

Exemplo 6.3.3 (Professor muito rigoroso)

Dado o estimador para a média M_2 , definido a seguir, determine se ele é um estimador viesado e, caso seja, calcule o viés.

$$M_2 = \frac{\sum_{i=1}^n X_i}{n+1}$$

Esse é um estimador em que, em vez de ser dividido pelo número de elementos da amostra, é dividido por um a mais. É como se, por exemplo, para a média final de três provas, as notas fossem somadas e divididas por quatro; ou, se fossem quatro provas, divididas por cinco. Claramente, esse procedimento *joga a média para baixo*.

Calcule a esperança de M_2 :

$$E(M_2) = E\left(\frac{\sum_{i=1}^n X_i}{n+1}\right)$$

$$E(M_2) = \frac{1}{n+1} E\left(\sum_{i=1}^n X_i\right)$$

$$E(M_2) = \frac{1}{n+1} E(X_1 + X_2 + \dots + X_n)$$

$$E(M_2) = \frac{1}{n+1} [E(X_1) + E(X_2) + \dots + E(X_n)]$$

$$E(M_2) = \frac{1}{n+1} [\mu + \mu + \dots + \mu]$$

$$E(M_2) = \frac{n\mu}{n+1} \neq \mu$$

Portanto, M_2 é um estimador viesado da média populacional μ e o viés é dado por:

$$\text{viés}(M_2) = E(M_2) - \mu$$

$$\text{viés}(M_2) = \frac{n\mu}{n+1} - \mu$$

$$\text{viés}(M_2) = \frac{n\mu - (n+1)\mu}{n+1}$$

$$\boxed{\text{viés}(M_2) = -\frac{n\mu}{n+1}}$$

O viés é negativo, pois, como já foi dito, esse estimador *joga para baixo* a média.

6.4 Variância de estimadores: estimadores eficientes

Não basta que um estimador acerte na média. É desejável que, além disso, ele seja o mais preciso possível, não disperse muito ou, em outras palavras, tenha a menor variância possível.

Um estimador é dito absolutamente eficiente, ou simplesmente **eficiente se:**

- for não viesado;
- entre os estimadores não viesados, apresentar a menor variância.

Portanto, para conhecer as propriedades de um estimador, convém que você saiba calcular sua variância. Para a média amostral, a variância é dada por:

$$\text{var}(\bar{X}) = \text{var}\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right)$$

Pelas propriedades da variância, sabe-se que:

$$\text{var}(\bar{X}) = \frac{1}{n^2} \text{var}(X_1 + X_2 + \dots + X_n)$$

Se supuser que cada um dos X_i é independente um do outro, o que é bastante razoável na maioria dos casos, tendo em vista que, se você, por exemplo, estiver calculando a média amostral das idades de algumas pessoas, a idade da primeira pessoa sorteada não afetará a idade da segunda, assim como a idade da segunda não afetará a da terceira, e assim sucessivamente. Nessa hipótese de independência, em que os X_i são independentemente distribuídos, as covariâncias entre X_i e X_j ($i \neq j$) são nulas e, assim, pode-se calcular a variância da soma como sendo a soma das variâncias.

$$\text{var}(\bar{X}) = \frac{1}{n^2} [\text{var}(X_1) + \text{var}(X_2) + \dots + \text{var}(X_n)]$$

Da mesma forma feita para a esperança, a variância que se espera de um elemento que será sorteado de uma população cuja variância é dada por σ^2 é o próprio σ^2 .

$$\text{var}(\bar{X}) = \frac{1}{n^2} [\sigma^2 + \sigma^2 + \dots + \sigma^2]$$

$$\text{var}(\bar{X}) = \frac{1}{n^2} \times n\sigma^2$$

$$\text{var}(\bar{X}) = \frac{\sigma^2}{n}$$

Portanto, a média amostral depende da variância da população, o que é lógico, como vai ver agora. Imagine que a população em questão seja formada pelas crianças matriculadas no segundo ano do ensino fundamental em uma cidade onde, por coincidência, todas as crianças têm a mesma idade. A variância populacional da idade é zero. Qualquer que seja o tamanho da amostra, o valor da média amostral será igual ao da média populacional, portanto, terá variância zero também.

A média amostral depende do tamanho da amostra. Se a amostra for de tamanho 1, o que significa, na prática, que a média é igual aos valores da variável em questão (idade, por exemplo), a variância da média amostral também será igual à variância populacional.

$$n = 1 \Rightarrow \text{var}(\bar{X}) = \frac{\sigma^2}{n} = \sigma^2$$

Por outro lado, se a amostra coincide com a população, o valor da média amostral também coincide com a média populacional (e é exato!) e, portanto, a variância é nula. Como se considerado que a população é muito grande (infinita), então, uma amostra que coincide com a população corresponde a um n tendendo a infinito.

$$n \rightarrow \infty \Rightarrow \text{var}(\bar{X}) = \lim_{n \rightarrow \infty} \frac{\sigma^2}{n} = 0$$

Exemplo 6.4.1

Dado o caso da cidade com as pessoas altas e baixas do Exemplo 6.3.1 e considerando uma média amostral obtida a partir de uma amostra de cinco elementos, verifique que é válida a expressão $\text{var}(\bar{X}) = \frac{\sigma^2}{n}$.

Nessa cidade, metade dos habitantes tem 1,60m e metade, 1,80m. A variância populacional é dada por:

$$\sigma^2 = \text{var}(X) = 0,5 \times (1,80 - 1,70)^2 + 0,5 \times (1,60 - 1,70)^2$$

$$\sigma^2 = 0,5 \times (0,10)^2 + 0,5 \times (-0,10)^2$$

$$\sigma^2 = 0,01$$

Considerando todas as médias amostrais obtidas no Exemplo 6.3.1, a variância da média amostral é dada por:

$$\text{var}(\bar{X}) = \frac{1 \times (1,60 - 1,70)^2 + 5 \times (1,64 - 1,70)^2 + 10 \times (1,68 - 1,70)^2 + 10 \times (1,72 - 1,70)^2 + 5 \times (1,76 - 1,70)^2 + 1 \times (1,80 - 1,70)^2}{32}$$

$$\text{var}(\bar{X}) = 0,002$$

Que é exatamente o valor de σ^2 dividido por cinco (o tamanho da amostra).

$$\text{var}(\bar{X}) = \frac{\sigma^2}{n} = \frac{0,01}{5} = 0,002$$

Exemplo 6.4.2

Determine a variância do estimador M_1 apresentado no Exemplo 6.3.2.

$$M_1 = \frac{2X_1 + 3X_2}{5}$$

Você viu, no Exemplo 6.3.2, que esse é um estimador não viesado, assim como a média amostral. Sua variância é dada por:

$$\text{var}(M_1) = \text{var} \frac{(2X_1 + 3X_2)}{5}$$

Pelas propriedades de variância, vale:

$$\text{var}(M_1) = \frac{1}{25} \text{var}(2X_1 + 3X_2)$$

Considerando que X é distribuído independentemente:

$$\text{var}(M_1) = \frac{1}{25} [\text{var}(2X_1) + \text{var}(3X_2)]$$

$$\text{var}(M_1) = \frac{1}{25} [4\text{var}(X_1) + 9\text{var}(X_2)]$$

$$\text{var}(M_1) = \frac{1}{25} [4\sigma^2 + 9\sigma^2]$$

$$\text{var}(M_1) = \frac{13}{25} \sigma^2 = 0,52\sigma^2$$

Repare que, para uma amostra de dois elementos (que é o caso desse estimador), a variância da média amostral é dada por:

$$\text{var}(\bar{X}) = \frac{\sigma^2}{2} = 0,5\sigma^2$$

Portanto, embora ambos os estimadores sejam não viesados, a média amostral é um estimador melhor do que M_1 , já que possui variância menor.

Não dá para afirmar, entretanto, que (\bar{X}) seja um estimador eficiente da média amostral. Para isso, seria preciso compará-lo com todos os estimadores não viesados da média populacional. É possível, porém, demonstrar que, se a variável X segue uma distribuição normal,⁵ a média amostral (\bar{X}) é um estimador eficiente da média populacional.

Se não se sabe nada sobre a distribuição de X , só dá para dizer que \bar{X} é relativamente mais eficiente do que M_1 .

Portanto, entre dois estimadores não viesados, o mais eficiente é o que apresenta menor variância. Mas e se se compararem dois estimadores quaisquer? Para isso, usa-se o erro quadrático médio.

Erro quadrático médio é a média da diferença entre o valor do estimador e do parâmetro ao quadrado. Assim, para um estimador $\hat{\theta}$, vale:

$$\text{EQM}(\hat{\theta}) = E(\hat{\theta} - \theta)^2$$

Desenvolvendo essa expressão:

$$\text{EQM}(\hat{\theta}) = E(\hat{\theta}^2 - 2q\hat{\theta} + \theta^2)$$

Usando as propriedades da esperança, vem:

$$\text{EQM}(\hat{\theta}) = E(\hat{\theta}^2) - 2E(q\hat{\theta}) + E(\theta^2)$$

Como q é o parâmetro populacional e é, portanto, uma constante:

$$\text{EQM}(\hat{\theta}) = E(\hat{\theta}^2) - 2qE(\hat{\theta}) + \theta^2$$

Somando e subtraindo $[E(\hat{\theta})]^2$, obtém:

$$\text{EQM}(\hat{\theta}) = E(\hat{\theta}^2) - [E(\hat{\theta})]^2 + [E(\hat{\theta})]^2 - 2qE(\hat{\theta}) + \theta^2$$

Os dois primeiros termos dessa expressão correspondem à variância de $\hat{\theta}$, enquanto os três últimos formam um quadrado perfeito:

$$\text{EQM}(\hat{\theta}) = \text{var}(\hat{\theta}) + [E(\hat{\theta}) - \theta]^2$$

⁵ Por meio da desigualdade de Cramer-Rao.

E a expressão entre colchetes é o viés do estimador $\hat{\theta}$. Assim:

$$\text{EQM}(\hat{\theta}) = \text{var}(\hat{\theta}) + [\text{viés}(\hat{\theta})]^2$$

Ou seja, o erro (ao quadrado) do estimador tem dois componentes: o estimador erra o valor do parâmetro em função do quanto varia (sua variância) e ainda, quando for o caso, pelo fato de não acertar na média (ser viesado).

Para dois estimadores quaisquer, $\hat{\theta}_1$ e $\hat{\theta}_2$, se $\hat{\theta}_1$ tem menor erro quadrático médio do que $\hat{\theta}_2$, então $\hat{\theta}_1$ é relativamente mais eficiente do que $\hat{\theta}_2$.

Note que, para dois estimadores não viesados, dizer que o erro quadrático médio é menor equivale a dizer que a variância é menor (já que o viés é nulo).

Exemplo 6.4.3

Determine qual dos estimadores da média dados a seguir é relativamente mais eficiente.

$$M_1 = \frac{2X_1 + 3X_2}{5}$$

$$M_3 = \frac{X_1 + X_2}{3}$$

Para saber qual dos estimadores é relativamente mais eficiente, é preciso calcular o erro quadrático médio de cada um.⁶ Para o estimador M_1 , você já sabe que ele não é viesado e sua variância foi determinada no Exemplo 6.4.2.

$$\text{EQM}(M_1) = \text{var}(M_1) + [\text{viés}(M_1)]^2$$

$$\text{EQM}(M_1) = \text{var}(M_1) + 0$$

$$\text{EQM}(M_1) = 0,52\sigma^2 + 0$$

$$\text{EQM}(M_1) = 0,52\sigma^2$$

Para o estimador M_3 , primeiro deve-se verificar se é um estimador não viesado:

$$E(M_3) = E\left(\frac{X_1 + X_2}{3}\right)$$

$$E(M_3) = \frac{1}{3}E(X_1 + X_2)$$

$$E(M_3) = \frac{1}{3}(\mu + \mu)$$

$$E(M_3) = \frac{2}{3}\mu$$

⁶ Repare que o estimador M_3 é um caso particular do estimador M_2 apresentado no Exemplo 6.3.3, bastando substituir n por 2^a.

Portanto, M_3 é um estimador viesado, e seu viés é dado por:

$$\text{viés}(M_3) = E(M_3) - \mu$$

$$\text{viés}(M_3) = \frac{2}{3}\mu - \mu$$

$$\text{viés}(M_3) = -\frac{1}{3}\mu$$

E sua variância é:

$$\text{var}(M_3) = \text{var}\left(\frac{X_1 + X_2}{3}\right)$$

$$\text{var}(M_3) = \frac{1}{9} \text{var}(X_1 + X_2)$$

$$\text{var}(M_3) = \frac{1}{9} (\sigma^2 + \sigma^2)$$

$$\text{var}(M_3) = \frac{2}{9} \sigma^2$$

Dessa forma, o erro quadrático médio do estimador M_3 é dado por:

$$\text{EQM}(M_3) = \text{var}(M_3) + [\text{viés}(M_3)]^2$$

$$\text{EQM}(M_3) = \frac{2}{9} \sigma^2 + \left[-\frac{1}{3} \mu\right]^2$$

$$\text{EQM}(M_3) = \frac{2}{9} \sigma^2 + \frac{1}{9} \mu^2$$

Como você pode ver, não dá para dizer qual dos dois é relativamente mais eficiente sem que se conheçam os verdadeiros valores de σ e μ .

Se, por exemplo, $\mu = 0$, vale:

$$\text{EQM}(M_3) = \frac{2}{9} \sigma^2 = 0,22 \dots \sigma^2 < \text{EQM}(M_1)$$

Portanto, *nesse caso*, M_3 seria um estimador relativamente mais eficiente do que M_1 . Porém, de modo geral, não se conhece o verdadeiro valor de σ^2 (variância populacional), assim como também se desconhece o valor correto de μ (média populacional). Para se estimar μ , pode-se utilizar a média amostral, que, como você já viu, é um estimador não viesado e eficiente (se a distribuição for normal) da média populacional.

Entretanto, não se tem ainda um estimador para a variância populacional σ^2 .

6.5 Estimador para a variância: variância amostral

Assim como você procedeu para a média, o óbvio seria que o estimador da variância fosse a variância calculada na amostra, isto é:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$$

A primeira questão que surge é: esse estimador ($\hat{\sigma}^2$) é um estimador não viesado da variância populacional (σ^2)?

$$E(\hat{\sigma}^2) = E\left[\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}\right]$$

$$E(\hat{\sigma}^2) = \frac{1}{n} E\left[\sum_{i=1}^n (X_i - \bar{X})^2\right]$$

Com um pequeno artifício: soma-se e se subtraia a média populacional (μ), chega-se a:

$$E(\hat{\sigma}^2) = \frac{1}{n} E\left[\sum_{i=1}^n (X_i - \mu + \mu - \bar{X})^2\right]$$

Tem-se aí um *quadrado da soma*, em que se considera o primeiro termo $X_i - \mu$ e o segundo, $\mu - \bar{X}$.

$$E(\hat{\sigma}^2) = \frac{1}{n} E\left[\sum_{i=1}^n (X_i - \mu)^2 + 2 \sum_{i=1}^n (X_i - \mu)(\mu - \bar{X}) + \sum_{i=1}^n (\mu - \bar{X})^2\right]$$

Como, para qualquer valor do índice i , μ e \bar{X} têm sempre o mesmo valor, pode-se escrever:

$$E(\hat{\sigma}^2) = \frac{1}{n} E\left[\sum_{i=1}^n (X_i - \mu)^2 + 2(\mu - \bar{X}) \sum_{i=1}^n (X_i - \mu) + n(\mu - \bar{X})^2\right]$$

E você sabe que:

$$\sum_{i=1}^n (X_i) = n \bar{X}$$

Portanto:

$$E(\hat{\sigma}^2) = \frac{1}{n} E\left[\sum_{i=1}^n (X_i - \mu)^2 + 2n(\mu - \bar{X})(\bar{X} - \mu) + n(\mu - \bar{X})^2\right]$$

Ou:

$$E(\hat{\sigma}^2) = \frac{1}{n} E\left[\sum_{i=1}^n (X_i - \mu)^2 - 2n(\mu - \bar{X})(\mu - \bar{X}) + n(\mu - \bar{X})^2 \right]$$

$$E(\hat{\sigma}^2) = \frac{1}{n} E\left[\sum_{i=1}^n (X_i - \mu)^2 - 2n(\mu - \bar{X})^2 + n(\mu - \bar{X})^2 \right]$$

$$E(\hat{\sigma}^2) = \frac{1}{n} E\left[\sum_{i=1}^n (X_i - \mu)^2 - n(\mu - \bar{X})^2 \right]$$

Como, numa expressão elevada ao quadrado, o sinal no interior dos parênteses não importa, podemos inverter o sinal da segunda expressão sem problemas:

$$E(\hat{\sigma}^2) = \frac{1}{n} E\left[\sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2 \right]$$

Aplicando a esperança na expressão, vem:

$$E(\hat{\sigma}^2) = \frac{1}{n} \{ E\left[\sum_{i=1}^n (X_i - \mu)^2 \right] - nE(\bar{X} - \mu)^2 \}$$

Como a esperança da soma é a soma das esperanças, vale:

$$E(\hat{\sigma}^2) = \frac{1}{n} \left[\sum_{i=1}^n (X_i - \mu)^2 - nE(\bar{X} - \mu)^2 \right]$$

Mas, pela própria definição de variância:

$$E(X_i - \mu)^2 = \text{var}(X) = \sigma^2 \text{ e}$$

$$E(\bar{X} - \mu)^2 = \text{var}(\bar{X}) = \frac{\sigma^2}{n}$$

Portanto:

$$E(\hat{\sigma}^2) = \frac{1}{n} \left[n\sigma^2 - \frac{n\sigma^2}{n} \right]$$

$$E(\hat{\sigma}^2) = \frac{1}{n} [n\sigma^2 - \sigma^2]$$

$$E(\hat{\sigma}^2) = \frac{1}{n} \sigma^2(n-1)$$

$$E(\hat{\sigma}^2) = \frac{n-1}{n} \sigma^2 \neq \sigma^2$$

Conclui-se, então, que o estimador $\hat{\sigma}^2$ é um estimador viesado da variância populacional σ^2 . Isso, entretanto, pode ser facilmente corrigido se for utilizado um estimador para a variância (chamado de s^2), tal que:

$$S^2 = \frac{n}{n-1} \hat{\sigma}^2$$

$$S^2 = \frac{n}{n-1} \times \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$$

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

Você pode verificar que s^2 é um estimador não viesado da variância populacional σ^2 , pois:

$$E(s^2) = \frac{n}{n-1} E(\hat{\sigma}^2)$$

$$E(s^2) = \frac{n}{n-1} \times \frac{n}{n-1} \sigma^2 = \sigma^2$$

Portanto, para se obter um estimador não viesado da média amostral, deve-se dividir por $n-1$, *não por n*. Qual é a razão disso? A resposta está no artifício que foi utilizado para a demonstração ao se somar e subtrair a média populacional (μ). Não se tem a média populacional, mas a média amostral, ou seja, a média que você utiliza no cálculo da variância é, ela própria, um estimador. Repare que, se a média verdadeira fosse conhecida, o estimador $\hat{\sigma}^2$ não seria viesado.

Imagine que foi escolhida uma amostra de apenas um elemento, o que é perfeitamente viável para a média (ainda que não muito aconselhável), mas isso tornaria impossível uma estimativa não viesada para a variância, pois o valor de $\hat{\sigma}^2$ seria sempre zero para qualquer amostra de qualquer população, o que claramente é viesado. Em outras palavras, só faz sentido estimar a variância em uma amostra que tem, no mínimo, dois elementos.

Assim, de agora em diante, quando falarmos de variância amostral, ou de estimador da variância, estaremos nos referindo a s^2 , a não ser que seja explicitamente dito o contrário.

Exemplo 6.5.1

Em uma fábrica na qual trabalham muitas pessoas, foi perguntado a cinco delas seu salário. As respostas foram R\$ 1.000,00, R\$ 2.000,00, R\$ 1.500,00, R\$ 800,00 e R\$ 700,00. Determine a média amostral, a variância amostral e a variância da média amostral.

A média amostral é dada por:

$$\bar{X} = \frac{1.000 + 2.000 + 1.500 + 800 + 700}{5} = R\$ 1.200$$

A variância amostral (s^2) é:

$$S^2 = \frac{(1.000 - 1.200)^2 + (2.000 - 1.200)^2 + (1.500 - 1.200)^2 + (800 - 1.200)^2 + (700 - 1.200)^2}{4}$$

$$S^2 = 295.000$$

A variância da média amostral seria dada por $\frac{\sigma^2}{n}$, mas, como não se conhece o valor de σ^2 , utiliza-se⁷ seu estimador S^2 .

$$\text{var}(\bar{X}) = \frac{S^2}{n} = \frac{295.000}{5} = 59.000$$

6.6 Melhor estimador linear não viesado

Uma terceira propriedade desejável de um estimador é que ele seja um MELiNV (melhor estimador linear não viesado).⁸

Para ser um MELiNV, o estimador tem de:

- ser não viesado;
- ser linear;
- entre os estimadores lineares e não viesados, apresentar a menor variância.

Um estimador é linear se for obtido por meio de uma combinação linear das observações da amostra. Por exemplo, o estimador mostrado a seguir é linear:

$$\tilde{X} = \sum_{i=1}^n a_i X_i = a_1 X_1 + a_2 X_2 + \dots + a_n X_n$$

Isso vale se cada um dos a_i for uma constante. Claramente, a média amostral é um estimador linear, pois é um caso particular do \tilde{X} exposto anteriormente, em que:

$$a_1 = a_2 = \dots = a_n = \frac{1}{n}$$

E, diga-se de passagem, é um MELiNV, pois não há outro estimador linear com menor variância.

Os conceitos de estimador eficiente e MELiNV são parecidos. De fato, se um estimador eficiente for linear, será um MELiNV, mas um estimador que seja MELiNV pode não ser eficiente se houver um estimador não viesado e não linear que apresente variância menor.

⁷ Portanto, a variância da média amostral a ser calculada é, na verdade, um estimador da variância da média amostral.

⁸ Há quem prefira a sigla MELiNT (trocando o *viesado* por *tendencioso*) ou mesmo a sigla em inglês BLUE (best linear unbiased estimator).

Pode-se dizer, entretanto, que um estimador MELiNV é um estimador eficiente *dentro da classe dos estimadores lineares* (isto é, apresenta menor variância entre os estimadores lineares, mas não necessariamente entre todos).

Resumindo as propriedades vistas até agora:

I. Estimador não viesado

É aquele que, na média, acerta: $E(\hat{\theta}^2) = \theta$.

II. Estimador eficiente

É aquele que, entre os estimadores não viesados, apresenta menor variância.

III. Melhor estimador linear não viesado (MELiNV)

É aquele que, entre os estimadores lineares e não viesados, apresenta menor variância.

6.7 Propriedades assintóticas: estimadores assintoticamente não viesados

Todas essas três propriedades se aplicam a qualquer tamanho de amostra e, em particular, a amostras pequenas.

Quando a amostra cresce (tende ao infinito), há propriedades desejáveis que seriam aplicáveis nesse caso. As propriedades dos estimadores quando o tamanho da amostra tende para o infinito são chamadas de **propriedades assintóticas**.

A primeira propriedade é a de que um estimador seja não viesado. Há estimadores que, embora viesados, quando a amostra cresce, tem seu viés diminuído, isto é, o viés vai desaparecendo à medida que o tamanho da amostra aumenta. Esses estimadores são chamados de **assintoticamente não viesados**.

Um estimador é dito assintoticamente não viesado se:

$$\lim_{n \rightarrow \infty} E(\hat{\theta}^2) = \theta$$

É claro que, se o estimador for não viesado, será assintoticamente não viesado. A recíproca não é verdadeira, como você poderá ver nos exemplos a seguir.

Exemplo 6.7.1

Verifique que o estimador M_2 do Exemplo 6.3.3 é assintoticamente não viesado.

$$M_2 = \frac{\sum_{i=1}^n X_i}{n+1}$$

Como você viu no Exemplo 6.3.3, esse estimador é viesado, pois sua esperança é dada por:

$$E(M_2) = \frac{n\mu}{n+1}$$

Mas, quando a amostra cresce, veja que:

$$\lim_{n \rightarrow \infty} E(M_2) = \lim_{n \rightarrow \infty} \frac{n\mu}{n+1} = \mu$$

Pois, quando n é muito grande, n é praticamente igual a $n + 1$.

Portanto, embora M_2 seja um estimador viesado da média, ele é um estimador assintoticamente não viesado. Isso equivale a dizer que, na prática, se a amostra é grande, tanto faz dividir por n ou $n + 1$, porque a diferença será muito pequena (nula, quando n tende a infinito).

Exemplo 6.7.2

Verifique que $\hat{\sigma}^2$ é um estimador assintoticamente não viesado da variância populacional.

Como você viu na Seção 6.5, σ^2 é um estimador viesado da variância, já que:

$$E(\hat{\sigma}^2) = \frac{n-1}{n} \sigma^2$$

Mas, se for tomado o limite para n tendendo ao infinito:

$$\lim_{n \rightarrow \infty} E(\hat{\sigma}^2) = \lim_{n \rightarrow \infty} \frac{n-1}{n} \sigma^2 = \sigma^2$$

Assim, $\hat{\sigma}^2$ é um estimador assintoticamente não viesado de σ^2 .

De novo, quando a amostra é grande, é praticamente irrelevante se a divisão é feita por n ou por $n - 1$.

6.8 Estimadores consistentes

Um estimador é dito **consistente** se, à medida que a amostra cresce, ele vai convergindo para o valor verdadeiro do parâmetro, ou seja, quando o tamanho da amostra vai aumentando, o viés (se existir) vai sumindo e a variância também. Pode-se dizer que um estimador consistente *colapsa* no valor verdadeiro do parâmetro quando o tamanho da amostra vai para o infinito.

Um *estimador* $\hat{\theta}$ é consistente se:

$$\lim_{n \rightarrow \infty} E(\hat{\theta}) = \theta \text{ e}$$

$$\lim_{n \rightarrow \infty} \text{var}(\hat{\theta}) = 0$$

A média amostral é um estimador consistente da média, pois é um estimador não viesado e:

$$\lim_{n \rightarrow \infty} \text{var}(\bar{X}) = \lim_{n \rightarrow \infty} \frac{\sigma^2}{n} = 0$$

Da mesma forma, você pode verificar que os estimadores dos Exemplos 6.7.1 e 6.7.2 são consistentes.

Uma maneira alternativa de verificar se um estimador é consistente é por meio do erro quadrático médio. Como o erro quadrático médio é composto da variância e do viés ao quadrado, o estimador $\hat{\theta}$ será consistente se:

$$\lim_{n \rightarrow \infty} \text{EQM}(\hat{\theta}) = 0$$

Essa é uma condição suficiente,⁹ mas não necessária, ou seja, se o erro quadrático médio tender a zero com o aumento da amostra, isso implica que o estimador é consistente, mas a recíproca não é verdadeira. Por sorte, os casos em que isso ocorre (o erro quadrático médio não vai para zero, mas o estimador é consistente) são raros.

Exemplo 6.8.1

Verifique se o estimador da média M_4 , dado a seguir, é não viesado e consistente.

$$M_4 = \frac{1}{2} X_1 + \frac{1}{2(n-1)} \sum_{i=1}^n X_i$$

Veja se ele é, ou não, viesado:

$$E(M_4) = E\left[\frac{1}{2} X_1 + \frac{1}{2(n-1)} \sum_{i=1}^n X_i\right]$$

$$E(M_4) = E\left(\frac{1}{2} X_1\right) + E\left[\frac{1}{2(n-1)} \sum_{i=1}^n X_i\right]$$

$$E(M_4) = \frac{1}{2} E(X_1) + \frac{1}{2(n-1)} E(X_2 + X_3 + \dots + X_n)$$

$$E(M_4) = \frac{1}{2} E(X_1) + \frac{1}{2(n-1)} [E(X_2) + E(X_3) + \dots + E(X_n)]$$

$$E(M_4) = \frac{1}{2} \mu + \frac{1}{2(n-1)} [\mu + \mu + \dots + \mu]$$

$$E(M_4) = \frac{1}{2} \mu + \frac{1}{2(n-1)} (n-1)\mu$$

$$E(M_4) = \frac{1}{2} \mu + \frac{1}{2} \mu = \mu$$

⁹ Também se diz que, quando essa condição é válida, o estimador apresenta consistência do erro quadrado. A consistência do erro quadrado implica consistência, mas nem sempre (embora quase sempre) um estimador consistente apresenta consistência do erro quadrado.

Portanto, M_4 é um estimador não viesado da média. Como ele é não viesado, o erro quadrático médio coincide com a variância.

$$\text{EQM}(M_4) = \text{var}(M_4) = \text{var}\left(\frac{1}{2}X_1 + \frac{1}{2(n-1)}\sum_{i=1}^n X_i\right)$$

$$\text{EQM}(M_4) = \text{var}\left(\frac{1}{2}X_1\right) + \text{var}\left(\frac{1}{2(n-1)}\sum_{i=1}^n X_i\right)$$

$$\text{EQM}(M_4) = \frac{1}{4}\text{var}(X_1) + \frac{1}{4(n-1)^2}\text{var}(X_2 + X_3 + \dots + X_n)$$

$$\text{EQM}(M_4) = \frac{1}{4}\sigma^2 + \frac{1}{4(n-1)^2}(\sigma^2 + \sigma^2 + \dots + \sigma^2)$$

$$\text{EQM}(M_4) = \frac{1}{4}\sigma^2 + \frac{1}{4(n-1)^2}(n-1)\sigma^2$$

$$\text{EQM}(M_4) = \frac{1}{4}\sigma^2 + \frac{1}{4(n-1)^2}\sigma^2$$

Quando se toma o limite para n tendendo ao infinito:

$$\lim_{n \rightarrow \infty} \text{EQM}(M_4) = \lim_{n \rightarrow \infty} \left[\frac{1}{4}\sigma^2 + \frac{1}{4(n-1)^2}\sigma^2 \right]$$

O segundo termo vai para zero, pois tem $n - 1$ no denominador, mas o mesmo não ocorre com o primeiro termo. Dessa forma:

$$\lim_{n \rightarrow \infty} \text{EQM}(M_4) = \frac{1}{4}\sigma^2$$

Portanto, M_4 não é consistente, ainda que seja não viesado. Isso poderia ser percebido sem a necessidade de cálculos, tendo em vista que o primeiro elemento a ser sorteado na amostra (X_1) tem peso 50%, não importando o tamanho da amostra. Portanto, ainda que o viés não exista, por maior que seja a amostra, a variância não desaparecerá, tendo em vista o peso desproporcional que tem o primeiro elemento da amostra (dependendo de quem cair primeiro, o valor de M_4 será diferente, ainda que a amostra seja muito grande). Você viu, então, duas propriedades assintóticas:

- I. Estimador assintoticamente não viesado:

$$\lim_{n \rightarrow \infty} E(\hat{\theta}) = \theta$$

- II. Estimador consistente:

Aquele que *colapsa* no verdadeiro valor do parâmetro quando a amostra aumenta.

Condição suficiente: se $\lim_{n \rightarrow \infty} \text{EQM}(\hat{\theta}) = 0$, então $\hat{\theta}$ é consistente.

6.9 Lei dos grandes números

A **lei dos grandes números (LGN)** diz que, quando a amostra cresce (tende a infinito), a média amostral converge para a média populacional. Isto é, quanto maior a amostra, mais o valor obtido pela média amostral estará próximo do valor “correto” da média.

Repare que a LGN equivale à afirmação de que a média amostral é um estimador consistente da média populacional.

6.10 Teorema do limite central

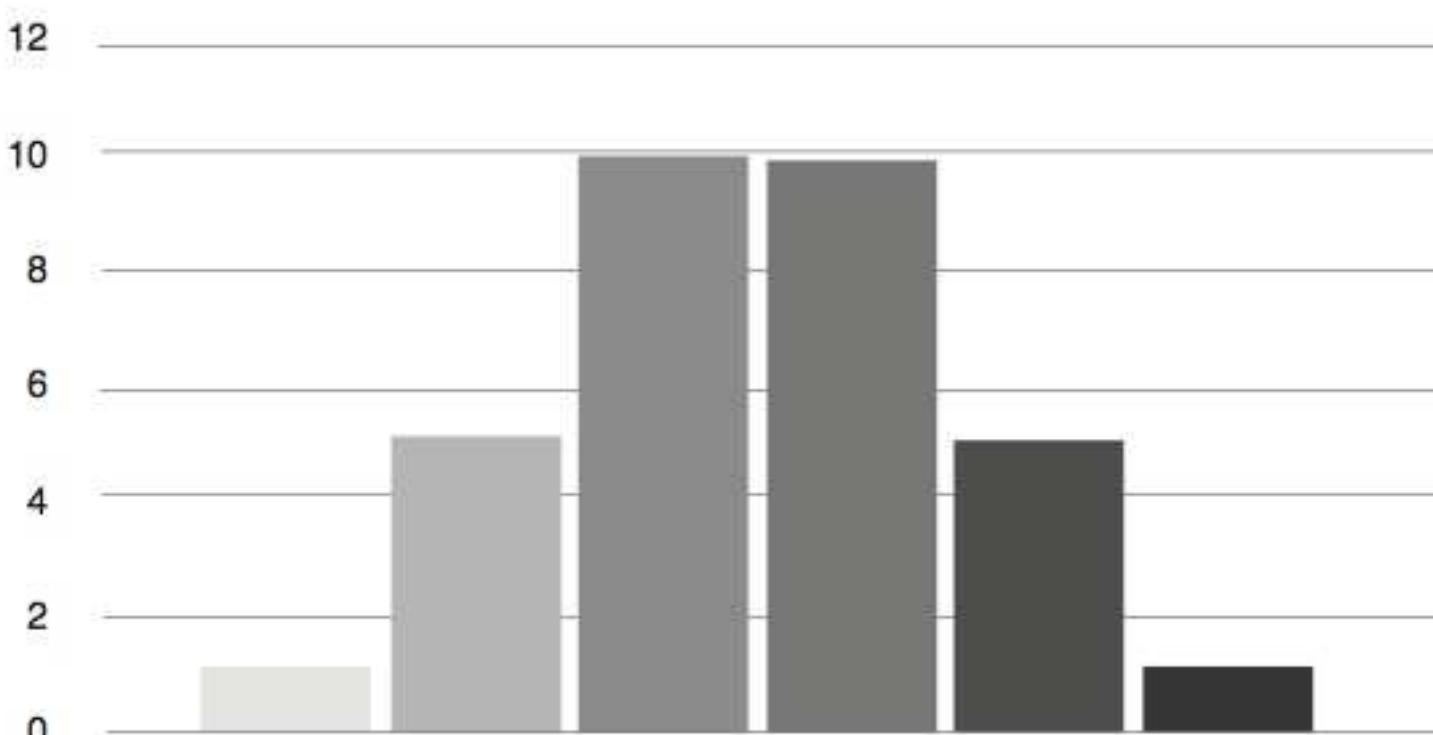
Retome o Exemplo 6.3.1 (aquele da cidade com pessoas altas e baixas). Com amostras de cinco elementos, há 32 possibilidades (já que só há dois resultados possíveis para cada elemento da amostra), sendo essas possibilidades listadas na Tabela 6.2.

Tabela 6.2 – Possibilidades na cidade com pessoas altas e baixas (resumo)

Média amostral obtida (m)	Número de possibilidades
1,60	1
1,64	5
1,68	10
1,72	10
1,76	5
1,80	1

Esses resultados podem ser representados num histograma:

Figura 6.1 – Histograma para a cidade com pessoas altas e baixas



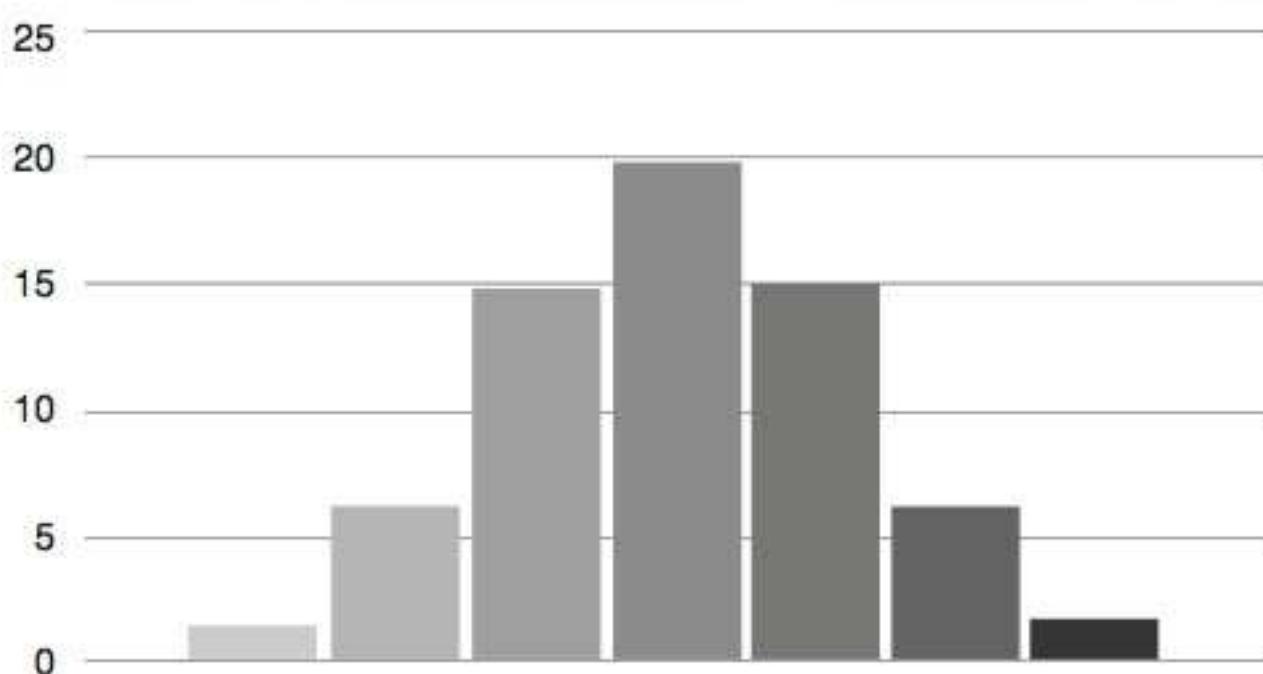
Se o tamanho da amostra for aumentado para seis, como mostra a Tabela 6.3, as possibilidades, num total de $64 = 2^6$, passam a ser (verifique!):

Tabela 6.3 – Possibilidades da cidade com pessoas altas e baixas (amostra de seis elementos)

Média amostral obtida (m)	Número de possibilidades
1,60	1
1,63	6
1,67	15
1,70	20
1,73	15
1,77	6
1,80	1

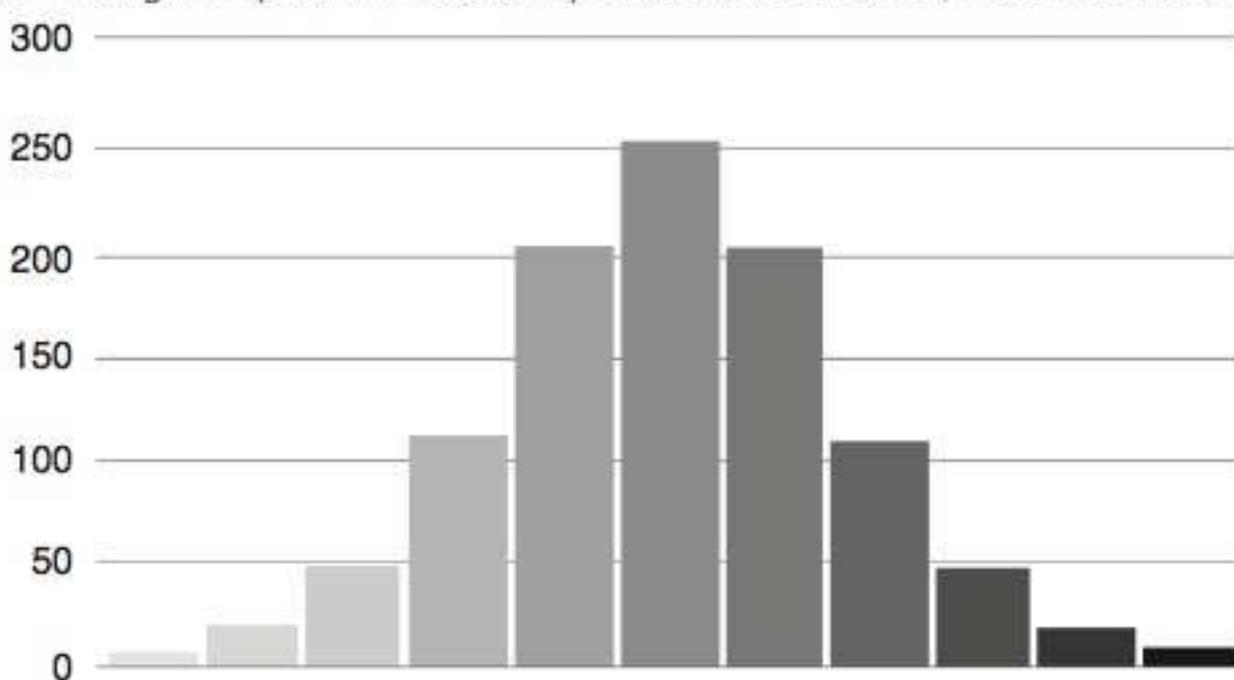
O histograma será, então, o apresentado na Figura 6.2.

Figura 6.2 – Histograma para a cidade com pessoas altas e baixas (amostra de seis elementos)



Se o tamanho da amostra for aumentado para, por exemplo, $n = 10$, agora há $1.024 (= 2^{10})$ possibilidades, o histograma passa a ser:

Figura 6.3 – Histograma para a cidade com pessoas altas e baixas (amostra de dez elementos)



Algo familiar? Pois é, à medida que o tamanho da amostra aumenta, mais o histograma que representa a distribuição da média amostral se aproxima de uma normal. De fato, é isso que diz o **teorema do limite central (TLC)**: dada uma variável X , *independente*¹⁰ e *identicamente*¹¹ distribuída (i.i.d.), com média μ e variância σ^2 , a média amostral \bar{X} segue (desde que a amostra seja suficientemente grande) uma distribuição normal, com média μ e variância $\frac{\sigma^2}{n}$, qualquer que seja a distribuição de X .

Se a variável \bar{X} for padronizada, ou seja, subtrai-se dela a média e é dividida pelo desvio padrão (lembrando que o desvio padrão é dado por $\sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$), obtém-se:

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \sqrt{n} \frac{(\bar{X} - \mu)}{\sigma}$$

Assim, pode-se escrever o TLC em uma única sentença matemática:

$$\sqrt{n} \frac{(\bar{X} - \mu)}{\sigma} \xrightarrow{D} N(0,1)$$

onde a seta com o D em cima se lê *converge em distribuição*. Portanto, essa sentença pode ser lida como $\sqrt{n} \frac{(\bar{X} - \mu)}{\sigma}$ converge em distribuição para uma normal com média 0 e desvio padrão 1.

Os histogramas foram montados baseando-se na cidade apresentada no Exemplo 6.3.1, mas o resultado seria o mesmo qualquer que fosse a distribuição utilizada. O TLC permite afirmar que, se for média, é normal.

Quanto ao tamanho de amostra suficientemente grande, é comum se utilizar uma “receita de bolo”: usar uma amostra de no mínimo 30 elementos. Na verdade, o que se deve levar em conta é que a distribuição da média amostral seja aproximadamente uma normal e que essa aproximação seja tão melhor quanto maior for a amostra. Se for usada uma amostra muito pequena, não é que a aproximação não seja válida, mas será muito grosseira.

Exemplo 6.10.1

Uma variável X tem média igual a 10 e variância igual a 144. Qual a probabilidade de que, numa amostra com 36 elementos, encontrar uma média amostral superior a 11?

¹⁰ Significa que os diversos X_i são independentes uns dos outros.

¹¹ Significa que os mesmos parâmetros da distribuição (seja ela qual for) se aplicam a todos os X_i .

Sabe-se que:

$$E(\bar{X}) = 10$$

$$\text{var}(\bar{X}) = \frac{144}{36} = 4$$

Pelo TLC, você sabe que a média amostral segue uma distribuição normal com média 10 e desvio padrão 2 ($=\sqrt{4}$). Você quer saber a probabilidade de \bar{X} ser maior do que 11. Padronizando (para que possa consultar a tabela do final deste livro):

$$Z = \frac{11 - 10}{2} = 0,5$$

Portanto:

$$P(\bar{X} > 11) = P(Z > 0,5) = 0,5 - 0,1915 = 0,3085 = 30,85\%$$

Exemplo 6.10.2

Recentemente uma mulher foi impedida de voar por uma companhia aérea norte-americana. Num voo com 100 passageiros, qual a probabilidade de que, tendo um passageiro pesando 120kg, o peso total dos passageiros ultrapasse 10% do esperado? Considere, para isso, que o peso médio das pessoas seja 70kg com desvio padrão de 15kg.

O teorema do limite central vale para a média amostral e, portanto, vale para a soma amostral também. Como a soma é simplesmente a média multiplicada por n (*o tamanho da amostra*), se a média amostral tem distribuição normal, a soma amostral também tem. A média e variância da soma amostral são:

$$E\left[\sum_{i=1}^n X_i\right] = E[X_1 + X_2 + \dots + X_n] = \mu + \mu + \dots + \mu = n\mu.$$

$$\text{var}\left[\sum_{i=1}^n X_i\right] = \text{var}[X_1 + X_2 + \dots + X_n] = \sigma^2 + \sigma^2 + \dots + \sigma^2 = n\sigma^2$$

Portanto, a soma amostral dos pesos tem distribuição normal (e você já sabe isso por causa do teorema do limite central) com média $n\mu$ e variância $n\sigma^2$.

O valor esperado da soma dos 100 passageiros é 7.000kg. Considerando a margem de segurança de 10%, o avião aguentaria até 7.700kg. Levando em conta que um passageiro tem 120kg, o peso dos restantes 99 não pode ultrapassar $7.700 - 120 = 7.580$ kg.

O valor esperado (média) dos 99 passageiros é $99 \times 70 = 6.930$ kg e a variância $99 \times 15^2 = 22.275$. O desvio padrão é, portanto, aproximadamente 149,25kg.

Você deve calcular a probabilidade de a soma (s) ser maior do que 7.580kg. Para utilizar a tabela normal, deve-se padronizar esse valor:

$$Z = \frac{7580 - 6930}{149,25} \cong 4,36.$$

$$P(s > 7580) = P(Z > 4,36) = 0,5 - 0,499993 = 0,000007 = 0,0007\%.$$

A **probabilidade é ínfima**. A truculência do funcionário foi absolutamente desnecessária. Claro que ela foi baseada em algumas suposições, exceto a média do peso das pessoas, informação tirada de vários elevadores espalhados pelos edifícios. Você pode refazer o cálculo supondo diferentes médias, desvios e margens de segurança. Por exemplo, em nosso cálculo consideramos que todos os passageiros são adultos. Se houver crianças, a média será menor, mas o desvio padrão será maior. Ah, e tudo foi feito considerando-se apenas o peso dos passageiros. Nada de bagagem, por exemplo.

6.11 População finita

Por **população finita** entende-se, na prática, uma população cujo tamanho é comparável com o da amostra a ser estudada.

No caso de uma pesquisa eleitoral em que mil, 2 mil eleitores são pesquisados em uma população de milhões, a amostra é muito pequena em relação à população. Essa não é, a rigor, infinita, mas, para efeitos práticos, é como se fosse.

O mesmo não ocorre se, por exemplo, em uma escola com mil alunos, toma-se uma amostra de 50 ou, em uma fazenda com 200 cabeças de gado, utiliza-se uma amostra de 20.

No primeiro caso, a amostra representa 5% da população; no segundo, 10%; é em casos como esses que se considera a população finita.

Mas qual é a diferença? É que, quando você calcula a variância da média amostral, assume que a variância esperada de cada elemento da amostra é igual à variância populacional σ^2 . Ocorre que, quando se retira o primeiro elemento da amostra, a variância dos que sobram foi alterada. Portanto, a variância esperada do segundo elemento da amostra (bem como de todos os outros) não será σ^2 . Se a população é “infinita” (na prática, se é muito maior do que a amostra), a retirada de um elemento não terá efeitos sobre a variância dos demais.

Repare que esse raciocínio da população finita não se aplica se a amostra for retirada com reposição. Portanto, se a população for infinita ou mesmo se for finita, desde que a amostra seja retirada *com reposição*, é válida a expressão:

$$\text{var}(\bar{X}) = \frac{\sigma^2}{n}$$

Agora, se a população for *finita* e a amostra retirada *sem reposição*, essa expressão precisa ser corrigida. Se a população tem tamanho igual a N, a variância da média amostral é dada por:

$$\text{var}(\bar{X}) = \frac{\sigma^2}{n} \times \frac{N-n}{N-1}$$

Repare que, se o tamanho da amostra (n) é muito pequeno em relação ao tamanho da população (N), o fator de correção $\frac{N-n}{N-1}$ é praticamente igual a 1 e, dessa forma, a expressão da variância da média amostral é praticamente a mesma da utilizada quando a população é infinita. Se o tamanho da amostra é igual ao da população ($n = N$), a média amostral é igual à média populacional e a variância de \bar{X} é nula.

Exemplo 6.11.1

Numa classe de 50 alunos, são escolhidos, ao acaso, cinco alunos para realizar um teste, cujas notas vão de 0 a 100, para aferir o aproveitamento da turma. Se o desvio padrão histórico dessa turma em testes desse tipo é 12, determine a variância e o desvio padrão da média amostral nesse teste.

Como se trata de uma população finita e a amostragem é feita sem reposição, e assumindo que o desvio padrão populacional se mantém no valor histórico, então:

$$\text{var}(\bar{X}) = \frac{\sigma^2}{n} \times \frac{N-n}{N-1}$$

$$\text{var}(\bar{X}) = \frac{12^2}{5} \times \frac{50-5}{50-1}$$

$$\text{var}(\bar{X}) = \frac{144}{5} \times \frac{45}{49}$$

$$\text{var}(\bar{X}) \cong 26,45$$

$$\text{dp}(\bar{X}) \equiv \hat{\sigma}_{\bar{x}} = \sqrt{\text{var}(\bar{X})}$$

$$\hat{\sigma}_{\bar{x}} = \sqrt{26,45}$$

$$\hat{\sigma}_{\bar{x}} \cong 5,14$$

6.12 Métodos de estimação

6.12.1 Método dos momentos

Ao se usar a média amostral como estimador da média populacional, isso, por si só, já é um método, conhecido como método dos momentos.

Esse método consiste em utilizar, como estimador de certo momento populacional, o respectivo momento amostral.

Assim, utiliza-se a média amostral como estimador da média populacional, a variância amostral como estimador da variância populacional e assim por diante.

6.12.2 Método dos mínimos quadrados

O método dos mínimos quadrados consiste em minimizar a soma dos quadrados dos desvios (diferenças). O princípio é semelhante ao adotado para definir a variância. Se forem somados simplesmente os desvios, os valores positivos e os negativos se anularão. Portanto, deve-se, de alguma forma, eliminar os sinais negativos. Uma possibilidade é tomar os desvios ao quadrado¹²,

Assim, dada uma amostra $\{x_1, x_2, x_3, \dots, x_n\}$, a soma dos quadrados dos desvios em relação à média é:

$$S = (x_1 - \mu)^2 + (x_2 - \mu)^2 + (x_3 - \mu)^2 + \dots + (x_n - \mu)^2$$

$$S = \sum_{i=1}^n (x_i - \mu)^2$$

A questão é: qual o estimador de μ que minimiza a soma dos quadrados dos desvios, ou seja, que minimiza S ? A resposta é obtida derivando-se S em relação a μ e igualando a zero:

$$\frac{\partial S}{\partial \mu} = -2 \sum_{i=1}^n (x_i - \hat{\mu}) = 0$$

Note que o símbolo μ foi trocado por $\hat{\mu}$, pois, se o valor que minimiza S é obtido da amostra, esse portanto será um **estimador** de μ .

$$\sum_{i=1}^n (x_i - \hat{\mu}) = 0$$

$$\sum_{i=1}^n x_i - \sum_{i=1}^n \hat{\mu} = 0$$

Como $\hat{\mu}$ é uma constante:

$$\sum_{i=1}^n x_i - n\hat{\mu} = 0$$

¹² Outra seria usar o módulo, ou seja, em vez dos desvios ao quadrado, os desvios absolutos.

$$-n\hat{\mu} = -\sum_{i=1}^n x_i$$

$$n\hat{\mu} = \sum_{i=1}^n x_i$$

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{X}.$$

O estimador que minimiza a soma dos quadrados dos desvios em relação à média – o estimador de mínimos quadrados da média – é a média amostral.

6.12.3 Estimação por máxima verossimilhança

O princípio da estimação por máxima verossimilhança é o seguinte: se for conhecida a distribuição de probabilidade da população¹³, os valores dos parâmetros estimados são os que maximizarão a chance (a probabilidade, a verossimilhança) de que os valores obtidos na amostra sigam, de fato, a distribuição em questão.

Suponha que uma variável aleatória x tenha uma função densidade de probabilidade dada por:

$$f.d.p. \text{ de } x = f(x_i; \theta_k)$$

Nessa notação, depois do ponto e vírgula estão os parâmetros da função, isto é, f é uma função dos valores de x_i (até aí, nenhuma novidade), dados os parâmetros da distribuição, θ_k , supostamente conhecidos.

Por exemplo, para uma distribuição normal, os parâmetros são a média e a variância (ou o desvio padrão). Se você conhecer ambos, dado um valor de x , é fácil calcular o valor de f .

E se os parâmetros não forem conhecidos? Você tem os valores de x , que obteve de uma amostra, e precisa estimar os parâmetros. A função agora depende dos parâmetros θ . Quando é assim, a função passa a ser chamada de função de verossimilhança:

$$\text{função de verossimilhança} = L(\theta_k; x_i)$$

A estimação por máxima verossimilhança consiste em achar os valores dos parâmetros θ_k que maximizem a função de verossimilhança ou, em outras palavras, que maximize a probabilidade de que a amostra pertença, de fato, a uma população cuja distribuição de probabilidade tem função de densidade¹⁴ dada por f .

¹³ Essa condição é absolutamente necessária para que se possa fazer uma estimação por máxima verossimilhança.

¹⁴ Note que a função de verossimilhança e a função densidade de probabilidade têm a mesma forma funcional. Invertendo-se a lógica: enquanto a f.d.p. é uma função dos valores da variável aleatória x , sendo dados os parâmetros, a função de verossimilhança é uma função dos parâmetros, sendo dados os valores de x .

Exemplo 6.12.1

Uma variável aleatória x tem distribuição normal (*independentemente distribuída*) com média e variância desconhecidas. Dada uma amostra $\{x_1, x_2, \dots, x_n\}$, determine os estimadores de máxima verossimilhança para a média e a variância.

Se a distribuição é normal, então a função de verossimilhança terá a mesma forma funcional de uma normal multivariada (veja Capítulo 5):

$$L(\mu, \sigma^2; x_i) = \frac{1}{(2\sigma^2)^{\frac{n}{2}}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right]$$

Em que $\exp(x) \equiv e^x$.

Os valores de μ e σ^2 serão obtidos pela maximização da função de verossimilhança L . Mas essa função é um pouquinho *complicada*. Para simplificar o trabalho, lembre que, quando uma função sofre uma transformação monotônica¹⁵ crescente, a função resultante tem os mesmos pontos de máximo e/ou mínimo.

Tire, então, o logaritmo de L :

$$l(\mu, \sigma^2; x_i) \circ \ln[L(\mu, \sigma^2; x_i)] = \ln \left\{ \frac{1}{(2\sigma^2)^{\frac{n}{2}}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right] \right\}$$

$$l(\mu, \sigma^2; x_i) = \ln \left(\frac{1}{(2\sigma^2)^{\frac{n}{2}}} - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right)$$

$$l(\mu, \sigma^2; x_i) = -\ln(2\sigma^2)^{\frac{n}{2}} - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

$$l(\mu, \sigma^2; x_i) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

Para encontrar o ponto de máximo dessa função, você deve encontrar as derivadas de l em relação a μ e σ^2 .

Derivando em relação a μ , vem:

$$\frac{\partial l}{\partial \mu} = -\frac{1}{2\sigma^2} 2 \sum_{i=1}^n (x_i - \hat{\mu}) = 0$$

$$\sum_{i=1}^n (x_i - \hat{\mu}) = 0$$

¹⁵ Sempre crescente ou sempre decrescente.

$$\sum_{i=1}^n x_i - \sum_{i=1}^n \hat{\mu} = 0$$

Como m é uma constante:

$$\sum_{i=1}^n x_i - n\hat{\mu} = 0$$

$$\hat{\mu} = \frac{\sum_{i=1}^n x_i}{n}$$

Ou seja, o estimador de máxima verossimilhança da média de uma distribuição normal é a própria média amostral \bar{X} .

Derivando em relação a σ^2 e já incluindo esse resultado, vem:

$$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2} \frac{1}{\hat{\sigma}^2} + \frac{1}{4\hat{\sigma}^2} \sum_{i=1}^n (x_i - \bar{x})^2 = 0$$

$$-n\hat{\sigma}^2 + \sum_{i=1}^n (x_i - \bar{x})^2 = 0$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

Portanto, o estimador de máxima verossimilhança para σ^2 é, como você já viu, viesado. Conclui-se que o fato de o estimador ser de máxima verossimilhança não garante que ele seja não viesado. Os estimadores de máxima verossimilhança têm, entretanto, algumas propriedades muito úteis:

- são consistentes;
- têm distribuição assintótica normal;
- são assintoticamente eficientes.¹⁶

Exemplo 6.12.2

Uma variável aleatória x tem distribuição uniforme. Dada a amostra $\{x_1, x_2, \dots, x_n\}$, determine os estimadores de máxima verossimilhança para os parâmetros da distribuição.

Uma distribuição uniforme apresenta uma função densidade $f(x) = \frac{1}{b-a}$, para $a \leq x \leq b$. Os parâmetros que precisam ser encontrados são justamente a e b , que são os valores mínimo e máximo, respectivamente, que a variável x pode apresentar.

¹⁶ Essa propriedade está mais bem explicada no Apêndice 6.B.2.

Os valores da amostra que têm a maior chance de ser esses valores são justamente o mínimo e o máximo valor encontrado na amostra. Assim, os estimadores de máxima verossimilhança para a e b são:

$$a = \min \{x_1, x_2, \dots, x_n\}$$

$$b = \max \{x_1, x_2, \dots, x_n\}$$

Exemplo 6.12.3

Uma variável aleatória Y tem distribuição binomial com parâmetro p . Em uma amostra de N elementos, x_i apresentaram o atributo sucesso. Determine o estimador de máxima verossimilhança para p .

Faça $Y = \sum_{i=1}^n x_i$, de tal modo que x tenha distribuição de Bernoulli com o mesmo parâmetro p . Então:

$$P(x_i = 1) = p$$

$$P(x_i = 0) = 1 - p$$

O que pode ser escrito, numa única linha, assim:

$$P(x_i) = p^x(1-p)^{1-x}.$$

A probabilidade conjunta de uma amostra de n desses x , independentes entre si, é dada por:

$$P(x_1, x_2, \dots, x_n) = p^{x_1}(1-p)^{1-x_1} \times p^{x_2}(1-p)^{1-x_2} \times \dots \times p^{x_n}(1-p)^{1-x_n}$$

$$P(x_1, x_2, \dots, x_n) = \prod_{i=1}^n p^{x_i}(1-p)^{1-x_i}.$$

A função de verossimilhança, como de costume, tem a mesma forma, mas o argumento passa a ser p .

$$L(p; x_i) = \prod_{i=1}^n p^{x_i}(1-p)^{1-x_i}.$$

Cujo logaritmo é dado por:

$$l(p; x_i) = \ln \left[\prod_{i=1}^n p^{x_i}(1-p)^{1-x_i} \right]$$

$$l(p; x_i) = \sum_{i=1}^n \ln [p^{x_i}(1-p)^{1-x_i}]$$

$$l(p; x_i) = \sum_{i=1}^n x_i \ln(p) + \sum_{i=1}^n (1-x_i) \ln(1-p)$$

$$l(p; x_i) = \ln(p) \sum_{i=1}^n x_i + \ln(1-p) \sum_{i=1}^n (1-x_i)$$

Derivando em relação a p e igualando a zero, chega-se a:

$$\frac{\partial l(p; x_i)}{\partial p} = \frac{1}{\hat{p}} \sum_{i=1}^n x_i - \frac{1}{1-\hat{p}} \sum_{i=1}^n (1-x_i) = 0$$

$$\frac{1}{\hat{p}} \sum_{i=1}^n x_i = \frac{1}{1-\hat{p}} \sum_{i=1}^n (1-x_i)$$

$$(1-\hat{p}) \sum_{i=1}^n x_i = \hat{p} \sum_{i=1}^n (1-x_i)$$

$$1 \sum_{i=1}^n x_i - \hat{p} \sum_{i=1}^n x_i = \hat{p} \sum_{i=1}^n 1 - \hat{p} \sum_{i=1}^n x_i$$

$$\sum_{i=1}^n x_i - \hat{p} \sum_{i=1}^n x_i = n\hat{p} - \hat{p} \sum_{i=1}^n x_i$$

$$\sum_{i=1}^n x_i = n\hat{p}$$

$$n\hat{p} = \sum_{i=1}^n x_i$$

$$\hat{p} = \frac{\sum_{i=1}^n x_i}{n} = \frac{Y}{n}.$$

O valor amostral para p que dá a maior chance de essa amostra pertencer a uma população com essas características é justamente a proporção amostral.

O estimador de máxima verossimilhança é, portanto:

$$\hat{p} = \frac{Y}{N}.$$

Exercícios

- Para as amostras dadas a seguir, determine a média amostral, a variância amostral e a variância da média amostral:
 - {2; 4; 6; 9; 12};
 - {1,6; 1,8; 1,9; 2,1; 1,5; 1,7};
 - {1.000; 1.200; 1.300; 1.600; 900; 700; 1.400}.

Enunciado para os Exercícios 2 a 6:

A variável aleatória X tem média μ e variância σ^2 . Um pesquisador resolve utilizar os seguintes estimadores para a média:

$$M_1 = \frac{X_1 + 2X_2}{4}$$

$$M_2 = \frac{3X_1 + 4X_2}{7}$$

2. Determine quais estimadores são viesados e o viés, se houver.
3. Determine a variância dos estimadores.
4. Determine o erro quadrático médio dos estimadores.
5. Suponha que $\mu = 0$. Qual dos estimadores é relativamente mais eficiente?
6. Suponha agora que $\mu = 10$ e $\sigma = 2$. Nessas condições, qual é o estimador relativamente mais eficiente?

Enunciado para os Exercícios 7 a 13:

A variável aleatória X tem média μ e variância σ^2 . Um pesquisador resolve utilizar os seguintes estimadores para a média:

$$M_3 = \frac{\sum_{i=1}^n X_i}{n-2}$$

$$M_4 = \frac{1}{2}x_1 + \frac{\sum_{i=2}^n X_i}{n-1}$$

7. Determine quais estimadores são viesados e o viés, se houver.
8. Determine a variância dos estimadores.
9. Determine o erro quadrático médio dos estimadores.
10. Suponha que $\mu = 0$ e $n = 10$. Qual dos estimadores é relativamente mais eficiente?
11. Suponha agora que $n = 10$, $\mu = 12$ e $\sigma = 3$. Nessas condições, qual é o estimador relativamente mais eficiente?
12. Determine quais estimadores são assintoticamente não viesados.
13. Determine se os estimadores apresentam consistência do erro quadrado.

14. Uma variável aleatória X tem média 12 e desvio padrão 6. Determine a média e a variância de uma variável Y definida a partir de uma amostra de 10 elementos da variável X, como se segue:

$$Y = \sum_{i=1}^{10} X_i$$

15. Uma variável aleatória X tem média 9 e desvio padrão 2. Determine a média e a variância de uma variável W definida a partir de uma amostra de 5 elementos da variável X, como se segue:

$$W = \frac{\sum_{i=1}^5 iX_i}{\sum_{i=1}^5 i}$$

16. Uma variável aleatória X tem média 20 e variância 64. Determine a probabilidade de que, em uma amostra de 49 elementos, a média amostral seja inferior a 18.

17. Uma variável aleatória X tem distribuição de Poisson com parâmetro 9. Determine a probabilidade de que, em uma amostra de 36 elementos, a média amostral esteja entre 8 e 10.

18. Uma variável aleatória X tem distribuição binomial em que a proporção de sucessos é 0,8. Determine a probabilidade de que, em uma amostra de 100 elementos, encontrem-se menos de 75 sucessos.

19. Em uma classe de 50 alunos, foi retirada uma amostra de 5. As notas desses alunos foram, respectivamente, 7, 5, 3, 8 e 5. Determine a média amostral, a variância amostral e a variância da média amostral.

Utilize a amostra a seguir, supondo que os valores são independentes, para os Exercícios 20 a 22:

{25, 30, 28, 29, 32, 35, 21, 33, 26, 27}

20. Suponha que essa amostra foi retirada de uma população cuja distribuição é normal. Estime os parâmetros da distribuição por máxima verossimilhança.

21. Suponha que essa amostra foi retirada de uma população cuja distribuição é uniforme. Estime os parâmetros da distribuição por máxima verossimilhança.

22. Suponha que essa amostra foi retirada de uma população cuja distribuição é exponencial. Estime os parâmetros da distribuição por máxima verossimilhança.

23. Seja x_1, x_2, \dots, x_n uma amostra aleatória simples proveniente de uma população com distribuição de Poisson. Determine o estimador de máxima verossimilhança do parâmetro λ .

24. Seja x_1, x_2, \dots, x_n uma amostra aleatória simples proveniente de uma população com distribuição cuja função densidade é dada por $f(x) = \theta(1+x)^{-\theta+1}$, $0 < x < \infty$, $\theta > 1$. Determine o estimador de máxima verossimilhança para θ .

25. Mostre que o estimador dado no exemplo 6.12.3 é, de fato, o estimador de máxima verossimilhança da distribuição binomial. (*Sugestão: faça $Y = \sum_{i=1}^n X_i$, onde cada elemento da amostra X_i tem distribuição de Bernoulli.*)
26. Assinale verdadeiro ou falso.
- A média amostral é um estimador viesado para a média populacional quando a amostra é muito pequena.
 - A média amostral é um estimador eficiente para a média populacional.
 - Embora σ^2 seja um estimador viesado para a variância populacional, sua variância é menor do que a de S^2 .
 - Todo estimador não viesado é consistente.
 - Todo estimador viesado é inconsistente.
 - Todo estimador consistente é não viesado.
 - Todo estimador eficiente é não viesado.
 - Dados dois estimadores, um deles viesado e outro não, esse último será sempre preferível.
 - Dados dois estimadores, um deles viesado e outro não, esse último terá sempre menor erro quadrático médio.
 - A variância da média em uma população finita é igual à de uma população infinita, desde que a amostragem tenha sido feita com reposição.
 - Para se fazer uma estimação por máxima verossimilhança, é necessário saber qual é a distribuição populacional.
 - Um estimador de máxima verossimilhança é sempre não viesado.
 - Um estimador de máxima verossimilhança é sempre consistente.
 - A lei dos grandes números garante que a média amostral segue uma distribuição assintótica normal.
 - A lei dos grandes números garante que a média amostral é um estimador consistente da média amostral.
 - A média amostral segue uma distribuição normal para qualquer tamanho de amostra.
27. Suponha que X_1 e X_2 sejam s independentes, com média μ e variâncias $V(Y_1) = 60$ e $V(Y_2) = 20$. O valor de μ é desconhecido e é proposto estimar μ por uma média ponderada de Y_1 e Y_2 , isto é, por: $\alpha X_1 + (1-\alpha)X_2$. Qual valor de α para que esse estimador seja MELiNV?

Apêndice 6.B – Convergências e mais propriedades de estimadores

6.B.1 Convergências

Dado um estimador $\hat{\theta}$ de um parâmetro populacional θ . Você viu que, se:

$$\lim_{n \rightarrow \infty} P(|\hat{\theta} - \theta| < \varepsilon) = 1$$

diz-se que $\hat{\theta}$ converge em probabilidade para θ ou:

$$\hat{\theta} \xrightarrow{p} \theta$$

Se o estimador $\hat{\theta}$ converge para θ de outra forma, como mostrado a seguir:

$$P(\lim_{n \rightarrow \infty} \hat{\theta} = \theta) = 1$$

Então se diz que $\hat{\theta}$ apresenta **convergência quase certa** para θ , ou **convergência com probabilidade 1** para θ , o que é representado por:

$$\hat{\theta} \xrightarrow{qc} \theta$$

Note que a convergência quase certa implica convergência em probabilidade, mas a recíproca não é verdadeira, isto é, a convergência quase certa é mais “forte” do que a convergência em probabilidade.

No caso da média amostral como estimador da média populacional: você viu que a lei dos grandes números estabelece que a média amostral converge para a média populacional à medida que a amostra cresce. Essa lei, entretanto, aparece em duas versões, de acordo com o tipo de convergência.

A **lei fraca dos grandes números** estabelece que a média amostral converge em probabilidade para a média populacional, enquanto a **lei forte dos grandes números** propõe que a média amostral converge quase certamente para a média populacional.

$$\text{LGN versão fraca: } \bar{X} \xrightarrow{p} \mu$$

$$\text{LGN versão forte: } \bar{X} \xrightarrow{qc} \mu$$

As condições para que se verifique a lei forte são mais restritas. Para que se verifique a lei fraca, basta que os X_i ($i = 1, 2, \dots, n$) sejam uma sequência de números aleatórios com variância finita, mas não necessariamente independentes. Para que se verifique a lei forte, é necessário que os X_i sejam IID (independentes e identicamente distribuídos).

6.B.2 Eficiência assintótica

Neste capítulo foram definidas duas propriedades assintóticas desejáveis de estimadores: ser assintoticamente não viesado e consistência.

Para um estimador $\hat{\theta}$ de um parâmetro populacional θ , define-se a variância assintótica como:

$$\text{var-ass}(\hat{\theta}) = \frac{1}{n} \lim_{n \rightarrow \infty} E[\sqrt{n} (\hat{\theta} - \lim_{n \rightarrow \infty} E(\hat{\theta}))]^2$$

O que, no caso de estimadores assintoticamente não viesados, se reduz a:

$$\text{var-ass}(\hat{\theta}) = \frac{1}{n} \lim_{n \rightarrow \infty} E[\sqrt{n} (\hat{\theta} - \theta)]^2$$

O estimador $\hat{\theta}$ tem a propriedade de **eficiência assintótica** se:

- apresenta distribuição assintótica com média e variância finitas;
- é consistente;
- entre os estimadores consistentes de θ , ele for o que apresentar menor variância assintótica.



Capítulo 7

INTERVALO DE CONFIANÇA E TESTES DE HIPÓTESES

7.1 Intervalo de confiança

Em geral, cada dois anos, as pessoas acompanham as pesquisas eleitorais. Elas costumam ser mostradas assim:

Candidato	Intenção de voto
João da Silva	35%
Maria Aparecida	32%
José Severino	16%

Na maioria das vezes essas informações são acompanhadas da famosa *margem de erro* da pesquisa. Suponha que, para o caso dessa pesquisa, ela seja de dois pontos percentuais para cima ou para baixo, o que vale dizer que o candidato João da Silva tem entre 33% e 37% das intenções de voto, enquanto Maria Aparecida tem entre 30% e 34%.

Portanto, embora o mais provável é que o candidato João da Silva esteja *ganhando*, é possível que ele tenha 33% dos votos enquanto sua adversária direta tenha 34%, estando assim ela, e não ele, na frente da corrida eleitoral. Em resumo, não dá para afirmar quem está na frente, é o famoso *empate técnico* entre os candidatos.

Porém dá para ter certeza de que João da Silva tem no mínimo 33% dos votos e no máximo 37%? Ora, essa informação foi obtida por meio de uma amostra que, ainda que grande, é pequena em relação ao total da população. Mesmo que a amostragem tenha sido feita de maneira correta, é possível (por mais que seja pouco provável) que a amostra contenha, por coincidência, um número exageradamente grande (ou pequeno) de eleitores do referido candidato. Assim, falta uma informação referente a quanto esses valores, mesmo incluindo a margem de erro, são confiáveis. Nem sempre essa informação é omitida quando da divulgação das pesquisas. Por vezes, ela pode ser encontrada na imprensa escrita (embora dificilmente na manchete).

Construir um **intervalo de confiança** nada mais é do que estabelecer uma *margem de erro* para um estimador e calcular o grau de confiança correspondente a essa margem. Ou, como é mais comum, estabelecido um grau de confiança, calcular a margem de erro que corresponda a essa confiança. Como se faz isso? É necessário que se conheça a distribuição de probabilidade do estimador.

Exemplo 7.1.1

Numa amostra de 100 estudantes foi encontrada uma idade média de 23,2 anos. Sabendo-se que a variância das idades é 25, construa um intervalo de 95% de confiança para a média.

Pelo teorema do limite central visto no capítulo anterior, sabe-se que a *média* segue uma distribuição que se aproxima da normal (e 100 é um tamanho de amostra suficientemente grande).

A variância da média amostral, vista no capítulo anterior, é dada por:

$$\text{var}(\bar{X}) = \frac{\text{var}(X)}{n}$$

Ou, abreviando mais a notação:

$$\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$$

O desvio padrão da média amostral pode ser calculado diretamente por:

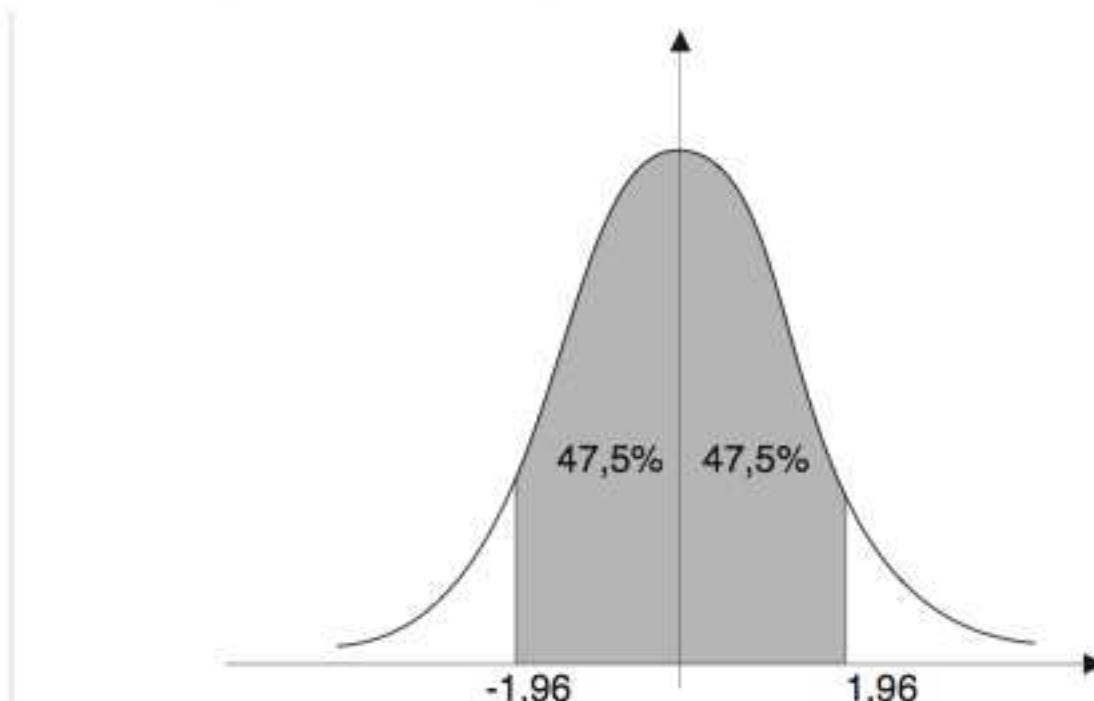
$$\sigma_{\bar{X}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

Cujo valor, nesse caso, é dado por:¹

$$\sigma_{\bar{X}} = \frac{5}{\sqrt{100}} = 0,5$$

O intervalo desejado é com 95% de confiança. Como a distribuição de probabilidade é a normal (que é simétrica), é preciso encontrar o valor na tabela correspondente à área de 47,5%.

Figura 7.1 – Distribuição normal com probabilidade de 95%



¹ Lembrando que, se a variância populacional é 25, o desvio padrão populacional é 5.

O valor (para z) de 1,96 na tabela de distribuição normal é 0,475002, portanto, bem próximo dos 47,5%. Lembrando que a tabela representa uma normal padronizada, isto é, com média 0 e desvio padrão igual a 1, para que os valores da média amostral fiquem compatíveis com os da tabela, deve-se subtrair a média e dividir pelo desvio padrão. Você já sabe que a *média da média amostral* é a própria média populacional (μ) e seu desvio padrão, conforme já calculado, é igual a 0,5. Portanto:

$$\frac{|\bar{X} - \mu|}{\sigma_{\bar{X}}} = 1,96$$

A diferença é em módulo porque o valor encontrado para a média amostral pode estar tanto abaixo como acima da média populacional. O valor encontrado para a média amostral foi 23,2. Substituindo, chega-se a:

$$\frac{|23,2 - \mu|}{0,5} = 1,96$$

$$|23,2 - \mu| = 0,5 \times 1,96$$

$$|23,2 - \mu| = 0,98$$

Como é em módulo, isto é, a média pode estar acima ou abaixo de 23,2, há duas possibilidades:

$$\begin{aligned} 23,2 - \mu &= 0,98 \\ -\mu &= 0,98 - 23,2 \\ \mu &= 23,2 - 0,98 \\ \mu &= 22,22 \end{aligned}$$

$$\begin{aligned} 23,2 - \mu &= -0,98 \\ -\mu &= -0,98 - 23,2 \\ \mu &= 23,2 + 0,98 \\ \mu &= 24,18 \end{aligned}$$

ou

Ou seja, a média populacional pode estar entre 22,22 e 24,18. Repare que esses valores foram obtidos somando-se e subtraindo-se 0,98 da média amostral inicialmente obtida (23,2). Vale dizer que 0,98 é a tal da *margem de erro*, obtida multiplicando-se o desvio padrão pelo valor encontrado na tabela.

Portanto, o intervalo de confiança é dado por:

$$IC_{95\%} = [22,22; 24,18]$$

Com 95% de confiança, como assinalado. Mas o que significa isso, afinal? Significa que, se a experiência for repetida (calcular a média de idade a partir de uma amostra de 100 pessoas) muitas (infinitas) vezes, *em 95% delas o intervalo conteria o valor verdadeiro da média populacional*.

Não é, a rigor, a probabilidade de que o intervalo, uma vez construído, conte a verdadeira média populacional, pois, se ele já foi construído, ou ele contém ou não contém o valor verdadeiro (seja ele qual for); a probabilidade seria 1 ou 0, respectivamente.

Exemplo 7.1.2

Após entrevistar 49 membros de uma categoria profissional, um pesquisador encontrou um salário médio de R\$ 820,00. O desvio padrão dos salários dessa categoria, conhecido, é R\$ 140,00. Construa um intervalo para a média:

a) com 80% de confiança.

Com 80% de confiança, procure na tabela a metade, isto é, 40%. O valor mais próximo é 0,399727, que corresponde ao valor de z de 1,28. Como a média amostral tem distribuição aproximadamente normal:

$$\frac{|\bar{X} - \mu|}{\sigma_{\bar{x}}} = 1,28$$

Em que:

$$\bar{X} = 820 \text{ e}$$

$$\sigma_{\bar{x}} = \frac{140}{\sqrt{49}} = 20$$

$$\frac{|820 - \mu|}{20} = 1,28$$

$$|820 - \mu| = 25,6$$

A chamada margem de erro é 25,6. Os pontos extremos do intervalo de confiança podem ser encontrados somando-se e subtraindo-se 25,6 da média amostral.

$$IC_{80\%} = [794,4; 845,6]$$

b) com 90% de confiança.

Agora procure na tabela o valor correspondente a 45%. Esse valor está entre 1,64 e 1,65. De fato, o valor de z é aproximadamente 1,645.

$$\frac{|820 - \mu|}{20} = 1,645$$

$$|820 - \mu| = 32,9$$

Portanto, o intervalo de confiança é:

$$IC_{90\%} = [787,1; 852,9]$$

Acontece aqui um problema de *cobertor curto* – quando se cobre o pescoço, descobrem-se os pés: se o grau de confiança for aumentado, a precisão do intervalo cai (a margem de erro aumenta).

Para aumentar tanto a precisão do intervalo como sua confiança (ou, pelo menos, aumentar uma sem diminuir a outra), é preciso *aumentar o pano do cobertor*, isto é, aumentar a amostra. Veja o exemplo seguinte.

Exemplo 7.1.3

Do exemplo anterior, qual é o tamanho de amostra necessário para que, mantidos os 90% de confiança, a margem de erro seja de, no máximo, 20?

Para 90% de confiança:

$$\frac{|\bar{X} - \mu|}{\sigma_{\bar{x}}} = 1,645$$

Em que:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Substituindo:

$$\frac{|\bar{X} - \mu|}{\frac{\sigma}{\sqrt{n}}} = 1,645$$

A margem de erro é dada por:

$$\frac{\sigma}{\sqrt{n}} \times 1,645 = 20$$

$$\frac{140}{\sqrt{n}} \times 1,645 = 20$$

$$\frac{230,3}{\sqrt{n}} = 20$$

$$\sqrt{n} = \frac{230,3}{20}$$

$$\sqrt{n} = 11,515$$

Elevando ao quadrado os dois lados da equação:

$$(\sqrt{n})^2 = (11,515)^2$$

$$n = 132,59$$

Como a pergunta é qual o tamanho mínimo da amostra (e esse deve ser um número inteiro), a resposta é 133 elementos.

Exemplo 7.1.4

Em uma pesquisa eleitoral com 1.000 eleitores, 240 declararam que pretendem votar no candidato A. Construa um intervalo com 95% de confiança para as intenções de voto para esse candidato.

Neste exemplo, a resposta pedida é exatamente o que é apresentado pelos meios de comunicação quando divulgam uma pesquisa eleitoral.

O valor (amostral) para a proporção de eleitores que desejam votar nesse candidato é:

$$\hat{p} = \frac{X}{n} = \frac{240}{1.000} = 0,24 = 24\%$$

(onde X é o número de pessoas que votam no candidato A. Note que X tem distribuição binomial).

Mas é preciso calcular a margem de erro para que o resultado (o intervalo de confiança) seja completo. Para isso, é preciso calcular a variância desse estimador.

Como fazê-lo? Suponha que 24% é o valor correto das intenções de voto. Isso significa que, para cada eleitor entrevistado, é como se fosse um jogo em que há 24% de chance de esse eleitor votar no candidato A e 76% de votar em outros candidatos (incluindo aí votos brancos e nulos). Da mesma forma que, quando se joga uma moeda, há 50% de chance de dar cara e 50% de não dar cara (dar coroa); ou, quando se joga um dado, há 1/6 de chances de cair certo número desejado e 5/6 de chances de não cair.

Portanto, é como se cada eleitor entrevistado fosse uma distribuição de Bernoulli, cuja variância é calculada, como já foi visto, por:

$$\sigma^2 = p(1 - p)$$

Sendo p a probabilidade de ocorrência de sucesso (dar cara na moeda, dar 6 no dado ou... encontrar um eleitor que vote no candidato A), e $(1 - p)$ a probabilidade de ocorrência do fracasso.

Como são n eleitores, a proporção encontrada é, na verdade, uma proporção média, cuja variância é dada, a exemplo da média amostral comum, por:²

$$\text{var}(\hat{p}) = \text{var}\left(\frac{X}{n}\right) = \frac{n\hat{p}(1 - \hat{p})}{n^2} = \frac{\hat{p}(1 - \hat{p})}{n}$$

Que, neste caso, é dada por:

$$\text{var}(\hat{p}) = \frac{0,24 \times 0,76}{1.000} = 0,0001824$$

E o desvio padrão:

$$\text{var}(\hat{p}) = \sqrt{0,0001824} \cong 0,0135 = 1,35\%$$

² Note que, também a exemplo da média amostral, essa variância é estimada, já que não se conhece o valor correto de p .

Você já tem o valor do estimador e seu desvio padrão. Pode, portanto, calcular o intervalo de confiança da proporção verdadeira (populacional) p (o valor tabelado para 95% é 1,96):

$$\frac{|\hat{p} - p|}{dp(\hat{p})} = 1,96$$

$$\frac{|24 - p|}{1,35} = 1,96$$

$$|24 - p| \approx 2,6\%$$

Portanto, o intervalo de 95% de confiança para as intenções de voto para o candidato A é:

$$IC_{95\%} = [21,4\%; 26,6\%]$$

Ou, como preferem os meios de comunicação, o candidato A tem 24% das intenções de voto, com margem de erro de 2,6 pontos percentuais, para cima ou para baixo... isso se forem considerados, evidentemente, 95% de confiança.

7.2 Testes de hipóteses

Todo mundo já os fez um dia na vida... talvez não com as ferramentas mais adequadas, mas já fez, sim. Imagine uma menina de uns 11, 12 anos que, no intervalo da aula, vai à lanchonete da escola e lá está aquele garoto que sempre olha estranho para ela. Ela vai à quadra e lá está o garoto de novo. Então, ela volta para a classe um pouco antes e adivinhe quem também voltou! Aí, a menina para e pensa: “É muita coincidência... esse garoto gosta de mim!”.

A menina estabeleceu duas hipóteses:

- primeira hipótese: o garoto não gosta dela;
- segunda hipótese: o garoto gosta dela.

Suponhamos que fosse verdade a primeira hipótese. Então, o garoto só estaria nos mesmos lugares que ela, quando isso ocorresse, por mera coincidência, não intencionalmente. Como ele esteve em três lugares diferentes próximo à menina, durante um curto período de tempo, isso não deve ser coincidência, portanto, a primeira hipótese deve ser rejeitada.

Duas observações devem ser feitas: uma é o critério do que é coincidência ou não. Esse é arbitrário. Uma menina que estivesse torcendo para que o garoto gostasse dela poderia ser menos rigorosa e aceitar que bastariam, digamos, dois lugares diferentes para que se considerasse muita coincidência. Outra poderia querer que o fenômeno se repetisse em outros dias para que se considerasse muita coincidência.

A outra observação é que, ainda que o raciocínio esteja correto, é possível que a conclusão seja incorreta, pois, embora pouco provável, não é impossível que o garoto estivesse em todos aqueles lugares por mera coincidência. Nesses casos, nunca dá para ter certeza absoluta.

Os testes que você vai fazer, entretanto, não lidam com coisas tão complexas como o coração humano (qualquer que seja a idade). Você vai escolher coisas que possam ser medida em números. O método, todavia, é parecido. O primeiro passo é estabelecer as duas hipóteses. A primeira hipótese, também conhecida como **hipótese nula** (H_0), geralmente é uma igualdade, isto é, supõe-se que determinado parâmetro é igual a um número. A segunda hipótese, a chamada **hipótese alternativa** (H_1), contradiz a hipótese nula de alguma forma, portanto, é uma desigualdade: pode ser “o parâmetro é diferente do número”, “maior do que o número” ou “menor do que o número”. Então, há três pares de hipóteses possíveis num teste para determinado parâmetro θ :

$$H_0: \theta = \theta_0$$

$$H_1: \theta \neq \theta_0$$

ou

$$H_0: \theta = \theta_0$$

$$H_1: \theta < \theta_0$$

ou

$$H_0: \theta = \theta_0$$

$$H_1: \theta > \theta_0$$

Sendo θ_0 um valor qualquer que o parâmetro θ pode assumir.

A segunda parte é estabelecer o que é muita coincidência, isto é, qual a probabilidade que será considerada muita coincidência. Essa probabilidade é conhecida como **significância** do teste.

Isso significa que a realização do teste depende do conhecimento da distribuição de probabilidade do parâmetro. Por isso mesmo, quando você usa o primeiro par dessas hipóteses, o teste se chama **bicaudal**, já que *diferente* pode ser maior ou menor, indicando que serão utilizadas as duas *caudas* da distribuição. Quando o teste é feito com um dos dois últimos pares de hipóteses, ele é conhecido como **monocaudal**.

Tome um exemplo bem simples: uma moeda que *insiste* em dar cara. Será que ela é viciada?

O primeiro passo é estabelecer as hipóteses: se ela não é viciada, a proporção populacional de caras é 0,5. Caso contrário, é diferente.³

$$H_0: p = 0,5$$

$$H_1: p \neq 0,5$$

³ Como já foi estabelecido que ela está dando mais caras, poderia ser utilizada a hipótese de ser maior do que 0,5. Aí é uma questão de critério.

O segundo passo é estabelecer a significância do teste ou, em outras palavras, definir o que é muita coincidência. Arbitrariamente escolhemos 10%.

A distribuição de probabilidade aqui é uma binomial. Suponha que, nas duas primeiras jogadas, o resultado tenha sido cara. Supondo que a moeda não é viciada, a probabilidade de isso ocorrer é:

$$P(2 \text{ caras}) = 0,5 \times 0,5 = 0,25 = 25\%$$

O que é bem possível de ocorrer, de acordo com o critério adotado. Nada indica que a moeda esteja viciada, foi um resultado absolutamente normal, é perfeitamente possível que a hipótese nula seja verdadeira. Nesse caso, costuma-se dizer que a *hipótese nula é aceita*.

Agora, imagine que tenha dado cara em três lançamentos da moeda:

$$P(3 \text{ caras}) = 0,5 \times 0,5 \times 0,5 = 0,125 = 12,5\%$$

Ou seja, uma moeda não viciada tem apenas 12,5% de chance de apresentar esse resultado. Mas 12,5% não é considerado muita coincidência pelo critério, que é de 10%. Então, continue acreditando na honestidade da moeda, isto é, continue aceitando a hipótese nula.

Mas suponha que sejam quatro caras seguidas:

$$P(4 \text{ caras}) = 0,5 \times 0,5 \times 0,5 \times 0,5 = 0,0625 = 6,25\%$$

Estabelecemos que 10% é muita coincidência. Mas uma moeda não viciada teria apenas 6,25% de dar esse resultado. Então, a conclusão é de que a moeda não pode ser honesta. Rejeita-se a hipótese nula de que a moeda tem proporção igual a 0,5, ou seja, ela é viciada.

Como no caso da menina, ainda que improvável, o resultado pode ocorrer (com 6,25% de chances) mesmo que se trate de uma moeda não viciada. Note-se que, se o critério fosse 5%, seria possível continuar acreditando na honestidade da moeda. Se a significância do teste fosse qualquer valor abaixo de 6,25%, você aceitaria a hipótese nula e, para qualquer valor acima, rejeitaria tal hipótese. Esse valor (no caso, 6,25%), que dá o limite entre a aceitação e a rejeição e nem sempre é muito fácil de ser calculado sem o auxílio de computadores ou calculadoras, é conhecido como p-valor ou valor p.

Exemplo 7.2.1

Afirma-se que a altura média dos jogadores de basquete que disputam determinada liga é 1,95 m. Numa amostra de 36 jogadores, foi encontrada uma média de 1,93 m. Sabe-se que o desvio padrão da altura dos jogadores é 12 cm. Teste, com um nível de significância de 10%, se a afirmação é verdadeira.

A hipótese nula deve ser a própria afirmação, isto é, que a média é 1,95 m. A hipótese alternativa é que a afirmação é falsa, ou seja, diferente de 1,95 m.

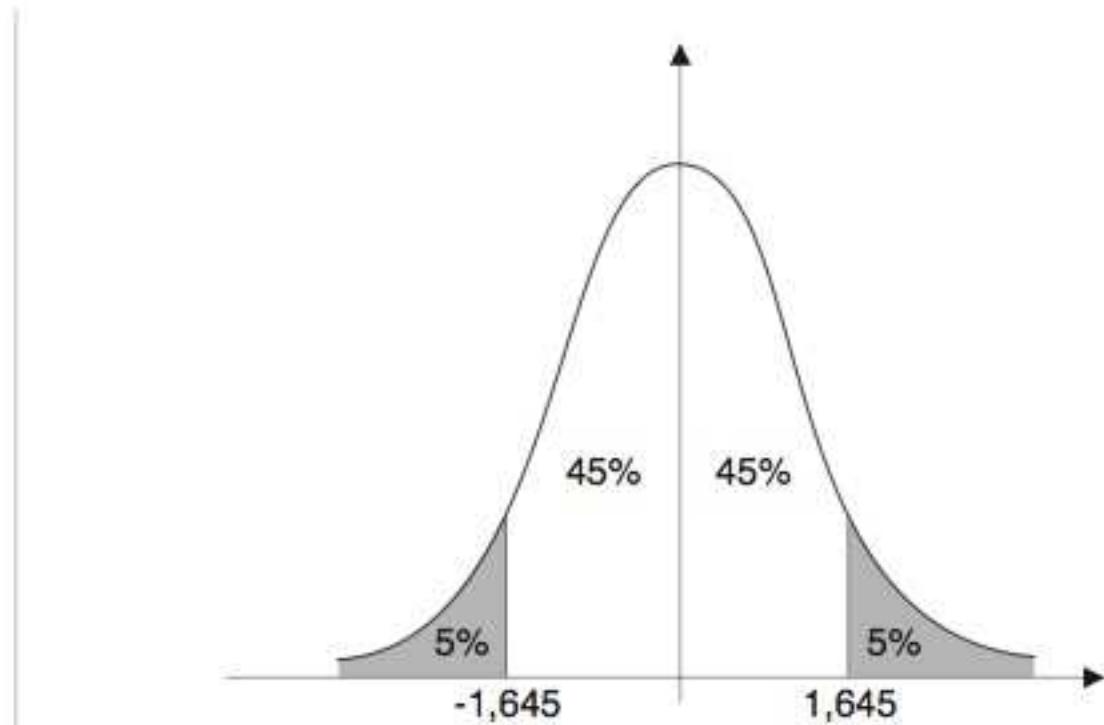
$$H_0: \mu = 1,95$$

$$H_1: \mu \neq 1,95$$

Trata-se de um teste bicaudal, portanto. Qual a distribuição de probabilidade a ser usada? Estamos falando de média, o que vale dizer, pelo teorema do limite central, que é uma variável cuja distribuição é normal.

Se a significância do teste é 10% e o teste é bicaudal, então isso equivale a 5% em cada *cauda*. Na tabela da distribuição normal padronizada, isso equivale a um valor de *z* de 1,645.

Figura 7.2 – Distribuição de probabilidades para a média de altura



Conhecida a distribuição de probabilidade, o procedimento é parecido com o intervalo de confiança: construa um intervalo, supondo que a hipótese nula seja verdadeira e contenha 90% dos possíveis valores amostrais. Fora desse intervalo, não é que seja impossível, mas a probabilidade é menor do que 10%, o que, pelo critério estabelecido (significância do teste), é muita coincidência.

Tem-se:

$$\frac{|\bar{X} - \mu|}{\sigma_{\bar{x}}} = 1,645$$

Em que μ (supostamente) é 1,95 e o desvio padrão da média ($\sigma_{\bar{x}}$) é dado por:

$$\sigma_{\bar{x}} = \frac{0,12}{\sqrt{36}} = 0,02$$

Substituindo, chega-se a:

$$\frac{|\bar{X} - 1,95|}{0,02} = 1,645$$

$$|\bar{X} - 1,95| \cong 0,033$$

Portanto, os valores que podem ocorrer numa amostra de 36 jogadores, com 90% de probabilidade, estão entre $1,95 + 0,033$ e $1,95 - 0,033$. Se o valor amostral estiver dentro desse intervalo, aceita-se a hipótese nula. Por isso, chamaremos este intervalo de **região de aceitação** (RA).⁴

$$RA = [1,917; 1,983]$$

O valor amostral foi 1,93, que está dentro da RA, portanto, *aceita-se a hipótese nula*. Aceitar a hipótese nula pode significar que se assume como se ela fosse verdade e, de fato, há respaldo para isso. Mas talvez o mais correto fosse dizer que não é possível rejeitar a hipótese nula. Na verdade, é isso que ocorre: pelo valor obtido na amostra, não é possível contestar a informação inicial, mas também é possível que o valor verdadeiro seja outro. Então, seria mais correto dizer que *não se rejeita a hipótese nula*. Note que é possível fazer o teste de uma outra maneira, totalmente equivalente, montando a RA por meio dos valores da normal padronizada.

A RA em termos dos valores da normal é:

$$RA = [-1,645; 1,645]$$

E o valor obtido na amostra (lembrando que $\bar{X} = 1,93$, $\mu = 1,95$) e $\sigma_{\bar{X}} = 0,02$ é:

$$\frac{\bar{X} - \mu}{\sigma_{\bar{X}}} = \frac{1,93 - 1,95}{0,02} = -1$$

Como o valor encontrado pertence à RA, aceita-se a hipótese nula. Como foi dito, essas duas formas são totalmente equivalentes e vão dar o mesmo resultado. Note que o módulo é desnecessário agora, já que foram incluídos os valores negativos na RA.

Exemplo 7.2.2

Em uma amostra com 100 famílias em uma cidade do interior, foi encontrada uma renda média de R\$ 580,00. Segundo o prefeito, essa pesquisa está errada, pois a renda média em sua cidade é de, *no mínimo*, R\$ 650,00. Teste a afirmação do prefeito com 10% de significância, sabendo-se que o desvio padrão da renda é de R\$ 120,00.

O prefeito não afirma que a renda é exatamente R\$ 650,00, mas que é no mínimo R\$ 650,00. Pode ser R\$ 700,00, R\$ 800,00 etc. A hipótese alternativa (contrária à do prefeito) deve ser que a renda média seja *menor* do que R\$ 650,00.

⁴ O conjunto dos pontos que não pertencem à região de aceitação é chamado de região de rejeição ou região crítica.

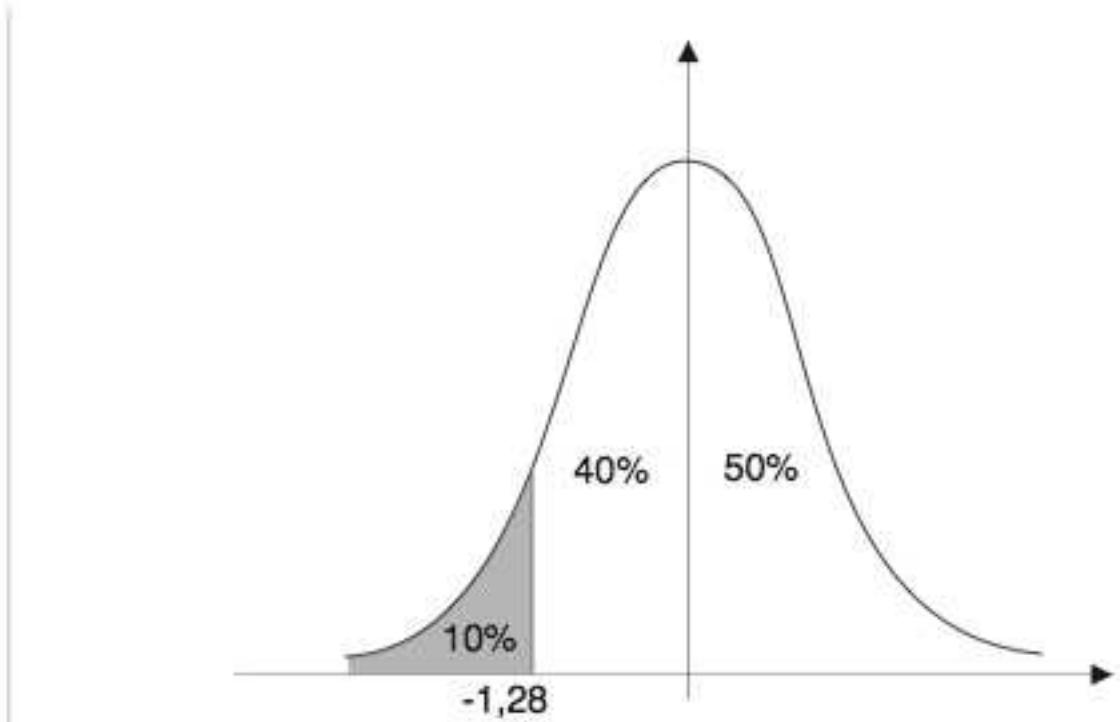
$$H_0: \mu = 650$$

$$H_1: \mu < 650$$

Ou seja, estamos falando aqui de um teste monocaudal. Os 10% devem estar concentrados na cauda esquerda da curva normal. Na verdade, como a normal é simétrica, tanto faz à direita ou à esquerda, o que importa é que os 10% estejam concentrados em um só lado.

Aqui vale uma observação: a rigor, a hipótese nula deveria ser $\mu \geq 650,00$, por dois motivos: primeiro, porque o prefeito disse “no mínimo”, ou seja, segundo ele a renda é R\$ 650,00 ou mais; segundo, porque a hipótese alternativa (menor) contradiz (é complementar a) “menor ou igual”, não propriamente a igualdade (veja Apêndice 7.A).

Figura 7.3 – Distribuição de probabilidades para a média dos salários



Assim, o valor a ser utilizado da tabela normal padronizada é 1,28 (em módulo). Portanto:

$$\frac{|\bar{X} - \mu|}{\sigma_{\bar{x}}} = 1,28$$

Levando-se em conta que:

$$\sigma_{\bar{x}} = \frac{120}{\sqrt{100}} = 12$$

$$\frac{|\bar{X} - 650|}{12} = 1,28$$

$$|\bar{X} - 650| = 15,36$$

Como estamos testando a hipótese alternativa de ser menor (se a amostra apresentasse um valor maior do que R\$ 650,00, o prefeito não teria feito nenhuma objeção), a RA inclui todos os valores maiores do que R\$ 650,00. O que realmente importa são os valores menores, que têm seu limite inferior dado por $650,00 - 15,36 = 634,64$. Portanto, a RA é dada por:

$$\text{RA} = [634,36; \infty]$$

O valor encontrado na amostra foi R\$ 580,00, que não pertence a esse intervalo. Vale dizer que, se a renda fosse realmente R\$ 650,00 no mínimo, a chance de encontrar R\$ 580,00 numa amostra de 100 elementos é inferior a 10%, então *rejeita-se a hipótese nula*, ou seja, conclui-se que o prefeito está equivocado.

Exemplo 7.2.3 (Novamente pesquisas eleitorais)

Uma pesquisa feita com 300 eleitores revelou que 23% votariam no candidato A. O candidato B, entretanto, afirma que seu oponente tem, *no máximo*, 20% dos votos. Teste a afirmação do candidato B, utilizando um nível de significância de 5%.

As hipóteses nesse caso são:

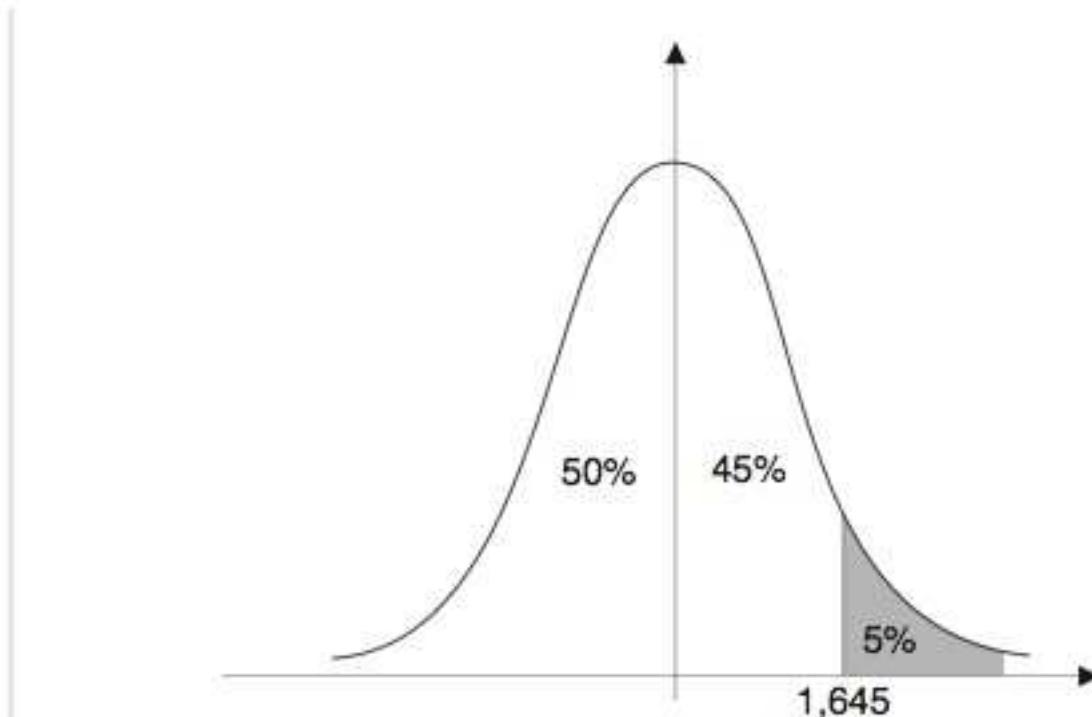
$$H_0: p = 0,2$$

$$H_1: p > 0,2$$

Uma vez que a alternativa à hipótese lançada pelo candidato B é a de que A tenha, de fato, mais do que 20% das intenções de voto.

De novo, é um teste monocaudal, desta vez sendo utilizada a cauda da direita.

Figura 7.4 – Distribuição de probabilidades para a intenção de voto no candidato A



A variância da proporção encontrada numa amostra de 300 eleitores é:

$$\text{var}(\hat{p}) = \frac{0,2 \times 0,8}{300} = 0,000533... \Leftrightarrow dp(\hat{p}) \cong 0,023 = 2,3\%$$

Então:

$$\frac{|\hat{p} - p|}{dp(\hat{p})} = 1,645$$

$$\frac{|\hat{p} - 20|}{2,3} = 1,645$$

$$|\hat{p} - 20| \cong 3,8$$

De novo, como o teste é monocaudal, só é preciso se preocupar com a parte superior do intervalo.

$$RA = [-\infty; 23,8\%]$$

Como o valor amostral foi 23%, o que está dentro da RA, então, *aceita-se (não se rejeita) a hipótese nula* (considerando 5% de significância) ou, em outras palavras, não é possível contestar a afirmação do candidato B (ainda que o candidato A tenha no máximo 20% dos votos, a probabilidade de que, numa amostra de 300 eleitores, encontrem-se 23% que votem em A é superior a 5%).

Exemplo 7.2.4

Fez-se um estudo sobre aluguéis em dois bairros, A e B. No primeiro, em 12 residências, o aluguel médio encontrado foi de R\$ 330,00. No segundo, em 19 residências, o aluguel médio foi de R\$ 280,00. Sabe-se que o desvio padrão dos aluguéis no bairro A é R\$ 50,00 e, no bairro B, R\$ 40,00. Afirma-se que os aluguéis médios são iguais nos dois bairros. Teste essa afirmação com 10% de significância.

Aqui não se trata de testar uma média como sendo igual ou não a determinado valor, mas sim comparar duas médias. Você quer saber se as médias são, ou não, iguais. As hipóteses são:

$$H_0: \mu_A = \mu_B$$

$$H_1: \mu_A \neq \mu_B$$

É um pouco diferente do que foi feito até agora, mas é possível, com uma simples transformação, deixar o problema na mesma forma, uma vez que dizer que a média é igual é a mesma coisa que dizer que a *diferença* das médias é zero. Portanto, as hipóteses anteriores são equivalentes a:

$$H_0: \mu_A - \mu_B = 0$$

$$H_1: \mu_A - \mu_B \neq 0$$

É como se fosse criada uma nova variável $Y = (X_A - X_B)$ e se fizesse o teste de hipóteses para a média de Y ser igual a zero.

Lembrando que:

$$\text{var}(Y) = \text{var}(X_A - X_B) = \text{var}(X_A) + \text{var}(X_B) - 2\text{cov}(X_A, X_B)$$

Mas supondo que os aluguéis em cada bairro sejam variáveis independentes:

$$\text{var}(Y) = \text{var}(X_A - X_B) = \text{var}(X_A) + \text{var}(X_B)$$

Já que a covariância é zero. O mesmo vale para a variância da média:

$$\text{var}(\bar{Y}) = \text{var}(\bar{X}_A) + \text{var}(\bar{X}_B)$$

Então:

$$\text{var}(\bar{X}_A) = \frac{50^2}{12} \cong 208,3$$

$$\text{var}(\bar{X}_B) = \frac{40^2}{19} \cong 84,2$$

Portanto, a variância da média (da diferença) é:

$$\text{var}(\bar{Y}) \cong 292,5$$

E o desvio padrão:

$$\sigma_{\bar{Y}} \cong \sqrt{292,5} \cong 17,1$$

Como se trata de um teste a 10% de significância, bicaudal, o valor encontrado na distribuição normal é 1,645. Então:

$$\frac{|\bar{Y} - 0|}{17,1} = 1,645$$

$$|\bar{Y} - 0| = 28,13$$

Portanto, a região de aceitação para a diferença é:

$$RA = [-28,13; 28,13]$$

Como a diferença amostral encontrada foi 50 ($= 330 - 280$), o que extrapola a RA, *rejeita-se a hipótese nula*, isto é, conclui-se que os aluguéis médios são diferentes nos dois bairros.

7.3 Testando a variância

Nos exemplos anteriores, foram feitos testes para a média porque, evidentemente, não se conhecia seu valor, havia o valor amostral e apenas algum tipo de suposição ou afirmação de alguém sobre o valor populacional. Entretanto, o desvio padrão (e, por tabela, a variância) era conhecido,⁵ o que é no mínimo um pouco estranho. Se não se sabe qual é a média, como se saberia a variância?

A única resposta plausível é que, em geral, não se sabe mesmo. A variância também é obtida pela amostra e, portanto, passível de teste. O próximo passo é testar a variância.

Quando obtida a amostra, a variância (amostral) é dada por:

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

Pode-se escrever:

$$(n-1)S^2 = \sum_{i=1}^n (X_i - \bar{X})^2$$

Se os dois lados forem divididos pela variância populacional σ^2 , tem-se:

$$(n-1)\frac{S^2}{\sigma^2} = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{\sigma^2}$$

Ou:

$$(n-1)\frac{S^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma} \right)^2$$

Repare que, se X for uma variável cuja distribuição é normal (e isso é importante!), a expressão dentro dos parênteses será quase uma normal padronizada, já que se subtrai a média e divide-se pelo desvio padrão. Para ser exatamente uma normal padronizada, seria necessário ter a média populacional e não a média amostral.

Do capítulo anterior, quando se procurou encontrar um estimador não viesado para a variância, você descobriu que:

$$\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2$$

Substituindo:

$$(n-1)\frac{S^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 - n \left(\frac{\bar{X} - \mu}{\sigma} \right)^2$$

⁵ Com exceção dos exemplos de proporção (pesquisas eleitorais). Isso será discutido mais adiante.

Ou ainda:

$$(n-1) \frac{S^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 - \left(\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \right)^2$$

Agora, no lado direito da equação há um somatório de n variáveis normais padronizadas, uma vez que se subtraiu a média populacional μ . Além disso, subtraiu-se outra variável normal padronizada, já que \bar{X} é uma variável com distribuição normal (teorema do limite central) com média μ e desvio padrão dado por σ / \sqrt{n} .

Portanto, chega-se a uma soma de $n - 1$ variáveis normais padronizadas. Como se conhece a distribuição normal padronizada, é possível obter os valores da distribuição dessa nova variável desde que se conheça também o valor de n . De fato, essa distribuição leva o nome de χ^2 (qui-quadrado).

A χ^2 é a distribuição de uma variável que é a soma de n variáveis normais padronizadas. Diz-se que essa variável tem distribuição χ^2 com n graus de liberdade.

Portanto, a expressão $(n - 1) \frac{S^2}{\sigma^2}$ segue uma distribuição χ^2 com $n - 1$ graus de liberdade (porque é uma soma de $n - 1$ variáveis normais padronizadas), desde que, é claro, S^2 tenha sido obtida de uma variável cuja distribuição é normal. Escreve-se, resumidamente, da seguinte forma:

$$(n-1) \frac{S^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

As curvas que representam a f.d.p. de variáveis com distribuição χ^2 são mostradas nas Figuras 7.5, 7.6 e 7.7.

Figura 7.5 – χ^2 com um grau de liberdade

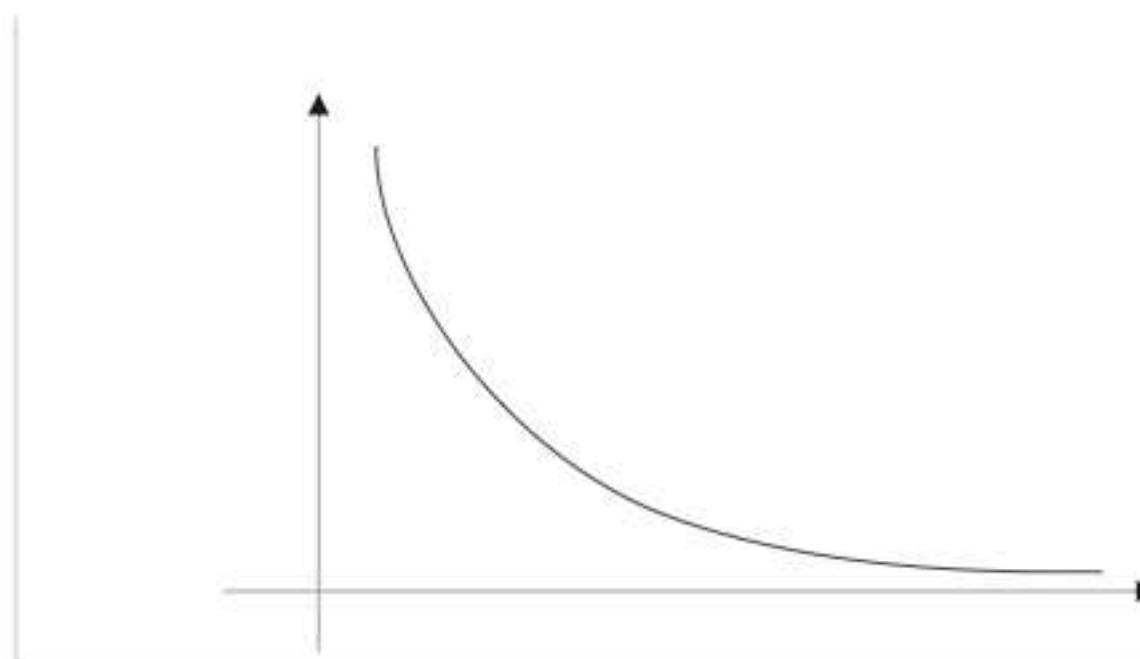
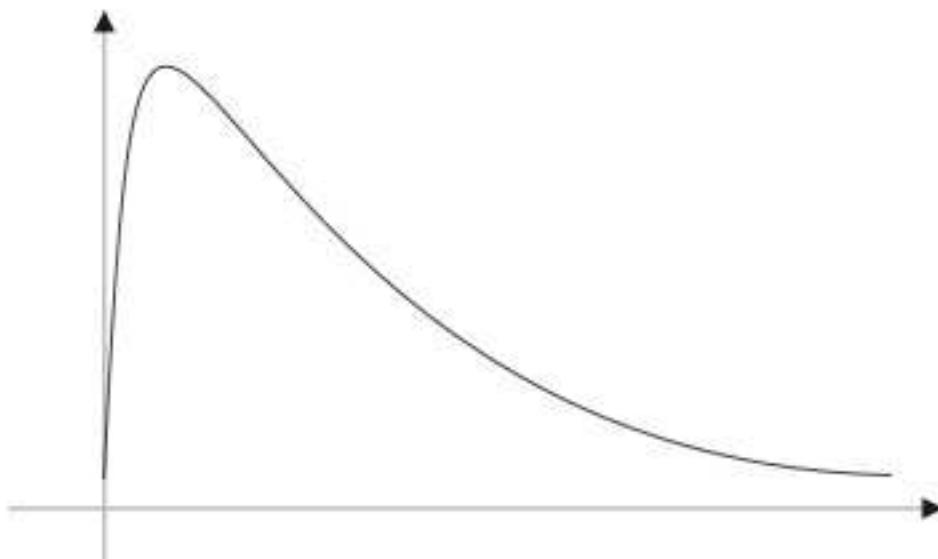
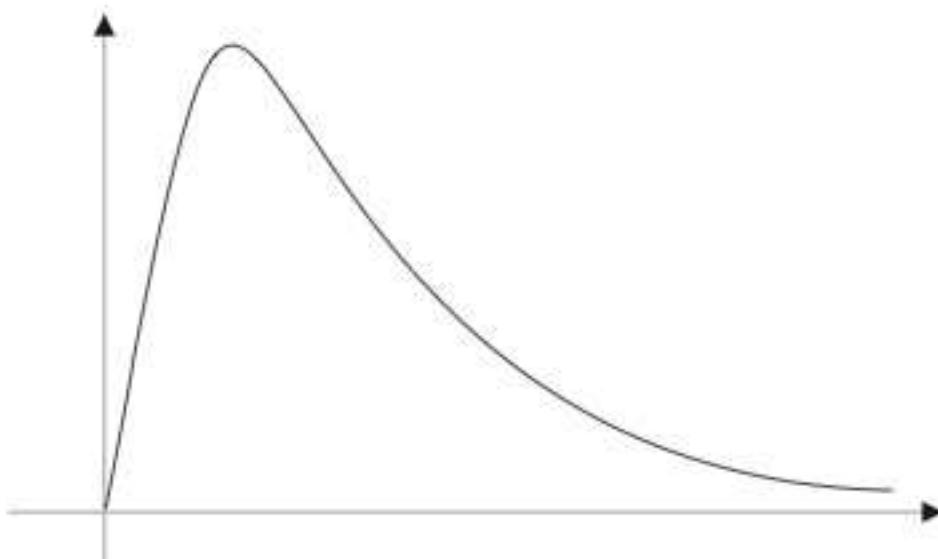


Figura 7.6 – χ^2 com 3 graus de liberdadeFigura 7.7 – χ^2 com 5 graus de liberdade

Repare que a distribuição vai se tornando mais simétrica à medida que se aumentam os graus de liberdade (na verdade, quando n é grande, a χ^2 se aproxima de uma normal), mas em geral ela não é simétrica, o que tem implicações para os testes, pois os valores nas caudas direita e esquerda serão diferentes.

Exemplo 7.3.1

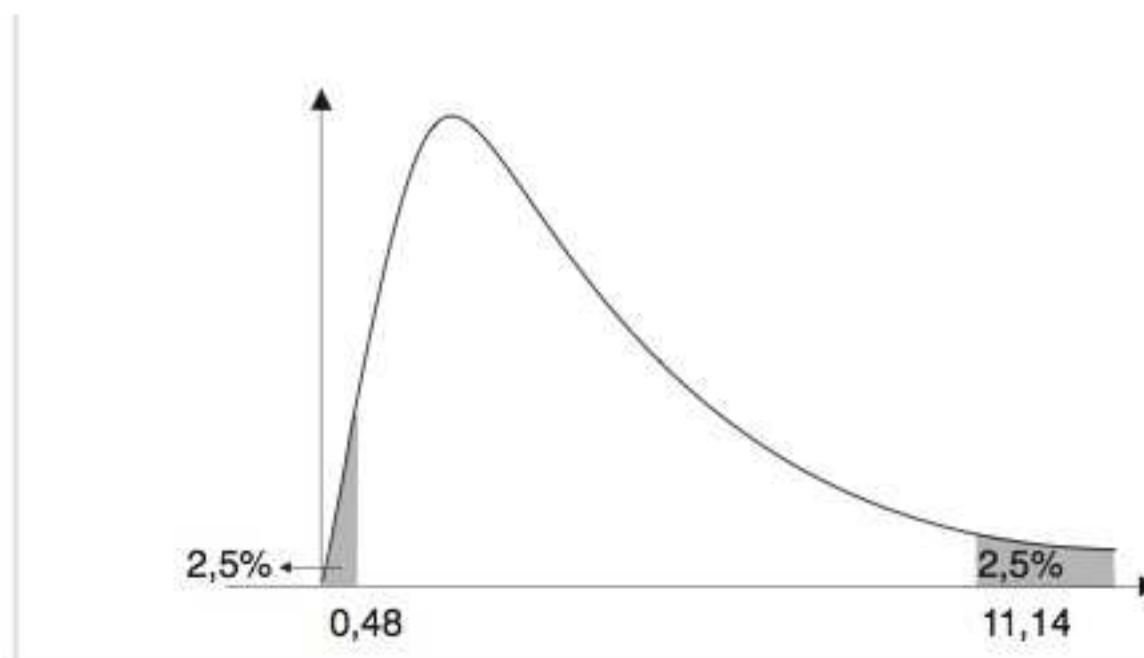
Em determinada empresa, empregados que desempenham a mesma função têm salários diferentes em função do tempo de casa e bonificações por desempenho. Segundo a empresa, o desvio padrão para o salário de certa função é R\$ 150,00. Entrevistando cinco funcionários que desempenham essa função, verificou-se que seus salários eram, respectivamente, R\$ 1.000,00, R\$ 1.200,00, R\$ 1.500,00, R\$ 1.300,00 e R\$ 900,00. Teste a afirmação da empresa com significância de 5%, supondo que os salários sejam normalmente distribuídos. A hipótese apresentada pela empresa é de que o desvio padrão é 150,00, portanto, a variância é $150^2 = 22.500$. As hipóteses nula e alternativa devem ser:

$$H_0: \sigma^2 = 22.500$$

$$H_1: \sigma^2 \neq 22.500$$

Como os salários seguem uma distribuição normal, sua variância amostral segue uma distribuição χ^2 com 4 graus de liberdade (já que há cinco elementos na amostra, $n - 1 = 5 - 1 = 4$) e o teste é bicaudal, o que vale dizer que uma área equivalente a 2,5% é tomada em cada cauda da distribuição. Na tabela da distribuição χ^2 , na linha correspondente aos 4 graus de liberdade, encontram-se os valores nas colunas 2,5% (que corresponde à cauda esquerda) e 97,5% (cauda direita).

Figura 7.8 – Distribuição de probabilidades para a variância dos salários



Os valores encontrados são 0,48 e 11,14. A região de aceitação, em termos dos valores tabelados, é:

$$RA = [0,48; 11,42]$$

Supõe-se que a variância verdadeira (populacional) é 22.500. Pela amostra, a variância obtida é:

$$S^2 = \frac{(1.000 - 1.180)^2 + (1.200 - 1.180)^2 + (1.500 - 1.180)^2 + (1.300 - 1.180)^2 + (900 - 1.180)^2}{4}$$

$$S^2 = 57.000$$

Verifique que a média amostral é 1.180.

Para fazer o teste, você precisa calcular a expressão:

$$(n - 1) \frac{S^2}{\sigma^2} = 4 \times \frac{57.000}{22.500} \cong 10,13$$

Esse valor está dentro da RA, portanto, *não se rejeita a hipótese nula* para um nível de 5% de significância. A afirmação da empresa não pode ser contestada.

Exemplo 7.3.2

Uma caixa de fósforos de certa marca vem com a inscrição: “Contém, em média, 40 palitos”. Segundo o fabricante, o desvio padrão é de, *no máximo*, dois palitos. Em uma amostra com 51 caixas, entretanto, foi encontrado um desvio padrão amostral de três palitos. Supondo que o número de palitos por caixa seja uma variável normal, teste a afirmativa do fabricante utilizando um nível de significância de 1%.

As hipóteses são:

$$H_0: \sigma^2 = 4$$

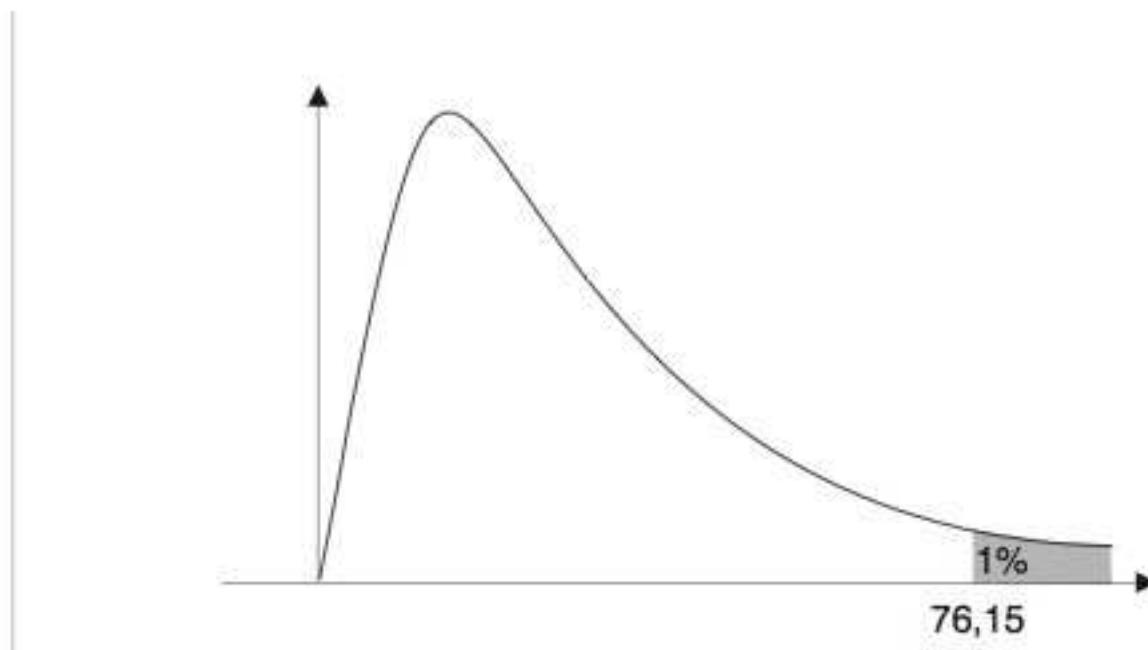
$$H_1: \sigma^2 > 4$$

A expressão é:

$$(n-1) \frac{s^2}{\sigma^2} = 50 \times \frac{9}{4} = 112,5$$

Que, você já sabe, segue uma distribuição χ^2 com 50 graus de liberdade. Para 1% de significância, num teste monocaudal, procure na tabela a coluna de 99% (você está testando a hipótese alternativa *maior*).

Figura 7.9 – Distribuição de probabilidades para a variância dos palitos de fósforo



O valor encontrado foi 76,15, o que significa que, em termos dos valores tabelados, a RA é:

$$RA = [0; 76,15]$$

Note que, como é um teste para a variância, o menor valor possível é zero, já que não existe variância negativa. Como o valor encontrado não pertence à RA, *rejeita-se a hipótese nula* quando o nível de significância é 1%. A afirmação do fabricante não é correta.

Exemplo 7.3.3

Do Exemplo 7.3.1, construa um intervalo de 90% de confiança para a variância. A exemplo de um intervalo de confiança para a média, para um intervalo de confiança de 90% para a variância você utilizará 45% abaixo e 45% acima da variância amostral encontrada. O que equivale, na tabela, às colunas 5% e 95% da linha correspondente aos 4 graus de liberdade que há no Exemplo 7.3.1. Os valores tabelados são 0,71 e 9,49.

Chamando de χ^2_t os valores tabelados encontrados, nas extremidades do intervalo de confiança, vale:

$$(n-1) \frac{S^2}{\sigma^2} = \chi^2_t$$

Rearranjando:

$$\sigma^2 = (n-1)S^2 / \chi^2_t$$

Para encontrar os valores limites do intervalo, basta substituir por cada um dos valores tabelados encontrados:

$$\sigma^2_1 = 4 \times 57.000 / 9,49 \cong 24.025,3$$

$$\sigma^2_2 = 4 \times 57.000 / 0,71 \cong 321.126,8$$

Portanto, o intervalo com 90% de confiança para a variância é:

$$IC_{90\%} = [24.025,3; 321.126,8]$$

Ou, se preferir, o intervalo de confiança para o desvio padrão:

$$IC_{90\%} = [155,0; 566,7]$$

Note que um intervalo alternativo poderia ser escolhido. Como a distribuição **não é simétrica**, poder-se-ia não “repartir” a confiança igualmente, isto é, não necessariamente 45% de cada lado e, assim, quem sabe, obter um intervalo menor com a mesma confiança. Então, um procedimento interessante seria escolher o menor intervalo possível com 90% de confiança. Isso, porém, aumentaria em complexidade a construção do intervalo. O procedimento adotado no exemplo tem a vantagem de ser mais simples.

7.4 Testando a média quando a variância é desconhecida e...

Agora que você conhece a distribuição da variância (pelo menos quando se trata de uma variável normal), pode retomar a questão do teste da média quando a variância também é obtida da amostra.

O cálculo da estatística, em vez de ser dado pela expressão:

$$\frac{|\bar{X} - \mu|}{\sigma / \sqrt{n}}$$

É calculado por:

$$\frac{|\bar{X} - \mu|}{S / \sqrt{n}}$$

Já que a variância populacional σ^2 não é conhecida e, portanto, só é possível obter a variância amostral S^2 .

A média amostral, já é sabido, segue uma distribuição normal. A expressão $(n-1)S^2/\sigma^2$ segue uma distribuição χ^2 com $n-1$ graus de liberdade, sendo n o tamanho da amostra. Isso, é claro, se S^2 foi obtido a partir de uma variável normal.

Portanto, a segunda expressão anterior é um quociente de uma variável que tem distribuição normal padronizada por uma variável que, ao quadrado, tem distribuição χ^2 , exceto pelo fator $(n-1)$. Para perceber isso, basta dividir por σ o numerador e o denominador:

$$\sqrt{n} \frac{\frac{|\bar{X} - \mu|}{\sigma}}{\frac{S}{\sigma}} = \frac{\frac{|\bar{X} - \mu|}{\sigma / \sqrt{n}}}{\frac{S}{\sigma}} = \frac{\frac{|\bar{X} - \mu|}{\sigma / \sqrt{n}}}{\sqrt{\frac{S^2}{\sigma^2}}} = \frac{\frac{|\bar{X} - \mu|}{\sigma / \sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2}}} = \frac{|\bar{X} - \mu|}{\sqrt{\frac{(n-1)S^2}{\sigma^2}}} = \frac{|\bar{X} - \mu|}{\sqrt{(n-1)S^2 / \sigma^2}}$$

Lembrando que:

$$\frac{|\bar{X} - \mu|}{\sigma / \sqrt{n}} \sim N(0,1)$$

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$$

Essa combinação, embora pareça complicada, vem de duas distribuições já conhecidas: no numerador, há uma distribuição normal padronizada; no denominador, uma variável que, ao quadrado e dividida pelos graus de liberdade, tem distribuição χ^2 com $n-1$ graus de liberdade. Então, é possível construir a distribuição dessa expressão, que é conhecida como t de Student.

A distribuição t , como vem (também) da χ^2 , depende dos mesmos graus de liberdade desta última, mas, como a normal padronizada, ela é simétrica e tem média zero. A f.d.p. de uma variável que se distribui como uma t de Student se assemelha a uma *normal achatada*. Portanto, diz-se que a última expressão segue uma distribuição t , de Student, com $n-1$ graus de liberdade. Ou:

$$\frac{|\bar{X} - \mu|}{S / \sqrt{n}} \sim t_{(n-1)}$$

Como para a distribuição χ^2 é preciso que a amostra seja extraída de uma população cuja distribuição é normal, o mesmo vale para a distribuição t de Student. Portanto, essa é uma condição necessária para se usar a distribuição t de Student em um teste de hipóteses.

Exemplo 7.4.1

No Exemplo 7.3.1, suponha que o empregador afirme ainda que o salário médio é, no mínimo, R\$ 1.250,00. Teste a afirmação do empregador utilizando um nível de 10% de significância.

As hipóteses são:

$$H_0: \mu = 1.250$$

$$H_1: \mu < 1.250$$

A média amostral obtida no Exemplo 7.3.1 foi 1.180 e a variância amostral, 57.000. Portanto, o desvio padrão amostral é:

$$S = \sqrt{57.000} \cong 238,75$$

O desvio padrão da média é:

$$S_{\bar{x}} = \frac{S}{\sqrt{n}} = \frac{238,75}{\sqrt{5}} \cong 106,8$$

Como esse desvio padrão foi obtido a partir de uma amostra (que, no caso do Exemplo 7.3.1, veio de uma população normalmente distribuída), a distribuição a ser utilizada é a t , de Student, com 4 ($= 5 - 1$) graus de liberdade.

Na distribuição t de Student, com 4 graus de liberdade e 10% de significância, monocaudal, o valor encontrado é 1,53.

$$\frac{|\bar{X} - \mu|}{S_{\bar{x}}} = 1,53$$

$$\frac{|\bar{X} - 1.250|}{106,8} = 1,53$$

$$|\bar{X} - 1.250| = 163,4$$

Como é um teste monocaudal, a RA será dada por:

$$RA = [1.086,6; \infty]$$

Como o valor encontrado na amostra (1.180) pertence à RA, aceita-se a hipótese nula, isto é, não é possível desmentir a afirmação do empregador.

Alternativamente, pode-se construir a RA em termos dos valores tabelados da distribuição de Student:

$$RA = [-1,53; \infty]$$

O valor é negativo porque se testa a hipótese alternativa de que a média é *menor* do que 1.250.

O cálculo da estatística é:

$$\frac{|\bar{X} - \mu|}{S_{\bar{x}}} = \frac{1.180 - 1.250}{106,8} \cong -0,655$$

Que, da mesma forma, pertence à RA, então, não se rejeita a hipótese nula.

Exemplo 7.4.2

Para verificar a informação de que a temperatura média de uma cidade, no verão, é de 35°C, um estudante coletou a temperatura durante dez dias e encontrou uma média amostral de 33°C, com desvio padrão de 0,7°C. Supondo que a temperatura se distribua normalmente no verão naquela cidade, teste a informação inicial com 10% de significância.

As hipóteses são:

$$H_0: \mu = 35$$

$$H_1: \mu \neq 35$$

O desvio padrão da média é:

$$S_{\bar{x}} = \frac{S}{\sqrt{n}} = \frac{0,7}{\sqrt{10}} \cong 0,22$$

Como o desvio padrão foi obtido da amostra (e você sabe que a distribuição é normal!), a distribuição a ser utilizada é a de Student, com 9 graus de liberdade. Com 10% de significância (teste bicaudal), o valor encontrado é 1,83.

$$\frac{|\bar{X} - \mu|}{S_{\bar{x}}} = 1,83$$

$$\frac{|\bar{X} - 35|}{0,22} = 1,83$$

$$|\bar{X} - 35| = 0,4$$

A região de aceitação é dada por:

$$RA = [34,6; 35,4]$$

Como o valor encontrado na amostra (33°C) não pertence à RA, *rejeita-se a hipótese nula* e, portanto, conclui-se que a temperatura média da cidade no verão *não é 35°C* . Repare na tabela t de Student, por exemplo, na coluna de 5% bicaudal. Se a variância fosse conhecida, o valor na distribuição normal a ser utilizado seria 1,96. Na t de Student, com 5 graus de liberdade, é 2,57; se os graus de liberdade são aumentados para 10, passa a ser 2,23; com 30 graus de liberdade, é 2,04 (diferença de menos de 0,1). À medida que se aumenta a amostra e, por conseguinte, os graus de liberdade, o valor encontrado na tabela t de Student se aproxima do valor da normal – o que faz todo sentido se você pensar em termos da consistência do estimador da variância ou mesmo em termos da lei dos grandes números. De fato, o valor na linha *inf* (infinitos graus de liberdade) é *exatamente* o valor encontrado na distribuição normal, o que vale dizer que a t de Student tende, assintoticamente, a uma distribuição normal. Portanto, se a variância for desconhecida, mas a amostra for grande, fará pouca diferença se você usar a normal ou a t de Student (e fará menos diferença quanto maior for a amostra).

Assim, o título completo desta seção seria *Testando a média quando a variância é desconhecida e... a amostra, pequena*.

Repare que, no Exemplo 7.1.4, a rigor você teria de usar a distribuição de Student para construir o intervalo de confiança, pois a variância também foi obtida da amostra. Isso, no entanto, é desnecessário, pois se trata de uma amostra de 1.000 eleitores.

7.5 Comparação de variâncias

No Exemplo 7.2.4, você fez um teste comparando duas médias. Isto é, a partir de médias obtidas de duas amostras diferentes, procurou-se testar se a média populacional em ambas era igual. E se quiser fazer a mesma coisa com variâncias obtidas de amostras diferentes?

Exemplo 7.5.1

Uma maneira (bem simples, diga-se) de ter uma ideia sobre distribuição de renda é calculando a variância. Suponha que, em duas comunidades, tomaram-se duas amostras, de nove famílias para a comunidade A e cinco famílias para comunidade B. Foram coletados os seguintes valores para as rendas mensais dessas famílias:

Comunidade A: 800, 600, 550, 400, 300, 250, 900, 600, 700.

Comunidade B: 700, 1.200, 300, 500, 1.000.

Teste, com 10% de significância, se a distribuição de renda (medida pela variância) é diferente nas duas comunidades. Suponha que, em ambas, a renda é normalmente distribuída. Verifique que a variância amostral da renda na comunidade A é cerca de 48.611, enquanto na comunidade B é 133.000.

A pergunta é: poderiam ser essas duas variâncias (populacionais) iguais, sendo a diferença obtida resultado de uma coincidência na extração da amostra? A resposta vem por meio do seguinte teste de hipóteses:

$$H_0: \sigma_A^2 = \sigma_B^2$$

$$H_1: \sigma_A^2 \neq \sigma_B^2$$

Como fazê-lo? Sabe-se que, como a distribuição é normal, a expressão $(n - 1)S^2/\sigma^2$ é uma distribuição χ^2 com $n - 1$ graus de liberdade para ambas as comunidades (8 para A e 4 para B).

Se se tomar a razão das variâncias amostrais e dividir pelas respectivas variâncias populacionais (que supostamente são iguais), chega-se a:

$$\frac{S_B^2}{S_A^2} = \frac{\frac{S_A^2}{\sigma^2}}{\frac{S_B^2}{\sigma^2}}$$

No numerador e no denominador uma estatística χ^2 é dividida pelos respectivos graus de liberdade. Essa distribuição resultante desse quociente recebe o nome de distribuição de Fisher-Snedecor ou distribuição F. Ela dependerá dos graus de liberdade do numerador e do denominador.

$$\frac{S_B^2}{S_A^2} = \frac{\frac{S_B^2}{\sigma^2}}{\frac{S_A^2}{\sigma^2}} = \frac{\frac{\chi_4^2}{4}}{\frac{\chi_8^2}{8}} \sim F_{4,8}$$

Diz-se, então, que o quociente das duas variâncias segue uma distribuição F com 4 graus de liberdade no numerador e 8 graus de liberdade no denominador. Note que, como a distribuição χ^2 vem, necessariamente, de uma população normal, a distribuição F terá de vir de duas populações normais também.

O gráfico da f.d.p de uma variável que tem como distribuição uma F é semelhante ao de uma variável cuja distribuição é uma χ^2 . Não é uma distribuição simétrica, portanto. Do ponto de vista de quem utiliza uma tabela, há uma limitação que advém de o papel ter só suas dimensões,⁶ então, as colunas ficam reservadas aos graus de liberdade do numerador, enquanto as linhas aos graus de liberdade do denominador (por exemplo). Não há como representar diferentes níveis de significância, portanto. Para cada nível de significância é necessária uma tabela.

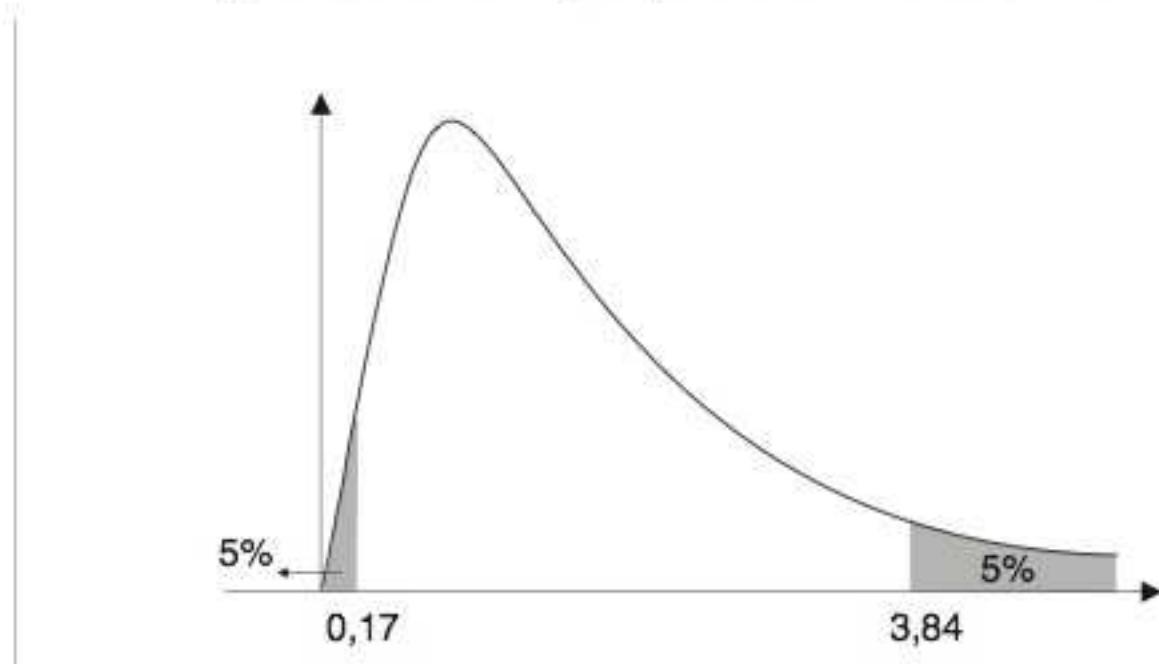
Na tabela F para significância de 10% bicaudal (que é a mesma de 5% moncaudal), o valor máximo da RA pode ser encontrado diretamente na coluna dos 4 graus de liberdade (numerador) e 8 graus de liberdade (denominador). Esse valor é 3,84.

⁶ O papel tem espessura, mas usualmente só usamos a altura e a largura para escrever.

O valor inferior do intervalo é o inverso do valor da distribuição quando se inverte a posição do numerador e do denominador. O valor da tabela para 8 graus de liberdade no numerador e 4 no denominador é 6,04. O limite inferior do intervalo é, então:

$$\frac{1}{F_{8,4}} = \frac{1}{6,04} \cong 0,17$$

Figura 7.10 – Distribuição F para as variâncias das comunidades



A região de aceitação é, então:

$$RA = [0,17; 3,84]$$

Dica: se *sempre* se dividir a maior variância amostral pela menor, esta última conta $\left(\frac{1}{F_{8,4}}\right)$ será desnecessária, pois você já está desconsiderando valores menores do que 1.

O valor calculado pela amostra é:

$$\frac{S_B^2}{S_A^2} = \frac{133.000}{48.125} \cong 2,76$$

Como o valor encontrado pertence à RA, *não se rejeita a hipótese nula* e, assim, não se pode afirmar que a distribuição de renda seja diferente nas duas comunidades.

Exemplo 7.5.2

A média e o desvio padrão amostral dos salários na empresa A são, respectivamente, R\$ 600,00 e R\$ 50,00, valores obtidos a partir de uma amostra de 20 trabalhadores. Na empresa B, utilizando uma amostra de 18 trabalhadores, a média e o desvio padrão amostral encontrados foram R\$ 500,00 e R\$ 80,00, respectivamente. Aparentemente, o desvio padrão é maior na empresa B. Teste essa hipótese com significância de 5%. O teste é, de novo, uma comparação entre variâncias, só que desta vez é monocaudal.

$$H_0: \sigma_A^2 = \sigma_B^2$$

$$H_1: \sigma_A^2 < \sigma_B^2$$

Como foram dados os desvios padrão, você precisa encontrar as variâncias amostrais:

$$S_A^2 = 50^2 = 2.500$$

$$S_B^2 = 80^2 = 6.400$$

A estatística a ser calculada é:

$$\frac{S_B^2}{S_A^2} = \frac{6.400}{2.500} \cong 2,6$$

Pela tabela, o valor limite da distribuição F, com 17 graus de liberdade no numerador e 19 no denominador, é:

$$F_{17,19} = 2,20$$

Então, *rejeita-se* a hipótese nula de que as variâncias são iguais (e, portanto, os desvios padrão) e se considera que, de fato, o desvio padrão da empresa B é maior.

7.6 Erros e poder de um teste

Imagine um julgamento: em países democráticos e/ou civilizados, costuma-se estabelecer uma regra de que todo mundo é inocente até prova em contrário. Quando se faz uma acusação, o acusador é que tem de provar a culpa do réu e, se não conseguir, este é considerado inocente. Dessa forma, procura-se eliminar (ou pelo menos minimizar) a possibilidade de se condenar um inocente.

O problema é que aí se aumenta a possibilidade de que um culpado acabe escapando da condenação. É um preço que se tem de pagar pois, se fosse o contrário (o acusado tivesse de provar sua inocência), embora certamente se reduzisse a chance de que um culpado escapasse, aumentaria a possibilidade de se condenarem inocentes.

Com testes de hipóteses acontece a mesma coisa, embora de forma menos dramática. O resultado de um teste de hipóteses sempre tem alguma chance de estar errado. Na verdade, há dois tipos de erro.

O erro do tipo I ocorre quando se rejeita a hipótese nula quando ela é verdadeira. O **erro do tipo II** significa que se aceita a hipótese nula quando ela é falsa.

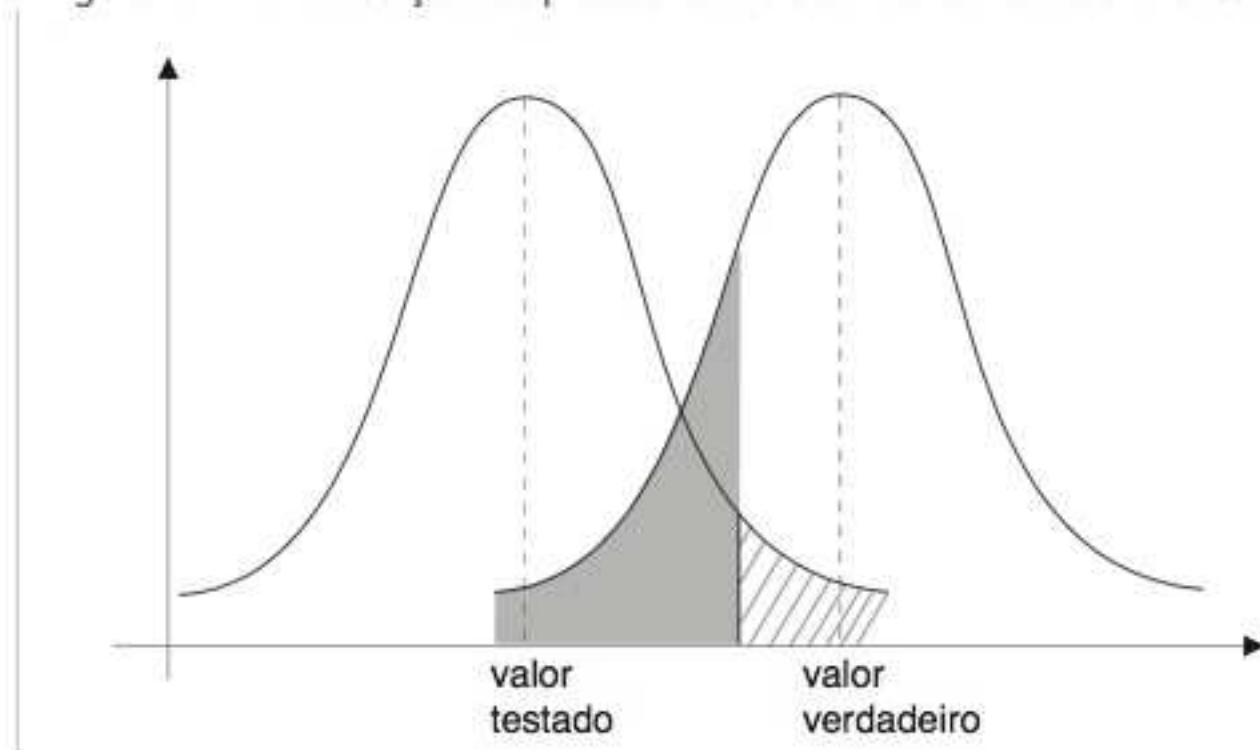
Fazendo a analogia com julgamentos, se você considerar a hipótese nula como sendo o *acusado é inocente* e, portanto, a hipótese alternativa sendo o *acusado é culpado*, o erro do tipo I seria condenar um inocente, enquanto o erro do tipo II seria análogo a absolver um culpado.

A probabilidade de cometer o erro do tipo I é a própria significância do teste, portanto, ela é definida *a priori*.

$$P(\text{erro do tipo I}) = \alpha = \text{significância do teste}$$

Suponha uma situação em que o valor a ser testado não é o valor verdadeiro. Evidentemente, o pesquisador que está fazendo o teste não sabe disso. A situação pode ser ilustrada no gráfico da Figura 7.11.

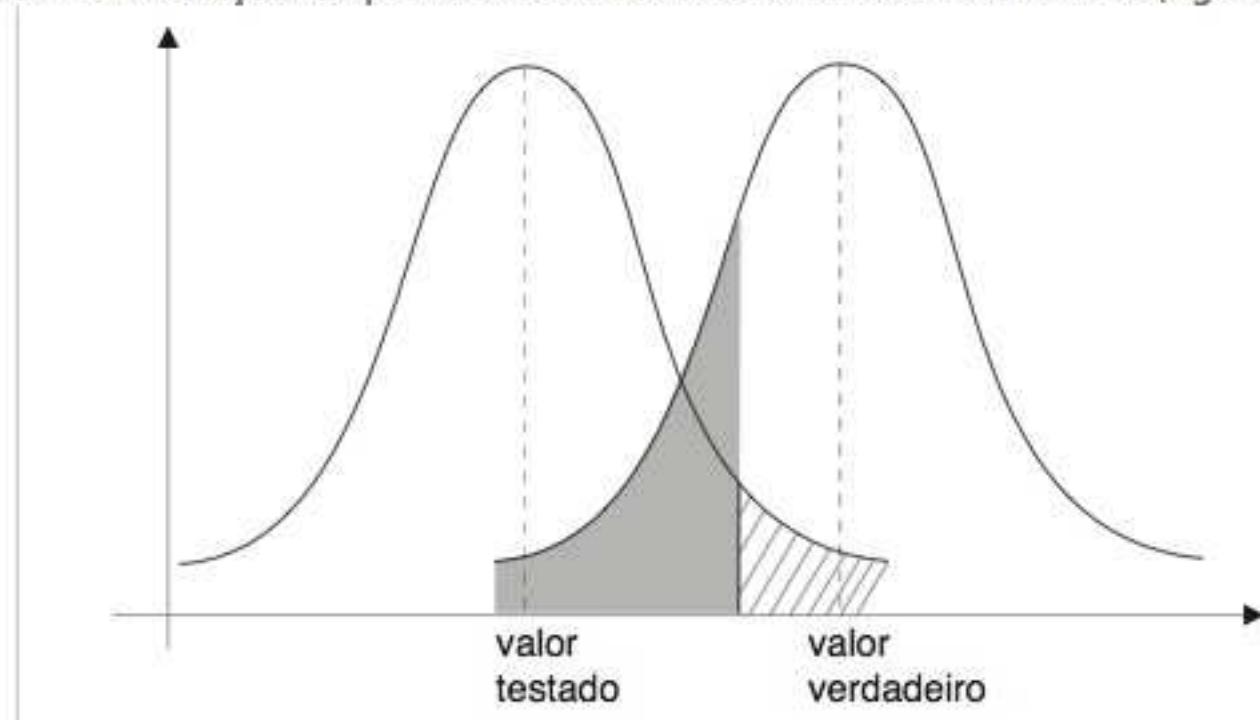
Figura 7.11 – Distribuições de probabilidade dos valores verdadeiro e testado



A área hachurada representa a significância do teste e, pelo menos do ponto de vista do pesquisador que não sabe qual é o valor verdadeiro, a probabilidade de cometer o erro do tipo I. A área cinzenta representa⁷ a probabilidade do erro do tipo II, pois, se o valor amostral cair na região cinzenta, aceita-se a hipótese nula de que o valor testado é o correto, o que não é verdade.

Repare que, se o pesquisador fizer um teste mais rigoroso, isto é, diminuir a significância, aumentará a probabilidade de cometer um erro do tipo II. Portanto, *mais rigoroso* aí significa que a possibilidade de rejeitar a hipótese nula quando ela é falsa é menor. Mas (não tem jeito) a chance de aceitar a hipótese nula, sendo ela falsa, aumenta, o que pode ser visto na Figura 7.12.

Figura 7.12 – Distribuições de probabilidade dos valores verdadeiro e testado (significância menor)



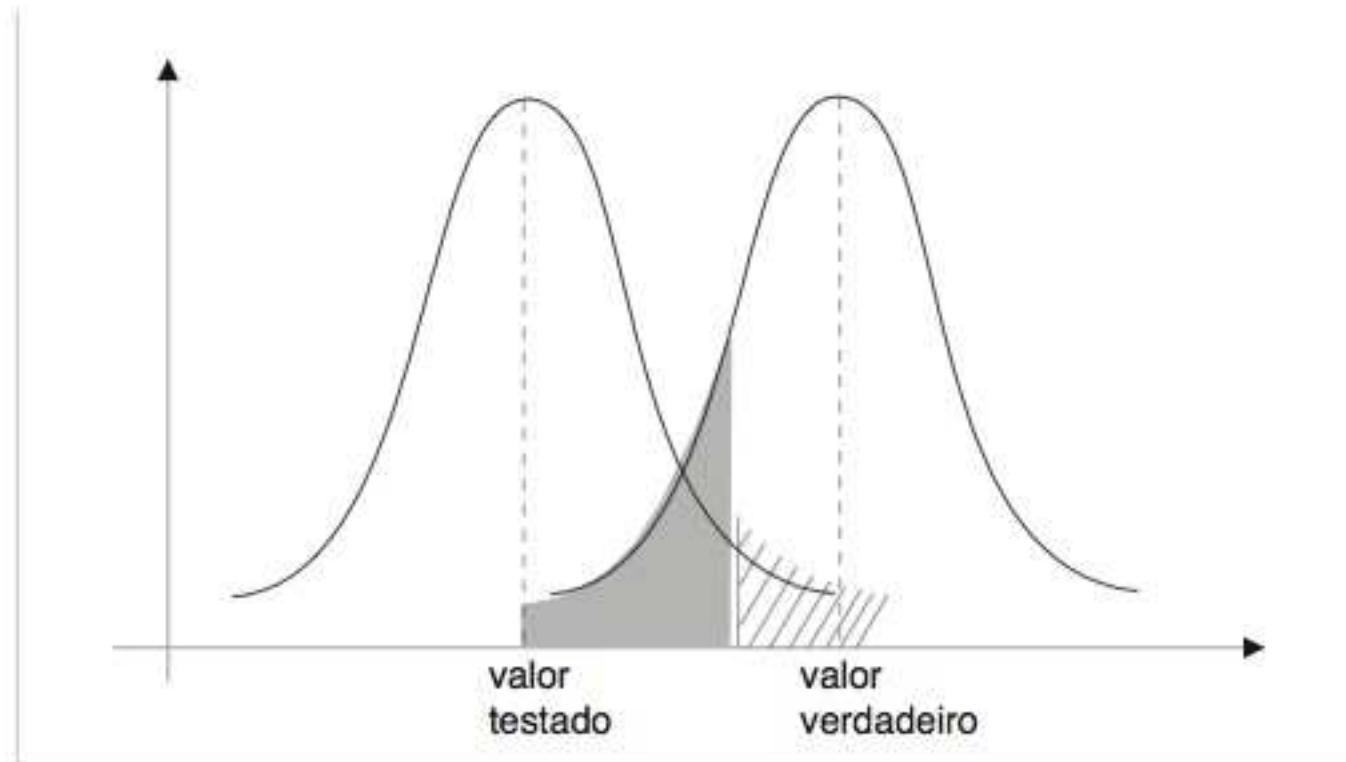
⁷ Na verdade, essas áreas vão até o infinito, se as distribuições forem normais, como é o caso do exemplo. Evidentemente, não é possível pintar um gráfico até o infinito, mas se deve ter isso em mente.

Ao diminuir a significância (área hachurada), aumenta a probabilidade de erro do tipo II.

Mas não tem jeito mesmo? Como num julgamento, um maior número de provas pode levar a um veredito mais correto; no caso de um teste de hipóteses, conseguir *mais provas* significa aumentar a amostra.

Aumentar a amostra significa que os valores amostrais (estimadores) apresentarão variância menor. Com variância menor, as curvas de distribuição se tornarão mais *fininhas* e, portanto, será possível reduzir a probabilidade dos dois erros, como pode ser visto na Figura 7.13.

Figura 7.13 – Distribuições de probabilidade dos valores verdadeiro e testado (amostra maior)



Chame a probabilidade de cometer o erro do tipo II de β .

$$P(\text{erro do tipo II}) = \beta$$

A probabilidade de cometer o erro do tipo II, entretanto, não é conhecida em geral, pois o valor verdadeiro não é conhecido.

Como a significância é previamente estabelecida, um teste de hipóteses será tão melhor quanto menor for a probabilidade de cometer o erro do tipo II. De fato, chama-se de **poder do teste** justamente a probabilidade de *não* cometer o erro do tipo II, isto é, a probabilidade de rejeitar a hipótese nula quando ela é falsa:

$$\text{Poder do teste} = 1 - \beta$$

7.7 O valor-p de um teste

Quando se realiza um teste de hipóteses, por exemplo, para a média, rejeita-se a hipótese nula quando é pouco provável que a média amostral seja encontrada se fosse verdadeira a hipótese sobre a média. Mais especificamente, essa probabilidade é menor

do que a significância previamente estabelecida. Não se rejeita a hipótese nula se essa probabilidade é maior do que a significância. Mas quanto é essa probabilidade? Ela é conhecida como valor-p (ou p-valor) de um teste.

Se a hipótese alternativa de um teste é, por exemplo, a média ser maior do que um valor μ_0 , e a média amostral encontrada for m_0 , o valor-p deste teste é a probabilidade de encontrar, para a média amostral, um valor de m_0 , ou **maior**, supondo que μ_0 é o valor correto.

$$\text{valor-p} = P(\bar{X} > m_0 \mid \mu = \mu_0).$$

Se a hipótese alternativa for, no entanto, a média ser menor do que um valor μ_0 , e a média amostral encontrada for m_0 , o valor-p deste teste é a probabilidade de encontrar, para a média amostral, um valor de m_0 , ou **menor**, supondo que μ_0 é o valor correto.

$$\text{valor-p} = P(\bar{X} < m_0 \mid \mu = \mu_0).$$

Num teste bicaudal, isso deve ser levado em conta na hora de calcular o valor-p. A conta é igual a uma das anteriores, dependendo se a média amostral encontrada m_0 for maior ou menor do que μ_0 e, por ser bicaudal, deve ser contada em dobro, isto é, devem ser levadas em conta as duas caudas da distribuição:

$$\text{valor-p} = 2P(\bar{X} > m_0 \mid \mu = \mu_0), \text{ se } m_0 > \mu_0.$$

$$\text{valor-p} = 2P(\bar{X} < m_0 \mid \mu = \mu_0), \text{ se } m_0 < \mu_0.$$

Exemplo 7.7.1

Calcule o valor-p do teste do exemplo 7.2.1.

A hipótese nula é a média ser 1,95 m (com desvio padrão de 12 cm). O valor médio encontrado na amostra foi 1,93 m, menor portanto. Então, qual a probabilidade de, numa amostra de 36 jogadores, encontrar uma média de 1,93 m, ou menor?

$$Z = \frac{\frac{193 - 195}{12}}{\sqrt{36}} = \frac{193 - 195}{2} = -1.$$

$$P(\bar{X} < 193 \mid \mu = 195) = P(Z < -1) = 0,5 - 0,3413 = 0,1587 = 15,87\%.$$

Como se trata de um teste bicaudal, o valor-p é o dobro disso, isto é, **31,74%**.

Veja que é possível conduzir o teste de outra maneira se, a primeira coisa que se fizer for calcular o valor-p. Como a significância é 10% e o valor-p é maior, não se rejeita a hipótese nula. Se o valor-p fosse menor, a hipótese nula seria rejeitada.

Exercícios

1. Tomando-se uma amostra de 30 alunos de uma faculdade, verificou-se que a nota média do provão foi 4,0. Sabendo-se que o desvio padrão das notas é de 1,5, determine:
 - a) Um intervalo que contenha 60% dos alunos dessa faculdade.
 - b) Um intervalo de 90% de confiança para a média obtida pela faculdade.
 - c) Você utilizou alguma hipótese adicional para resolver os itens anteriores? Se sim, qual(is) hipótese(s) em qual(ns) item(ns)?
2. Num estudo sobre a renda em determinada cidade com uma amostra de 36 habitantes, encontrou-se uma renda média de R\$ 830,00. Estudo anterior encontrou um valor de R\$ 800,00. Teste se o primeiro estudo continua válido com nível de significância de 2%, sendo conhecida a variância da renda de 9.600.
3. Estudo feito sobre a mortalidade infantil em 40 cidades em um estado encontrou um valor de 80 por mil crianças nascidas. O governador afirma, no entanto, que a mortalidade infantil não passa de 70 por mil. Teste essa afirmação utilizando significância de 10%, sabendo que o desvio padrão da mortalidade infantil é 20.
4. Numa pesquisa entre 500 eleitores, 100 declararam intenção de votar no candidato A.
 - a) Construa um intervalo de confiança de 95% para as intenções de voto em A.
 - b) O candidato A afirma que possui, no mínimo, 25% das intenções de voto. Teste a afirmação do candidato com 5% de significância.
 - c) Quantos deveriam ser os eleitores pesquisados de tal modo que a “margem de erro” do item a) seja 2% (dois pontos percentuais). (Considere que a proporção amostral se mantenha com o aumento da amostra.)
5. O valor médio dos aluguéis em um bairro, obtido de uma amostra de 30 imóveis, é R\$ 290,00. Em outro bairro, numa amostra de 26 imóveis, foi obtido um valor de R\$ 310,00. Teste a afirmação de que o valor médio do aluguel é idêntico nos dois bairros, com significância de 5%, sabendo-se que os desvios padrão são iguais a 50 e 40, respectivamente.
6. O fabricante de uma máquina de empacotar afirma que o desvio padrão máximo dos pacotes embalados por ela é de 9g. Numa amostra de 15 pacotes, o desvio padrão encontrado foi de 10g. Teste a afirmação do fabricante com um nível de significância de 5%, admitindo que a distribuição seja normal.
7. Imagina-se que o desvio padrão das idades de uma classe seja 2 anos. Tomando-se 5 pessoas aleatoriamente, as idades foram de: 30, 27, 25, 29 e 22. Teste com 10% de significância a validade da afirmação inicial, supondo distribuição normal para as idades.

8. Numa pesquisa com 20 economistas, os valores da média e do desvio padrão dos salários foram R\$ 2.000,00 e R\$ 500,00. Se os salários são distribuídos normalmente, teste a afirmação de que o salário médio dos economistas é, no mínimo, R\$ 2.250,00, utilizando um nível de 5% de significância.
9. Com os dados do exercício 7, teste, com 1% de significância, a afirmação de que a média de idade da classe é de 30 anos.
10. Na cidade X, de uma amostra de 26 habitantes, foi obtida uma renda média de R\$ 600,00 com desvio padrão de R\$ 200. Na cidade Y, com uma amostra de 20 habitantes, foi obtida a mesma renda média, mas com desvio padrão de R\$ 300,00. Afirma-se que a distribuição de renda na cidade Y é pior do que a da cidade X. Teste essa afirmação com 5% de significância, admitindo que a distribuição da renda é normal nas duas cidades.
11. Foi feito um estudo em duas fábricas para investigar a uniformidade da produção em ambas (durante cinco dias na fábrica I e quatro na fábrica II). Teste com 10% de significância se as duas fábricas variam a produção da mesma forma, admitindo que a distribuição seja normal em ambos os casos.

Tabela 7.1 – Produção nas fábricas I e II

Fábrica	Produção				
	dia 1	dia 2	dia 3	dia 4	dia 5
I	100	120	90	95	105
II	105	104	96	94	-

12. A média de uma variável aleatória é 120. Sem saber disso, um pesquisador usa uma amostra de 15 elementos para testar a hipótese de que a média é igual a 100 (teste bicultural). Sabendo-se que a variância dessa variável é 400 (e isso também é sabido pelo pesquisador), se o nível de significância utilizado é 10%, qual é o poder do teste? E se o nível de significância for 5%? Qual será o poder do teste se ele for para a média igual a 90?
13. Uma caixa contém bolas brancas e pretas. Quer-se testar a hipótese de que a proporção seja de metade para cada cor. Para isso, retiram-se 50 bolas (com reposição). O critério adotado é o seguinte: se o número de bolas brancas retiradas for de 20 a 30 (inclusive), aceita-se a hipótese nula de que as proporções são iguais. Determine a significância deste teste.
14. Para pesquisar o gasto médio mensal em cinema em uma comunidade foram pesquisadas 5 famílias. O gasto delas em um mês foi de 40, 50, 30, 20 e 30 reais, respectivamente.
 - Afirma-se que o gasto médio mensal desta comunidade é de 40 reais. Teste esta afirmação para 10% de significância.
 - Afirma-se que o desvio padrão do gasto é de R\$ 10/mês. Teste esta afirmação para 10% de significância.
 - É necessária alguma hipótese adicional para a resolução dos itens anteriores? Justifique.

15. Em uma prova, um aluno afirma que o professor não deu a matéria cobrada em uma questão de múltipla escolha com 5 alternativas. O professor argumenta que isso é impossível, porque, em uma classe com 50 alunos, 16 acertaram a questão. Teste, com 5% de significância, a hipótese de que os alunos tenham acertado a questão no “chute”.
16. O responsável pelo controle de qualidade de uma fábrica afirma que, no máximo, 1% de seus produtos é defeituoso. Numa amostra de 200 produtos, foram encontrados 4 com defeito. Teste a hipótese do responsável com 8% de significância.
17. Duas fábricas, A e B, produzem determinado tipo de lâmpada, sendo que, as da fábrica A duram em média 90 horas, enquanto as da fábrica B, 110 horas. Um comprador dessas lâmpadas decide verificar a origem de seu estoque. Para isso, seleciona uma amostra aleatória de 10 lâmpadas e verifica a duração de cada uma delas. Se a duração média for maior do que 100 horas, conclui que a lâmpada foi fabricada pela fábrica B; caso contrário, conclui que veio da fábrica A. Considere que a duração das lâmpadas das duas fábricas segue distribuição normal com desvio padrão de 10 horas.
- Quais as hipóteses nula e alternativa desse teste?
 - Qual a significância desse teste?
 - Qual o poder do teste?
 - Qual seria o critério de decisão se a significância do teste fosse, como é usual, estabelecida previamente em 5%? E em 10%?
18. Assinale verdadeiro ou falso:
- Num teste para a média, podemos **sempre** utilizar a distribuição normal.
 - Dada a variância amostral S^2 , obtida numa amostra de n elementos, sabe-se que a expressão $(n-1)S^2/\sigma^2$ segue uma distribuição χ^2 com $n - 1$ graus de liberdade.
 - A distribuição χ^2 com $n - 1$ graus de liberdade é a distribuição de uma variável que é a soma de $n - 1$ variáveis normais.
 - A distribuição χ^2 com $n - 1$ graus de liberdade é a distribuição de uma variável que é a soma de n variáveis normais padronizadas.
 - Não é possível realizar testes de comparação de variâncias se as médias são diferentes.
 - A média de uma variável, cuja distribuição é a t de Student, é zero.
 - Um teste é realizado com 5% de significância. Se o mesmo teste for repetido, com a mesma amostra, a 1% de significância, terá um poder maior.
 - Um teste é realizado com 5% de significância. Se for utilizada uma amostra maior, mantidos os 5% de significância, a probabilidade de erro do tipo I será menor.
 - Um teste é realizado com 5% de significância. Se for utilizada uma amostra maior, mantidos os 5% de significância, a probabilidade de erro do tipo II será menor.
 - O valor-p de um teste é a menor significância pela qual a hipótese nula seria rejeitada.
19. Determine o valor-p dos testes dos Exemplos 7.2.2, 7.2.3 e 7.2.4.

Apêndice 7.B – Propriedades e conceitos adicionais de testes de hipóteses

7.B.1 Caso geral dos testes de hipóteses

Ao longo deste capítulo os testes sempre foram do tipo variável = valor, ou variável 1 = variável 2, sempre sendo essas variáveis e valores escalares.

No caso mais geral, a hipótese nula seria que o parâmetro θ pertence a um conjunto $\bar{\omega}$. A hipótese alternativa que θ pertence, na verdade, ao complementar de $\bar{\omega}$:

$$H_0: \theta \in \bar{\omega}$$

$$H_1: \theta \in \omega$$

Nesse sentido, os testes de hipótese monocaudais apresentados neste capítulo estariam mais bem representados se a hipótese nula também fosse uma desigualdade, de modo que a hipótese alternativa representasse de fato o complementar, desta forma:

$$H_0: \theta \leq \theta_0$$

$$H_1: \theta > \theta_0$$

Para a hipótese alternativa “maior que”. Ou:

$$H_0: \theta \geq \theta_0$$

$$H_1: \theta < \theta_0$$

Para a hipótese alternativa “menor que”. Ao longo do texto, entretanto, foi mantida a convenção da maioria dos livros de que a hipótese nula deve ser sempre representada por uma igualdade.

7.B.2 Propriedades desejáveis de testes de hipóteses

Assim como estimadores, testes de hipóteses também devem ter algumas propriedades.

Um teste de hipóteses é dito **não viesado** se a probabilidade de rejeitar a hipótese nula quando ela é falsa é maior do que a de rejeitar a hipótese nula quando ela é verdadeira. Em outras palavras, ele será não viesado se o poder do teste for maior do que sua significância.

Um teste T_1 com significância α_1 e tendo β_1 como a probabilidade de cometer o erro do tipo II é dito **inadmissível** se houver um teste T_2 de tal modo que $\alpha_2 \leq \alpha_1$ e $\beta_2 \leq \beta_1$ (com a desigualdade estrita valendo em pelo menos um dos casos).

Finalmente, um teste é dito **o mais poderoso** se, para dado nível de significância, for o teste que apresentar o maior poder, isto é, a maior probabilidade de rejeitar a hipótese nula quando ela é falsa.

7.B.3 Teste de comparação de médias quando a variância é desconhecida

Este teste tem as seguintes hipóteses, no caso bicaudal:

$$H_0: \mu_A = \mu_B$$

$$H_1: \mu_A \neq \mu_B$$

Ou, como vimos, alternativamente:

$$H_0: \mu_A - \mu_B = 0$$

$$H_1: \mu_A - \mu_B \neq 0$$

As duas populações são normalmente distribuídas. O tamanho das amostras são n_A e n_B ; as médias amostrais são \bar{X}_A e \bar{X}_B ; e as variâncias amostrais são S_A^2 e S_B^2 .

Há duas possibilidades: a primeira é a de que, embora as variâncias amostrais sejam diferentes, elas sejam estimadores de uma mesma variância populacional.

O estimador dessa variância é dado por uma média ponderada das variâncias amostrais:

$$S^2 = \frac{(n_A - 1)S_A^2 + (n_B - 1)S_B^2}{n_A + n_B - 2}$$

A estatística do teste é dada por:

$$\frac{|\bar{X}_A - \bar{X}_B|}{\sqrt{\frac{S^2}{n_A} + \frac{S^2}{n_B}}} = \frac{|\bar{X}_A - \bar{X}_B|}{S \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}}$$

Que, sob a hipótese nula, segue uma distribuição *t de Student* com $n_A + n_B - 2$ graus de liberdade.

Outra possibilidade é a de que as variâncias sejam, na verdade, diferentes. Então a estatística é dada por:

$$\frac{|\bar{X}_A - \bar{X}_B|}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}}$$

Que é possível demonstrar que segue (aproximadamente) uma distribuição *t de Student* com η graus de liberdade, onde η é dado por:

$$\eta = \frac{\left(\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B} \right)}{\left(\frac{S_A^2}{n_A} \right)^2 + \left(\frac{S_B^2}{n_B} \right)^2} \cdot \frac{(n_A - 1)}{(n_A - 1) + (n_B - 1)}$$

7.B.4 Quadro resumindo algumas das principais distribuições contínuas

Distribuição	Função densidade – $f(x)$	Média	Variância
Normal	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2
χ^2 com n graus de liberdade	$\frac{(x/2)^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{2\Gamma(n/2)}$	n	$2n$
t de Student	$\frac{1}{\sqrt{n}} \frac{\Gamma[(n+1)/2]}{\Gamma(n/2)\Gamma(1/2)} \left 1 + \frac{x^2}{n}\right ^{-(n+1)/2}$	0 ($n > 1$)	$\frac{n}{(n-2)}, n > 2$
Fisher-Snedecor	$\left[\frac{m}{n}\right]^{m/2} \frac{\Gamma[(m+n)/2]}{\Gamma(m/2)\Gamma(n/2)} \frac{x^{(m-2)/2}}{[1 + (m/n)x]^{(m+n)/2}}$	$\frac{n}{(n-2)}$ ($n > 2$)	$\frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$ ($n > 4$)

Onde $\Gamma(\alpha) = \int_0^\infty e^{-x} x^{\alpha-1} dx$, sendo que para α inteiro positivo, $\Gamma(\alpha) = (\alpha - 1)!$

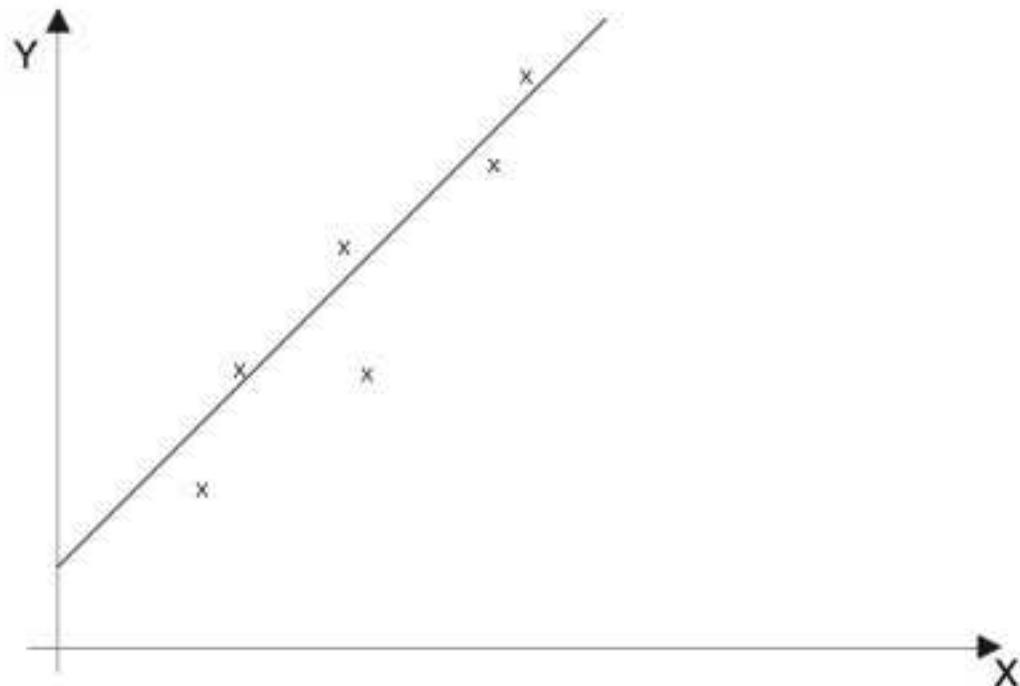


Capítulo 8

REGRESSÃO LINEAR

Imagine duas variáveis, que serão chamadas genericamente de Y e X – mas que poderiam ser consumo e renda; salários e anos de estudo; pressão de um gás e sua temperatura; vendas e gastos em propaganda; enfim, quaisquer duas variáveis que, supostamente, tenham relação entre si. Suponha, ainda, que X é a variável independente e Y é a variável dependente, isto é, Y que é afetado por X , e não o contrário.

Figura 8.1 – Relação entre as variáveis X e Y (pontos)



No gráfico da Figura 8.1, verifica-se que existe, sim, uma dependência entre Y e X . O processo de encontrar a relação entre Y e X é chamado de **regressão**. Se esse processo é uma reta (como parece ser o caso), é uma **regressão linear**. Se houver apenas uma variável independente (só tem um X), é uma **regressão linear simples**.

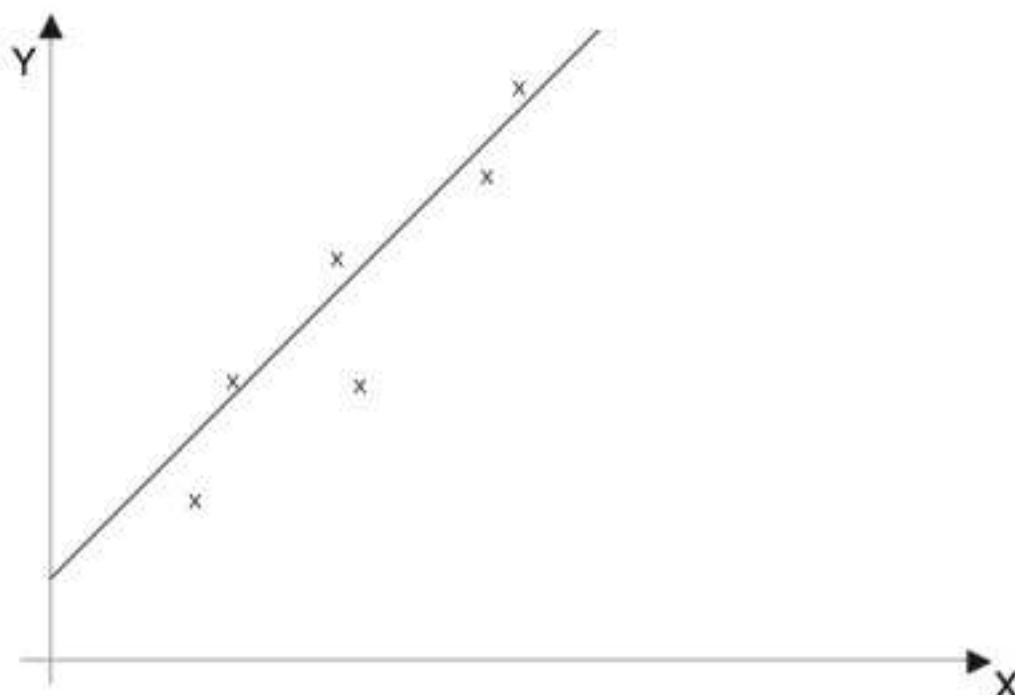
8.1 Regressão linear simples

Como a relação expressa pelo gráfico da Figura 8.1 é, aparentemente, uma função afim (*linear*), cada Y pode ser escrito em função de cada X da seguinte forma:

$$Y_i = \alpha + \beta X_i + \epsilon_i$$

Sendo $\alpha + \beta X$ a equação da reta, e ϵ o termo de erro. Este último termo tem de ser incluído porque, como você pode ver, o valor de Y não será dado exatamente pelo ponto da reta a ser encontrada, como pode ser observado no gráfico da Figura 8.2.

Figura 8.2 – Relação entre as variáveis X e Y (pontos e reta)



Qual a razão de existir esse erro? (Repare que ainda não estamos falando de estimadores; essa relação é, supostamente, exata!) Bom, uma razão seria a existência de imprecisões em medidas, o que é o mais comum em experimentos de laboratório – por mais preciso que seja um instrumento de medida, sempre haverá um limite para essa precisão. No caso de modelos econômicos ou que envolvam qualquer tipo de ciência social, esse erro é um componente importante.

Imagine que Y seja o preço de um imóvel e X sua área. Suponha, ainda, que, em todos os casos analisados, o bairro seja o mesmo, o padrão de construção também, de modo que a única variável (conhecida) que influencia o preço do imóvel é sua área. Ainda assim, haveria pontos acima e abaixo da reta.

Um ponto abaixo poderia ser o da dona Maricota, simpática senhora aposentada e viúva que, precisando com urgência de dinheiro para um tratamento médico e não estando informada a respeito do mercado imobiliário da região, vendeu uma casa que seu marido deixou de herança por um preço abaixo do que seria o de mercado.

Um ponto acima poderia ser o do seu João, antigo morador do bairro que, depois de se tornar comerciante bem-sucedido, fez questão de voltar às suas origens e fez uma oferta irrecusável por uma casa do bairro.

Note que é impossível num emaranhado de pontos se conhecerem todas as *histórias*. Mesmo que fossem conhecidas, essas variáveis seriam muito difíceis de medir. Como seria igualmente difícil medir a euforia causada por uma grande conquista esportiva ou

militar (ou a depressão pela derrota) que faria com que o consumo, naquele ano, fosse proporcionalmente maior (ou menor) em relação à renda.

Enfim, o erro dá conta de todos esses eventos que são difíceis de medir, mas que são (supostamente) aleatórios. Mais do que isso, se o modelo (a reta) estiver corretamente especificado, pode-se supor que o erro, em média, será zero. Traduzindo: a probabilidade de o erro ser x unidades acima da reta é a mesma de ser x unidades abaixo.

Essa é a primeira hipótese a ser feita sobre o erro – em média, ele é zero, isto é:

$$E(\varepsilon_i) = 0$$

Bom, o próximo passo é encontrar ou, melhor dizendo, *estimar* a reta de regressão, já que sempre se está trabalhando com uma amostra, o que implica que não se têm os valores verdadeiros de α e β , mas seus estimadores.

8.2 Método dos mínimos quadrados

Encontrar (estimar, na verdade) a reta de regressão significa encontrar estimadores para α e β . Com um pequeno *truque* é possível tornar esse trabalho mais fácil.

Vamos definir as variáveis x e y da seguinte forma:

$$x = X - \bar{X}$$

$$y = Y - \bar{Y}$$

As variáveis x e y são ditas centradas na média.

Assim, como a média dos erros é zero, tomando as médias da equação da reta, chega-se a:

$$\bar{Y}_i = \alpha + \beta \bar{X}_i + \varepsilon_i$$

$$\bar{Y} = \alpha + \beta \bar{X} + 0$$

Subtrai-se a segunda equação da primeira:

$$\bar{Y}_i - \bar{Y} = (\alpha - \alpha) + \beta(X_i - \bar{X}) + \varepsilon_i$$

$$y_i = \beta x_i + \varepsilon_i$$

Ou seja, se as variáveis forem consideradas centradas na média, em vez das variáveis originais, reduz-se o trabalho no que se refere ao número de parâmetros a ser estimado.

O método a ser utilizado pressupõe que se queira estimar uma reta que tenha *menos erro*, mas somar os erros, pura e simplesmente, não acrescenta muita informação, pois haverá erros positivos e negativos (de pontos acima e abaixo da reta), que se *cancelarão* numa soma simples.

Um problema parecido foi resolvido quando foi definida a variância: basta elevar ao quadrado, eliminando assim os números negativos. Então, a *melhor reta* será aquela cuja soma dos quadrados dos erros for mínima. Daí o nome **método dos mínimos quadrados**.

Da equação da reta usando as variáveis centradas, o(s) erro(s) é(são) dado(s) por:

$$\varepsilon_i = y_i - \beta x_i$$

A soma dos quadrados dos erros é:

$$\sum_{i=1}^n (\varepsilon_i)^2 = \sum_{i=1}^n (y_i - \beta x_i)^2$$

Omitindo, por mera economia de notação, os índices $i = 1$ a n :

$$\Sigma \varepsilon^2 = \Sigma (y_i - \beta x_i)^2$$

$$\Sigma \varepsilon^2 = \Sigma (y^2 + \beta^2 x^2 - 2\beta x y)$$

Utilizando as propriedades da soma, vem:

$$\Sigma \varepsilon^2 = \Sigma y^2 + \Sigma \beta^2 x^2 - 2 \Sigma \beta x y$$

Como β é uma constante em todo o somatório:

$$\Sigma \varepsilon^2 = \Sigma y^2 + \beta^2 \Sigma x^2 - 2 \beta \Sigma x y$$

Para encontrar o valor de β que dê o mínimo dessa soma, o procedimento é derivar e igualar a zero. Como esse valor de β é um estimador, a partir de agora utilizaremos $\hat{\beta}$. Derivando em relação a β :

$$2 \hat{\beta} \Sigma x^2 - 2 \Sigma x y = 0$$

Note que, em vez de β , utiliza-se $\hat{\beta}$, já que o valor que minimiza a expressão anterior vai ser retirado da amostra. É, portanto, um estimador de β .

Dividindo por dois ambos os lados:

$$\hat{\beta} \Sigma x^2 - \Sigma x y = 0$$

$$\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2}$$

O estimador para α pode ser facilmente encontrado na equação da reta para as médias:

$$\bar{Y} = \alpha + \beta \bar{X}$$

Substituindo pelos respectivos estimadores:

$$\bar{Y} = \hat{\alpha} + \hat{\beta} \bar{X}$$

Portanto:

$$\hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X}$$

Exemplo 8.2.1

Dados os valores de Y e X na Tabela 8.1, estime a reta que exprime a relação entre Y e X .

Tabela 8.1 – Valores de X e Y

X	Y	X	Y	X	Y	X	Y	X	Y
103	160	189	256	141	193	197	273	107	161
123	167	211	290	156	219	204	272	135	195
145	207	178	237	166	235	125	181	144	201
126	173	155	209	179	234	112	166	188	255

O primeiro passo é calcular a média de Y e X e encontrar as variáveis centradas, como mostra a Tabela 8.2.

Tabela 8.2 – Valores de X e Y (centradas na média)

	X	Y	x	y
	103	160	-51,2	-54,2
	123	167	-31,2	-47,2
	145	207	-9,2	-7,2
	126	173	-28,2	-41,2
	189	256	34,8	41,8
	211	290	56,8	75,8
	178	237	23,8	22,8
	155	209	0,8	-5,2
	141	193	-13,2	-21,2
	156	219	1,8	4,8
	166	235	11,8	20,8
	179	234	24,8	19,8
	197	273	42,8	58,8
	204	272	49,8	57,8
	125	181	-29,2	-33,2
	112	166	-42,2	-48,2
	107	161	-47,2	-53,2
	135	195	-19,2	-19,2
	144	201	-10,2	-13,2
	188	255	33,8	40,8
Soma	3.084	4.284	0	0
Média	154,2	214,2	0	0

Note que, se a variável é centrada na média, sua soma (e, por conseguinte, sua média), é zero.

Agora, determine x^2 , y^2 e xy , como mostra a Tabela 8.3.

Tabela 8.3 – Quadrados e produtos das variáveis centradas de X e Y

	X	Y	x	y	x^2	y^2	xy
	103	160	-51,2	-54,2	2.621,44	2.937,64	2.775,04
	123	167	-31,2	-47,2	973,44	2.227,84	1.472,64

	X	Y	x	y	x^2	y^2	xy
	145	207	-9,2	-7,2	84,64	51,84	66,24
	126	173	-28,2	-41,2	795,24	1.697,44	1.161,84
	189	256	34,8	41,8	1.211,04	1.747,24	1.454,64
	211	290	56,8	75,8	3.226,24	5.745,64	4.305,44
	178	237	23,8	22,8	566,44	519,84	542,64
	155	209	0,8	-5,2	0,64	27,04	-4,16
	141	193	-13,2	-21,2	174,24	449,44	279,84
	156	219	1,8	4,8	3,24	23,04	8,64
	166	235	11,8	20,8	139,24	432,64	245,44
	179	234	24,8	19,8	615,04	392,04	491,04
	197	273	42,8	58,8	1.831,84	3.457,44	2.516,64
	204	272	49,8	57,8	2.480,04	3.340,84	2.878,44
	125	181	-29,2	-33,2	852,64	1.102,24	969,44
	112	166	-42,2	-48,2	1.780,84	2.323,24	2.034,04
	107	161	-47,2	-53,2	2.227,84	2.830,24	2.511,04
	135	195	-19,2	-19,2	368,64	368,64	368,64
	144	201	-10,2	-13,2	104,04	174,24	134,64
	188	255	33,8	40,8	1.142,44	1.664,64	1.379,04
Soma	3.084	4.284	0	0	21.199,2	31.513,2	25.591,2
Média	154,2	214,2	0	0	1.059,96	1.575,66	1.279,56

Agora, pode-se facilmente estimar a reta de regressão:

$$\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{25.591,2}{21.199,2} \cong 1,207$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X} = 214,2 - 1,207 \times 154,2 \cong 28,05$$

Portanto, a reta estimada é dada por:

$$\hat{Y} = 28,05 + 1,207X$$

Isso quer dizer que, se X for igual a 300, um valor estimado (médio) para Y é dado por:

$$\hat{Y} = 28,05 + 1,207 \times 300 \cong 390,2$$

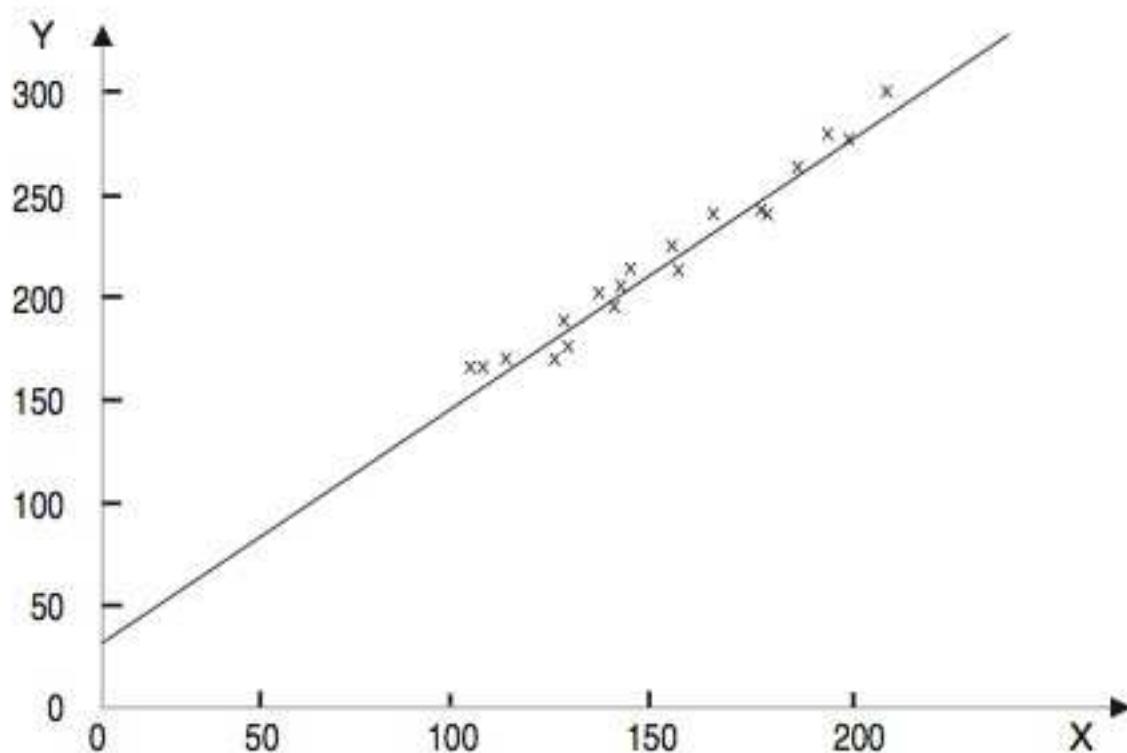
Mas fica uma questão: essa previsão é confiável? Ou essa regressão é *boa*? Veja o exemplo seguinte.

Exemplo 8.2.2

Teste a validade da regressão do Exemplo 8.2.1.

Embora não seja muito rigorosa, uma inspeção gráfica, na base do *olhômetro*, é sempre útil. Se você colocar no mesmo plano cartesiano os pontos dados na Tabela 8.2 e a reta obtida pela regressão, tem-se a Figura 8.3.

Figura 8.3 – Relação entre (novos) X e Y



Visualmente, pode-se constatar que, de fato, a relação é uma reta e que a reta de regressão prevê com boa precisão os valores verdadeiros de Y.

Como se pode verificar isso de maneira mais rigorosa? A primeira coisa é calcular a diferença entre os Y dados no exemplo e os calculados pela reta de regressão (\hat{Y}) na Tabela 8.4.

Tabela 8.4 – Diferença entre Y e \hat{Y}

	X	Y	\hat{Y}	Y - \hat{Y}
	103	160	152,39	7,61
	123	167	176,54	-9,54
	145	207	203,09	3,91
	126	173	180,16	-7,16
	189	256	256,21	-0,21
	211	290	282,77	7,23
	178	237	242,93	-5,93
	155	209	215,17	-6,17
	141	193	198,27	-5,27
	156	219	216,37	2,63
	166	235	228,44	6,56
	179	234	244,14	-10,14
	197	273	265,87	7,13
	204	272	274,32	-2,32
	125	181	178,95	2,05
	112	166	163,26	2,74
	107	161	157,22	3,78
	135	195	191,02	3,98
	144	201	201,89	-0,89
	188	255	255,00	0,00
Soma	3.084	4.284	4.284	0
Média	154,2	214,2	214,2	0

De fato, você pode verificar que as diferenças são bem pequenas quando comparadas com os valores de Y .

Essas diferenças, aliás, podem ser precipitadamente confundidas com os erros. É quase isso. Os erros são as diferenças entre os valores de Y e a reta *verdadeira*, isto é, a reta dada pelos valores populacionais de α e β (que não são conhecidos). As diferenças encontradas são entre os valores de Y e os dados pela reta com os valores estimados (amostrais) de α e β . São, portanto, não os erros, mas os estimadores dos erros, ou simplesmente os **resíduos** da regressão.

Faça, agora, uma análise com os quadrados dos resíduos e com sua variância. Essa análise, mostrada na Tabela 8.5, é conhecida como **análise de variância** ou, pela sigla em língua inglesa, **ANOVA**.

Tabela 8.5 – ANOVA para X e Y

	X	Y	\hat{Y}	Resíduos	Quadrados dos resíduos
	103	160	152,39	7,61	57,87
	123	167	176,54	-9,54	90,94
	145	207	203,09	3,91	15,26
	126	173	180,16	-7,16	51,23
	189	256	256,21	-0,21	0,04
	211	290	282,77	7,23	52,31
	178	237	242,93	-5,93	35,17
	155	209	215,17	-6,17	38,02
	141	193	198,27	-5,27	27,72
	156	219	216,37	2,63	6,90
	166	235	228,44	6,56	42,97
	179	234	244,14	-10,14	102,78
	197	273	265,87	7,13	50,88
	204	272	274,32	-2,32	5,37
	125	181	178,95	2,05	4,20
	112	166	163,26	2,74	7,52
	107	161	157,22	3,78	14,28
	135	195	191,02	3,98	15,82
	144	201	201,89	-0,89	0,79
	188	255	255,00	0,00	0,00
Soma	3.084	4.284	4.284	0	620,08
Média	154,2	214,2	214,2	0	31,004

A análise de variância envolve dividir a variável Y em duas partes: a parte explicada pela regressão e a não explicada (resíduos). Então, o primeiro passo é calcular a soma dos quadrados da variável Y e de suas partes explicada e não explicada. Como se trata de variância, trata-se aqui da variável menos a média, isto é, das variáveis centradas na média. Calcule, então, a soma dos quadrados totais (SQT) de Y (centrado), a soma dos quadrados explicados (SQE), isto é, do Y estimado, e a soma dos quadrados dos resíduos (SQR).

A soma dos quadrados totais já foi calculada no Exemplo 8.2.1.

$$SQT = \sum y_i^2 = 31.513,2$$

Para o cálculo da soma dos quadrados explicados, há duas maneiras: ou se calcula um a um, tira-se a média e se eleva ao quadrado, ou pode-se utilizar a equação da reta:

$$\hat{y}_i = \hat{\beta}x_i$$

$$SQE = \sum \hat{y}_i^2 = \sum (\hat{\beta}x_i)^2 = \sum \hat{\beta}^2 x_i^2 = \hat{\beta}^2 \sum x_i^2 = 30.893,12$$

A soma dos quadrados dos resíduos foi calculada já nesse exemplo, na Tabela 8.5:

$$SQR = 620,08$$

Repare que:

$$SQT = SQE + SQR$$

Portanto, não seria necessário calcular as três, bastariam duas e a terceira sairia por essa relação.

Comece, então, a preencher a Tabela 8.6 pelas somas de quadrados:

Tabela 8.6 – Soma de quadrados	
	Soma de quadrados
SQE	= 30.893,12
SQR	= 620,08
SQT	= 31.513,2

Com essas informações, já é possível tirar uma conclusão a respeito da regressão, já que a soma dos quadrados dos resíduos é uma parcela bem pequena do total ou, o que é equivalente, a soma dos quadrados explicados é uma parcela importante. Essa proporção é conhecida como **poder explicativo, coeficiente de determinação**, ou simplesmente **R²**:

$$R^2 = \frac{SQE}{SQT} = \frac{30.893,12}{31.512,2} \cong 0,9803 = 98,03\%$$

Repare que é impossível que a SQE seja maior do que a SQT e, como se trata de uma soma de quadrados, ela não pode ser negativa. Então, em qualquer regressão, $0 \leq R^2 \leq 1$, sendo válido expressá-lo como um percentual.

Como o R² encontrado foi 98,03%, diz-se que 98,03% da variância de Y é explicada pela variável X, o que indica que a regressão de Y por X apresentou um resultado (muito) bom.

Aliás, o termo R² é, assim, ao quadrado, por uma razão muito simples: o coeficiente de correlação entre X e Y é 0,9901 (verifique!), que ao quadrado é 0,9803. O R², numa regressão simples, é o quadrado do coeficiente de correlação amostral entre X e Y. Ou ainda, o que dá na mesma (por quê?), o quadrado do coeficiente de correlação amostral entre \hat{Y} e Y. Esta última vale para regressões que não forem simples também.

A análise continua. Na coluna seguinte, colocam-se os graus de liberdade. Para a SQT, os graus de liberdade são os mesmos de uma variância amostral normal, isto é, $n - 1 (= 20 - 1 = 19)$.

Para a soma de quadrados dos resíduos, é preciso lembrar que são resíduos de uma reta. Para uma reta, você sabe, são necessários dois pontos. Porém, com apenas dois pontos não teríamos variação alguma (portanto, nenhum resíduo). Os graus de liberdade em relação aos resíduos são, dessa forma, $n - 2 (= 20 - 2 = 18)$.

Quanto à SQE, há dois raciocínios: ou a diferença ($19 - 18 = 1$) ou o fato de que há *apenas uma* variável explicativa (afinal, é uma regressão simples). Portanto, chega-se à Tabela 8.7.

Tabela 8.7 – Soma de quadrados e graus de liberdade

Soma de quadrados	g. l.
SQE = 30.893,12	1
SQR = 620,08	18
SQT = 31.513,2	19

Agora, resta calcular as variâncias propriamente ditas ou, como preferem alguns, os quadrados médios, dividindo as somas de quadrados pelos respectivos graus de liberdade (Tabela 8.8).

Tabela 8.8 – Acrescentando quadrados médios

Soma de quadrados	g. l.	Quadrados médios
SQE = 30.893,12	1	30.893,12
SQR = 620,08	18	34,45
SQT = 31.513,2	19	1.658,59

O que você vai testar, agora, é se, estatisticamente falando, a variância explicada é maior do que a variância dos resíduos, isto é, fará um teste de comparação de variâncias. Se rejeitar a hipótese nula de que as variâncias são iguais, a regressão *explica mais do que não explica* e, então, considera-se a regressão válida.

O teste F é feito dividindo-se uma variância pela outra. Porém, para realizar tal teste, é necessário que as variáveis das quais foram obtidas as variâncias sejam normais. Portanto, para realizar esse teste é preciso que a variável Y seja normalmente distribuída. Como ela é composta de uma reta (fixa) mais um erro aleatório, a variância de Y será dada pela variância do erro. Por isso, cria-se uma hipótese adicional sobre o erro, a de que ele segue uma distribuição normal.

Faça, então, o teste F na Tabela 8.9.

Tabela 8.9 – Tabela ANOVA completa (com teste F)

Soma de quadrados	g. l.	Quadrados médios	Teste F
SQE = 30.893,12	1	30.893,12	896,75
SQR = 620,08	18	34,45	
SQT = 31.513,2	19	1.658,59	

Consultando a Tabela 8.9, obtém-se o valor limite da distribuição para o teste, com um grau de liberdade no numerador e 18 graus de liberdade no denominador, com 5% de significância, que é:

$$F_{1,18} = 4,41$$

Como o F calculado é maior do que o tabelado (nesse caso, bem maior), *rejeita-se a hipótese nula*, isto é, a regressão é válida a 5% de significância.

Exemplo 8.2.3

Teste a significância dos parâmetros da regressão obtida no Exemplo 8.2.1.

Testar a significância dos parâmetros significa testar a hipótese nula de que α e β são, na verdade, iguais a zero. Isto é, será que α ou β de fato não existem, e o valor que você encontrou é apenas resultado da amostra?

Isso equivale a testar as seguintes hipóteses para β (e depois também para α):

$$H_0: \beta = 0$$

$$H_1: \beta \neq 0$$

Como são variáveis normalmente distribuídas (mantendo-se a hipótese do Exemplo 8.2.2), cuja variância *não* é conhecida ao certo, a distribuição a ser utilizada é a t, de Student. Os valores tabelados com 18 ($= n - 2$) graus de liberdade, com 1%, 5% e 10% (bicaudais), são:

$$t_{(18,10\%)} = 1,73$$

$$t_{(18,5\%)} = 2,10$$

$$t_{(18,1\%)} = 2,88$$

O valor calculado da estatística é dado por:

$$\frac{\hat{\beta} - 0}{S_{\hat{\beta}}} = \frac{\hat{\beta}}{S_{\hat{\beta}}}$$

Isto é, basta dividir o coeficiente encontrado pelo seu desvio padrão. A questão, agora, é encontrar o desvio padrão de $\hat{\beta}$. Você sabe que:

$$\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2}$$

Então, pode escrever:

$$\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\sum x_i (\beta x_i + \epsilon_i)}{\sum x_i^2} = \frac{\sum (\beta x_i^2 + x_i \epsilon_i)}{\sum x_i^2} = \frac{\beta \sum x_i^2}{\sum x_i^2} + \frac{\sum x_i \epsilon_i}{\sum x_i^2} = \beta + \frac{\sum x_i \epsilon_i}{\sum x_i^2}$$

$$\text{var}(\hat{\beta}) = \text{var}\left(\beta + \frac{\sum x_i \epsilon_i}{\sum x_i^2}\right)$$

Como a soma de uma constante não altera a variância, vale:

$$\text{var}(\hat{\beta}) = \text{var}\left(\frac{\sum x_i \varepsilon_i}{\sum x_i^2}\right)$$

$$\text{var}(\hat{\beta}) = \frac{\sum x_i^2}{(\sum x_i^2)^2} \text{var}(\varepsilon_i) = \frac{1}{\sum x_i^2} \text{var}(\varepsilon_i) = \frac{\sigma^2}{\sum x_i^2}$$

Note que nessa passagem supõe-se que a variância dos erros é constante e que a variância da soma dos erros é igual à soma das variâncias, ou seja, foi suposto que não há correlação entre os erros.

O estimador dessa variância (valor amostral) é:

$$S_{\hat{\beta}}^2 = \frac{1}{\sum x_i^2} \text{var}(\text{resíduos}) = \frac{S^2}{\sum x_i^2}$$

Já que a variância de Y dado X , isto é, a variância de Y no modelo de regressão, é a própria variância dos resíduos, calculada na tabela ANOVA, então, é igual a 34,45 e foi obtida por meio da expressão $SQR/(n - 2)$:

$$S_{\hat{\beta}}^2 = \frac{SQR/(n - 2)}{\sum x_i^2} \text{var}(\text{resíduos})$$

$$S_{\hat{\beta}}^2 = \frac{34,45}{21.199,2} \cong 0,0016 \Rightarrow S_{\hat{\beta}} \cong 0,04$$

O cálculo da estatística é, então:

$$\frac{\hat{\beta}}{S_{\hat{\beta}}} = \frac{1,207}{0,04} \cong 30,2$$

Como o valor calculado é superior aos valores tabelados (inclusive para 1%), rejeita-se a hipótese nula de que β é igual a zero. Dizemos, então, que β é estatisticamente diferente de zero a 1% de significância ou, simplesmente, é *significante a 1%*.

O procedimento para α é quase o mesmo. A diferença está no cálculo de seu desvio padrão.

Você sabe que:

$$\hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X}$$

$$\text{var}(\hat{\alpha}) = \text{var}(\bar{Y} - \hat{\beta} \bar{X})$$

$$\text{var}(\hat{\alpha}) = \text{var}(\bar{Y}) + \text{var}(\hat{\beta} \bar{X})$$

$$\text{var}(\hat{\alpha}) = \frac{\sigma^2}{n} + \bar{X}^2 \text{var}(\hat{\beta})$$

$$\text{var}(\hat{\alpha}) = \frac{\sigma^2}{n} + \bar{X}^2 \sum_{i=1}^n x_i^2 \text{var}(\hat{\beta})$$

$$\text{var}(\hat{\alpha}) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum x_i^2} \right]$$

Cujo estimador é dado por:

$$S_{\hat{\alpha}}^2 = \frac{\text{SQR}}{n-2} \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum x_i^2} \right]$$

$$S_{\hat{\alpha}}^2 = 34,45 \times \left(\frac{1}{20} + \frac{154,2^2}{21.199,2} \right) \cong 40,36 \quad S_{\hat{\alpha}} \cong 6,4$$

O cálculo da estatística é, então:

$$\frac{\hat{\alpha}}{S_{\hat{\alpha}}} = \frac{28,05}{6,4} \cong 4,4$$

Que é superior aos valores tabelados, portanto, α também é *significante a 1%*.

Exemplo 8.2.4

Com uma amostra contendo 16 observações de duas variáveis Y e X , foram obtidos os seguintes resultados:

$$\Sigma X^2 = 57.751 \quad \Sigma Y = 1.918,1$$

$$\Sigma Y^2 = 288.511,35 \quad \Sigma x^2 = 10.553,4375$$

$$\Sigma XY = 127.764,4 \quad \Sigma y^2 = 58.567,124375$$

$$\Sigma X = 869 \quad \Sigma xy = 23.587,59375$$

Sendo $x = X - \bar{X}$ e $y = Y - \bar{Y}$.

Estime os parâmetros da reta de regressão e teste sua significância, assim como a validade da regressão.

Os parâmetros da regressão são dados por:

$$\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{23.587,59375}{210.553,4375} \cong 2,235$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X} = \frac{1.918,1}{16} - 2,235 \times \frac{869}{16} \cong -1,51$$

O modelo encontrado é, então:

$$\hat{Y} = -1,51 + 2,235X$$

Para testar a validade da regressão, monta-se uma tabela ANOVA. Para isso, são calculadas as somas dos quadrados:

$$SQT = \sum y^2 = 58.567,12$$

$$SQE = \hat{\beta}_2 \sum x^2 = 52.719,75$$

$$SQR = SQT - SQE = 5.847,37$$

Tabela 8.10 – Tabela ANOVA para o Exemplo 8.2.4

Soma de quadrados	g. l.	Quadrados médios	Teste F
SQE = 52.719,75	1	52.719,75	126,22
SQR = 5.847,37	14	417,67	
SQT = 58.567,12	15	3.904,47	

Na Tabela 8.10, com 1 grau de liberdade no numerador e 14 no denominador, a 5%, o valor encontrado é 4,60. De novo, o valor encontrado é (bem) maior do que o tabelado, portanto, aceita-se a validade da regressão (com folga).

De quebra, pode-se calcular o poder explicativo (R^2):

$$R^2 = \frac{52.719,75}{58.567,12} = 0,9002$$

Quanto à significância de cada um dos parâmetros, verifique que os desvios padrão são iguais a:

$$S_{\hat{\alpha}} = 11,95$$

$$S_{\hat{\beta}} = 0,199$$

As estatísticas t são, portanto:

$$\frac{\hat{\alpha}}{S_{\hat{\alpha}}} = \frac{-1,51}{11,95} \cong 0,13$$

$$\frac{\hat{\beta}}{S_{\hat{\beta}}} = \frac{2,235}{0,199} \cong 11,2$$

Os valores críticos para a distribuição t , de Student, com 14 graus de liberdade, são:

$$t_{(14,10\%)} = 1,76$$

$$t_{(14,5\%)} = 2,14$$

$$t_{(14,1\%)} = 2,98$$

Como o valor encontrado para β é superior a todos esses valores, ele é significante a 1%. Já para α , ocorre o contrário, portanto, conclui-se que α não é significante, o que vale dizer que não se pode rejeitar a hipótese de que α é zero. É possível, também, dizer simplesmente que o intercepto não existe (do ponto de vista estatístico).

O procedimento agora seria, portanto, retirar o intercepto, isto é, estimar de novo a regressão sem o coeficiente α , o que é feito no exemplo seguinte.

Exemplo 8.2.5

Tendo em vista que o intercepto da regressão do Exemplo 8.2.4 se mostrou estatisticamente insignificante, estime novamente a regressão, desta vez sem o intercepto.

Trata-se, portanto, de estimar os parâmetros de uma reta que passa pela origem, isto é:

$$Y_i = \beta X_i + \epsilon_i$$

Quando o estimador de mínimos quadrados foi encontrado, havia sido utilizado um *truque* de substituir as variáveis originais (X e Y) pelas variáveis centradas. O objetivo era, exatamente, eliminar o intercepto da equação. Como ele de fato não existe, o estimador de mínimos quadrados é o mesmo, exceto pelo fato de que não serão usadas mais variáveis centradas.

$$\hat{\beta} = \frac{\sum X_i Y_i}{\sum X_i^2}$$

Substituindo pelos valores dados no Exemplo 8.2.4:

$$\hat{\beta} = \frac{127.764,4}{57.751} \cong 2,212$$

O modelo é, então, dado por:

$$\hat{Y} = 2,212X$$

Para o teste do coeficiente encontrado, você precisa de seu desvio padrão. A soma dos quadrados explicados pela regressão é dada por:

$$SQE = \hat{\beta}^2 \sum X^2 \cong 282.657,3$$

A soma dos quadrados dos resíduos é, portanto:

$$SQR = SQT - SQE = \sum Y^2 - \hat{\beta}^2 \sum X^2 = 288.511,35 - 282.657,3 = 5.854,05$$

Assim, você pode encontrar a variância dos resíduos (que é a própria variância da regressão):

$$\text{var(resíduos)} = S^2 = \frac{SQR}{n-1} = \frac{5.854,05}{15} = 390,27$$

Repare que foi usado $n-1$, e não $n-2$, como feito quando a regressão incluía o intercepto. Isso é fácil de entender uma vez que, ao excluir o intercepto, supõe-se conhecer a existência de pelo menos um ponto da reta, que é a origem, o que faz ganhar 1 grau de liberdade.

Para calcular a variância (e o desvio padrão) do coeficiente $\hat{\beta}$, usa-se a mesma fórmula já empregada, apenas trocando o x (centrado) pelo X :

$$S_{\hat{\beta}}^2 = \frac{SQR / (n - 1)}{\sum X_i^2} = \frac{390,27}{57.751} \cong 0,00676 \Rightarrow S_{\hat{\beta}} \cong 0,082$$

Portanto, a estatística t é:

$$\frac{\hat{\beta}}{S_{\hat{\beta}}} = \frac{2,212}{0,082} \cong 27$$

O que, evidentemente, é maior do que os valores tabelados. Em todo caso, esses valores, para 15 graus de liberdade, são:

$$t_{(15,10\%)} = 1,75$$

$$t_{(15,5\%)} = 2,13$$

$$t_{(15,1\%)} = 2,95$$

O valor encontrado, 27, é (bem) maior do que os valores tabelados; logo, o coeficiente é significante a 1%.

Até o R^2 tem de ser visto com reservas quando se trata de uma regressão sem intercepto. Isso porque, na medida em que forem usadas variáveis não centradas, ele é diferente do R^2 usual, e ambos não podem ser comparados.¹ Esse R^2 especial para modelos sem intercepto é conhecido como R^2 não centrado ou R^2 bruto. Nesse caso, ele é:

$$R^2_{NC} = \frac{282.657,3}{288.511,35} = 0,9797$$

Quando se comparam os resultados obtidos nos dois modelos (com e sem intercepto), verifica-se que as diferenças entre os coeficientes β são muito pequenas. O desvio padrão, quando a estimativa foi realizada sem intercepto, foi menor (o que é uma vantagem). De fato, se a reta realmente passa pela origem, é razoável que uma estimativa que leve isso em conta seja mais precisa.

Há de se ressaltar, no entanto, que uma estimativa sem o intercepto tem implícita a hipótese de que a reta passa pela origem, o que pode, em alguns casos, ser uma hipótese um pouco exagerada. Além disso, como se viu, os resultados não são tão diferentes, o que faz com que, na maioria dos casos, os benefícios não compensem os custos (de um possível erro na especificação e das peculiaridades na avaliação do modelo). Assim, a estimativa sem o intercepto só é recomendável se existir uma razão muito forte para acreditar que a reta passe mesmo pela origem.

¹ Repare que, se usar R^2 com as variáveis centradas, o resultado poderá ser negativo.

8.3 A hipótese de normalidade

Até agora foram feitas duas hipóteses sobre o modelo de regressão: a de que os erros têm média zero e de que eles são normalmente distribuídos, hipótese essa que foi utilizada para a realização dos testes de hipótese acerca da regressão e de seus parâmetros.

As hipóteses vistas até agora podem ser resumidas assim:

- I. $E(\epsilon_i) = 0$ (*erros têm média zero*).
- II. Erros são normalmente distribuídos.

É razoável assumir que os erros sejam normalmente distribuídos? Sim, se parte do significado do termo de erro é uma soma de fatores que não foram incluídos no modelo (até porque não é possível). Se você imaginar que são muitos os fatores, a soma deles seguirá uma distribuição normal, pelo teorema do limite central. Se a média segue uma distribuição normal, basta multiplicar por n e tem-se a soma, que será, portanto, normalmente distribuída também.

Entretanto, se isso não for considerado satisfatório, é sempre possível testar a hipótese de que os resíduos sejam normais e, portanto, sejam originados de erros também normais e, assim, você tem maior segurança em relação aos testes de hipóteses².

Um teste muito utilizado para isso é o de **Jarque-Bera**, que utiliza os resultados para os momentos da distribuição normal (veja o Apêndice 3.B), em particular os coeficientes de assimetria (que é 0 para a distribuição normal) e de curtose (que vale 3).

O coeficiente de assimetria para os resíduos é dado por:

$$A = \frac{1}{n} \sum_{i=1}^n \left(\frac{\hat{\epsilon}_i}{\sigma} \right)^3$$

E o de curtose:

$$C = \frac{1}{n} \sum_{i=1}^n \left(\frac{\hat{\epsilon}_i}{\sigma} \right)^4$$

O teste de Jarque-Bera é feito por meio da seguinte estatística:

$$JB = \frac{n}{6} \left[A^2 + \frac{1}{4}(C - 3)^2 \right]$$

Demonstra-se que, sob a hipótese nula de que os resíduos sejam normalmente distribuídos, a estatística JB converge assintoticamente para uma distribuição χ^2 com 2 graus de liberdade.

Exemplo 8.3.1

Na Tabela 8.11, são mostrados os resíduos da regressão do Exemplo 8.2.4. Teste sua normalidade.

² Isso vale para amostras pequenas, já que é possível mostrar que a razão entre o coeficiente e seu desvio padrão converge para uma distribuição normal padrão sob a hipótese nula de que o coeficiente seja zero.

Tabela 8.11 – Resíduos da regressão			
22,304	-18,453	32,047	-23,521
30,918	-18,729	11,233	11,033
-20,167	16,519	-7,946	-9,839
-22,239	-16,424	-2,926	16,190

Calcule a variância desse conjunto de valores (independentemente de saber que se trata de resíduos de uma regressão, isto é, divide-se por n e não por $n - 2$) e depois o desvio padrão:

$$\sigma^2 \approx 365,46 \Rightarrow \sigma \approx 19,12$$

O coeficiente de assimetria é dado por:

$$A = \frac{1}{n} \sum_{i=1}^n \left(\frac{\hat{\epsilon}_i}{\sigma} \right)^3 = 0,3051$$

E o de curtose:

$$C = \frac{1}{n} \sum_{i=1}^n \left(\frac{\hat{\epsilon}_i}{\sigma} \right)^4 = 1,6056$$

A estatística de Jarque-Bera é dada, então, por:

$$JB = \frac{n}{6} \left[A^2 + \frac{1}{4}(C - 3)^2 \right] = 1,5443$$

Na tabela de χ^2 , verificamos que, para dois graus de liberdade, o valor crítico (para 10% de significância) é 4,61. Como o valor encontrado para a estatística JB é inferior, aceita-se a hipótese nula de que os resíduos são normais. Ou, em outras palavras, não é possível, estatisticamente falando, rejeitar a hipótese de que a distribuição desses resíduos seja normal.

8.4 Propriedades dos estimadores de mínimos quadrados

8.4.1 O estimador de β é não viésado?

A resposta a essa pergunta remete à esperança do estimador:

$$E(\hat{\beta}) = E\left(\frac{\sum x_i y_i}{\sum x_i^2} \right)$$

$$E(\hat{\beta}) = E\left[\frac{\sum x_i (\beta x_i + \varepsilon_i)}{\sum x_i^2} \right]$$

$$E(\hat{\beta}) = E\left[\frac{\sum(\beta x_i^2 + \epsilon_i x_i)}{\sum x_i^2}\right]$$

Como a esperança da soma é a soma das esperanças:

$$E(\hat{\beta}) = E\left[\frac{\beta \sum x_i^2}{\sum x_i^2}\right] + E\left[\frac{\sum \epsilon_i x_i}{\sum x_i^2}\right]$$

E β é uma constante, tem-se que:

$$E(\hat{\beta}) = E\left[\frac{\beta \sum x_i^2}{\sum x_i^2}\right] + E\left[\frac{\sum \epsilon_i x_i}{\sum x_i^2}\right]$$

$$E(\hat{\beta}) = E(\beta) + E\left[\frac{\sum \epsilon_i x_i}{\sum x_i^2}\right]$$

$$E(\hat{\beta}) = \beta + E\left[\frac{\sum \epsilon_i x_i}{\sum x_i^2}\right]$$

Observe o termo dentro da esperança: considere que os valores x_i são fixos ou, para ser mais preciso, fixos em amostras repetidas. O que significa que, se a amostra é de imóveis, e um dado imóvel é sorteado na amostra, ele tem certa área. Se uma nova amostragem for feita, e esse imóvel for sorteado de novo, apresentará exatamente o mesmo valor para área. Esse valor é fixo, não depende de probabilidade; portanto, a área de um imóvel se enquadra nessa hipótese.

Isso não se aplicaria, por exemplo, se a variável fosse a nota de um aluno em um teste. O mesmo aluno, fazendo um mesmo teste (ou tipo de teste) uma segunda vez, não necessariamente tiraria a mesma nota. Isso depende de uma distribuição de probabilidade, e x é, nesse caso, uma variável estocástica.

Se a variável x for fixa em amostras repetidas (como a área de um imóvel), cada x_i pode ser tratado como uma constante:

$$E(\hat{\beta}) = \beta + \frac{\sum E(\epsilon_i x_i)}{\sum x_i^2}$$

$$E(\epsilon_i x_i) = x_i E(\epsilon_i) = 0$$

Já que $E(\epsilon_i) = 0$. Portanto:

$$E(\hat{\beta}) = \beta + \frac{\sum E(\epsilon_i x_i)}{\sum x_i^2} = \beta$$

Dessa forma, $\hat{\beta}$ é um estimador não viesado do coeficiente β .

Adiciona-se, então, uma terceira hipótese:

- I. $E(\epsilon_i) = 0$ (*erros têm média zero*).
- II. Erros são normalmente distribuídos.
- III. Os x_i são fixos (não estocásticos).

Isso significa que, se a variável x for estocástica, o coeficiente será necessariamente viesado? Não, mas para isso seria preciso manter a condição de que $E(\epsilon_i x_i) = 0$, o que equivale a dizer que a correlação (e a covariância) entre ϵ_i e x_i é nula. Se não, veja:

$$\text{cov}(\epsilon_i, x_i) = E(\epsilon_i x_i) - E(\epsilon_i)E(x_i) = E(\epsilon_i x_i)$$

Já que $E(\epsilon_i) = 0$. Assim, pode-se garantir que o estimador é não viesado com uma hipótese mais fraca. O conjunto de hipóteses seria, nesse caso:

- I. $E(\epsilon_i) = 0$ (*erros têm média zero*).
- II. Erros são normalmente distribuídos.
- III.* $E(\epsilon_i x_i) = 0$ (x_i não são correlacionados com os erros).

8.4.2 Eficiência e MELINV

Se, além das hipóteses I e III, os erros tiverem *variância constante* e não forem *autocorrelacionados* (o erro de uma observação não é correlacionado com o de outra, isto é, os erros são independentes), o **teorema de Gauss-Markov** mostra que o estimador de mínimos quadrados $\hat{\beta}$ apresenta a menor variância entre todos os estimadores de β , que são lineares e não viesados, sendo, portanto, um MELINV (veja a demonstração no Apêndice 8.B).

Acrescentam-se, então, mais duas hipóteses:

- I. $E(\epsilon_i) = 0$ (*erros têm média zero*).
- II. Os erros são normalmente distribuídos.
- III. x_i são fixos (não estocásticos).
- IV. $\text{var}(\epsilon_i) = \sigma^2$ (constante).
- V. $E(\epsilon_i \epsilon_j) = 0, i \neq j$ (*erros não são autocorrelacionados*).

As hipóteses I, II, IV e V podem ser sintetizadas por $\epsilon_i \sim N(0, \sigma^2)$, isto é, os erros são normais e independentemente distribuídos com média zero e variância σ^2 . Se ainda se levar em conta a hipótese de normalidade, é possível demonstrar, por meio da desigualdade de Cramer-Rao, que o estimador $\hat{\beta}$ tem a menor variância entre todos os estimadores não viesados de β , ou seja, é um estimador eficiente.

8.5 Modelos não lineares nas variáveis

Muitos modelos não lineares são facilmente *linearizáveis* (também porque o modelo permanece linear nos **coeficientes**). Por exemplo, o modelo a seguir:

$$Y = \alpha + \beta X_i^2 + \epsilon_i$$

Pode-se tornar um modelo linear por meio da seguinte transformação:

$$Z_i \equiv X_i^2$$

Dessa forma:

$$Y = \alpha + \beta Z_i + \epsilon_i$$

É um modelo linear e pode ser estimado da mesma maneira que vínhamos fazendo.

Dos muitos modelos que podem ser transformados em lineares, dois se destacam.

Um deles é o modelo **multiplicativo**:

$$Y = \alpha X_i^\beta \epsilon_i$$

Aplicando logaritmo nos dois lados da equação:

$$\log Y = \log (\alpha X_i^\beta \epsilon_i)$$

$$\log Y = \log \alpha + \log X_i^\beta + \log \epsilon_i$$

$$\log Y = \log \alpha + \beta \log X_i + \log \epsilon_i$$

Fazendo:

$$Y' = \log Y$$

$$\alpha' = \log \alpha$$

$$X' = \log X$$

$$\mu = \log \epsilon$$

Chega-se a um modelo linear:

$$Y' = \alpha' + \beta X'_i + \mu_i$$

Em que as variáveis estão em logaritmos, por isso mesmo esse modelo é também conhecido como **log-log**.

É interessante notar o significado do coeficiente β nesse tipo de modelo. Isso pode ser feito derivando Y em relação a X :

$$\frac{\partial Y}{\partial X} = \alpha \beta X^{\beta-1} \epsilon = \frac{1}{X} \alpha \beta X^\beta \epsilon = \frac{1}{X} \beta Y$$

Portanto, β é dado por:

$$\beta = \frac{\partial Y}{\partial X} \times \frac{X}{Y}$$

Aproximando a derivada pela taxa de variação discreta:

$$\beta \equiv \frac{\Delta Y}{\Delta X} \times \frac{X}{Y} = \frac{\frac{\Delta Y}{Y}}{\frac{\Delta X}{X}} = \frac{\text{variação percentual de } Y}{\text{variação percentual de } X}$$

Ou seja, quando o modelo é estimado com as variáveis em logaritmo, o coeficiente β significa a razão entre as variações relativas (percentuais) das variáveis Y e X , em vez das absolutas, quando a regressão é feita com os valores originais das variáveis. Essa razão também é conhecida como **elasticidade** de Y em relação a X .

Outro tipo de modelo importante é o **exponencial**:

$$Y = \alpha e^{\beta X_i} \epsilon_i$$

De novo, aplicando logaritmo³ nos dois lados da equação:

$$\log Y = \log (\alpha e^{\beta X_i} \varepsilon_i)$$

$$\log Y = \log \alpha + \log e^{\beta X_i} + \log \varepsilon_i$$

$$\log Y = \log \alpha + \beta X_i + \log \varepsilon_i$$

E, de novo, fazendo as transformações:

$$Y' = \log Y$$

$$\alpha' = \log \alpha$$

$$\mu = \log \varepsilon$$

Temos novamente um modelo linear:

$$Y' = \alpha' + \beta X_i + \mu_i$$

Em que uma das variáveis foi transformada em seu logaritmo e, por isso mesmo, esse modelo é conhecido como **log-linear**.

Da mesma forma, derivou-se Y' em relação a X para encontrar o significado do coeficiente β :

$$\frac{\partial Y}{\partial X} = \beta \alpha e^{\beta X} \varepsilon = \beta Y$$

Portanto:

$$\beta = \frac{1}{Y} \frac{\partial Y}{\partial X}$$

Repetindo a aproximação:

$$\beta = \frac{1}{Y} \frac{\Delta Y}{\Delta X} = \frac{\Delta Y}{\Delta X} = \frac{\text{variação percentual de } Y}{\text{variação percentual de } X}$$

Se a variável X representar o tempo, o coeficiente β representa a taxa de crescimento (médio) da variável Y ao longo do tempo.

Exemplo 8.5.1

A Tabela 8.12 fornece o volume de vendas em uma empresa ao longo do tempo. Determine sua taxa de crescimento anual médio.

Tabela 8.12 – Volume de vendas

Ano	Vendas	Ano	Vendas
1986	1.020	1993	5.300
1987	1.200	1994	6.640

³ Embora nesse caso seja mais prático aplicar o logaritmo natural (base e), é importante ressaltar que tanto faz qual é a base do logaritmo, pois o valor do coeficiente β será o mesmo.

Ano	Vendas	Ano	Vendas
1988	1.450	1995	7.910
1989	1.800	1996	8.405
1990	2.550	1997	9.870
1991	3.320	1998	11.530
1992	4.250	1999	13.320

Para determinar a taxa de crescimento médio, deve-se fazer uma regressão do tipo log-linear, em que a variável Y é o logaritmo das vendas e X é a variável tempo.

Tabela 8.13 – Valores para a regressão log-linear

X	Y	X	Y
1	6,9276	8	8,5755
2	7,0901	9	8,8009
3	7,2793	10	8,9759
4	7,4955	11	9,0366
5	7,8438	12	9,1973
6	8,1077	13	9,3527
7	8,3547	14	9,4970

Note que a mudança na variável tempo (X), que, em vez de começar por 1986, começa por 1, não afeta a taxa de crescimento. (Por quê?)

O resultado da regressão é:

$$\begin{aligned} Y &= 6,77 + 0,2073X \\ &\quad (0,07) \quad (0,008) \end{aligned}$$

Os números entre parênteses são os desvios padrão dos coeficientes.

A taxa média de crescimento anual é, portanto, 0,2073 ou 20,73% *ao ano*.

8.6 Regressão múltipla

E se a variável dependente (Y) depender (com o perdão da redundância) de mais de uma variável? É preciso, então, colocar *mais* X (variáveis independentes) na equação. O modelo, então, de modo geral, seria como o dado a seguir:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + \epsilon_i$$

Como há mais de uma variável independente, esse modelo é conhecido como de **regressão múltipla**. Para estimar os coeficientes β , segue-se o que foi feito com a regressão simples, ou seja, utiliza-se o método dos mínimos quadrados.

Mas, se for feito *exatamente* como anteriormente, dá para perceber que será um pouco complicado e tão mais complicado quanto mais variáveis independentes houver. Um pequeno *truque* deixará esse modelo com uma forma similar à da regressão simples.

Admita que há n observações:

$$Y_1 = \beta_1 + \beta_2 X_{21} + \beta_3 X_{31} + \dots + \beta_k X_{k1} + \epsilon_1$$

$$Y_1 = \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + \dots + \beta_k X_{1k} + \epsilon_1$$

.....

$$Y_n = \beta_0 + \beta_1 X_{n1} + \beta_2 X_{n2} + \dots + \beta_k X_{nk} + \epsilon_n$$

Essas n equações podem ser reescritas em forma de matrizes:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1k} \\ 1 & X_{21} & X_{22} & \dots & X_{2k} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & X_{n1} & X_{n2} & \dots & X_{nk} \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

($n \times 1$) ($n \times k$) ($k \times 1$) ($n \times 1$)

Os valores entre parênteses são as dimensões das matrizes. Repare que, fazendo as respectivas operações com as matrizes, chega-se exatamente ao mesmo conjunto de equações, que podem ser reduzidas a:

$$Y = X\beta + \epsilon$$

Sendo Y um vetor (matriz linha) que contém as observações da variável dependente Y ; X , uma matriz que inclui as diversas observações das variáveis independentes, além de uma coluna de números 1, que correspondem ao intercepto; β , um vetor com os coeficientes a serem estimados; e ϵ , o vetor dos termos de erro.

Exceto por ser uma equação com matrizes, essa equação é muito parecida com a de regressão simples. Melhor ainda, é parecida com a equação de regressão simples sem intercepto. O estimador de mínimos quadrados⁴ para o vetor β será muito parecido com o da regressão simples:

$$\hat{\beta} = (X'X)^{-1}(X'Y)$$

Repare que o produto $X'Y$ é análogo a $\sum xy$ da regressão simples, enquanto o produto $X'X$ é análogo a $\sum x^2$. Como não existe divisão de matrizes, a multiplicação pela matriz inversa *faz o papel* da divisão.

Uma condição para a existência de $\hat{\beta}$ é a de que a matriz $X'X$ seja inversível. Para que isso ocorra, é necessário que nenhuma coluna da matriz X seja combinação linear de outras. Em outras palavras, não é possível que X_2 seja exatamente o dobro de X_1 , ou que X_4 seja igual a $2X_1 + 3X_3$, por exemplo.

Assim, adiciona-se ao conjunto de hipóteses mais uma, esta específica de regressões múltiplas:

- I. $E(\epsilon_i) = 0$ (*erros têm média zero*).
- II. Erros são normalmente distribuídos.
- III. Os x_i são fixos (não estocásticos).
- IV. $\text{var}(\epsilon_i) = \sigma^2$ (constante).
- V. $E(\epsilon_i \epsilon_j) = 0, i \neq j$ (*erros não são autocorrelacionados*).
- VI. Cada variável independente X_i não pode ser combinação linear das demais.

⁴ A derivação do estimador é feita no Apêndice 8.B.

Em notação matricial, as hipóteses IV e V podem ser sintetizadas como segue:

$$\text{var}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}$$

A matriz definida por $\text{var}(\boldsymbol{\varepsilon})$ é também chamada de matriz de variância e covariância dos erros. Nela, a diagonal principal contém as variâncias dos erros, e os demais elementos da matriz são as covariâncias entre os erros. Assim, o termo $\sigma^2 \mathbf{I}$ cobre as duas hipóteses, já que é o mesmo σ^2 que multiplica os “1” da matriz identidade, e as covariâncias entre os erros (autocovariâncias) valem zero, pois na matriz identidade os elementos fora da diagonal principal são zero.

Exemplo 8.6.1

Com os dados da Tabela 8.14, estime a regressão de Y em função de X_2 e X_3 , e faça os testes da regressão e de cada um dos parâmetros.

Tabela 8.14 – Valores de Y , X^2 e X^3

Y	X_2	X_3	Y	X_2	X_3
800	2	0,8	2.070	11	0,8
1.160	4	0,7	1.890	10	0,7
1.580	6	0,5	1.830	9	0,6
2.010	8	0,4	1.740	8	0,1
1.890	7	0,2	1.380	6	0,5
2.600	12	0,2	1.060	4	0,4

O modelo a ser estimado é:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$$

A matriz \mathbf{X} é dada por:

$$\mathbf{X} = \begin{bmatrix} 1 & 2 & 0,8 \\ 1 & 4 & 0,7 \\ 1 & 6 & 0,5 \\ 1 & 8 & 0,4 \\ 1 & 7 & 0,2 \\ 1 & 12 & 0,2 \\ 1 & 11 & 0,8 \\ 1 & 10 & 0,7 \\ 1 & 9 & 0,6 \\ 1 & 8 & 0,1 \\ 1 & 6 & 0,5 \\ 1 & 4 & 0,4 \end{bmatrix}$$

Em que a coluna preenchida pelo número 1 se refere à variável “ X_1 ”, que na verdade não é uma variável, é o intercepto.

A matriz $\mathbf{X}'\mathbf{X}$ será dada por:

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 12 & 87 & 5,9 \\ 87 & 731 & 41 \\ 5,9 & 41 & 3,53 \end{bmatrix}$$

E sua inversa:

$$(\mathbf{X}'\mathbf{X})^{-1} \cong \begin{bmatrix} 1,25 & -0,09 & -1,04 \\ -0,09 & 0,01 & 0,03 \\ -1,04 & 0,03 & 1,67 \end{bmatrix}$$

A matriz $\mathbf{X}'\mathbf{Y}$ é:

$$\mathbf{X}'\mathbf{Y} = \begin{bmatrix} 20010 \\ 160810 \\ 9309 \end{bmatrix}$$

O estimador \hat{b} é dado, então, por:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \begin{bmatrix} 789,33 \\ 149,56 \\ -419,26 \end{bmatrix}$$

Assim, o valor de cada um dos parâmetros é:

$$\hat{\beta}_1 = 789,33$$

$$\hat{\beta}_2 = 149,56$$

$$\hat{\beta}_3 = -419,26$$

Portanto, o modelo estimado é:

$$\hat{Y} = 789,33 + 149,56X_2 - 419,26X_3$$

Se os valores de X_2 e X_3 forem substituídos na equação anterior, podem-se encontrar os valores de Y explicados pela regressão (\hat{Y}), daí os resíduos mostrados na Tabela 8.15.

Tabela 8.15 – Resíduos (Exemplo 8.6.1)

46,9571	137,6067	-53,8093
65,9128	99,8102	-203,8783
102,9429	-29,0766	-97,0571
191,8987	-101,4430	-159,8641

Considerando a forma matricial, os valores da Tabela 8.15 são os componentes do vetor de resíduos e . A soma dos quadrados dos resíduos é dada por:

$$SQR = e'e = 173.444,02$$

Considerando \mathbf{y} o vetor das variáveis Y centradas, a soma dos quadrados totais é dada por $\mathbf{y}'\mathbf{y}$.

$$SQT = \mathbf{y}'\mathbf{y} = 2.749.025$$

A soma dos quadrados explicados pode ser calculada como:

$$SQE = SQT - SQR = 2.749.025 - 173.444,02 = 2.575.580,98$$

Com isso, pode-se construir uma tabela ANOVA para essa regressão, como foi feito para a regressão simples.

Tabela 8.16 – Tabela ANOVA para o Exemplo 8.6.1

Soma de quadrados	g. l.	Quadrados médios	Teste F
SQE = 2.575.580,98	2	1.287.790,49	66,82
SQR = 173.444,02	9	19.271,56	
SQT = 2.749.025	11	249.911,36	

Os graus de liberdade dos quadrados explicados são agora dois (em vez de um, como na regressão simples), tendo em vista que há duas variáveis explicativas (independentes), X_2 e X_3 . Os graus de liberdade dos quadrados dos resíduos são, dessa forma, nove ($= n - 3$). Para o modelo geral apresentado:

$$Y_i = \beta_0 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + \epsilon_i$$

Com $k - 1$ variáveis explicativas, portanto, os graus de liberdade são, respectivamente, $k - 1$ e $n - k$.

Há autores que chamam o intercepto de β_0 . Nesse caso, o número de variáveis explicativas seria representado por k , e os graus de liberdade seriam k e $n - k - 1$, respectivamente. Há de se tomar cuidado com possíveis confusões: basta lembrar que o número de graus de liberdade dos quadrados explicados é o número de variáveis explicativas. O teste F é feito comparando-se o valor calculado com o valor tabelado para dois graus de liberdade no numerador e nove no denominador. Para 5% de significância, esse valor é 4,26. Como o valor calculado (66,82) é maior, a regressão é válida.

O R^2 é calculado da mesma forma:

$$R^2 = \frac{2.575.580,98}{2.749.025} = 0,9369$$

Para testar a validade de cada um dos parâmetros, é preciso encontrar a variância de cada um deles. A variância do vetor de parâmetros $\hat{\beta}$ é dada por:

$$\text{var}(\hat{\beta}) = \text{var}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}]$$

$$\text{var}(\hat{\beta}) = \text{var}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \epsilon)]$$

$$\text{var}(\hat{\beta}) = \text{var}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon]$$

$$\text{var}(\hat{\beta}) = \text{var}[\beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon]$$

O raciocínio é o mesmo para a variância de um escalar. O termo $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ é uma constante, assim como β . A soma de uma constante não altera a variância, logo:

$$\text{var}(\hat{\beta}) = \text{var}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon]$$

Se fosse um escalar, o termo $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ seria extraído da variância elevando ao quadrado. Como é uma matriz, usa-se a forma quadrática.

$$\text{var}(\hat{\beta}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{var}(\epsilon)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

Além disso, sabe-se que a variância de ϵ é $\sigma^2\mathbf{I}$:

$$\text{var}(\hat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

Como $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}$ é igual à identidade (matriz multiplicada pela sua inversa):

$$\text{var}(\hat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

Cujo estimador é dado por:

$$S_{\hat{\beta}}^2 = S^2(\mathbf{X}'\mathbf{X})^{-1}$$

Que, para este exemplo, é dado por:

$$S_{\hat{\beta}}^2 = 19.271,56(\mathbf{X}'\mathbf{X})^{-1}$$

$$S_{\hat{\beta}}^2 \cong \begin{bmatrix} 24.104,99 & -1.747,65 & -19.990,34 \\ -1.747,65 & 202,34 & 570,85 \\ -19.990,34 & 570,85 & 32.240,76 \end{bmatrix}$$

Os valores da diagonal principal são as variâncias dos parâmetros, enquanto os demais valores representam as *covariâncias*. Por exemplo, a covariância entre os estimadores de β_2 e β_3 é $-19.990,34$.

Desse modo, as variâncias (e os desvios padrão) de cada parâmetro são:

$$S_{\hat{\beta}_1}^2 = 24.104,99 \Rightarrow S_{\hat{\beta}_1} = 155,26$$

$$S_{\hat{\beta}_2}^2 = 202,34 \Rightarrow S_{\hat{\beta}_2} = 14,22$$

$$S_{\hat{\beta}_3}^2 = 32.240,76 \Rightarrow S_{\hat{\beta}_3} = 179,56$$

Assim, pode-se calcular as estatísticas t para cada parâmetro:

$$t_{\hat{\beta}_1} = \frac{789,33}{155,26} = 5,08$$

$$t_{\hat{\beta}_2} = \frac{149,56}{14,22} = 10,51$$

$$t_{\hat{\beta}_3} = \frac{419,26}{179,56} = 2,33$$

Os valores tabelados para a distribuição t , de Student, com 9 graus de liberdade são:

$$t_{(9,10\%)} = 1,83$$

$$t_{(9,5\%)} = 2,26$$

$$t_{(9,1\%)} = 3,25$$

Como os valores calculados para o intercepto β_1 e para β_2 são superiores a todos os valores, aqueles são significantes a 1%. O valor para β_3 é inferior ao valor tabelado para 1%, mas superior ao tabelado a 5%, portanto, diz-se que ele é significante a 5%.

Exemplo 8.6.2

A partir dos dados do Exemplo 8.6.1, faça regressões simples de Y em função de X_2 e depois de X_3 .

Se fizer as regressões simples, você encontrará os seguintes resultados (os valores entre parênteses são os desvios padrão):

$$\hat{Y} = 529,38 + 156,98X_2 \quad R^2 = 0,8987 \\ (130,09) \quad (16,67)$$

$$\hat{Y} = 2.081,09 - 841,19X_3 \quad R^2 = 0,1619 \\ (328,2) \quad (605,12)$$

Como se vê, os coeficientes encontrados são diferentes daqueles que foram calculados na regressão múltipla. Por que isso acontece? Imagine que você queira estudar o volume de vendas de determinado bem: logicamente, se o preço cai, as vendas devem aumentar (o coeficiente da regressão deve ser negativo). Mas e se estiver ocorrendo uma recessão? Mesmo com o preço caindo, as vendas podem cair também. Se for feita uma regressão simples com quantidades e preços, podem ser encontrados resultados estranhos (coeficiente positivo). Isso seria evitado se se incluísse na regressão uma variável como a renda, assim, a influência da renda seria computada nesse modelo.

8.7 Variáveis dummy

Uma **variável dummy** serve para representar a influência de uma característica ou atributo *qualitativo*. Por exemplo, se você quer saber se o sexo influencia o salário, usa-se esta última variável como dependente e inclui-se uma série de variáveis que explicam o salário (anos de estudo, tempo de empresa etc.), além de uma variável D com as seguintes características:

$$D = \begin{cases} 0, & \text{se o trabalhador for homem} \\ 1, & \text{se o trabalhador for mulher} \end{cases}$$

Dessa forma, o coeficiente da variável D representa quanto as mulheres ganham a mais (ou a menos). Assim, se o coeficiente da variável D for -100 , por exemplo, isso significa que as mulheres, em média, ganham 100 reais a menos do que os homens.

Isso também pode ser feito com uma variável qualitativa que possua três estados possíveis. Por exemplo, o padrão de construção de um imóvel pode ser alto, médio ou baixo. Nesse caso, seriam necessárias duas variáveis *dummy*, que poderíamos definir assim:

$$D_1 = \begin{cases} 0, \text{ se for baixo ou alto} \\ 1, \text{ se for médio} \end{cases}$$

$$D_2 = \begin{cases} 0, \text{ se for baixo ou médio} \\ 1, \text{ se for alto} \end{cases}$$

Ou, alternativamente, assim:

$$D_1 = \begin{cases} 0, \text{ se for baixo} \\ 1, \text{ se for médio ou alto} \end{cases}$$

$$D_2 = \begin{cases} 0, \text{ se for baixo ou médio} \\ 1, \text{ se for alto} \end{cases}$$

Exemplo 8.7.1

Do Exemplo 8.6.1, adiciona-se uma variável qualitativa, que representa a existência ou não de determinado atributo.

Tabela 8.17 – Dados do Exemplo 8.6.1 com uma variável qualitativa

Y	X_1	X_2	Atributo
800	2	0,8	sim
1.160	4	0,7	sim
1.580	6	0,5	sim
2.010	8	0,4	sim
1.890	7	0,2	sim
2.600	12	0,2	sim
2.070	11	0,8	não
1.890	10	0,7	não
1.830	9	0,6	não
1.740	8	0,1	não
1.380	6	0,5	não
1.060	4	0,4	não

Estime a regressão de Y em função das três variáveis e faça as análises pertinentes.

Para incluir essa variável qualitativa no modelo, cria-se a variável *dummy D* definida por:

$$D = \begin{cases} 0, \text{ se não existir atributo} \\ 1, \text{ se existir o atributo} \end{cases}$$

As variáveis estão representadas na Tabela 8.18.

Tabela 8.18 – Dados do Exemplo 8.6.1 com uma variável *dummy*

Y	X ₁	X ₂	D
800	2	0,8	1
1.160	4	0,7	1
1.580	6	0,5	1
2.010	8	0,4	1
1.890	7	0,2	1
2.600	12	0,2	1
2.070	11	0,8	0
1.890	10	0,7	0
1.830	9	0,6	0
1.740	8	0,1	0
1.380	6	0,5	0
1.060	4	0,4	0

Deve-se estimar o modelo:

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 D + \epsilon$$

Cujos resultados são:

$$\hat{Y} = 536,09 + 161,87X_2 - 327,78X_3 + 238,08D$$

$$(64,35) \quad (5,34) \quad (65,48) \quad (30,26)$$

Como de costume, os desvios padrão estão entre parênteses. Verifique que todos os coeficientes são significantes a 1%. O resultado encontrado indica que a presença do atributo aumenta o valor de Y em 238,08 (na média).

A ANOVA está representada na Tabela 8.19.

Tabela 8.19 – Tabela ANOVA para o Exemplo 8.7.1

Soma de quadrados	g.l.	Quadrados médios	Teste F
SQE = 2.729.170,78	3	909.723,59	366,56
SQR = 19.854,22	8	2.481,78	
SQT = 2.749.025	11	249.911,36	

A regressão é válida (já que o valor tabelado para a distribuição F a 5% é 4,07) e o R² é 0,9928.

Exemplo 8.7.2

Suponha que, numa regressão para o preço de um imóvel (medido em 1.000 reais), levam-se em conta a área de tal imóvel (X_1), um índice que mede a qualidade dos serviços disponíveis no bairro (X_2) e duas variáveis *dummy* que representam o padrão de construção do imóvel, assim definidas:

$$D_1 = \begin{cases} 0, \text{ se for baixo} \\ 1, \text{ se for médio ou alto} \end{cases}$$

$$D_2 = \begin{cases} 0, \text{ se for baixo ou médio} \\ 1, \text{ se for alto} \end{cases}$$

Os resultados obtidos foram:

$$\hat{Y} = 16,34 + 1,27X_2 + 0,78X_3 + 12,04D_1 + 18,21D_2$$

$$(27,88) \quad (0,44) \quad (0,23) \quad (5,16) \quad (4,77)$$

Qual a diferença (em média) entre o preço de um imóvel de padrão baixo e um de padrão médio? E entre um imóvel de padrão médio e um de padrão alto?

Para um imóvel de baixo padrão, vale $D_1 = D_2 = 0$, enquanto para padrão médio, $D_1 = 1$ e $D_2 = 0$. Portanto, o coeficiente da variável D_1 representa a diferença média no preço de imóveis de padrão baixo e médio, que é, portanto, 12.040 reais.

Se o padrão for alto, então $D_1 = D_2 = 1$. Portanto, a diferença entre imóveis de padrão alto e médio é representada pelo coeficiente da variável D_2 , que é 18.210 reais.

Um cuidado especial deve ser tomado se a variável dependente for qualitativa. Como essa variável deve ter o mesmo tipo de distribuição que o erro, se ela for 0 ou 1, não poderá ser, por exemplo, uma variável normal. Quando esse for o caso, alguns métodos especiais devem ser utilizados para sua estimação, métodos esses que são encontrados em textos mais avançados de econometria.

8.8 Seleção de modelos

8.8.1 R^2 ajustado

Ao observar os Exemplos 8.6.1 e 8.7.1, verifica-se que houve aumento do R^2 quando se acrescenta a variável *dummy*. Isso, entretanto, não significa que o modelo estimado no Exemplo 8.7.1 seja necessariamente *melhor*, já que, se forem acrescentadas variáveis explicativas, o R^2 aumentará ou, muito raramente, ficará na *mesma*, mas jamais cairá. Você pode dizer em que circunstância ele ficará igual?

O R^2 é uma razão entre a soma dos quadrados explicados e a soma dos quadrados totais. Esta última será a mesma, não importando quantas (ou quais) variáveis explicativas são utilizadas. A soma dos quadrados explicados, justamente por ser uma soma de quadrados, quando se acrescenta uma variável explicativa, sempre terá agregada uma parcela positiva a seu total.

Assim, o R^2 , se dá uma medida interessante do ajuste de certo modelo, não serve como comparação entre modelos que têm número de variáveis explicativas diferente. Para fazer essa comparação, há de se usar uma medida diferente.

O R^2 pode ser calculado de duas maneiras:

$$R^2 = \frac{SQE}{SQT} = 1 - \frac{SQR}{SQT}$$

Partindo da segunda forma, se dividir o numerador e o denominador pelos respectivos graus de liberdade, obtém-se um *novo* R^2 , ajustado pelos graus de liberdade, chamado simplesmente de **R^2 ajustado** ou, ainda, \bar{R}^2

$$\bar{R}^2 = 1 - \frac{SQR / (n - k)}{SQT / (n - 1)}$$

Ao fazer esse ajuste pelos graus de liberdade, encontra-se um valor que pode ser usado para comparar modelos com número de variáveis diferente. Ele não tem as mesmas propriedades do R^2 , entretanto: ele será 1, no máximo (que corresponde ao caso em que *não há* resíduos), mas pode ser negativo.

Exemplo 8.8.1.1

Compare os modelos dos Exemplos 8.6.1 e 8.7.1 pelo critério do R^2 ajustado.

Para o modelo do Exemplo 8.6.1:

$$\bar{R}^2 = 1 - \frac{173.444,02 / 9}{2.749.025 / 11} = 0,9229$$

Enquanto, para o modelo do Exemplo 8.7.1:

$$\bar{R}^2 = 1 - \frac{19.854,22 / 8}{2.749.025 / 11} = 0,9901$$

Como o R^2 ajustado é maior para o modelo do Exemplo 8.7.1 (com a variável *dummy*), esse modelo é melhor por esse critério.

8.8.2 Critérios de informação

Há quem considere que o R^2 ajustado não *pune* suficientemente os graus de liberdade. Uma série de autores propõe critérios alternativos, chamados **critérios de informação**, e os mais conhecidos são os de **Schwarz (CIS)** e de **Akaike (CIA)**:

$$CIS = 1 + \ln 2\pi + \ln \frac{SQR}{n} + \frac{k \ln n}{n}$$

$$CIA = 1 + \ln 2\pi + \ln \frac{SQR}{n} + \frac{2k}{n}$$

A parcela $1 + \ln 2\pi$ é constante e pode ser omitida para efeito de comparação. Sua presença decorre do logaritmo da verossimilhança (veja o Apêndice 8.B). O processo de comparação é o mesmo, exceto que, para os critérios de informação, quanto *menor* o valor calculado, *melhor* o modelo.

Exemplo 8.8.2.1

Compare os modelos dos Exemplos 8.6.1 e 8.7.1 pelo critério de informação de Schwarz. Calculando para o modelo do Exemplo 8.6.1:

$$\text{CIS} = 13,04$$

Para o modelo do Exemplo 8.7.1 (com a variável *dummy*):

$$\text{CIS} = 11,08$$

E, de novo, o melhor modelo é o do Exemplo 8.7.1, pois apresentou menor CIS.

Exemplo 8.8.2.2

Compare os modelos dos Exemplos 8.6.1 e 8.7.1 pelo critério de informação de Akaike.

Calculando para o modelo do Exemplo 8.6.1:

$$\text{CIA} = 12,92$$

Para o modelo do Exemplo 8.7.1:

$$\text{CIA} = 10,92$$

De novo, o modelo do Exemplo 8.7.1 apresentou menor CIA e deve ser considerado o melhor entre os dois.

8.8.3 Usando o teste F para selecionar modelos

Outra maneira de escolher entre dois modelos, quando se acrescentam ou se retiram variáveis, é utilizando o teste F. Isso é feito pela comparação da soma dos quadrados dos resíduos entre os dois modelos.

O modelo com maior número de variáveis é chamado de **não restrito**, enquanto o que tem menos, de **restrito** (já que, neste modelo, é como se fosse imposta a restrição de que algumas das variáveis têm coeficiente zero). As somas dos quadrados dos resíduos em cada modelo serão SQNR e SQRR, respectivamente.

A estatística é calculada da seguinte forma:

$$F = \frac{\text{SQRR} - \text{SQNR}}{\frac{m}{n - k}}$$

Sendo m o número de variáveis que a equação não restrita tem a mais. Número esse que, sob a hipótese nula de que não há melhoria no modelo, segue uma distribuição F com m graus de liberdade no numerador e $n - k$ graus de liberdade no denominador.

Exemplo 8.8.3.1

Compare os modelos dos Exemplos 8.6.1 e 8.7.1 pelo teste F.

Nesse caso, o modelo com a variável *dummy* (Exemplo 8.7.1) é o não restrito, e o que não a tem (Exemplo 8.6.1) é o restrito.

$$SQRR = 173.444$$

$$SQRNR = 19.854,22$$

$$m = 1$$

O cálculo da estatística é dado por:

$$F = \frac{\frac{173.444 - 19.854,22}{1}}{19.854,22} = 61,89$$

8

Como o valor tabelado para a distribuição F com 1 grau de liberdade no numerador e 8 no denominador, a 5% de significância, é 5,32, rejeita-se a hipótese nula e, portanto, o modelo que contém a variável *dummy* deve ser considerado o melhor entre os dois. Nesse caso, uma única restrição foi testada, a de que o coeficiente da variável *dummy* é zero. Isso equivale ao teste de significância tradicional, feito pela estatística *t*. Aliás, repare que a estatística F (61,89) é o quadrado da estatística *t* selecionada ($238,08/30,26 = 7,87$). Os testes são equivalentes.

A vantagem do procedimento usando o teste F é que com ele é possível testar simultaneamente várias restrições. Suponha um modelo com duas ou mais variáveis *dummy*. A significância das duas (ou mais) poderia ser testada de uma vez – daí esse teste ser conhecido como teste de significância conjunta – e, nesse caso, o número de restrições, $m = 2$ (ou mais).

Na verdade, o teste F pode ser usado para testar qualquer tipo de restrição linear, como no exemplo a seguir.

Exemplo 8.8.3.2

Testou-se um modelo para a variável Y :

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \varepsilon_i$$

Argumentou-se que o modelo correto não usaria as variáveis X_{2i} e X_{3i} separadamente, o mesmo ocorrendo com as variáveis X_{4i} e X_{5i} . Em ambos os casos, segundo esse argumento, a diferença deveria ser utilizada. O modelo correto, então, deveria ser: $Y_i = \beta_1 + \beta_2 (X_{2i} - X_{3i}) + \beta_4 (X_{4i} - X_{5i}) + \varepsilon_i$

Como testar esse argumento? A segunda regressão, que o incorpora, pode ser vista como uma versão restrita da primeira, como pode ser observado se os termos em parênteses forem expandidos:

$$Y_i = \beta_1 + \beta_2 X_{2i} - \beta_2 X_{3i} + \beta_4 X_{4i} - \beta_4 X_{5i} + \varepsilon_i$$

Portanto, veja que as restrições que foram aplicadas ao primeiro modelo são $\beta_2 = -\beta_3$ e $\beta_4 = -\beta_5$. Ou ainda, se preferir, $\beta_2 + \beta_3 = 0$ e $\beta_4 + \beta_5 = 0$.

O teste a ser conduzido tem as seguintes hipóteses:

$$H_0: \beta_2 = -\beta_3 \text{ e } \beta_4 = -\beta_5$$

$$H_1: \beta_2 \neq -\beta_3 \text{ ou } \beta_4 \neq -\beta_5$$

Estimando-se a primeira regressão (com as variáveis separadas), pode-se obter a soma dos quadrados dos resíduos da regressão não restrita (SQNR). Estimando-se a segunda regressão (com as diferenças), obtém-se a soma dos quadrados dos resíduos da regressão restrita (SQRR). O número de restrições (m), nesse caso, é 2. Assim tem-se a estatística F:

$$\frac{\text{SQRR} - \text{SQNR}}{2}$$

$F = \frac{2}{\text{SQNR}}$ que, se for maior que o valor tabelado da distribuição F com $n - k$

2 graus de liberdade no numerador e $n - k$ no denominador, levará à rejeição da hipótese nula de que o argumento está correto.

Exercícios

1. Dados os valores de X e Y na Tabela 8.20:

Tabela 8.20 – Valores de X e Y

X	Y
2	6,9
3	8,7
-2	-5,8
1	3,4
3	8,2
4	10,8
-1	-1,6
2	6

- a) Estime os parâmetros da reta de regressão.
 - b) Construa a tabela ANOVA.
 - c) Calcule R^2 .
 - d) Faça os testes t e F .
2. Dados os valores de X e Y na Tabela 8.21:

Tabela 8.21 – Valores de X e Y

X	Y
6	104
7	122
8	202
9	193
5	76
4	32

X	Y
7	67
9	103
11	189

- a) Estime os parâmetros, calcule o R^2 e faça os testes t e F .
- b) Refaça os cálculos do item (a) utilizando, em vez dos valores originais, os logaritmos.
- c) Compare os resultados e explique.
3. Após uma regressão simples, em que se utilizou uma amostra com 20 elementos, foram tabulados os dados da Tabela 8.22.

Tabela 8.22 – ANOVA

Soma dos quadrados			
$SQE = 123$			
$SQT = 189$			

- a) Complete a tabela ANOVA.
- b) Calcule o R^2 .
- c) Faça o teste F .
4. Para uma amostra de 10 observações de X e Y foram obtidos:
- $$\Sigma x^2 = 697.440$$
- $$\Sigma y^2 = 1.003.620$$
- $$\Sigma xy = -828.110$$
- $$\bar{X} = 464$$
- $$\bar{Y} = 447,2$$
- a) Estime os parâmetros da reta de regressão.
- b) Construa a tabela ANOVA.
- c) Calcule R^2 .
- d) Faça os testes t e F .
5. Os resultados de uma regressão entre o preço de imóveis e suas áreas foram os seguintes:

$$\text{PREÇO} = 200 + 1,2 \text{ ÁREA}$$

$$(150) \quad (0,3)$$

em que os valores entre parênteses são os desvios padrão.

Teste a significância dos parâmetros, sabendo que foi utilizada uma amostra de 20 observações.

6. Mostre que:

$$\Sigma x^2 = \Sigma X^2 - n \bar{X}^2$$

$$\Sigma y^2 = \Sigma Y^2 - n \bar{Y}^2$$

$$\Sigma xy = \Sigma XY - n \bar{X} \bar{Y}$$

7. Mostre que $\Sigma xy = \Sigma xY = \Sigma Xy$.

8. Dado que as médias amostrais das variáveis X , Y e Z são, respectivamente, 1, 2 e -1, as variâncias 1, 4 e 4 e as covariâncias são $\text{cov}(X,Y) = 1$, $\text{cov}(X,Z) = -0,8$ e $\text{cov}(Z,Y) = -2$, estime as regressões (e calcule o R^2) das regressões de Y por X , Z por X e Z por Y . (Sugestão: confira o Apêndice 8.B.1)

9. Em que condições o estimador de mínimos quadrados ordinários é não viesado? Encontre exemplos em que isso não ocorre.

10. Em que condições o estimador de mínimos quadrados ordinários é eficiente ou, pelo menos, é o MELiNV? Encontre exemplos em que isso não ocorre.

11. Os resultados de uma regressão para o PIB de um país são dados abaixo:

$$\text{PIB} = 1,4 + 0,024t$$

Em que t é o tempo medido em anos e o PIB é anual, medido em logaritmos. Qual o significado dos coeficientes encontrados?

12. Para cada conjunto de observações da Tabela 8.23, estime os parâmetros da regressão com e sem intercepto, fazendo os testes relevantes. Comente os resultados.

Tabela 8.23 – Valores de X e Y

a)

Y	X
1,9	2,0
2,6	3,5
3,3	5,0
4,9	6,0
2,6	4,4

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

Y	X
6,3	7,0
5,4	7,7
7,3	8,3
6,0	6,8
4,9	5,9

b) $\bar{X} = 24,24$

$$\Sigma X^2 = 11.340,95$$

$$\Sigma Y^2 = 16.614,45$$

$$\Sigma XY = 12.226,63$$

$$n = 15$$

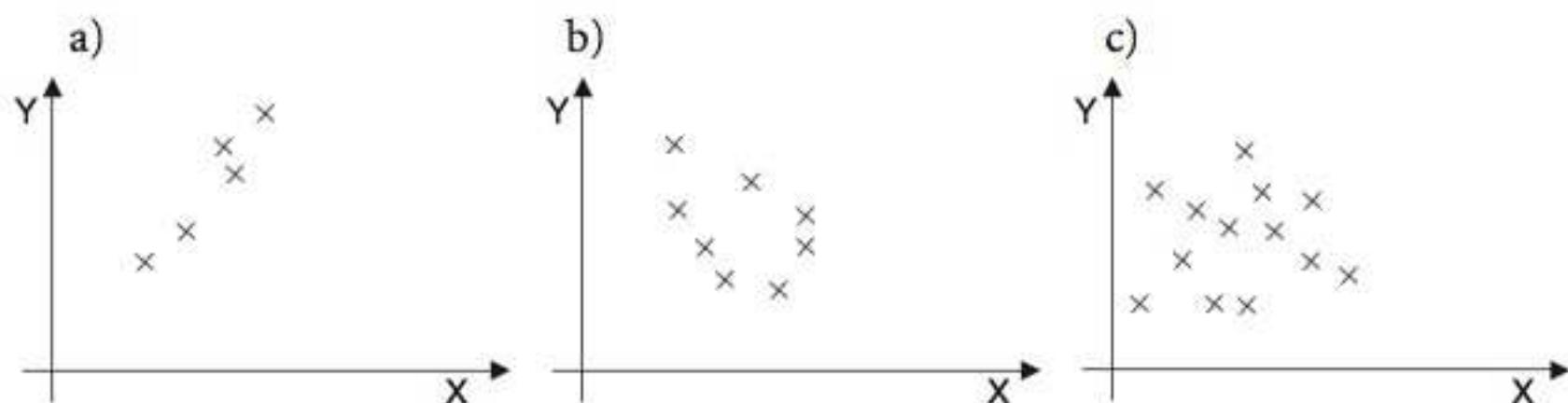
$$\bar{Y} = 27,79$$

13. Após uma regressão com 5 variáveis explicativas, em que se utilizou uma amostra com 30 observações, foram tabulados os dados da Tabela 8.24.

Tabela 8.24 – ANOVA

Soma dos quadrados		
SQE = 2.309,7		
SQT = 3.450,8		

- a) Complete a tabela ANOVA.
 b) Calcule o R^2 e o R^2 ajustado.
 c) Faça o teste F.
14. Numa regressão com 4 variáveis explicativas e uma amostra de 26 observações, a soma dos quadrados explicados foi 1.788,56 e a soma dos quadrados dos resíduos, 567,34. Ao se acrescentarem 2 variáveis ao modelo, a soma dos quadrados explicados aumentou para 1.895,28. Verifique se este modelo é melhor do que o anterior, usando o R^2 ajustado, os critérios de informação e o teste F.
15. Dados os gráficos a seguir, qual o resultado esperado para o sinal de β e o valor de R^2 ?



16. Na Tabela 8.25 são dados, para vários imóveis, a área (em m^2), o padrão de construção (alto, médio ou baixo), o número de dormitórios, de banheiros, de vagas na garagem, se há ou não piscina e o preço do imóvel (em 1.000 reais). Faça uma regressão do preço em função dessas características. Em seguida, teste a significância dos parâmetros e, se for o caso, elimine um ou mais e refaça a estimativa. Use os critérios vistos no texto e compare os dois modelos. Repita o procedimento até encontrar o modelo que melhor explique o preço dos imóveis. Interprete os resultados obtidos.

Tabela 8.25 – Informações sobre imóveis

área	padrão	dorm.	vagas	piscina	banheiros	preço
100	médio	2	1	sim	2	88,9
150	alto	3	1	sim	2	149,1
200	médio	3	2	sim	3	194,4
180	médio	3	1	não	2	153,5

área	padrão	dorm.	vagas	piscina	banheiros	preço
130	médio	2	1	não	1	121,7
89	médio	1	1	não	1	85,9
95	baixo	2	0	não	1	73,5
50	baixo	2	0	não	1	39,9
200	médio	4	3	sim	2	189,7
210	médio	3	2	sim	3	186,3
250	médio	6	3	sim	3	229,7
280	alto	4	2	sim	4	272,0
350	alto	5	2	sim	4	339,5
150	alto	3	1	não	2	155,2
240	alto	3	1	não	2	232,7
70	baixo	2	0	não	2	68,7
135	alto	2	1	sim	2	157,0
140	alto	3	2	sim	2	151,0

17. Teste a normalidade dos resíduos das regressões feitas no Exercício 16.
18. Encontre, em notação matricial, as expressões para a SQE.
19. Foi estimada uma regressão simples para as vendas de uma empresa em função da renda (mais uma constante) dos consumidores com dados trimestrais entre os anos 2000 e 2010 (inclusive). A soma dos quadrados dos resíduos encontrada foi 3.654.279. No intuito de testar a existência de sazonalidade das vendas, foram introduzidas no modelo variáveis *dummy* para os trimestres. A soma dos quadrados dos resíduos encontrada foi para 4.035.786. Quantas variáveis *dummy* foram introduzidas? Como deve ser conduzido o teste? Qual o resultado?
20. Suponha o modelo $Y_i = \beta_1 X_{2i}^{\beta_2} X_{3i}^{\beta_3} \varepsilon_i$. Como estimar os coeficientes por regressão linear? Como testar a hipótese de que $\beta_2 + \beta_3 = 1$?
21. Num modelo estimado por mínimos quadrados com 4 variáveis (mais a constante) e 80 observações, o R^2 encontrado foi 0,65. A seguir, estimou-se uma regressão em que 3 variáveis foram acrescentadas ao modelo e o R^2 aumentou para 0,73. Teste a significância conjunta dessas três novas variáveis (isto é, teste as restrições em que os coeficientes de todas elas é zero).
22. Considere o modelo a seguir:

$$\widehat{\ln(\text{sal})} = 12,3 + 0,06 \text{ escol} + 0,03 \text{ exper} - 0,0005 \text{ exper}^2 - 0,08 \text{ mulher} - 0,002 \text{ mulher} \times \text{escol}$$

$$(6,5) (0,02) (0,01) (0,0003) (0,03) (0,001)$$

$$R^2 = 0,4, n = 1246$$

Em que sal é o salário medido em reais, $escol$ é a escolaridade medida em anos, $exper$ é a experiência na função, também medida em anos, e $mulher$ é uma variável *dummy* que vale 1, se for mulher e 0, se for homem. Desvios padrão entre parênteses.

- Teste a validade da regressão (teste F).
 - Teste a significância dos parâmetros.
 - Qual a razão da inclusão da variável $exper^2$?
 - Qual o efeito de um ano de experiência sobre o salário quando se está no início de carreira? E com 10 anos de experiência?
 - Qual a razão da inclusão da variável $mulher \times escol$?
 - Qual o efeito da escolaridade sobre o salário dos homens? E das mulheres?
 - Interprete o significado do coeficiente da variável $mulher$.
 - Em que difere o efeito da experiência para homens e mulheres? Por quê?
 - Compare o aumento salarial, em média, de uma mulher com 10 anos de experiência que faz um curso de dois anos sem parar de trabalhar com o de um homem com 8 anos de experiência que para de trabalhar para fazer um curso de três anos.
23. Considere o seguinte modelo para a população: $Y = 1,8 + 2,3X_2 - 4,5X_3 + \epsilon$, em que ϵ é o termo aleatório que segue todas as hipóteses usuais. A partir de uma amostra de n observações, o modelo foi estimado omitindo-se a variável X_3 . Ou seja, o modelo estimado foi: $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i}$. Suponha que o seguinte resultado foi obtido na amostra:

$$\frac{\sum_{i=1}^n (X_{3i} - \bar{X}_3)(X_{2i} - \bar{X}_2)}{\sum_{i=1}^n (X_{2i} - \bar{X}_2)^2} = 0,9$$

Determine o viés do estimador $\hat{\beta}_2$ decorrente da omissão de uma variável.

24. Uma regressão utilizando dados de 6 empresas durante 10 anos tentou explicar o resultado (lucro) delas em função de 4 variáveis colhidas nos respectivos balanços, medidas em reais. O R^2 encontrado foi 0,42. A seguir, foi estimada uma nova regressão incluindo variáveis *dummy* referentes às empresas, e o R^2 aumentou para 0,62.
- Em que circunstâncias o estimador de MQO do primeiro modelo é viesado (e inconsistente)?
 - Teste a significância conjunta das variáveis *dummy*.
25. Em 1969, em um artigo publicado numa importante revista de economia, num estudo sobre o efeito da composição etária sobre a poupança em 74 países, os seguintes modelos foram apresentados:

$$\ln S/Y = 7.3439 + 0.1596 \ln Y/N + 0.0254 \ln G - 1.3520 \ln D_1 - 0.3990 \ln D_2$$

$$\ln S/N = 8.7851 + 1.1486 \ln Y/N + 0.0265 \ln G - 1.3438 \ln D_1 - 0.3966 \ln D_2$$

Em que S é a poupança, Y é o PIB, N é a população, G é a taxa de crescimento da renda per capita, D_2 é o percentual da população acima de 64 anos de idade e D_1 o percentual da população abaixo dos 15 anos de idade.

Esses resultados estão corretos? Justifique.

26. Três variáveis Y , X e Z têm média zero e variância igual a 1. O coeficiente angular da regressão de Y por X é dado por 0,2, o coeficiente angular da regressão de Y por Z é dado por 0,4, enquanto o da regressão de X por Z é dado por 0,3. Defina $W = X + Z$ e $V = X - Z$. Determine os coeficientes de determinação (R^2) das regressões de Y por W e de Y por V .

27. Mostre que numa regressão em que a constante é incluída, o ponto que inclui as médias amostrais faz parte da equação de regressão, isto é, $\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}_2 + \hat{\beta}_3 \bar{X}_3 + \dots + \hat{\beta}_k \bar{X}_k$.

28. Assinale verdadeiro ou falso:

- a) Se os resíduos não forem normais, os testes de hipóteses não serão válidos para qualquer tamanho de amostra.
- b) Numa regressão $Y_i = \alpha + \beta X_i + \varepsilon_i$, o significado de β é a elasticidade.
- c) O modelo log-linear pode ser usado para encontrar taxas de crescimento.
- d) Se a reta verdadeira passa pela origem, a estimativa sem o intercepto fornecerá estimadores mais precisos para β .
- e) O teste F para a regressão múltipla tem as seguintes hipóteses nula e alternativa:

$$H_0: \beta_2 = \beta_3 = \dots = \beta_k = 0$$

$$H_1: \text{todos os } \beta_i \neq 0$$

- f) Se a quantidade de variáveis explicativas for aumentada, o R^2 nunca será menor.
- g) Se as variáveis independentes X_i forem estocásticas, o estimador de β será viesado.
- h) Numa regressão $Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$, se $X_{1i} = 2X_{2i} + 3$, é possível obter estimativas para β_1 e β_2 .
- i) A omissão de variáveis relevantes necessariamente torna o estimador de mínimos quadrados ordinários viesado e inconsistente.
- j) A inclusão de variáveis irrelevantes pode tornar o estimador de mínimos quadrados ordinários viesado e inconsistente.
- k) A expressão $SQT = SQE + SQR$ só é válida para a soma dos quadrados quando se utilizam as variáveis centradas na média.
- l) Se, numa regressão, $Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$, a $\sum_{i=1}^n X_{1i} \hat{\varepsilon}_i = \sum_{i=1}^n X_{2i} \hat{\varepsilon}_i = 0$, isso pode indicar que as variáveis independentes não são correlacionadas com o erro.
- m) A soma dos resíduos de uma regressão é sempre zero.

Apêndice 8.A – Matrizes

Uma matriz é uma “tabela” de números, como a matriz **A** mostrada a seguir:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & -1 \\ 0 & 3 & 2 \end{bmatrix}$$

Esta matriz **A** tem 2 linhas e 3 colunas, diz-se que ela tem dimensões 2×3 . Se uma matriz **B** tiver as mesmas dimensões:

$$\mathbf{B} = \begin{bmatrix} 0 & 3 & 1 \\ 4 & -1 & -2 \end{bmatrix}$$

Pode-se definir:

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} 1 & 5 & 0 \\ 4 & 2 & 0 \end{bmatrix}$$

$$\mathbf{A} - \mathbf{B} = \begin{bmatrix} 1 & -1 & -2 \\ -4 & 4 & 4 \end{bmatrix}$$

E ainda é possível definir o produto de uma matriz por uma constante:

$$3 \times \mathbf{A} = \begin{bmatrix} 3 & 6 & -3 \\ 0 & 9 & 6 \end{bmatrix}$$

A transposta da matriz **A**, denominada **A'** ou **A^t**, é uma matriz cujas linhas equivalem às colunas de **A** e vice-versa.

$$\mathbf{A}' = \begin{bmatrix} 1 & 0 \\ 2 & 3 \\ -1 & 2 \end{bmatrix}$$

O produto de duas matrizes também é definido. Ele é feito multiplicando um a um os números de cada linha de uma matriz pelos da coluna da outra. Assim, para uma matriz **C**, de dimensões 3×2 :

$$\mathbf{C} = \begin{bmatrix} 1 & 0 \\ 1 & -1 \\ 2 & 1 \end{bmatrix}$$

o produto **AC** é dado por:

$$\mathbf{AC} = \begin{bmatrix} 1 & 2 & -1 \\ 0 & 3 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & -1 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 \times 1 + 2 \times 1 - 1 \times 2 & 1 \times 0 + 2 \times (-1) - 1 \times 1 \\ 0 \times 1 + 3 \times 1 + 2 \times 2 & 0 \times 0 + 3 \times (-1) + 2 \times 1 \end{bmatrix} = \begin{bmatrix} 1 & -3 \\ 7 & -1 \end{bmatrix}$$

Note que a ordem dos fatores altera o produto quando se trata de matrizes. Veja que só é possível efetuar o produto de matrizes se o número de colunas da primeira for igual ao número de linhas da segunda, e a matriz resultante terá o número de linhas da primeira e o número de colunas da segunda.

A matriz resultante do produto \mathbf{AC} é uma matriz que tem o mesmo número de linhas e colunas. Quando isso ocorre, trata-se de uma matriz quadrada. A matriz $\mathbf{P} = \mathbf{AC}$ é uma matriz quadrada de ordem 2.

Uma matriz quadrada especial é a identidade, cujos valores da diagonal principal são iguais a 1 e os demais valores são 0.

$$\mathbf{I}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \mathbf{I}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

É fácil verificar que a identidade é o elemento neutro na multiplicação de matrizes. Para uma matriz quadrada \mathbf{M} , vale:

$$\mathbf{IM} = \mathbf{MI} = \mathbf{M}$$

Não se define divisão de matrizes, mas, para matrizes quadradas é possível definir a inversa, assim:

$$\mathbf{MM}^{-1} = \mathbf{M}^{-1}\mathbf{M} = \mathbf{I}$$

Por exemplo, para a matriz \mathbf{P} calculada anteriormente, verifique que sua inversa é dada por:

$$\mathbf{P}^{-1} = \frac{1}{20} \begin{bmatrix} -1 & 3 \\ -8 & 1 \end{bmatrix}$$

Com o conceito de matriz inversa é possível, por exemplo, resolver a seguinte equação matricial:

$$\mathbf{AX} = \mathbf{B}$$

Bastando, para isso, multiplicar os dois lados da equação pela inversa de \mathbf{A} :

$$\mathbf{A}^{-1}\mathbf{AX} = \mathbf{A}^{-1}\mathbf{B}$$

$$\mathbf{X} = \mathbf{A}^{-1}\mathbf{B}$$

Vale a seguinte propriedade: a transposta da inversa é igual à inversa da transposta:

$$(\mathbf{M}')^{-1} = (\mathbf{M}^{-1})'$$

O determinante é um número associado à matriz quadrada. Para uma matriz quadrada de ordem 2:

$$\det(\mathbf{P}) = \begin{vmatrix} 1 & -3 \\ 7 & -1 \end{vmatrix} = 1 \times (-1) - (-3) \times 7 = 20$$

Para uma matriz de ordem 3, toma-se cada número da primeira linha, eliminam-se a coluna e a linha correspondentes e calcula-se o determinante da matriz de ordem 2 resultante, somando-se os três resultados:

$$\begin{bmatrix} 1 & -1 & 2 \\ 0 & 1 & 0 \\ 2 & 3 & 1 \end{bmatrix} = 1 \times \begin{vmatrix} 1 & 0 \\ 3 & 1 \end{vmatrix} + (-1) \times \begin{vmatrix} 0 & 0 \\ 1 & 2 \end{vmatrix} + 2 \times \begin{vmatrix} 0 & 1 \\ 2 & 3 \end{vmatrix} = 1 \times 1 - 1 \times 0 + 2 \times (-2) = -3$$

Para matrizes de ordens superiores, o procedimento é invertido. Note que não é possível inverter uma matriz cujo determinante é zero.

Se uma matriz apresentar uma linha (ou coluna) que seja uma combinação linear de outra(s) linha(s) (ou colunas), seu determinante é zero. Assim:

$$Q = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 0 & -1 \\ 4 & 2 & 2 \end{bmatrix} \quad R = \begin{bmatrix} 2 & 1 & 4 & 3 \\ 3 & 1 & 6 & 0 \\ 1 & 0 & 2 & -1 \\ 2 & 5 & 4 & 1 \end{bmatrix}$$

Tanto a matriz **Q** como a matriz **R** apresentam determinante nulo, pois, na matriz **Q**, a terceira linha é a soma das demais e, na matriz **R**, a terceira coluna é o dobro da primeira.

Só matrizes quadradas podem ser multiplicadas por elas mesmas, ou seja, serem elevadas ao quadrado (ou a qualquer potência), em função do problema das dimensões. Portanto, a operação:

$$M^2 = MM$$

só é possível se **M** for uma matriz quadrada. Entretanto, uma matriz **X** qualquer definida por:

$$X = \begin{bmatrix} x & y & z \\ w & v & t \end{bmatrix}$$

apresenta as chamadas formas quadráticas:

$$XX' = \begin{bmatrix} x^2 + y^2 + z^2 & xw + yv + zt \\ zw + yv + zt & w^2 + v^2 + t^2 \end{bmatrix} \quad \text{e} \quad X'X = \begin{bmatrix} x^2 + w^2 & xy + wv & xz + wt \\ xy + wv & y^2 + v^2 & yz + vt \\ xz + wt & yz + vt & z^2 + t^2 \end{bmatrix}$$

Uma particular forma quadrática é quando **X** é uma matriz coluna (vetor), isto é, de dimensões $n \times 1$:

$$X = \begin{bmatrix} x \\ y \end{bmatrix}$$

$$X'X = [x^2 + y^2] = x^2 + y^2$$

Isto é, a forma quadrática é um escalar (número), que é a própria soma dos quadrados. É possível encontrar derivadas matriciais. Dada uma matriz (variável), 1×2 , **X** e um vetor coluna (constante), 2×1 , **b**:

$$Xb = [x \ y] \begin{bmatrix} b \\ c \end{bmatrix} = [xb + yc]$$

A derivada de Xb é dada por:

$$\frac{\partial Xb}{\partial X} = \frac{\partial}{\partial X} [xb + yc] = \begin{bmatrix} \frac{\partial}{\partial x} & \\ \frac{\partial}{\partial y} & \end{bmatrix} [xb + yc] = \begin{bmatrix} b \\ c \end{bmatrix} = b$$

O operador $\frac{\partial}{\partial X}$, embora sozinho não signifique nada, é tratado como uma matriz qualquer, composta de operadores que são as derivadas em relação a x, y, z e w , que são multiplicados pela matriz Xb como se fossem números normais.

A derivada da forma quadrática $X'X$ é dada por:

$$\frac{\partial X'X}{\partial X} = \frac{\partial}{\partial X} [x^2 + y^2] = \begin{bmatrix} \frac{\partial}{\partial x} & \\ \frac{\partial}{\partial y} & \end{bmatrix} [x^2 + y^2] = \begin{bmatrix} 2x \\ 2y \end{bmatrix} = 2X'$$

Como se vê, a derivada de matrizes é análoga à derivada em escalares.

Quanto aos operadores esperança e variância aplicados a vetores coluna:

$$E(X) = E \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} E(x) \\ E(y) \end{bmatrix}$$

A variância de um escalar é dada por $E(x - \mu)^2$. Em notação matricial, usa-se a forma quadrática:

$$\text{var}(X) = E(X - \mu)(X - \mu)' = E \begin{bmatrix} x - \mu_x \\ y - \mu_y \end{bmatrix} [x - \mu_x \ y - \mu_y]$$

$$\text{var}(X) = E \begin{bmatrix} (x - \mu_x)^2 & (x - \mu_x)(y - \mu_y) \\ (x - \mu_x)(y - \mu_y) & (y - \mu_y)^2 \end{bmatrix}$$

Se for aplicado o operador esperança em cada um dos elementos dessa matriz, tem-se:

$$\text{var}(X) = \begin{bmatrix} \text{var}(x) & \text{cov}(x, y) \\ \text{cov}(x, y) & \text{var}(y) \end{bmatrix}$$

Por isso a matriz $\text{var}(X)$ é também chamada de matriz de variância e covariância de X .

Apêndice 8.B – Mais sobre regressão linear

8.B.1 Mais sobre R^2 e $\hat{\beta}$ na regressão simples

$$R^2 = \frac{\text{SQE}}{\text{SQT}} = \frac{\hat{\beta}^2 \sum_{i=1}^n x_i^2}{\sum_{i=1}^n y_i^2} = \left(\frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2} \right)^2 \frac{\sum_{i=1}^n x_i^2}{\sum_{i=1}^n y_i^2} = \frac{\left(\sum_{i=1}^n x_i y_i \right)^2}{\left(\sum_{i=1}^n x_i^2 \right)^2} \frac{\sum_{i=1}^n x_i^2}{\sum_{i=1}^n y_i^2} = \frac{\left(\sum_{i=1}^n x_i y_i \right)^2}{\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i^2}$$

Lembrando que as letras minúsculas representam as variáveis centradas, então:

$$R^2 = \frac{\left[\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) \right]^2}{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}$$

Dividindo (duas vezes) por n o numerador e o denominador:

$$R^2 = \frac{\left[\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) / n \right]^2}{\left[\sum_{i=1}^n (X_i - \bar{X})^2 / n \right] \left[\sum_{i=1}^n (Y_i - \bar{Y})^2 / n \right]} = \frac{[\text{cov}(X, Y)]^2}{\text{var}(X) \text{var}(Y)} = \left[\frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X) \text{var}(Y)}} \right]^2$$

$$= [\text{corr}(X, Y)]^2.$$

Ou seja, demonstra-se aqui a afirmação do capítulo de que o R^2 é, numa regressão simples, o quadrado do coeficiente de correlação entre X e Y .

$$\hat{\beta} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) / n}{\sum_{i=1}^n (X_i - \bar{X})^2 / n} = \frac{\text{cov}(X, Y)}{\text{var}(X)}$$

8.B.2 Demonstração do teorema de Gauss-Markov

A demonstração será feita para o caso da regressão simples, sendo o da regressão múltipla análogo.

Imagine um estimador de β qualquer, linear e não viesado. Para que seja linear, ele deve ser obtido por meio de uma função linear das observações de y_i , o que é feito por meio dos *pesos* w_i :

$$\beta^* = \sum w_i Y_i$$

Para que ele seja não viesado, além da condição usual sobre X_i , é necessário que valham as condições:

$$\sum w_i = 0 \quad \text{e} \quad \sum w_i X_i = \sum w_i \bar{X}_i = 1$$

Se não:

$$E(\beta^*) = E(\sum w_i X_i) = E[\sum w_i (\alpha + \beta X_i + \varepsilon_i)] = E(\alpha \sum w_i + \beta \sum w_i X_i + \sum w_i \varepsilon_i) = \beta + \sum w_i E(\varepsilon_i) = \beta$$

Para o caso específico do estimador de mínimos quadrados, o conjunto de pesos é dado por:

$$m_i = \frac{x_i}{\sum x_i^2}$$

Que segue as propriedades especificadas para w_i (verifique!), além de uma adicional:

$$\sum m_i^2 = \frac{\sum x_i^2}{(\sum x_i^2)^2} = \frac{1}{\sum x_i^2}$$

Estabelecido que $\hat{\beta}^*$ é um estimador não viesado, calcule sua variância:

$$\text{var}(\hat{\beta}^*) = \text{var}(\sum w_i Y_i)$$

Você sabe que a variância de Y_i é a própria variância do termo de erro. Admitindo que ela seja *constante* e que os erros sejam *independentes* (portanto, a variância da soma é a própria soma das variâncias), valem:

$$\text{var}(\hat{\beta}^*) = \sum w_i^2 \text{var}(Y_i)$$

$$\text{var}(\hat{\beta}^*) = \sum w_i^2 \sigma^2$$

$$\text{var}(\hat{\beta}^*) = \sigma^2 \sum w_i^2$$

Usando um pequeno truque:

$$w_i = w_i + m_i - m_i = m_i + (w_i - m_i)$$

Portanto:

$$\sum w_i^2 = \sum m_i^2 + \sum (w_i - m_i)^2 + 2 \sum m_i (w_i - m_i)$$

$$\sum w_i^2 = \sum m_i^2 + \sum (w_i - m_i)^2 + 2 \sum m_i w_i - 2 \sum m_i^2$$

$$\sum w_i^2 = \sum (w_i - m_i)^2 + 2 \sum m_i w_i - \sum m_i^2$$

$$\sum w_i^2 = \sum (w_i - m_i)^2 + 2 \frac{\sum x_i w_i}{\sum x_i^2} - \frac{1}{\sum x_i^2}$$

$$\sum w_i^2 = \sum (w_i - m_i)^2 + 2 \frac{1}{\sum x_i^2} - \frac{1}{\sum x_i^2}$$

$$\sum w_i^2 = \sum (w_i - m_i)^2 + \frac{1}{\sum x_i^2}$$

Substituindo, vem:

$$\text{var}(\hat{\beta}^*) = \sigma^2 \sum w_i^2$$

$$\text{var}(\hat{\beta}^*) = \sigma^2 \sum (w_i - m_i)^2 + \frac{\sigma^2}{\sum x_i^2}$$

O segundo termo é a própria variância do estimador de mínimos quadrados, assim:

$$\text{var}(\hat{\beta}^*) = \text{var}(\hat{\beta}) + \sigma^2 \sum (w_i - m_i)^2$$

Como o outro termo é uma soma de quadrados, ele é necessariamente não negativo. Assim, a variância de um estimador linear e não viesado qualquer $\hat{\beta}^*$ é, no mínimo, igual à variância de $\hat{\beta}$. Portanto, a variância de $\hat{\beta}$ é a menor entre as variâncias de todos os estimadores lineares e não viesados, ou seja, $\hat{\beta}$ é um MELiNV.

8.B.3 Estimação por máxima verossimilhança

Faremos aqui a estimação por máxima verossimilhança de uma regressão simples. As conclusões para a regressão múltipla são análogas. O modelo para a regressão simples com as variáveis centradas é dado a seguir:

$$y_i = \beta x_i + \varepsilon_i$$

O termo de erro é, portanto:

$$\epsilon_i = y_i - \beta x_i$$

Se a distribuição dos erros é normal e eles são independentes, ou seja:

$$\epsilon_i \sim NID(0, \sigma^2)$$

Então, a função de verossimilhança terá a mesma forma funcional de uma normal multivariada (veja o Capítulo 5).

$$L(\beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta x_i)^2\right]$$

Sendo $\exp(x) \equiv e^x$.

Tome, então, o logaritmo de L:

$$l(\beta, \sigma^2) = \ln L(\beta, \sigma^2) = \ln \left\{ \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta x_i)^2\right] \right\}$$

$$l(\beta, \sigma^2) = -\ln(2\pi\sigma^2)^{\frac{n}{2}} - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta x_i)^2$$

$$l(\beta, \sigma^2) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta x_i)^2$$

Para encontrar o ponto de máximo dessa função, deve-se encontrar as derivadas de l em relação a β e σ^2 .

Encontra-se os seguinte resultados:

$$\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2}$$

Portanto, o estimador de máxima verossimilhança de β coincide com o estimador de mínimos quadrados quando a distribuição dos erros é *normal*.

O estimador de máxima verossimilhança de σ^2 é dado por:

$$\hat{\sigma}^2 = \frac{SQR}{n}$$

Divide-se SQR por n , e não por $n - k$ como na estimação por mínimos quadrados. Repetindo o resultado do Capítulo 5, o estimador de máxima verossimilhança de σ^2 é viesado.

Voltando ao logaritmo da função de verossimilhança:

$$l(\beta, \sigma^2) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta x_i)^2$$

Substituindo σ^2 pelo seu estimador e lembrando que $\sum_{i=1}^n (y_i - \beta x_i)^2$ é a soma dos quadrados dos erros (cujo estimador é SQR), o valor do logaritmo da verossimilhança é dado por:

$$l(\beta, \sigma^2) = -\frac{n}{2} \ln \left(2\pi \frac{\text{SQR}}{n} \right) - \frac{n}{2\text{SQR}} \text{SQR}$$

$$l(\beta, \sigma^2) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \frac{\text{SQR}}{n} - \frac{n}{2}$$

$$l(\beta, \sigma^2) = -\frac{n}{2} \left[\ln 2\pi + \ln \frac{\text{SQR}}{n} + 1 \right]$$

Assim, os critérios de informação de Schwarz e Akaike podem ser reescritos da seguinte forma:

$$\text{CIS} = -\frac{2}{n} l(\beta, \sigma^2) + \frac{k \ln n}{n}$$

$$\text{CIA} = -\frac{2}{n} l(\beta, \sigma^2) + \frac{2k}{n}$$

8.B.4 Estimador de mínimos quadrados da regressão múltipla

O modelo de regressão múltipla é dado por:

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon$$

Portanto, o vetor de erros é dado por:

$$\epsilon = \mathbf{Y} - \mathbf{X}\beta$$

A soma dos quadrados dos erros, em notação matricial, é dada pela forma quadrática, que é feita pela multiplicação da matriz pela sua transposta.

$$\epsilon' \epsilon = (\mathbf{Y} - \mathbf{X}\beta)' (\mathbf{Y} - \mathbf{X}\beta)$$

$$\epsilon' \epsilon = \mathbf{Y}' \mathbf{Y} - \mathbf{Y}' \mathbf{X} \beta - \beta' \mathbf{X}' \mathbf{Y} + \beta' \mathbf{X}' \mathbf{X} \beta$$

$$\epsilon' \epsilon = \mathbf{Y}' \mathbf{Y} - 2\beta' \mathbf{X}' \mathbf{Y} + \beta' \mathbf{X}' \mathbf{X} \beta$$

Derivando em relação a β e igualando a zero:

$$-2\mathbf{X}' \mathbf{Y} + 2\mathbf{X}' \mathbf{X} \hat{\beta} = 0$$

$$2\mathbf{X}' \mathbf{X} \hat{\beta} = 2\mathbf{X}' \mathbf{Y}$$

Pré-multiplicando por $(\mathbf{X}' \mathbf{X})^{-1}$

$$(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{X} \hat{\beta} = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Y}$$

Portanto:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

Retomemos a expressão encontrada algumas linhas atrás:

$$\mathbf{X}'\mathbf{X}\hat{\beta} = \mathbf{X}'\mathbf{Y}$$

Ou

$$\mathbf{X}'\mathbf{Y} - \mathbf{X}'\mathbf{X}\hat{\beta} = 0$$

Colocando a matriz \mathbf{X}' em evidência:

$$\mathbf{X}'(\mathbf{Y} - \mathbf{X}\hat{\beta}) = 0$$

Sendo que, entre parênteses, claramente há o vetor dos resíduos da regressão:

$$\mathbf{X}'\mathbf{e} = 0$$

A matriz \mathbf{X}' é dada por:

$$\mathbf{X}' = \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_{21} & X_{22} & \dots & X_{2n} \\ \dots & \dots & \dots & \dots \\ X_{k1} & X_{k2} & \dots & X_{kn} \end{bmatrix}$$

Assim sendo, a matriz $\mathbf{X}'\mathbf{e}$ será dada por:

$$\mathbf{X}'\mathbf{e} = \begin{bmatrix} \sum_{i=1}^n \hat{\epsilon}_i \\ \sum_{i=1}^n X_{2i} \hat{\epsilon}_i \\ \dots \\ \sum_{i=1}^n X_{ki} \hat{\epsilon}_i \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \dots \\ 0 \end{bmatrix}$$

Donde se conclui que, pela primeira linha da matriz $\mathbf{X}'\mathbf{e}$, a soma dos resíduos de uma regressão é zero, desde que nela seja incluída uma constante, isto é, exista uma coluna de “1” na matriz \mathbf{X} (e, portanto, uma linha de “1” na matriz \mathbf{X}').

Das outras linhas conclui-se que a soma do produto de qualquer variável independente pelo resíduo é zero. Isso vale sempre!

8.B.5 Consistência do estimador de mínimos quadrados

Verificaremos a consistência do estimador de mínimos quadrados para a regressão simples, sendo a da regressão múltipla análoga.

Tem-se:

$$\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2}$$

Para que $\hat{\beta}$ seja consistente, é suficiente que:

$$\lim_{n \rightarrow \infty} E(\hat{\beta}) = \beta$$

e

$$\lim_{n \rightarrow \infty} \text{var}(\hat{\beta}) = 0$$

Para o primeiro limite, se são válidas as hipóteses básicas do modelo de regressão linear, será não viésado mesmo para as amostras pequenas, portanto, ele se verificará quando n crescer também.

Resta o segundo limite. Lembrando que:

$$\text{var}(\hat{\beta}) = \frac{\sigma^2}{\sum x^2}$$

Como σ^2 tende a ser menor à medida que a amostra cresce, realmente o $\lim_{n \rightarrow \infty} \text{var}(\hat{\beta}) = 0$ (e, portanto), $\hat{\beta}$ é um estimador consistente de β .



Capítulo 9

VIOLANDO AS HIPÓTESES BÁSICAS

No capítulo anterior, foram estabelecidas algumas hipóteses básicas sobre o modelo de regressão linear, apresentadas de novo a seguir:

- I. $E(\epsilon_i) = 0$ (*erros têm média zero*).
- II. Erros são normalmente distribuídos.
- III. Os x_i são fixos (não estocásticos).
- IV. $\text{var}(\epsilon_i) = \sigma^2$ (constante).
- V. $E(\epsilon_i \epsilon_j) = 0, i \neq j$ (*erros não são autocorrelacionados*).
- VI. Cada variável independente X_i não pode ser combinação linear das demais.

O número de hipóteses pode variar de um autor para outro, bem como a ordem em que são apresentadas. É possível sintetizar as hipóteses I, II, IV e V em uma só ($\epsilon \sim N(0, \sigma^2 I)$). Alguns autores adicionam hipóteses que, embora necessárias, podem ser consideradas óbvias, como a de que o número de observações tem de ser maior do que o número de variáveis.

Em muitas situações, entretanto, essas hipóteses não são verificadas, especialmente quando o objeto de estudo é uma relação social (como as relações econômicas, por exemplo), em que os dados não são produto de um experimento controlado (mas não só nesses casos).

Particularmente, as quatro últimas hipóteses muitas vezes não se verificam em relações desse tipo. Durante o restante do capítulo, vamos nos dedicar às consequências, à identificação e, se for o caso, ao *tratamento* a ser seguido quando cada uma dessas quatro hipóteses é violada.¹

¹ As consequências de que as duas primeiras hipóteses sejam violadas já foram discutidas no capítulo anterior.

9.1 Violando a hipótese VI: a multicolinearidade

A violação da hipótese VI é um caso extremo, que, em termos estatísticos, pode ser descrita como “há correlação exatamente igual a 1 (ou -1) entre duas (ou mais) variáveis explicativas (independentes)”.

Quando ocorre isso, 100% da variação de uma delas é decorrente da variação de outra, isto é, como enunciado na hipótese pode-se escrever a primeira como combinação linear da segunda, como nos exemplos a seguir – note a ausência de qualquer termo aleatório, ao contrário do que acontece no modelo de regressão.

$$X_3 = 2X_2$$

$$X_3 = X_2 + 3$$

$$X_3 = 4X_2 - 5$$

Ou, ainda, envolvendo mais de duas variáveis:

$$X_4 = 2X_2 + 3X_3 + 4$$

Tome um desses exemplos – o raciocínio será idêntico para todos –, o primeiro, em que uma variável é (*exatamente*) o dobro da outra: qualquer variação da segunda implicará variação proporcionalmente idêntica da primeira. É impossível distinguir qual é a influência de uma ou de outra para a variável dependente Y . Por isso mesmo, é *impossível estimar um modelo de regressão linear em que há multicolinearidade*, pelo menos como entendida até aqui.

No início, definiu-se multicolinearidade para a relação entre variáveis explicativas como a descrita anteriormente. Com o passar do tempo, o termo foi estendido, e essa situação passou a ser denominada **multicolinearidade exata ou perfeita**.

Multicolinearidade passou a designar a **alta correlação** (alta, mas não necessariamente 1, em módulo), situação em que é possível estimar o modelo, mas há alguma *dificuldade* associada.

Exemplo 9.1.1

Suponha que você queira obter a função consumo de determinada economia e que o consumo seja função da renda e da taxa real de juros. Se assumir ainda que essa relação é linear, então, a especificação do modelo econométrico a ser estimado pode ser dada por:

$$C_t = \beta_0 + \beta_1 Y_t + \beta_2 r_t + \mu_t$$

Sendo C o consumo, Y a renda nacional disponível e r a taxa real de juros de determinada economia. Os dados estão na Tabela 9.1.

Tabela 9.1 – Função-consumo

Ano/trimestre	Consumo (US\$ bilhões)	Renda (US\$ bilhões)	Taxa de juros real (% a.a.)
1990/1	72,2	105,6	12,00
1990/2	75,6	97,4	12,50
1990/3	89,6	112,0	11,00
1990/4	93,7	128,0	10,00
1991/1	92,2	120,2	10,50
1991/2	84,6	115,3	10,75
1991/3	90,8	105,4	11,25
1991/4	82,9	103,6	12,00
1992/1	65,8	102,7	12,25
1992/2	70,9	93,2	13,00
1992/3	63,1	98,3	12,50
1992/4	86,3	108,1	11,75
1993/1	87,2	115,8	11,50
1993/2	79,3	99,8	11,00
1993/3	87,4	110,5	10,50
1993/4	100,6	127,8	10,25

Os resultados da estimação do modelo são dados na Tabela 9.2.

Tabela 9.2 – Resultados da estimação (variável dependente: consumo)

	Coeficiente	Desvio padrão	Estatística t
Constante	111,487	66,840	1,667
Renda	0,374	0,288	1,298
Taxa de juros real	-6,097	3,314	-1,840

Estatística F = 17,645; R² = 0,7308.

Repare que o valor tabelado da estatística *t*, considerando-se 10% de significância e 13 graus de liberdade, é 1,771, ou seja, apenas o coeficiente da taxa de juros é significante; se considerar 5% (2,160 como valor tabelado), *todos* os coeficientes não são significantes. Esse resultado é, no mínimo, estranho. Imaginar que o nível de consumo não depende da renda disponível² é algo que surpreenderia não só aqueles familiarizados com a teoria econômica, mas qualquer pessoa de bom senso.

O pesquisador precipitado chegaria à rápida e fácil (porém equivocada) conclusão de que a economia de que trata o exemplo é muito peculiar. Se fosse rigoroso com relação à significância dos parâmetros, eliminaria as duas variáveis do modelo, ou formularia um novo modelo, ou assumiria que o consumo nessa economia não pode ser explicado racionalmente; se, entretanto, não fosse tão rigoroso e aceitasse os 10% de significância, ficaria com uma função consumo dependendo apenas da taxa de juros.

² Poder-se-ia argumentar que uma especificação mais adequada da função-consumo utilizaria não a renda presente, mas a renda passada, visto que o consumidor tomaria suas decisões em períodos anteriores; ou, ainda, que se deveria utilizar a renda permanente. Nenhum desses argumentos, no entanto, explicaria a não significância da renda presente, pois essa certamente guarda forte correlação tanto com valores passados como com a renda permanente.

Quem está mais atento, todavia, vai notar um detalhe nos resultados apresentados na Tabela 9.2: a estatística F. Observe que o valor tabelado de F (com 2 graus de liberdade no numerador e 13 no denominador) a 5% de significância é 3,81! Como o valor encontrado foi em torno de 17,6, pelo teste F conclui-se que o *modelo de regressão é válido!* Se a regressão foi validada pelo teste F, a pergunta que fica é: por que os dois parâmetros não são significantes (pelo menos a 5%)? O que deu errado com o teste *t*?

A resposta, nesse caso, pode ser encontrada na própria natureza das variáveis – nem sempre isso é possível, mas frequentemente o é –, se for lembrado que há uma forte influência (e, portanto, correlação) da taxa de juros real sobre a renda.

De fato, se se calcular a correlação amostral entre a taxa de juros real e a renda – e isso *sempre* é possível –, encontra-se o valor de – 0,86 (verifique!).

A correlação entre as variáveis do modelo é, portanto, *muito* alta (em valores absolutos). De fato, essa é a causa do problema (e não a loucura dos consumidores dessa economia) e é o que se chama, usualmente, de multicolinearidade.

Multicolinearidade é a (alta) correlação entre duas (ou mais) variáveis em um modelo de regressão múltipla.

O ideal seria, então, que não houvesse nenhuma correlação entre as variáveis? Cuidado! Ainda que não exista correlação **populacional** entre as variáveis do modelo, é pouco provável (quase impossível, na verdade) que não exista nenhuma correlação amostral. Ademais, se não houvesse nenhuma correlação entre as variáveis, nem sequer seria preciso utilizar a regressão múltipla, pois os resultados das regressões simples, separadamente, seriam os mesmos. Esse é um caso típico de experimentos controlados, em que as demais variáveis são controladas, de modo que é possível verificar a relação da variável dependente com cada uma das variáveis separadamente. Sabe-se que experimentos controlados não são, em geral, possíveis em ciências sociais. Além disso, num modelo econômico, interações entre as variáveis explicativas são um fato da vida. A preocupação deve se limitar a quando essa correlação fica em valores próximos a 1 (ou –1).

9.1.1 Consequências da multicolinearidade

No Exemplo 9.1.1, você viu uma das consequências da multicolinearidade: os testes *t* podem resultar insignificantes, ainda que as variáveis sejam relevantes. Isso ocorre porque a variância dos coeficientes das variáveis explicativas ($\hat{\beta}_1, \hat{\beta}_2$ etc.) aumenta quando ocorre multicolinearidade, por isso os testes *t* apresentam baixa significância (ou mesmo não são significantes). Se não, vejamos:

As variâncias dos coeficientes na regressão múltipla são dadas por:

$$S_{\hat{\beta}}^2 = S^2(\mathbf{X}'\mathbf{X})^{-1}$$

Se o coeficiente de correlação for próximo de 1 (ou -1), o valor do determinante da matriz³ $\mathbf{X}'\mathbf{X}$ será muito pequeno e, portanto, as variâncias de $\hat{\beta}_1$, $\hat{\beta}_2$, etc. serão muito grandes, daí os valores encontrados nos testes t .

Porém, note: isso *não significa* que os testes t sejam inválidos. A variância dos coeficientes estimados de fato é muito grande na presença de multicolinearidade. É até possível ser levado a conclusões erradas do ponto de vista econômico, mas, do ponto de vista estatístico, o valor do coeficiente, se insignificante, não pode ser considerado diferente de zero em função de sua alta variância. A não significância dos parâmetros, neste caso, resulta da **dificuldade que há em distinguir os efeitos de cada variável**, já que, dada a correlação alta, quando uma variável se “mexe”, a outra se “mexe” também.

Como a variância dos $\hat{\beta}$ é muito grande, pode ocorrer ainda que os sinais dos coeficientes ($\hat{\beta}$) sejam o inverso dos esperados; além do mais, seus valores ficam muito sensíveis (mudam demais) quando se acrescenta ou se retira uma variável do modelo ou quando há pequenas mudanças na amostra.

Com relação às propriedades dos estimadores, no entanto, mesmo na presença de multicolinearidade, são mantidas as propriedades usuais do estimador de mínimos quadrados, isto é, continuam não viesados, eficientes e consistentes. Em consequência, as previsões feitas a partir de um modelo com multicolinearidade também têm essas mesmas propriedades.

9.1.2 Como identificar a multicolinearidade?

Retomando o Exemplo 9.1.1, uma maneira de identificar a multicolinearidade, ou, pelo menos, suspeitar fortemente que ela exista, é obter um teste F bastante significante (ou um R^2 alto), acompanhado de estatísticas t para os coeficientes pouco significantes, ou até mesmo não significantes.

Sinais dos coeficientes diferentes do esperado, especialmente quando ele é *muito* esperado (sinal do preço na função demanda e/ou oferta, ou como no Exemplo 9.1.1, o sinal da renda e da taxa de juros⁴ na função consumo), já são, pelo menos, uma evidência de multicolinearidade.

No próprio Exemplo 9.1.1, verificou-se que o cálculo direto da correlação entre as variáveis também é uma forma de identificar a presença de multicolinearidade. No caso do Exemplo 9.1.1, o coeficiente de correlação entre renda e taxa de juros é -0,8602, o que, de fato, é bem alto.

O cálculo da correlação, no entanto, pode não funcionar muito bem quando há mais do que duas variáveis no modelo. Quando se calcula a correlação entre as variáveis, duas a duas, se for encontrada correlação próxima de 1, em valores absolutos, para qualquer par de variáveis, há multicolinearidade. A recíproca, no entanto, não é verdadeira, porque

³ No caso de multicolinearidade exata, o determinante da matriz $\mathbf{X}'\mathbf{X}$ seria zero e, portanto, ela não poderia ser invertida.

⁴ Claro que, como foi visto no próprio exemplo, o fato de os sinais serem de acordo com o esperado não exclui a possibilidade de multicolinearidade.

pode haver não um par de variáveis correlacionadas entre si, mas três (ou mais) variáveis correlacionadas simultaneamente, cujo valor de correlação, tomando-as duas a duas, não indique um valor muito alto.

Nesse caso, uma solução é observar o comportamento dos coeficientes quando se adicionam ou retiram variáveis ou quando há mudanças na amostra. Outra solução, nesse caso, seria fazer *sub-regressões*, combinando as variáveis explicativas do modelo e observando seu valor de R^2 . Se este fosse alto, a multicolinearidade seria identificada. Esse procedimento seria muito trabalhoso, especialmente quando houvesse muitas variáveis, a não ser que, seja pela teoria, por bom senso, ou conhecimento específico do assunto, existisse uma *pista* de quais são os grupos de variáveis correlacionadas entre si. Se ocorrerem mudanças muito drásticas, inclusive nos sinais de tais coeficientes, existe aí uma evidência de que há multicolinearidade no modelo.

Como decorrência de tudo isso, pode-se notar que um modelo que inclua muitas variáveis não é aconselhável, pois é maior a probabilidade de ocorrência de correlações altas entre diversas variáveis, tornando seu resultado bem pouco confiável.

9.1.3 O que fazer quando há multicolinearidade?

A providência óbvia é retirar variáveis correlacionadas do modelo. No caso do Exemplo 9.1.1, que só tem duas variáveis explicativas, uma delas seria retirada. A escolha, em princípio, recairia em qualquer uma. Como o mais *tradicional* é considerar a função consumo tendo como argumento apenas a renda, retira-se a taxa de juros, o que a reduziria à conhecida função keynesiana de consumo.

Exemplo 9.1.3.1

A Tabela 9.3 ilustra o resultado da estimação do modelo $C_t = \beta_0 + \beta_1 Y_t + \mu_t$.

Tabela 9.3 – Resultados da estimação (variável dependente: consumo)

	Coeficiente	Desvio padrão	Estatística t
Constante	-7,859	17,405	0,452
Renda	0,830	0,159	5,221

Estatística F = 27,264; $R^2 = 0,6607$.

Nesse caso, a multicolinearidade foi eliminada, pois sobrou apenas uma variável explicativa. Mesmo que não fosse esse o caso, a alta significância apresentada pelo coeficiente da renda não deixa dúvidas. O valor encontrado para a propensão marginal a consumir, 0,83, é bem mais confiável que o anterior, tendo em vista sua menor variância, ou, em outros termos, um intervalo de confiança construído para esse coeficiente (a um nível de confiança dado) é menor do que um construído para o coeficiente obtido no Exemplo 9.1.1. O critério por trás da retirada de variáveis é que, sendo elas altamente correlacionadas

com a variável (ou variáveis) restante(s), ela já capta o efeito das alterações na variável retirada, tornando-se esta desnecessária para o modelo.

A solução pode, entretanto, não ser satisfatória àquele pesquisador que pretendia obter também a influência direta das taxas de juros sobre o consumo. O pesquisador pode considerar, por exemplo, que, além do efeito sobre a renda, há o efeito da troca de consumo presente por consumo futuro. A retirada de uma variável relevante de um modelo pode trazer mais problemas do que aquele que pretendia resolver, isso porque torna o estimador de mínimos quadrados viesado.

Muitas vezes é possível reduzir os efeitos da multicolinearidade por meio do aumento da amostra. Isso porque a correlação alta observada pode ser decorrente da própria amostra, isto é, essa correlação pode não existir na população e um aumento das observações poderia refletir melhor esse fato; ou, ainda, a correlação pode ser resultado de algum tipo de política econômica transitória e, se a amostra incluir observações de períodos em que essa política não foi adotada, a correlação obtida será bem menor. No caso do Exemplo 9.1.1, isso provavelmente não aconteceria, pois a relação entre renda e taxa de juros não é resultado de nenhuma coincidência amostral, nem de algum tipo de política, mas de algo que se supõe existir sempre. Ainda assim, haveria uma chance de que, em uma amostra maior, essa correlação fosse pelo menos menor do que a obtida no Exemplo 9.1.1.

Em alguns casos, seria possível reespecificar o modelo. Imagine um que relaciona o preço de apartamentos a diversas características, entre elas o número de dormitórios e a área útil. Se esse estudo é realizado em um bairro ou em uma pequena cidade onde o padrão dos imóveis não varia muito, é possível que o tamanho dos aposentos também não varie, fazendo com que a área útil dos apartamentos esteja altamente correlacionada com o número de dormitórios. Nesse caso, talvez seja melhor substituir o preço total dos apartamentos pelo preço por metro quadrado (obtido pela simples divisão do preço total pela área útil). Procedimento semelhante poderia ser adotado no caso de um modelo que explicasse o preço de um produto agrícola em função da área plantada (ou colhida) e da produção, entre outras variáveis. Certamente haverá forte correlação entre a área plantada e a produção. É possível, então, substituí-las por uma única variável, a produtividade (que seria a razão entre a produção e a área plantada).

Existe ainda a alternativa de não fazer *nada*. Lembre que o estimador de mínimos quadrados mantém as propriedades desejáveis de um estimador (não viés, eficiência e consistência), mesmo na presença de multicolinearidade. Se o objetivo for, por exemplo, fazer previsões a respeito da variável explicada, a retirada de variáveis correlacionadas só vai reduzir a eficiência das previsões. Na previsão de valores do consumo na economia dos Exemplos 9.1.1 e 9.1.3.1, os resultados obtidos em 9.1.1 trarão melhores previsões, ainda que os valores dos coeficientes, em função de sua alta variância, reflitam muito pouco sua real relação.

De toda essa discussão pode-se concluir que a multicolinearidade é muito mais uma questão numérica do que um *problema*. De fato, há quem argumente que existe até certo exagero em atribuir um *nome* a uma questão como essa. Em seu livro, Goldberger (1991) chega a literalmente fazer gozação com o termo multicolinearidade, inventando a expressão *micronumerosidade*, que seria o *problema* que decorre de usar uma amostra pequena. Se a amostra é pequena, a variância dos estimadores é grande, portanto, não se alcança uma estimativa precisa e, no caso de *micronumerosidade perfeita*, isto é, quando o número de observações numa amostra é zero (!), não seria possível fazer a estimação.

9.2 Violando a hipótese V: a autocorrelação

Autocorrelação significa a correlação de uma variável com valores defasados (com diferenças, em geral, no tempo) dela mesma. Se a variável x_t (t medido em anos) tem correlação sistematicamente com seu valor no ano anterior (a correlação entre x_t e x_{t-1} não é nula), diz-se que x_t é uma variável autocorrelacionada. Note que as variáveis são distribuídas no tempo. De fato, usualmente, autocorrelação está associada a séries de tempo, mas não necessariamente. O problema é que, no tempo, só há dois *vizinhos* imediatos, a variável no período imediatamente anterior e a no período imediatamente posterior. No caso de variáveis distribuídas no espaço, o número de *vizinhos* pode ser maior, o que complica a análise, embora ela seja possível de ser feita.

A hipótese V faz menção à autocorrelação dos erros. Supõe-se que não existam, o que é bastante razoável, pois o erro não é uma variável específica, mas um conjunto de diversas influências que, por sua natureza, são difíceis de ser medidas, mas não exercem influência uma sobre a outra.

E se exercerem? E por que exerceriam? Imagine, por exemplo, que uma variável relevante esteja sendo omitida. A omissão dessa variável *joga* sua influência, sistemática, para o termo de erro, que supostamente é um conjunto de influências não sistemáticas na variável dependente. A omissão de uma variável relevante pode, portanto, fazer com que haja autocorrelação nos erros.

Outro tipo de erro que poderia levar à autocorrelação é a má especificação funcional. Se, por exemplo, admite-se que uma relação é linear, quando é quadrática, o erro apresenta um padrão sistemático.

A autocorrelação pode ocorrer também pela própria natureza do processo: por exemplo, a produção agrícola. A decisão de produzir não é simultânea à formação do preço, isto é, decide-se quanto se vai produzir no momento do plantio, mas só quando se der a colheita é que o produtor saberá qual o preço que poderá obter por seu produto. Portanto, o preço que influencia a quantidade produzida é o do período anterior, não o atual. Porém, se se produzir demais (ou de menos) em certo período, isso influenciará a decisão de produzir no período seguinte (se o

preço estiver muito baixo, o agricultor produzirá menos). Assim, esse é um processo em que a autocorrelação é parte integrante, mesmo sem haver algum erro de especificação.

Uma maneira de representar um modelo de regressão em que a autocorrelação esteja presente é a seguinte:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \varepsilon_t$$

Em que:

$$\varepsilon_t = \rho \varepsilon_{t-1} + \mu_t$$

Sendo ρ o coeficiente de correlação e μ_t um termo de erro com as características das hipóteses do modelo de regressão (isto é, entre outras coisas, sem autocorrelação).

Se o erro segue um processo como o descrito anteriormente, é dito um processo **autoregressivo de ordem 1**, ou simplesmente **AR(1)**. Nada impede que o processo seja, na verdade, de ordem 2, isto é, algo assim:

$$\varepsilon_t = \rho \varepsilon_{t-2} + \mu_t$$

Ou assim:

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \mu_t$$

Nesse caso, seria um **AR(2)**.

9.2.1 Consequências da autocorrelação

Como você viu no capítulo anterior, a hipótese de não existência de autocorrelação nos erros é um pré-requisito para a demonstração do Teorema de Gauss-Markov, com o qual se mostra que o estimador de mínimos quadrados de uma regressão linear é um MELiNV. Portanto, na presença de autocorrelação, o estimador de mínimos quadrados ordinários⁵ não é mais aquele que tem a menor variância possível entre todos os estimadores.

Isso, sim, já pode ser considerado um problema de fato, algo a ser *tratado*, uma vez que o estimador não é o mais preciso que se poderia obter.

Há de se notar, entretanto, que a hipótese necessária para que o estimador seja não viesado e consistente (que é a de que os regressores, os X , não sejam correlacionados com o erro) não seja violada e, portanto, que não tenha a menor variância, o estimador continua, em geral, não viesado e consistente, mesmo na presença de autocorrelação. Mas há exceções!

As exceções são os modelos que incluem, entre as variáveis independentes (regressores), defasagens da variável dependente, como no caso mostrado a seguir:

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 Y_{t-1} + \varepsilon_t \quad (9.2.1.1)$$

Suponha que o erro ε_t apresente autocorrelação, com um processo do tipo AR(1):

$$\varepsilon_t = \rho \varepsilon_{t-1} + \mu_t$$

⁵ Mínimos quadrados ordinários é como se chamam o método e o estimador usual de mínimos quadrados. É uma tradução no mínimo discutível da expressão em inglês *ordinary least squares*.

Para que o estimador seja não viesado, é preciso que $E(Y_{t-1}\varepsilon_t) = 0$, o que não ocorre, pois:

$$E(Y_{t-1}\varepsilon_t) = E[Y_{t-1}(\rho\varepsilon_{t-1} + \mu_t)] = E(\rho Y_{t-1}\varepsilon_{t-1} + Y_{t-1}\mu_t) = \rho E(Y_{t-1}\varepsilon_{t-1}) + E(Y_{t-1}\mu_t)$$

Embora, por hipótese, Y_{t-1} e μ_t não sejam correlacionados, o mesmo não ocorre com Y_{t-1} e ε_{t-1} , o que fica claro se se tomar uma defasagem da Expressão 9.2.1.1:

$$Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + \beta_3 Y_{t-2} + \varepsilon_{t-1}$$

Portanto, Y_{t-1} e ε_{t-1} são correlacionados e, assim, $E(Y_{t-1}\varepsilon_{t-1}) \neq 0$ e, consequentemente, $E(Y_{t-1}\varepsilon_t) \neq 0$. Como Y_{t-1} é uma variável independente no modelo representado pela Expressão 9.2.1.1, esse é um caso em que a existência de autocorrelação implica o viés do estimador de mínimos quadrados ordinários.

Além disso, é preciso lembrar que os estimadores para a variância dos coeficientes foram calculados supondo que não há autocorrelação entre os erros, isto é, assumindo que (em notação matricial), $\text{var}(\mathbf{\varepsilon}) = \sigma^2 \mathbf{I}$, o que não é verdade. Os estimadores das *variâncias* serão (sempre!) viesados, o que invalida os testes de hipóteses realizados na presença de autocorrelação.

9.2.2 Como identificar a autocorrelação?

A maneira mais comum de identificar a existência de autocorrelação é por meio do **teste de Durbin-Watson**, cuja estatística é calculada por:

$$DW = \frac{\sum_{t=2}^n (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=1}^n \hat{\varepsilon}_t^2}$$

Para entender seu significado, vamos desenvolver essa expressão:

$$DW = \frac{\sum_{t=2}^n (\hat{\varepsilon}_t - 2\hat{\varepsilon}_t \hat{\varepsilon}_{t-1} + \hat{\varepsilon}_{t-1}^2)^2}{\sum_{t=1}^n \hat{\varepsilon}_t^2}$$

$$DW = \frac{\sum_{t=2}^n \hat{\varepsilon}_t^2 - 2 \sum_{t=2}^n \hat{\varepsilon}_t \hat{\varepsilon}_{t-1} + \sum_{t=2}^n \hat{\varepsilon}_{t-1}^2}{\sum_{t=1}^n \hat{\varepsilon}_t^2}$$

Se a amostra for suficientemente grande, a diferença entre a soma de $\hat{\epsilon}_t^2$ e $\hat{\epsilon}_{t-1}^2$ é muito pequena, assim como é muito pequena a diferença entre somar de 1 a n ou de 2 a n . Então, pode-se dizer que essas somas são (quase) iguais:

$$DW \cong \frac{2 \sum_{t=1}^n \hat{\epsilon}_t^2 - 2 \sum_{t=2}^n \hat{\epsilon}_t \hat{\epsilon}_{t-1}}{\sum_{t=1}^n \hat{\epsilon}_t^2}$$

$$DW \cong 2 \left(\frac{\sum_{t=1}^n \hat{\epsilon}_t^2}{\sum_{t=1}^n \hat{\epsilon}_t^2} - \frac{\sum_{t=2}^n \hat{\epsilon}_t \hat{\epsilon}_{t-1}}{\sum_{t=1}^n \hat{\epsilon}_t^2} \right)$$

O primeiro termo é obviamente igual a 1. O segundo é um estimador para o coeficiente de correlação dos erros.

$$DW \cong 2(1 - \hat{\rho})$$

Se não há autocorrelação ($\rho = 0$), o valor de $\hat{\rho}$ deve ser em torno de 0 e, portanto, o valor de DW deve ser próximo de 2. Um valor próximo de 2 para DW implica, dessa forma, a não existência de autocorrelação.

Havendo autocorrelação, essa pode ser positiva ou negativa. Os casos extremos seriam $\rho = 1$ ou $\rho = -1$. Se o valor de $\hat{\rho}$ for próximo de 1, o valor de DW fica próximo de 0. Portanto, valores de DW (razoavelmente) abaixo de 2 indicam autocorrelação positiva. Da mesma forma, se $\hat{\rho}$ for próximo de -1 , DW é próximo de 4, isto é, valores (razoavelmente) acima de 2 indicam autocorrelação negativa.

Mas quão distante de 2 deve estar o valor da estatística DW para que se possa concluir que existe, de fato, autocorrelação? Isso foi resolvido por meio de simulações que resultaram numa tabela semelhante àquelas que foram utilizadas até agora, com a diferença de que ela não vem de fórmula analítica, como era o caso das distribuições derivadas da distribuição normal.

Observando essa tabela ao final do livro, verifica-se que o teste de Durbin-Watson apresenta uma limitação (não é a única!). Existe um intervalo de valores em que o teste é inconclusivo. Se, por exemplo, for testado um modelo com duas variáveis explicativas, com 20 observações, para um nível de significância de 5%, encontram-se os valores $d_l = 1,10$ e $d_u = 1,54$. Se o valor de DW for abaixo de 1,10, rejeita-se a hipótese nula de não autocorrelação, isto é, conclui-se que existe autocorrelação. Se DW estiver entre 1,54 e 2, conclui-se que *não* há autocorrelação (aceita-se a hipótese nula). Se, entretanto, o valor de DW cair entre 1,10 e 1,54, o teste é inconclusivo, não dá para dizer se há ou não autocorrelação.

Note que a tabela é montada para autocorrelações positivas ($DW < 2$). Para DW maior do que 2, o que indicaria autocorrelação negativa, basta fazer $DW^* = 4 - DW$, e o valor de DW^* poderá ser comparado normalmente com os valores da tabela.

Exemplo 9.2.2.1

Na Tabela 9.4, são apresentados os dados de consumo e renda trimestrais de um país durante sete anos e meio. Estime a função consumo (consumo como função da renda) e teste a existência de autocorrelação, com 5% de significância.

Tabela 9.4 – Consumo e renda de 1994 a 2001

Ano/trimestre	Consumo (US\$ bilhões)	Renda (US\$ bilhões)
1994/3	757,6	970,0
1994/4	745,2	988,5
1995/1	673,4	866,5
1995/2	652,2	812,4
1995/3	676,2	845,3
1995/4	709,1	891,9
1996/1	704,7	899,3
1996/2	691,8	911,2
1996/3	696,6	903,2
1996/4	667,6	904,5
1997/1	667,2	906,7
1997/2	671,0	920,2
1997/3	716,9	958,4
1997/4	698,4	934,1
1998/1	676,7	944,4
1998/2	661,4	956,3
1998/3	686,8	971,7
1998/4	685,2	958,9
1999/1	684,9	961,9
1999/2	675,1	966,4
1999/3	663,1	977,5
1999/4	672,8	988,5
2000/1	675,2	1.001,2
2000/2	693,1	996,7
2000/3	721,6	1.005,6
2000/4	747,5	1.011,2
2001/1	742,4	1.004,2
2001/2	740,5	997,4
2001/3	741,5	1.000,4
2001/4	722,6	1.006,6

Os resultados da estimação estão na Tabela 9.5.

Tabela 9.5 – Resultados da estimação (variável dependente: consumo)

	Coeficiente	Desvio padrão	Estatística t
Constante	402,672	87,676	4,59
Renda	0,311	0,092	3,37

Estatística F = 11,32.

Os resultados foram os esperados: o coeficiente da renda foi significante (a 1%) e a regressão foi válida (*aprovada* pelo teste F, a 1%). Antes de cometer a precipitação de afirmar que já sabe como a renda influencia o consumo, convém, especialmente porque se trata de dados em séries de tempo, testar a existência de autocorrelação.

Os resíduos foram obtidos dos resultados anteriores e estão na primeira coluna da Tabela 9.6. Nas colunas seguintes, são feitos os cálculos necessários para a obtenção da estatística DW.

Tabela 9.6 – Estatística de Durbin-Watson

Ano/trimestre	Resíduos ($\hat{\epsilon}_t$)	$\hat{\epsilon}_t - \hat{\epsilon}_{t-1}$	$(\hat{\epsilon}_t - \hat{\epsilon}_{t-1})^2$	$(\hat{\epsilon}_t)^2$
1994/3	53,70998			2.884,7624
1994/4	35,5651	-18,1449	329,2369	1.264,8761
1995/1	1,650302	-33,9148	1.150,2133	2,7235
1995/2	-2,749784	-4,4001	19,3608	7,5613
1995/3	11,03363	13,7834	189,9826	121,7410
1995/4	29,46273	18,4291	339,6315	868,0522
1996/1	22,76477	-6,6980	44,8626	518,2348
1996/2	6,169411	-16,5954	275,4060	38,0616
1996/3	13,45369	7,2843	53,0607	181,0017
1996/4	-15,95001	-29,4037	864,5773	254,4028
1997/1	-17,03318	-1,0832	1,1733	290,1294
1997/2	-17,4254	-0,3922	0,1538	303,6445
1997/3	16,61218	34,0376	1.158,5571	275,9647
1997/4	5,658172	-10,9540	119,9904	32,0149
1998/1	-19,24033	-24,8985	619,9356	370,1904
1998/2	-38,23569	-18,9954	360,8237	1.461,9683
1998/3	-17,61792	20,6178	425,0924	310,3913
1998/4	-15,24308	2,3748	5,6399	232,3516
1999/1	-16,47469	-1,2316	1,5168	271,4153
1999/2	-27,67209	-11,1974	125,3819	765,7447
1999/3	-43,11902	-15,4469	238,6077	1.859,2502
1999/4	-36,8349	6,2841	39,4902	1.356,8101
2000/1	-38,37869	-1,5438	2,3833	1.472,9239
2000/2	-19,08129	19,2974	372,3898	364,0955
2000/3	6,654957	25,7362	662,3542	44,2885
2000/4	30,81596	24,1610	583,7543	949,6237
2001/1	27,88971	-2,9263	8,5630	777,8357
2001/2	28,10134	0,2116	0,0448	789,6853
2001/3	28,16974	0,0684	0,0047	793,5341
2001/4	7,344423	-20,8253	433,6937	53,9405
Soma	0		8.425,8821	18.917,2199

Portanto, a estatística DW é dada por:

$$DW = \frac{\sum_{t=2}^n (\hat{\epsilon}_t - \hat{\epsilon}_{t-1})^2}{\sum_{t=1}^n \hat{\epsilon}_t^2} = \frac{8.425,8821}{18.917,2199} = 0,4454$$

Como o limite inferior da tabela de DW é, para 5% de significância, 30 observações e uma variável explicativa, $d_l = 1,35$, ou, para 1% de significância, 1,20 (em ambos os casos, maior do que 0,4454), conclui-se que *existe* autocorrelação (rejeita-se a hipótese nula de não autocorrelação).

Como foi dito, o teste de Durbin-Watson apresenta algumas limitações.⁶ Além da existência de um intervalo em que o teste é inconclusivo, o teste *não* é válido se:

- a regressão *não* incluir o intercepto (termo constante);
- a regressão incluir, como variáveis explicativas, defasagens da variável dependente.

Além disso, como é claro pela própria formulação do teste, ele é feito para testar apenas correlações de primeira ordem.

Uma alternativa ao teste Durbin-Watson sem essas limitações é o **teste de Breusch-Godfrey**. Esse teste consiste em regredir os resíduos da regressão com seus valores passados, da seguinte forma:

$$\hat{\epsilon}_t = \rho \hat{\epsilon}_{t-1} + \mu_t$$

A ideia é que, se essa regressão for significante, isso significa que os erros passados são relevantes para determinar o erro presente, o que quer dizer autocorrelação. Os resultados do Exemplo 9.1.3.1 estão na Tabela 9.7.

Tabela 9.7 – Resultados do teste Breusch-Godfrey (variável dependente: $\hat{\epsilon}_t$ – resíduos)

	Coeficiente	Desvio padrão	Estatística t
$\hat{\epsilon}_{t-1}$	0,7016	0,1130	6,21

Estatística F = 38,54; $R^2 = 0,5792$.

Verifica-se que a regressão é significante, o que, de novo, indica autocorrelação. A estatística mais usada para esse teste, no entanto, é $(n - 1)R^2 = 29 \times 0,5792 \cong 16,80$, que, sob a hipótese nula (não existência de autocorrelação), segue uma distribuição qui-quadrado com 1 grau de liberdade. Pela tabela, encontram-se os valores 3,84, para 5% de significância, e 6,63, para 1% de significância. Rejeita-se a hipótese nula, portanto. O $n - 1$, ou 29, no caso, decorre do fato de testar autocorrelação de 1ª ordem e, ao conduzir o teste, perde-se um elemento da amostra em função da defasagem utilizada.

⁶ Em textos mais avançados de econometria, é possível encontrar outros testes para autocorrelação.

É possível realizar o teste de Breusch-Godfrey para qualquer defasagem desejada, por exemplo, até a p -ésima defasagem. Nesse caso, a estatística do teste é $(n - p)R^2$ e deve ser comparada com o valor crítico da distribuição qui-quadrado com p graus de liberdade.

9.2.3 O que fazer quando há autocorrelação?

Primeiro, há a questão de qual é a causa da autocorrelação. Se o problema é de especificação, ele pode ser corrigido com a inclusão de mais variáveis ou com a alteração da forma funcional.

Se não é esse o caso, ou seja, a autocorrelação é *parte integrante* do modelo estimado, a correção passa pelo conhecimento prévio de como é a estrutura da autocorrelação. Suponha que seja um modelo com uma variável explicativa como mostrado a seguir.

$$Y_t = \beta_1 + \beta_2 X_t + \varepsilon_t \quad (9.2.3.1)$$

Em que existe autocorrelação e ela é de primeira ordem (é um AR(1)), ou seja:

$$\varepsilon_t = \rho \varepsilon_{t-1} + \mu_t$$

Suponha ainda que o coeficiente ρ seja conhecido. Se a Expressão 9.2.3.1 defasada for multiplicada por ρ :

$$\rho Y_{t-1} = \rho \beta_1 + \rho \beta_2 X_{t-1} + \rho \varepsilon_{t-1} \quad (9.2.3.2)$$

Subtraindo a Expressão 9.2.3.2 da Expressão 9.2.3.1:

$$Y_t - \rho Y_{t-1} = \beta_1 - \rho \beta_1 + \beta_2 (X_t - \rho X_{t-1}) + (\varepsilon_t - \rho \varepsilon_{t-1})$$

Mas você sabe que:

$$\varepsilon_t - \rho \varepsilon_{t-1} = \mu_t$$

Se for feito o seguinte:

$$Y_t^* = Y_t - \rho Y_{t-1}$$

$$\beta_1^* = \beta_1 - \rho \beta_1$$

$$X_t^* = X_t - \rho X_{t-1}$$

Chega-se ao seguinte modelo:

$$Y_t^* = \beta_1^* + \beta_2 X_t^* + \mu_t$$

Que é um modelo *sem* autocorrelação (que pode, portanto, ser estimado sem problemas por mínimos quadrados ordinários) e, o que é importante, apresenta o *mesmo* coeficiente β_2 do modelo original.

Mesmo assim, ainda resta o problema de como conhecer o coeficiente ρ . Uma estimativa pode ser encontrada, entretanto, por meio do próprio valor de DW, já que:

$$DW \cong 2(1 - \hat{\rho})$$

Então:

$$\hat{\rho} \cong 1 - \frac{DW}{2}$$

Exemplo 9.2.3.1

Refaça a estimação do Exemplo 9.2.2.1, corrigindo o problema da autocorrelação. O primeiro passo é encontrar uma estimativa para o coeficiente de correlação, o que, como você viu, pode ser feito pela própria estatística DW:

$$\hat{\rho} \cong 1 - \frac{DW}{2} = 1 - \frac{0,4454}{2} \cong 0,7773$$

Alternativamente, seria possível utilizar o coeficiente encontrado no teste de Breusch-Godfrey, 0,7016.

Se o consumo é a variável Y_t e renda é a variável X_t , as variáveis *corrigidas*, isto é, aquelas cuja regressão não apresentará autocorrelação (pelo menos assim se espera), são dadas por:

$$Y_t^* = Y_t - 0,7773Y_{t-1}$$

$$X_t^* = X_t - 0,7773X_{t-1}$$

e estão mostradas na Tabela 9.8.

Tabela 9.8 – Variáveis corrigidas

Ano/trimestre	Consumo (Y_t)	Y_t^*	Renda (X_t)	X_t^*
1994/3	757,6		970,0	
1994/4	745,2	156,3175	988,5	234,519
1995/1	673,4	94,15604	866,5	98,13895
1995/2	652,2	128,7662	812,4	138,8696
1995/3	676,2	169,2449	845,3	213,8215
1995/4	709,1	183,4897	891,9	234,8483
1996/1	704,7	153,5166	899,3	206,0261
1996/2	691,8	144,0367	911,2	212,1741
1996/3	696,6	158,8639	903,2	194,9242
1996/4	667,6	126,1328	904,5	202,4426
1997/1	667,2	148,2745	906,7	203,6322
1997/2	671	152,3854	920,2	215,4221
1997/3	716,9	195,3317	958,4	243,1285
1997/4	698,4	141,1536	934,1	189,1357
1998/1	676,7	133,8337	944,4	218,3241

Ano/trimestre	Consumo (Y)	Y*	Renda (X)	X*
1998/2	661,4	135,4011	956,3	222,2179
1998/3	686,8	172,6938	971,7	228,368
1998/4	685,2	151,3504	958,9	203,5976
1999/1	684,9	152,294	961,9	216,547
1999/2	675,1	142,7272	966,4	218,7151
1999/3	663,1	138,3448	977,5	226,3173
1999/4	672,8	157,3724	988,5	228,6893
2000/1	675,2	152,2326	1.001,2	232,839
2000/2	693,1	168,267	996,7	218,4672
2000/3	721,6	182,8534	1.005,6	230,8651
2000/4	747,5	186,6003	1.011,2	229,5471
2001/1	742,4	161,3683	1.004,2	218,1942
2001/2	740,5	163,4325	997,4	216,8353
2001/3	741,5	165,9094	1.000,4	225,121
2001/4	722,6	146,2321	1.006,6	228,9891

Os resultados obtidos agora são os mostrados na Tabela 9.9.

Tabela 9.9 – Resultados da variável dependente Y*

	Coeficiente	Desvio padrão	Estatística t
Constante	45,6158	20,2617	2,25
X*	0,5105	0,0945	5,39

Estatística F = 29,07.
DW = 1,4393.

O coeficiente da renda foi maior, e com um desvio padrão menor (repare que essa última comparação nem sequer era necessária, já que se sabe que o estimador do desvio padrão do Exemplo 9.2.2.1 era viesado em função da autocorrelação).

Para ficarmos satisfeitos com esse novo resultado, no entanto, é preciso prestar atenção na estatística de Durbin-Watson. Se o valor encontrado (1,4393) for comparado com a tabela para 29 observações (sim, há uma observação a menos agora), nota-se que, para 5% de significância, $d_l = 1,34$ e $d_s = 1,48$, portanto, o teste é inconclusivo, o que não é uma notícia maravilhosa, mas pelo menos não se pode afirmar que há autocorrelação. A 1% de significância, entretanto, os valores tabelados são $d_l = 1,12$ e $d_s = 1,25$, então, aceita-se a hipótese de não existência de autocorrelação com essa significância.

9.3 Violando a hipótese IV: a heterocedasticidade

A hipótese IV estabelece que a variância dos erros deve ser constante (o que é conhecido como **homocedasticidade**).

Imagine uma regressão em que a variável dependente seja a altura das pessoas. Medindo a altura com uma régua comum, podem-se cometer erros em função da medição dessa altura, em função da precisão da régua ou mesmo da precisão de como a medida é feita. Não há por quê, entretanto, acreditar que a variância desse erro de medição será diferente para diferentes grupos de pessoas (altas ou baixas, por exemplo). A hipótese IV, nesse caso, é bem razoável.

Agora imagine uma pesquisa de salários em função dos anos de estudo. A relação certamente existe, pois pessoas com vários anos de escolaridade ganham, em média, mais do que pessoas pouco escolarizadas, mas a situação muda muito no que se refere ao erro: para aqueles com pouco ou nenhum estudo, os salários não deverão variar muito (pelo menos para a grande maioria), fazendo com que a variância seja muito pequena. No caso de pessoas com muitos anos de estudo (nível superior, pós-graduação etc.), embora se espere que ganhem mais, as possibilidades são bem mais amplas: é possível que uma pessoa desse grupo tenha problemas em avançar na carreira ou se torne presidente de uma grande empresa, o que torna a variância dos salários, nesse caso, muito alta.

Há outros exemplos, como a poupança das famílias em função da renda: famílias com renda muito baixa pouparão valores muito próximos entre si (um valor muito pequeno, por sinal, até porque não há muito para poupar), enquanto, entre famílias mais ricas, existe ampla gama, que vai desde famílias bastante perdulárias a outras muito poupadoras.

9.3.1 Consequências da heterocedasticidade

A hipótese IV (assim como a hipótese V) é necessária para a demonstração do Teorema de Gauss-Markov. Dessa forma, as consequências são basicamente as mesmas da presença da autocorrelação:⁷ os estimadores de mínimos quadrados ordinários continuam não viesados, mas já não são os de menor variância. As variâncias dos estimadores são viesadas, invalidando, assim, os testes de hipóteses.

9.3.2 Como identificar a heterocedasticidade?

Existem vários testes na literatura que têm como objetivo identificar a presença de heterocedasticidade, mas vamos apresentar dois.

O teste de **Goldfeld e Quandt** consiste em separar a regressão em duas: uma com valores menores de X e outra com valores maiores, e aí fazer um teste para comparar a variância em cada regressão (um teste comum de comparação de variâncias, isto é, um teste F). Havendo diferença

⁷ Exceto quando há autocorrelação nos casos em que se usam defasagens da variável dependente como variáveis explicativas, o que torna o estimador de mínimos quadrados ordinários viesado, o que não ocorre na presença de heterocedasticidade.

nas variâncias das duas regressões, a hipótese nula de homocedasticidade é rejeitada e, sendo esse o caso, conclui-se que há presença de heterocedasticidade, a qual deverá ser corrigida.

Exemplo 9.3.2.1

São dados, na Tabela 9.10, os salários de 20 trabalhadores e os anos de estudo de cada um. Faça uma regressão dos salários em função dos anos de estudo e verifique a existência de heterocedasticidade utilizando o teste de Goldfeld e Quandt.

Tabela 9.10 – Anos de estudo x salários

Anos de estudo	Salários (R\$)
1	410,00
2	508,90
3	857,70
2	551,30
3	789,20
4	935,50
7	1.529,30
8	1.497,50
9	2.317,70
11	2.169,50
11	2.596,80
13	2.844,60
13	3.391,00
14	2.671,20
16	2.653,80
16	2.939,10
17	3.437,00
18	4.583,30
19	3.559,30
19	4.896,70

Os resultados da regressão tendo o salário como variável dependente são dados na Tabela 9.11.

Tabela 9.11 – Resultados da regressão (variável dependente: salários)

	Coeficiente	Desvio padrão	Estatística t
Constante	139,074	184,155	0,755
Anos de estudo	205,621	15,400	13,35

F = 178,28.

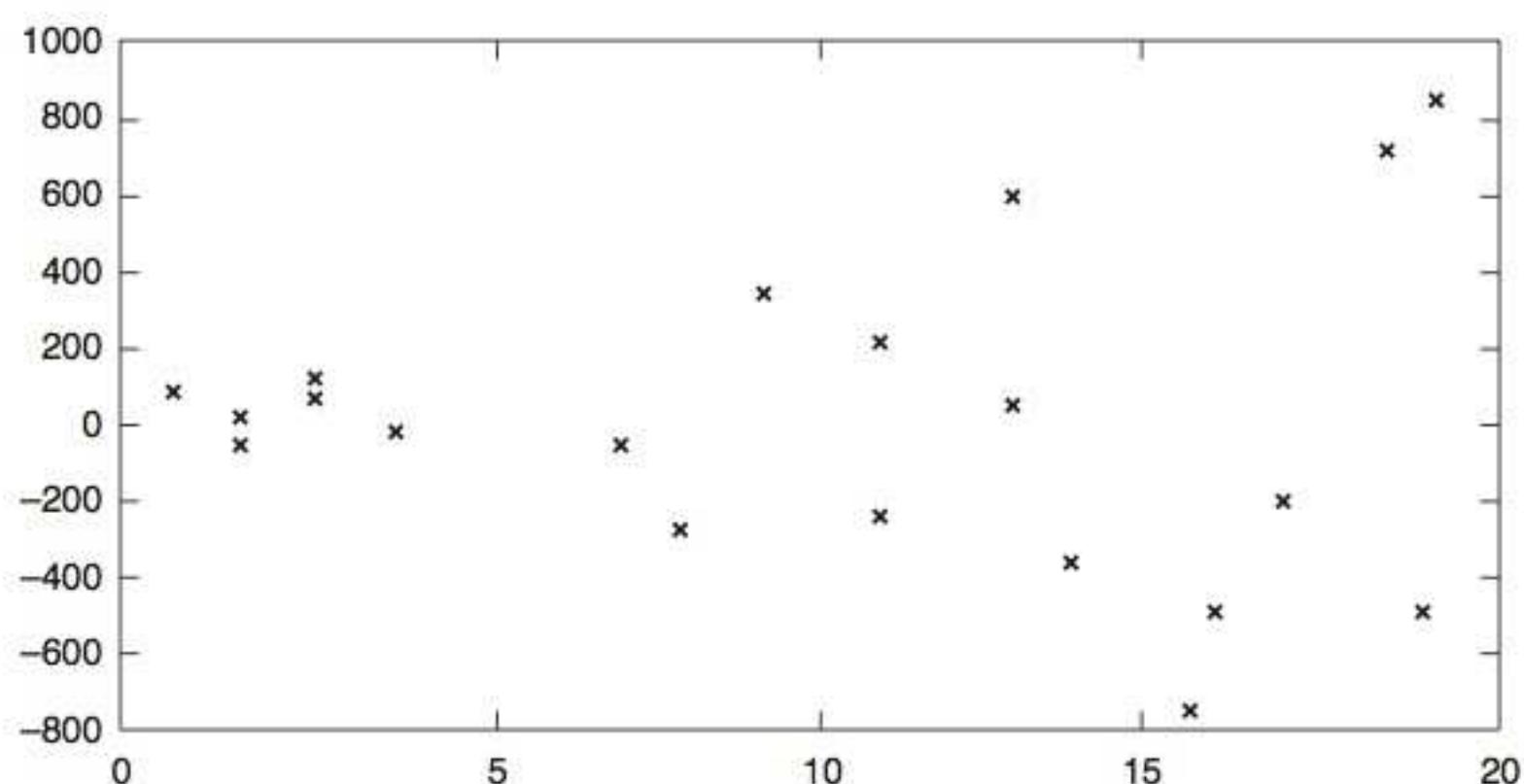
Os resíduos dessa regressão estão na Tabela 9.12.

Tabela 9.12 – Resíduos da regressão

65,30477	-26,05806	195,8953	-489,9094
-41,41617	-49,12089	32,45345	-197,6303
101,7629	-286,5418	578,8535	743,0487
0,983826	328,0372	-346,5675	-486,5722
33,26288	-231,4047	-775,2094	850,8278

A Figura 9.1 mostra o comportamento dos resíduos.

Figura 9.1 – Resíduos



O gráfico da Figura 9.1 dá um indício de que os resíduos realmente são mais *espalhados* quando os salários são maiores.

Para testar a heterocedasticidade, dividem-se os dados em dois grupos, seguindo os procedimentos do teste de Goldfeld e Quandt. Essa divisão é arbitrária, mas o teste tende a ser mais eficiente se se omitirem os dados do *meio*, isto é, toma-se um grupo com os valores de *X* menores (um a quatro anos de estudo) e outro com os valores de *X* maiores (14 ou mais anos de estudo), como os apresentados na Tabela 9.13.

Tabela 9.13 – Grupo com os menores salários

Anos de estudo	Salários (R\$)
1	410,00
2	508,90
3	857,70
2	551,30
3	789,20
4	935,50

Tabela 9.14 – Resíduos da regressão

	Coeficiente	Desvio padrão	Estatística t
Constante I	183,797	69,187	2,66
Anos de estudo I	196,655	25,844	7,61

$$F_I = 57,9.$$

$$SQR_I = 14.694,4.$$

$$S^2_I = \frac{SQR_I}{n - 2} = \frac{14.694,4}{4} = 3.673,60.$$

Tabela 9.15 – Grupo com os maiores salários

Anos de estudo	Salários (R\$)
14	2.671,20
16	2.653,80
16	2.939,10
17	3.437,00
18	4.583,30
19	3.559,30
19	4.896,70

Tabela 9.16 – Resultados da regressão (maiores salários)

	Coeficiente	Desvio padrão	Estatística t
Constante II	-3.171,137	2.246,672	-1,41
Anos de estudo II	394,44	131,509	2,99

$$F_{II} = 8,996.$$

$$SQR_{II} = 1.729.453,67.$$

$$S^2_{II} = \frac{SQR_{II}}{n - 2} = \frac{1.729.453,67}{5} = 345.890,73.$$

Compara-se, então, a variância das duas regressões num teste F e, para isso, divide-se uma variância pela outra:

$$\frac{S^2_{II}}{S^2_I} = \frac{345.890,73}{3.673,60} = 94,16$$

Como o valor limite na tabela F, com 5% de significância, para cinco graus de liberdade no numerador e quatro graus de liberdade no denominador é 6,26, rejeita-se a hipótese de que as variâncias sejam iguais (vale a hipótese de que a variância da segunda regressão é maior) e, portanto, rejeita-se a hipótese nula de homocedasticidade. Concluiu-se, então, que o modelo de regressão estimado é heterocedástico.

Outro teste que pode ser usado para detecção do problema de heterocedasticidade é o **teste de White**, que consiste em partir de um modelo de regressão qualquer. Neste caso, usa-se um com duas variáveis apenas por simplificação.

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \epsilon_i$$

É feita uma regressão auxiliar em que a variável dependente é o resíduo ao quadrado e os regressores são os próprios regressores da regressão original, seus quadrados e os produtos cruzados, desta forma:

$$\hat{\epsilon}_i^2 = \gamma_1 + \gamma_2 X_{2i} + \gamma_3 X_{3i} + \gamma_4 X_{2i}^2 + \gamma_5 X_{3i}^2 + \gamma_6 X_{2i} X_{3i} + \mu_i$$

Um R^2 elevado nessa regressão auxiliar é um indício de que há heterocedasticidade. Mais precisamente, pode-se demonstrar que o produto nR^2 , nesse caso o R^2 não centrado, assim como no teste de Breusch-Godfrey sendo n o número de observações, segue uma distribuição de χ^2 com o número de graus de liberdade equivalente ao número de regressores da regressão auxiliar (menos o intercepto).

Exemplo 9.3.2.2

Na Tabela 9.17, estão apresentados os dados de consumo de energia elétrica médio por residência para 17 cidades. Cada cidade possui uma tarifa diferente e também é dada a renda familiar mensal média. Estime o consumo de energia em função da tarifa e da renda e verifique se há heterocedasticidade pelo teste de White.

Tabela 9.17 – Consumo de energia elétrica médio

Cidade	Consumo (kwh/mês)	Tarifa (\$/kwh)	Renda (\$/mês)
A	355,7	1,50	600
B	393,8	1,80	400
C	429,1	2,00	700
D	250,5	1,20	300
E	484,9	1,30	600
F	377,1	1,60	700
G	194,3	3,00	500
H	328,2	2,50	600
I	498,6	2,20	850
J	444,5	1,90	550
K	217,1	0,90	300
L	279,8	1,10	700
M	300,9	1,50	800
N	199,8	1,40	650
O	798,2	1,30	900
P	483,4	1,80	500
Q	518,9	2,40	400

Os resultados da regressão são os mostrados na Tabela 9.18.

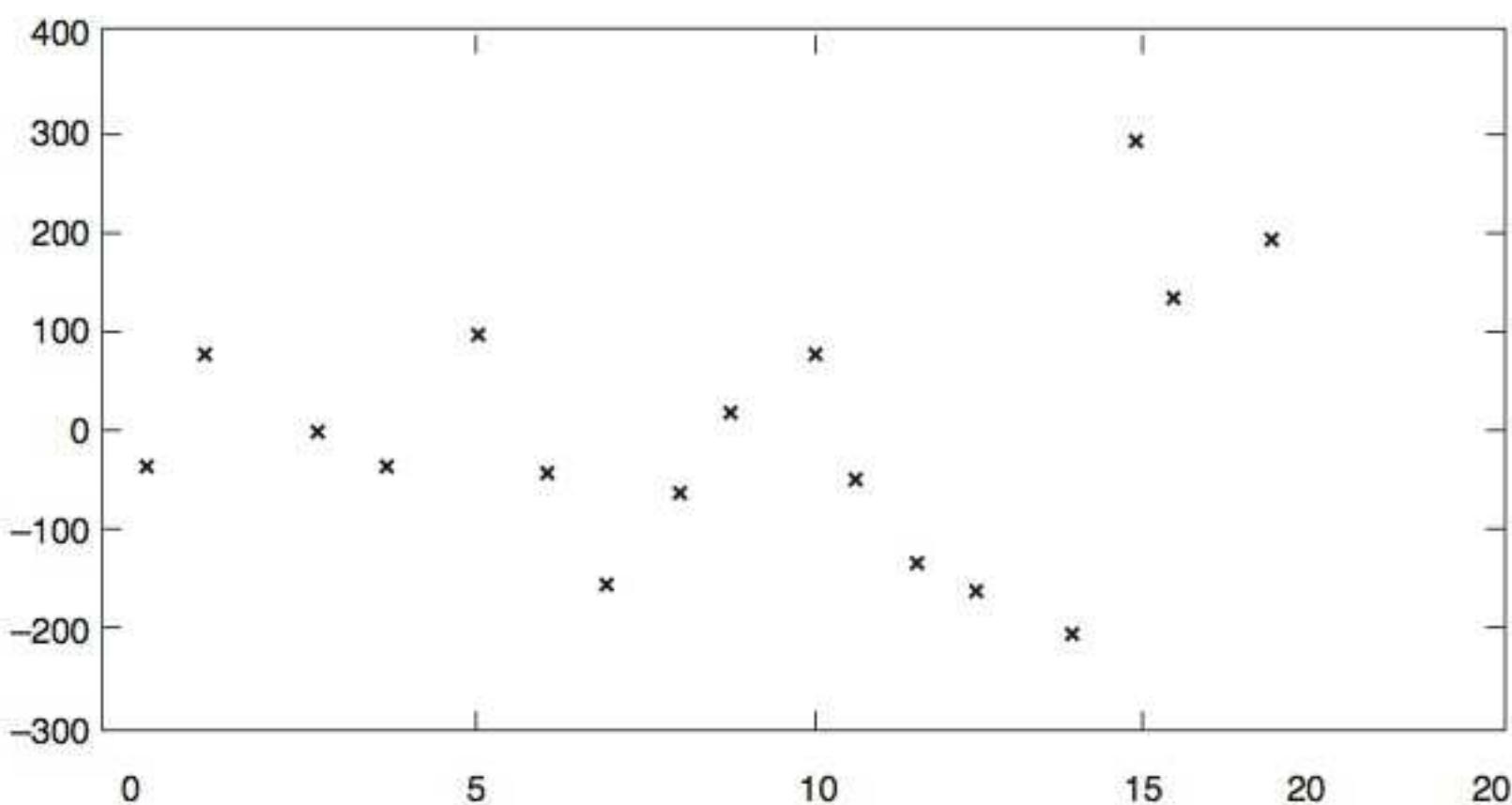
Tabela 9.18 – Resultados da regressão (consumo de energia)			
	Coeficiente	Desvio padrão	Estatística t
Constante	154,457	169,422	0,91
Renda	0,371	0,204	1,82
Tarifa	6,719	65,326	0,10

Estatística F = 1,65.

O coeficiente da renda foi significante apenas a 10%, o coeficiente da tarifa (assim como o intercepto) *não* foi significante (ainda bem, pois o sinal do coeficiente da tarifa supostamente seria negativo). Além disso, o teste F indica que a regressão *não* é válida, mas essas conclusões só são válidas se não existir heterocedasticidade, o que ainda não se sabe.

Uma análise da Figura 9.2 é sempre útil nesses casos.

Figura 9.2 – Resíduos (consumo de energia)



No eixo horizontal, o número 1 corresponde à cidade A, o 2 à B e assim sucessivamente. Novamente, é possível visualizar uma discrepância na dispersão dos erros: ela parece maior para as últimas cidades da tabela do que para as primeiras. Para ter uma ideia mais precisa, use o teste de White. Os dados para a regressão auxiliar são mostrados na Tabela 9.19.

Tabela 9.19 – Dados para a regressão auxiliar

Cidade	Resíduos	Resíduos ao quadrado	Tarifa (\$/kwh)	Renda (\$/mês)	Tarifa ao quadrado	Renda ao quadrado	Renda x tarifa
A	-31,611	999,26	1,50	600	2,25	360.000	900
B	78,731	6.198,64	1,80	400	3,24	160.000	720
C	1,300	1,69	2,00	700	4,00	490.000	1.400
D	-23,408	547,92	1,20	300	1,44	90.000	360
E	98,933	9.787,70	1,30	600	1,69	360.000	780
F	-48,012	2.305,17	1,60	700	2,56	490.000	1.120
G	-165,961	27.543,06	3,00	500	9,00	250.000	1.500
H	-65,830	4.333,65	2,50	600	6,25	360.000	1.500
I	13,762	189,41	2,20	850	4,84	722.500	1.870
J	73,066	5.338,60	1,90	550	3,61	302.500	1.045
K	-54,792	3.002,16	0,90	300	0,81	90.000	270
L	-141,952	20.150,50	1,10	700	1,21	490.000	770
M	-160,669	25.814,64	1,50	800	2,25	640.000	1.200
N	-205,404	42.190,68	1,40	650	1,96	422.500	910
O	300,845	90.507,94	1,30	900	1,69	810.000	1.170
P	131,202	17.214,03	1,80	500	3,24	250.000	900
Q	199,800	39.919,95	2,40	400	5,76	160.000	960

Tabela 9.20 – Regressão auxiliar

	Coeficiente	Desvio padrão	Estatística t
Constante	-41.106,88	46.462,86	-0,88
Renda	-67,308	127,155	-0,53
Tarifa	81.023,92	46.659,85	1,74
Renda ao quadrado	0,380	0,110	3,46
Tarifa ao quadrado	9.511,886	10.013,56	0,95
Renda x tarifa	-212,428	40,447	-5,25

O valor encontrado para o R^2 foi alto, o que indica que há mesmo heterocedasticidade. Entretanto, o teste definitivo é feito multiplicando-se o R^2 pelo número de observações.

$$n \times R^2 = 17 \times 0,7942 \approx 13,5$$

Como o valor limite⁸ da distribuição χ^2 com 5 graus de liberdade e 5% de significância é 11,07, rejeita-se a hipótese nula de homocedasticidade, ou seja, conclui-se que o modelo estimado apresenta, sim, heterocedasticidade.

⁸ Limite superior, bem entendido. Portanto, na tabela no final deste livro, olha-se a coluna dos 95% se se quiser 5% de significância.

9.3.3 O que fazer quando há heterocedasticidade?

Havendo heterocedasticidade, o procedimento de *correção* é mais simples se souber (ou, pelo menos, suspeitar) qual é o padrão da heterocedasticidade.

Tome o modelo de regressão a seguir e suponha que exista heterocedasticidade.

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$$

Admita que seja conhecido que a variância dos erros é dada por:

$$\text{var}(\varepsilon_i) = \sigma^2_i = z_i \sigma^2$$

Ou seja, a variância, que não é constante, é uma variável z_i multiplicada por uma constante. Se conseguir eliminar a variável z da variância, você tem, então, uma variância constante e aí estaria livre do problema da heterocedasticidade.

Você já sabe, do Capítulo 2, que, para transformar uma variável cuja variância é $z_i \sigma^2$ em outra cuja variância é simplesmente σ^2 , deve-se dividi-la por $\sqrt{z_i}$ (ressaltando que variância lembra quadrados). A solução, então, é dividir todo o modelo de regressão por $\sqrt{z_i}$:

$$\frac{Y_i}{\sqrt{z_i}} = \beta_1 \frac{1}{\sqrt{z_i}} + \beta_2 \frac{X_{2i}}{\sqrt{z_i}} + \beta_3 \frac{X_{3i}}{\sqrt{z_i}} + \mu_i$$

Então, a variância desse novo termo de erro μ_i é dada por:

$$\text{var}(\mu_i) = \text{var}\left(\frac{\varepsilon_i}{\sqrt{z_i}}\right) = \frac{1}{z_i} \text{var}(\varepsilon_i) = \frac{1}{z_i} \sigma^2_i = \frac{1}{z_i} z_i \sigma^2 = \sigma^2$$

Que é constante e, portanto, esse modelo transformado será homocedástico (se, é claro, a variância seguir de fato o padrão já indicado).

Quando se estima o modelo transformado por mínimos quadrados, o método é chamado de método dos **mínimos quadrados ponderados**.

Claro que o método dos mínimos quadrados ponderados também pode ser usado quando o padrão conhecido é o do desvio padrão. Suponha que o desvio padrão dos erros seja dado por:

$$\text{dp}(\varepsilon_i) = \sigma_i = z_i \sigma$$

Nesse caso, a solução é simplesmente dividir o modelo por z_i :

$$\frac{Y_i}{z_i} = \beta_1 \frac{1}{z_i} + \beta_2 \frac{X_{2i}}{z_i} + \beta_3 \frac{X_{3i}}{z_i} + \mu_i$$

E o desvio padrão do erro desse modelo é dado por:

$$\text{dp}(\mu_i) = \text{dp}\left(\frac{\varepsilon_i}{z_i}\right) = \frac{1}{z_i} \text{dp}(\varepsilon_i) = \frac{1}{z_i} z_i \sigma = \sigma$$

O desvio padrão será, então, uma constante e, obviamente, a variância também, eliminando a heterocedasticidade.

Exemplo 9.3.3.1

Estime de novo a regressão do Exemplo 9.3.2.1, corrigindo o problema da heterocedasticidade.

Supostamente, a causa da heterocedasticidade naquele exemplo é a de que a variação dos salários é maior para maior tempo de estudo. Seria possível imaginar que a variância ou o desvio padrão sejam proporcionais ao tempo de estudo.

Se se considerar o desvio padrão proporcional aos anos de estudo, a solução indicada é dividir toda a equação pelos anos de estudo. Nesse caso, entretanto, a variável a ser dividida é a própria variável independente do modelo. Ou seja, o modelo inicial:

$$Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i$$

Sendo Y os salários e X os anos de estudo, torna-se:

$$\frac{Y_i}{X_i} = \beta_1 \frac{1}{X_i} + \beta_2 + \mu$$

$$\frac{Y_i}{X_i} = \beta_1 \frac{1}{X_i} + \beta_2 + \mu$$

Então, para estimar os coeficientes β_1 e β_2 sem o problema da heterocedasticidade, deve-se estimar uma regressão simples em que a variável independente é a razão salário/anos de estudo e a variável dependente é o inverso dos anos de estudo. Os dados estão na Tabela 9.21.

Tabela 9.21 – Regressão simples (anos de estudo x salários)

Anos de estudo (X)	Salários (Y)	1/X	Y/X
1	410,00	1,000000	410,0000
2	508,90	0,500000	254,4500
3	857,70	0,333333	285,9000
2	551,30	0,500000	275,6500
3	789,20	0,333333	263,0667
4	935,50	0,250000	233,8750
7	1.529,30	0,142857	218,4714
8	1.497,50	0,125000	187,1875
9	2.317,70	0,111111	257,5222
11	2.169,50	0,090909	197,2273
11	2.596,80	0,090909	236,0727
13	2.844,60	0,076923	218,8154
13	3.391,00	0,076923	260,8462
14	2.671,20	0,071429	190,8000
16	2.653,80	0,062500	165,8625
16	2.939,10	0,062500	183,6937
17	3.437,00	0,058824	202,1765
18	4.583,30	0,055556	254,6278
19	3.559,30	0,052632	187,3316
19	4.896,70	0,052632	257,7211

Os resultados dessa nova regressão estão resumidos na Tabela 9.22.

Tabela 9.22 – Resultados da nova regressão (anos de estudo x salários)

	Coeficiente	Desvio padrão	Estatística t
$\hat{\beta}_2$	198,869	9,126	21,79
$\hat{\beta}_1$	188,745	29,716	6,35

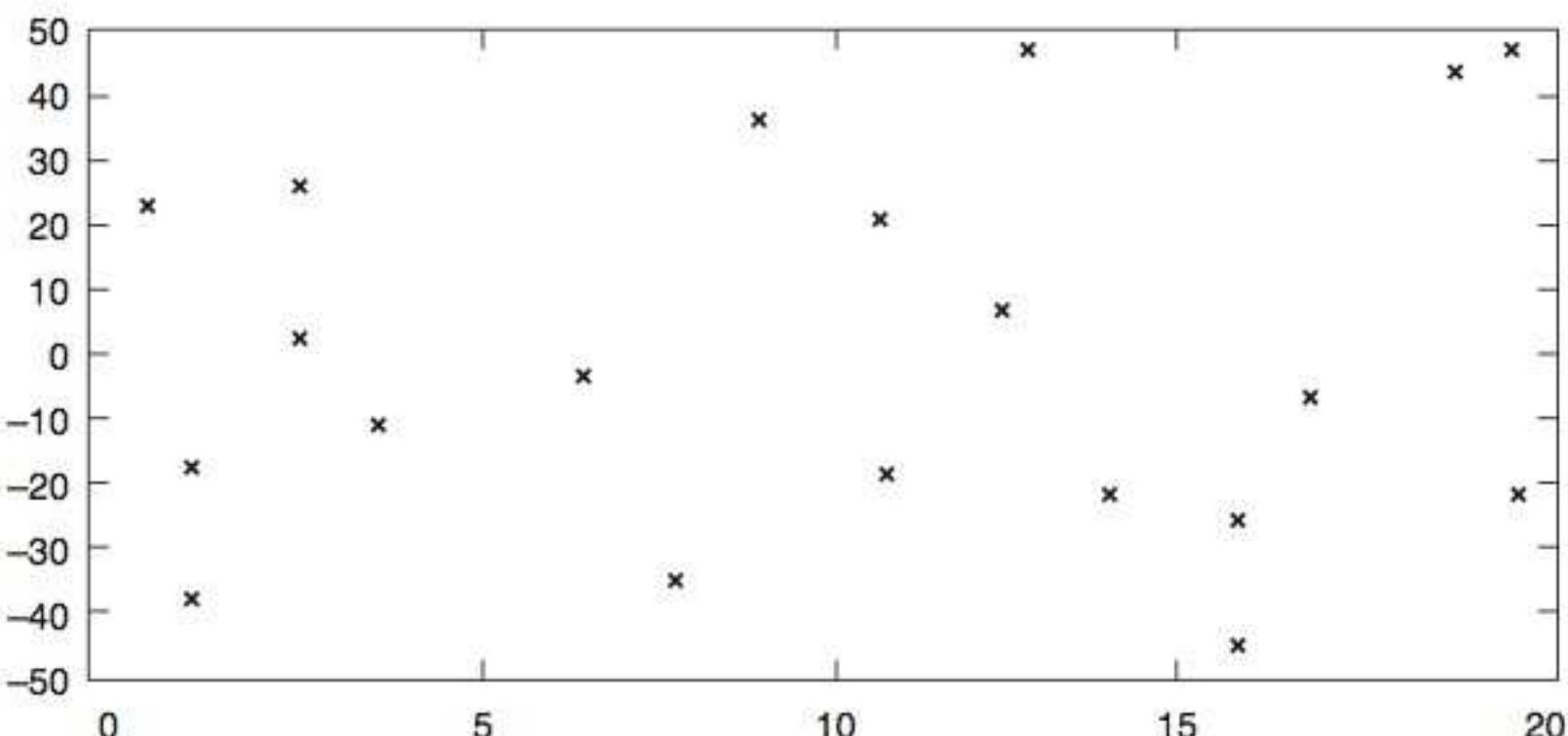
 $F = 40,34$.

Os valores de $\hat{\beta}_1$ e $\hat{\beta}_2$ obtidos agora, por mínimos quadrados ponderados, representam uma estimativa mais precisa dos dois coeficientes; além disso, é possível confiar nos testes de hipóteses tendo em vista que não há heterocedasticidade. Bom, isso se não houver realmente. Para ter certeza disso, usa-se um dos testes vistos, por exemplo, o teste de White. Antes disso, seria interessante observar os resíduos num gráfico (veja a Figura 9.3), depois de os dados da Tabela 9.23 serem tabulados.

Tabela 9.23 – Novos resíduos (anos de estudo x salários)

22,38656	-12,18006	20,04521	-26,97171
-38,79119	-7,361003	5,427646	-7,795078
24,11623	-35,27449	47,45842	45,27304
-17,59119	37,68168	-21,55068	-21,47127
1,282899	-18,80025	-44,80296	48,91820

Figura 9.3 – Novos resíduos (anos de estudo x salários)



Como se vê, pelo menos aparentemente, os resíduos se mostram mais *equilibrados* no que se refere a sua dispersão. De fato, isso pode ser visto no resultado do teste de White mostrado na Tabela 9.24.

Tabela 9.24 – Resultados da regressão auxiliar do teste de White			
	Coeficiente	Desvio padrão	Estatística t
Constante	1.172,566	337,864	3,47
Variável independente	-2.546,962	2.502,224	-1,02
Variável independente ao quadrado	2.026,627	2.615,736	0,77

 $R^2 = 0,0758$. $n \times R^2 = 20 \times 0,0758 \approx 1,52$.

Como o valor limite, a 5% de significância, com 2 graus de liberdade, na distribuição χ^2 é 5,99, aceita-se a hipótese nula de homocedasticidade para esse modelo.

Quando não se conhece o padrão da heterocedasticidade, as formas de correção são um pouco mais complexas. Há uma possibilidade, entretanto, que já foi até discutida no capítulo anterior: muitas vezes (mas nem sempre), quando o modelo nas variáveis originais apresenta heterocedasticidade, ela não ocorre se essas variáveis estão em logaritmo. Esta é uma possibilidade, então: calculam-se os logaritmos das variáveis envolvidas na regressão e se testa de novo para a heterocedasticidade. Eis, assim, um terceiro motivo para o uso de modelos com o logaritmo das variáveis. Os outros motivos seriam um eventual melhor ajuste com logaritmos e a possibilidade de estimação direta das elasticidades.

9.4 Violando a hipótese III: o problema da simultaneidade

A hipótese III estabelece que as variáveis independentes, os regressores – os X , enfim –, num modelo de regressão devem ser fixos, isto é, não estocásticos, não aleatórios. Uma versão mais branda dessa hipótese, vista no capítulo anterior, estabelece que, se uma (ou mais) variável independente for estocástica, é preciso que, pelo menos, ela não tenha correlação com o termo de erro. E se tiver?

Isso remete a outra questão: *o que* levaria uma variável supostamente⁹ independente a ter correlação com o termo de erro? A resposta a essa pergunta lembra uma antiga propaganda de um biscoito, em que se discutia a relação de causa e efeito: ele vende mais porque está sempre fresquinho ou está sempre fresquinho porque vende mais?

Note que, no *modelo teórico* proposto pela propaganda, há duas *funções*: a quantidade de biscoitos vendidos é função da probabilidade de que encontremos biscoitos fresquinhos; por outro lado, o número de unidades fresquinhos será maior se as vendas forem maiores, já que os biscoitos não ficarão em estoque por muito tempo. Há, portanto, duas **equações simultâneas**, em que as variáveis *estar sempre fresquinho* e *quantidade de vendas* se determinam mutuamente.

⁹ Note que, se ela tem correlação com o erro, ela não é tão independente assim.

Em economia e em outras ciências sociais, essas situações ocorrem com frequência. Em particular, o modelo de determinação de preços básico na economia, de *oferta* e *demand*a, é um desses casos: na oferta, o produtor produzirá maior quantidade quanto *maior* for o preço; na demanda, o consumidor comprará maiores quantidades quanto *menor* for o preço.

Assim, se o preço estiver muito baixo, muitos consumidores vão querer adquirir o produto, mas a produção será pequena, o que fará com que o preço suba; da mesma forma, se a quantidade produzida for muito grande, os produtores serão obrigados a baixar o preço para vender toda sua produção. Preços e quantidades, portanto, determinam-se mutuamente.

Suponha que a quantidade a ser produzida, chamada de quantidade ofertada, seja função única e exclusivamente do preço:

$$Q_i^O = \alpha_0 + \alpha_1 P_i + \mu_i$$

Sendo $\alpha_1 > 0$.

Já quanto aos consumidores, assuma que, além do preço, eles levem em conta a renda em sua decisão de consumir. Então, para a quantidade demandada:

$$Q_i^D = \beta_0 + \beta_1 P_i + \beta_2 R_i + \nu_i$$

Sendo $\beta_1 < 0$.

Como no equilíbrio de mercado, $Q^O = Q^D$, e o que é observado são quantidades de equilíbrio (já que o que é consumido tem de ser igual ao que é vendido), não há ambiguidade em chamar ambas simplesmente de Q . Então, chega-se a um sistema de duas equações:

$$Q_i = \alpha_0 + \alpha_1 P_i + \mu_i \quad (\text{oferta})$$

$$Q_i = \beta_0 + \beta_1 P_i + \beta_2 R_i + \nu_i \quad (\text{demanda})$$

As variáveis Q e P se determinam mutuamente nesse modelo, por isso são chamadas de variáveis **endógenas**. Já R é uma variável realmente independente no modelo; seu valor já é *predeterminado*, é uma variável **exógena**.

A regressão por mínimos quadrados ordinários dessas equações levará a estimadores *viesados* e *inconsistentes*, já que um dos regressores é uma variável endógena, determinada pelo próprio modelo descrito pelas equações e, portanto, está correlacionada com o termo de erro. Repare que é a mesma situação do biscoito, pois, imagine que a renda dos consumidores aumente: haverá maior procura pelo produto, aumentando o preço; mas o preço maior estimula maior produção. Quantidade afeta o preço, que afeta a quantidade.

9.4.1 A questão da identificação

Partindo do sistema de equações anterior, *isole* as variáveis endógenas. Se igualar os Q das equações de oferta e demanda (omitindo os índices i por simplicidade de notação), chega-se a:

$$Q = Q$$

$$\alpha_0 + \alpha_1 P + \mu = \beta_0 + \beta_1 P + \beta_2 R + v$$

$$\alpha_1 P - \beta_1 P = \beta_0 - \alpha_0 + \beta_2 R + v - \mu$$

$$P = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \frac{\beta_2}{\alpha_1 - \beta_1} R + \frac{v - \mu}{\alpha_1 - \beta_1}$$

Você encontra uma equação que coloca o preço em função apenas de variáveis exógenas (uma só, nesse caso). Observando essa equação, fica mais clara a correlação do preço com os (dois) termos de erro.

Substituindo a equação do preço que acabou de encontrar na equação de oferta:

$$Q = \alpha_0 + \alpha_1 P + \mu$$

$$Q = \alpha_0 + \alpha_1 \left(\frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \frac{\beta_2}{\alpha_1 - \beta_1} R + \frac{v - \mu}{\alpha_1 - \beta_1} \right) + \mu$$

Fazendo as operações adequadas, chega-se a:

$$Q = \frac{\alpha_1 \beta_0 - \alpha_0 \beta_1}{\alpha_1 - \beta_1} + \frac{\alpha_1 \beta_2}{\alpha_1 - \beta_1} R + \frac{\alpha_1 v - \beta_1 \mu}{\alpha_1 - \beta_1}$$

Essa equação também coloca uma das variáveis endógenas (Q) em função da variável exógena R . Há um novo sistema de equações, que *isola* as variáveis endógenas em cada equação, e essas equações são chamadas de equações na **forma reduzida**. O sistema original de equações é chamado **forma estrutural** do modelo.

As equações na forma reduzida são, então:

$$P = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \frac{\beta_2}{\alpha_1 - \beta_1} R + \frac{v - \mu}{\alpha_1 - \beta_1}$$

$$Q = \frac{\alpha_1 \beta_0 - \alpha_0 \beta_1}{\alpha_1 - \beta_1} + \frac{\alpha_1 \beta_2}{\alpha_1 - \beta_1} R + \frac{\alpha_1 v - \beta_1 \mu}{\alpha_1 - \beta_1}$$

Sistema que pode ser escrito de maneira mais simples como:

$$P = \pi_1 + \pi_2 R + \tau$$

$$Q = \pi_3 + \pi_4 R + \xi$$

Em que:

$$\pi_1 = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}$$

$$\pi_2 = \frac{\beta_2}{\alpha_1 - \beta_1}$$

$$\pi_3 = \frac{\alpha_1 \beta_0 - \alpha_0 \beta_1}{\alpha_1 - \beta_1}$$

$$\pi_4 = \frac{\alpha_1 \beta_2}{\alpha_1 - \beta_1}$$

$$\tau = \frac{v - \mu}{\alpha_1 - \beta_1}$$

$$\xi = \frac{\alpha_1 v - \beta_1 \mu}{\alpha_1 - \beta_1}$$

Note que as equações na forma reduzida não têm mais o problema de que um ou mais regressores são correlacionados com o termo de erro e, então, elas podem perfeitamente ser estimadas por mínimos quadrados ordinários. Só que, estimando as equações na forma reduzida, encontram-se os π , e não os α e β . Fica o problema de, dados os parâmetros da forma reduzida, encontrar os da forma estrutural. Da equação de oferta:

$$Q = \alpha_0 + \alpha_1 P + \mu$$

Substituindo pelas equações da forma reduzida e omitindo os termos de erro (já que estamos falando dos estimadores), vêm:

$$\hat{\pi}_3 + \hat{\pi}_4 R = \hat{\alpha}_0 + \hat{\alpha}_1 (\hat{\pi}_1 + \hat{\pi}_2 R)$$

$$\hat{\pi}_3 + \hat{\pi}_4 R = \hat{\alpha}_0 + \hat{\alpha}_1 \hat{\pi}_1 + \hat{\alpha}_1 \hat{\pi}_2 R$$

Lembrando que os estimadores já foram obtidos das equações na forma reduzida por mínimos quadrados ordinários, as incógnitas são os $\hat{\alpha}$. Para manter a igualdade, é preciso ter os coeficientes *puros* iguais em cada lado, bem como os coeficientes da renda:

$$\hat{\pi}_3 = \hat{\alpha}_0 + \hat{\alpha}_1 \hat{\pi}_1$$

$$\hat{\pi}_4 = \hat{\alpha}_1 \hat{\pi}_2$$

Que é um sistema de duas equações e duas incógnitas que, não só tem solução, como nesse caso é até fácil encontrá-la, pois, da segunda equação:

$$\hat{\alpha}_1 = \frac{\hat{\pi}_4}{\hat{\pi}_2}$$

E aí, substituindo na primeira:

$$\hat{\pi}_3 = \hat{\alpha}_0 + \hat{\alpha}_1 \hat{\pi}_1$$

$$\hat{\pi}_3 = \hat{\alpha}_0 + \frac{\hat{\pi}_4}{\hat{\pi}_2} \hat{\pi}_1$$

$$\hat{\alpha}_0 = \hat{\pi}_3 - \frac{\hat{\pi}_4}{\hat{\pi}_2} \hat{\pi}_1$$

Portanto, é perfeitamente possível encontrar os coeficientes da oferta a partir dos coeficientes obtidos da estimação na forma reduzida. Veja se o mesmo ocorre para a demanda:

$$Q = \beta_0 + \beta_1 P + \beta_2 R + v$$

Fazendo o mesmo procedimento, isto é, substituindo pelas equações da forma reduzida e omitindo os termos de erro:

$$\hat{\pi}_3 + \hat{\pi}_4 R = \hat{\beta}_0 + \hat{\beta}_1 (\hat{\pi}_1 + \hat{\pi}_2 R) + \hat{\beta}_2 R$$

$$\hat{\pi}_3 + \hat{\pi}_4 R = \hat{\beta}_0 + \hat{\beta}_1 \hat{\pi}_1 + (\hat{\beta}_1 \hat{\pi}_2 + \hat{\beta}_2) R$$

Que geram as equações:

$$\hat{\pi}_3 = \hat{\beta}_0 + \hat{\beta}_1 \hat{\pi}_1$$

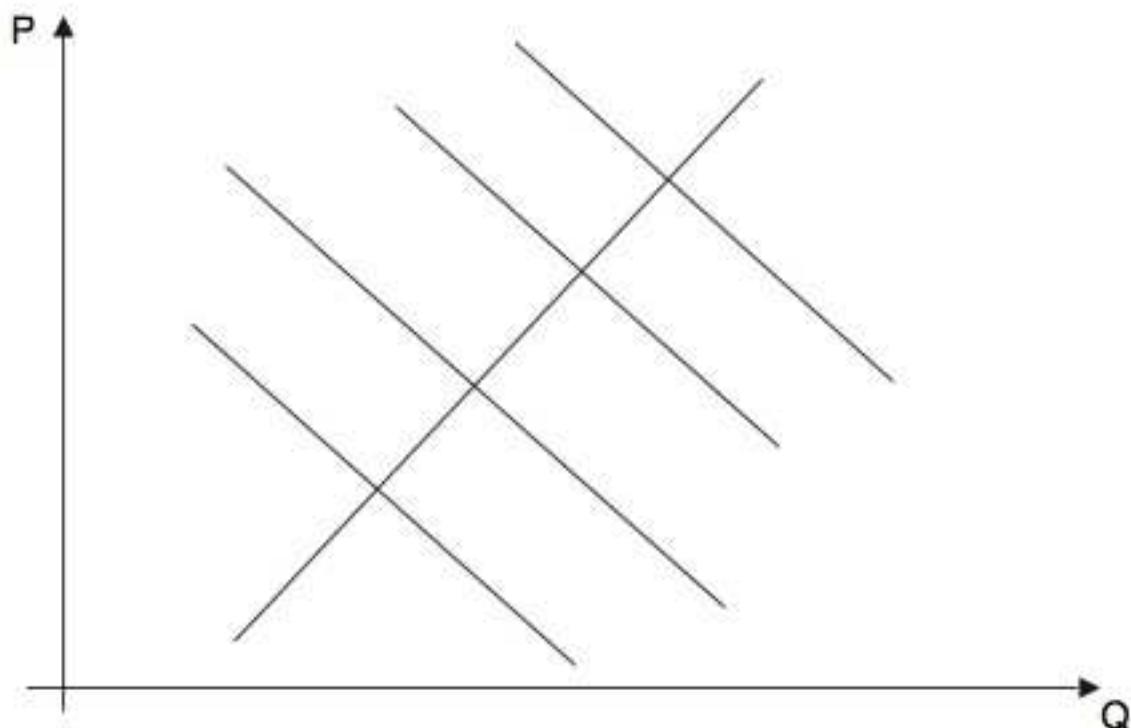
$$\hat{\pi}_4 = \hat{\beta}_1 \hat{\pi}_2 + \hat{\beta}_2$$

Há agora *três* incógnitas ($\hat{\beta}_0$, $\hat{\beta}_1$ e $\hat{\beta}_2$) e apenas *duas* equações. Não é possível encontrar os coeficientes da demanda a partir dos coeficientes estimados na forma reduzida. Outra maneira de dizer isso é que não se pode *identificar* a equação de demanda ou, simplesmente, que a equação da demanda apresentada no modelo é **subidentificada**.

A equação de oferta, ao contrário, é possível de ser identificada. Diz-se que a equação de oferta é **exatamente identificada** (você já descobrirá o motivo desse *exatamente*).

Para quem está familiarizado com a teoria econômica, a analogia é clara. Como existe a renda na equação da demanda, mudanças naquela implicam deslocamento da curva de demanda. Deslocando a curva de demanda, podem-se encontrar vários pontos na curva de oferta e, assim, é possível identificá-la.

Figura 9.4 – Uma curva de oferta e diferentes curvas de demanda



Note que, se, além da renda, a equação da demanda contemplasse também o preço de um bem substituto como variável, seria mais uma variável que poderia *deslocar* a demanda e identificar a oferta. Nesse caso, a equação de oferta estaria **superidentificada** (daí o motivo de se usar o *exatamente* para qualificar a identificação da oferta).

Qual é a regra? Com duas variáveis endógenas em cada equação, para a equação ser identificada, é preciso haver *uma* variável exógena *fora* da equação. Dá para estender o raciocínio para três variáveis endógenas, seriam necessárias duas exógenas fora e assim por diante. Pode-se generalizar da seguinte forma:

Se número de variáveis endógenas incluídas – 1 = número de variáveis exógenas excluídas, a equação é exatamente identificada.

Se número de variáveis endógenas incluídas – 1 > número de variáveis exógenas excluídas, a equação é subidentificada.

Se número de variáveis endógenas incluídas – 1 < número de variáveis exógenas excluídas, a equação é superidentificada.

Mas atenção: isso se refere apenas à *condição necessária* para a identificação, também conhecida como questão de **ordem**. Veja que, no exemplo anterior da oferta e demanda, a equação de oferta é exatamente identificada, desde que a renda de fato exista na equação da demanda, isto é, que o coeficiente β_2 seja diferente de zero.

A condição de ordem é uma condição necessária, mas não suficiente. No modelo de oferta e demanda, a presença da renda na equação de demanda faz com que a oferta se torne identificada. Mas isso se esse coeficiente não for nulo. Essa seria a condição suficiente, nesse caso.

O modo geral de verificar a condição suficiente é visto no exemplo a seguir.

Exemplo 9.4.1.1

Dado o seguinte modelo:

$$(1) Y_t = C_t + I_t + G_t$$

$$(2) C_t = \alpha_0 + \alpha_1 Y_t + \alpha_2 Y_{t-1} + \alpha_3 r_t + \varepsilon_{1t}$$

$$(3) I_t = \beta_0 + \beta_1 r_t + \beta_2 Y_t + \varepsilon_{2t}$$

$$(4) r_t = \gamma_0 + \gamma_1 m_t + \gamma_2 Y_t + \varepsilon_{3t}$$

Sendo Y a renda nacional, C o consumo, I o investimento, G os gastos governamentais, r a taxa de juros e m a quantidade de moeda emitida. O governo controla seus gastos e a emissão de moeda. Verifique a condição de identificação para cada uma das equações. A equação (1) é uma identidade, não tem coeficientes a serem estimados, portanto, não cabe a questão da identificação para ela. Para as seguintes, sim, mas será mostrada somente equação (4), ficando as demais como exercício.

O governo estipula quais serão seus gastos e a emissão de moeda, portanto, essas são variáveis exógenas. As demais são endógenas, mas quando se tomam valores defasados de tais variáveis, elas já estão, obviamente, predeterminadas (elas vêm do passado, afinal); assim, do ponto de vista do modelo no período atual, elas têm o mesmo comportamento que as variáveis exógenas. Então:

variáveis endógenas: Y_t, C_t, I_t, r_t

variáveis exógenas: G_t, m_t, Y_{t-1}

No caso da equação (4):

variáveis endógenas incluídas = 2

variáveis exógenas excluídas = 2

A equação, pela condição de ordem, é superidentificada. Mas é preciso verificar a condição suficiente, o que é mais complicado agora, porque há várias equações. Para isso, monte uma tabela (como a tabela 9.25) com as várias equações, preenchida com os coeficientes das variáveis incluídas em cada equação.

Tabela 9.25 – Coeficientes das variáveis incluídas ou não em cada equação

Equação	Y_t	C_t	I_t	G_t	r_t	m_t	Y_{t-1}
(1)	-1	1	1	1	0	0	0
(2)	α_1	-1	0	0	α_3	0	α_2
(3)	β_2	0	-1	0	β_1	0	0
(4)	γ_2	0	0	0	-1	γ_1	0

Note que os coeficientes das variáveis que estão à direita das equações têm o coeficiente igual a -1 porque todas as variáveis foram consideradas como se estivessem do outro lado da equação. O Y da equação (1) passa para o outro lado, então $0 = -Y + C + I + G$, daí o coeficiente -1 para Y .

Monte uma matriz a partir da Tabela 9.25 com a seguinte regra: excluir a linha correspondente à equação estudada e incluir as colunas correspondentes às variáveis excluídas da equação (4) (C_t , I_t , G_t e Y_{t-1}). Chega-se à seguinte matriz 3×4 .

$$\begin{bmatrix} 1 & 1 & 1 & 0 \\ -1 & 0 & 0 & \alpha_2 \\ 0 & -1 & 0 & \beta_1 \end{bmatrix}$$

Se forem tomadas, por exemplo, as três primeiras colunas da matriz, tem-se uma matriz 3×3 cujo determinante é igual a -1 , diferente de zero, portanto. O posto da matriz então é 3, e é essa a condição de posto, que o posto da matriz construída na forma acima seja igual a $G - 1$ (número total de variáveis endógenas - 1).

9.4.2 Como estimar um modelo de equações simultâneas

Um método já foi explicitado na seção anterior: estimam-se os parâmetros da forma reduzida. Conhecida a relação entre os parâmetros da forma reduzida e da forma estrutural, podem-se encontrar estes últimos. Esse método é conhecido como dos **mínimos quadrados indiretos**. É importante ressaltar que, em métodos de equações simultâneas, não é possível, em geral, obter estimadores não viesados; o que se consegue é eliminar a inconsistência.

Porém, isso só pode ser feito para equações exatamente identificadas. Se a equação for subidentificada, não dá para estimar mesmo. Se a equação for superidentificada – o que, em princípio, é bom, pois há mais informação –, não dá para encontrar uma relação “de um para um” entre os parâmetros da forma estrutural e reduzida que dê uma única solução.

Um método que pode ser estendido a equações superidentificadas é o dos **mínimos quadrados de dois estágios**, que consiste em estimar as equações da forma reduzida. Aí, encontrar os valores estimados para as variáveis endógenas. Como são valores estimados, não incluem os resíduos e, portanto, não têm correlação com o termo de erro. Então, usam-se esses valores estimados como substitutos das variáveis endógenas, que, no modelo estrutural, aparecem no lado direito das equações.

Exemplo 9.4.2.1

Dado o modelo estrutural para o mercado de um bem:

$$Q_i = \alpha_0 + \alpha_1 P_i + \alpha_2 M_i + \alpha_3 S_i + \mu_i \quad (\text{oferta})$$

$$Q_i = \beta_0 + \beta_1 P_i + \beta_2 R_i + v_i \quad (\text{demanda})$$

Sendo Q a quantidade comercializada, P o preço, R a renda média dos consumidores, M o preço da matéria-prima e S os salários médios pagos aos trabalhadores que trabalham na produção desse bem. Com os dados da Tabela 9.26, estime os parâmetros do modelo.

Tabela 9.26 – Dados para o Exemplo 9.4.2.1

Q	P	R	M	S
98	10,00	399,20	200,00	410,00
99	10,40	480,80	195,00	405,00
102	10,30	473,60	189,00	405,00
101	10,50	485,60	185,00	410,00
104	9,80	498,40	181,00	350,00
103	9,90	504,00	176,00	360,00
104	10,10	525,60	169,00	370,00
100	10,50	562,40	165,00	350,00
100	9,60	472,80	160,00	355,00
102	9,10	411,20	154,00	395,00
95	9,30	300,80	152,00	495,00
92	9,88	315,20	144,00	555,00
94	10,30	376,80	140,00	545,00
98	9,90	424,80	135,00	495,00
105	9,50	524,80	131,00	390,00
100	9,85	540,80	126,00	375,00
103	8,60	471,20	120,00	345,00
100	10,40	535,20	115,00	435,00
105	10,55	585,60	112,00	455,00

Há duas variáveis endógenas (Q e P) e três variáveis exógenas (S , M e R). É fácil verificar que a equação de demanda é superidentificada e a de oferta é exatamente identificada.

As equações na forma reduzida são:

$$P = \pi_1 + \pi_2 R_i + \pi_3 M_i + \pi_4 S_i + \tau_i$$

$$Q = \pi_5 + \pi_6 R_i + \pi_7 M_i + \pi_8 S_i + \xi_i$$

Os resultados da estimativa por mínimos quadrados ordinários das equações na forma reduzida foram:

$$\hat{P} = -0,683 + 0,00867R_i + 0,0148M_i + 0,0102S_i$$

$$(0,850) \quad (0,00075) \quad (0,0017) \quad (0,0009)$$

$$\hat{Q} = 103,062 + 0,0215R_i - 0,0107M_i - 0,0269S_i$$

$$(10,561) \quad (0,0093) \quad (0,0207) \quad (0,0116)$$

Os valores entre parênteses são os desvios padrão.

A partir dessas equações, calcule as estimativas de Q e P , incluídas na Tabela 9.27.

Tabela 9.27 – Dados para o Exemplo 9.4.2.1, com estimativas de Q e P

\hat{Q}	\hat{P}	R	M	S
98,4643	9,9287	399,20	200,00	410,00
100,4062	10,5109	480,80	195,00	405,00
100,3157	10,3597	473,60	189,00	405,00
100,4819	10,4557	485,60	185,00	410,00
102,4148	9,8940	498,40	181,00	350,00
102,3195	9,9708	504,00	176,00	360,00
102,5895	10,1567	525,60	169,00	370,00
103,9616	10,2121	562,40	165,00	350,00
101,9549	9,4125	472,80	160,00	355,00
99,6186	9,1986	411,20	154,00	395,00
94,5756	9,2343	300,80	152,00	495,00
93,3558	9,8542	315,20	144,00	555,00
94,9917	10,2268	376,80	140,00	545,00
97,4227	10,0577	424,80	135,00	495,00
102,4409	9,7919	524,80	131,00	390,00
103,2421	9,7033	540,80	126,00	375,00
102,6180	8,7044	471,20	120,00	345,00
101,6246	10,1053	535,20	115,00	435,00
102,2015	10,7023	585,60	112,00	455,00

Como o preço é a única variável endógena que aparece do lado direito da equação, utilize seus valores estimados para a estimação do modelo estrutural, cujos resultados são mostrados a seguir:

$$\hat{Q}_i = 104,756 + 2,479 \hat{P}_i - 0,0523S_i - 0,0474M_i \quad (\text{oferta})$$

$$(11,575) \quad (1,254) \quad (0,0097) \quad (0,023)$$

$$\hat{Q}_i = 101,225 - 2,0568 \hat{P}_i + 0,0416R_i \quad (\text{demanda})$$

$$(9,085) \quad (0,984) \quad (0,0063)$$

Note que os sinais obtidos foram os esperados e os coeficientes encontrados foram significantes a, pelo menos, 10% (verifique!).

Exercícios

Enunciado para os Exercícios 1 a 3: dados os modelos estimados a seguir, verifique (baseado em intuição ou teoria) se os sinais obtidos são adequados, bem como outras evidências de *multicolinearidade*, e identifique as possíveis causas e eventuais correções.

1. CONSENER = $234 - 0,8 \text{ POP} + 0,2 \text{ CASAS} + 1,2 \text{ RENDA} - 12,1 \text{ PREÇO}$
 (176) (0,7) (0,12) (0,7) (9,3)

$R^2 = 0,92$

n = 20 observações

CONSENER = consumo de energia elétrica

POP = população

CASAS = número de residências

RENDAM = renda média da população

PREÇO = preço do kwh de energia elétrica

2. SALÁRIO = $23,5 - 1,89 \text{ PONTOS} + 8,9 \text{ REB} + 1,4 \text{ ASSIST} + 0,89 \text{ ROUB} + 12,1 \text{ PERC}$
 (18,7) (2,03) (4,0) (0,4) (0,75) (10,8)

F = 45,21

SALÁRIO = salário pago em uma liga profissional de basquete

PONTOS = número de pontos por jogo

REB = número de rebotes por jogo

ASSIST = número de assistências por jogo

ROUB = número de *roubadas* de bola por jogo

PERC = aproveitamento percentual dos arremessos à cesta

3. CRIME = $18,9 - 2,91 \text{ ÁREA} + 0,31 \text{ RENDA} + 0,78 \text{ POP} - 3,1 \text{ ESCOLA}$
 (11,2) (1,76) (0,20) (0,49) (2,1)

$R^2 = 0,86$

CRIME = índice de criminalidade em uma cidade

ÁREA = área total da região urbana em km²

RENDAM = renda *per capita* da cidade

POP = população da cidade

ESCOLA = número médio de anos de escolaridade da população

4. Dados os valores de Y, X, Z e W mostrados na Tabela 9.28:

Tabela 9.28 – Valores de Y, X, Z e W

Y	X	Z	W
13,0	17,16	2,3	0,56
14,0	8,14	4,5	0,34
12,0	10,67	6,7	0,67
11,5	-3,39	8,9	0,21
16,0	-2,01	10,1	0,39
17,0	0,31	12,3	0,71
18,8	-15,02	14,4	0,18
15,4	-6,83	16,5	0,77
13,9	-16,57	17,8	0,43
16,2	-20,32	18,1	0,28

- a) Calcule os coeficientes de correlação simples entre X , W e Z .
- b) É possível estimar o modelo de regressão $Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + \beta_3 W_i + \mu_i$? Justifique. (Sugestão: faça regressões utilizando as variáveis X , Z e W .)
5. Em uma cidade, foram obtidos os valores da Tabela 9.29. Faça uma regressão que tome como variável dependente o preço do imóvel e como variáveis explicativas as variáveis *distância ao centro, número de dormitórios, área do imóvel e renda mensal do chefe da família*. Feita essa estimativa, calcule as correlações amostrais entre as variáveis explicativas; com estes últimos resultados, faça alterações no modelo que você julgar relevantes e discuta os resultados obtidos.

Tabela 9.29 – Variáveis dos imóveis

Preço (R\$)	Distância (km)	Dormitórios	Área (m ²)	Renda mensal (R\$)
107.135	1	2	94	3.537
107.750	2	2	96	3.174
108.573	2	3	116	3.072
99.151	3	4	149	2.683
85.663	3	2	98	2.512
80.614	3	3	115	2.580
74.624	4	2	93	2.031
64.195	5	3	119	1.549
40.950	6	4	142	1.104
82.479	4	2	93	2.119
41.926	6	3	122	1.068
20.386	7	1	72	549
48.141	6	1	72	1.043
30.062	7	2	97	671
65.520	5	4	148	1.521

6. Dados os resultados da estimativa do modelo de regressão da Tabela 9.30, realizada com uma amostra com 25 observações:

Tabela 9.30 – Modelo de regressão para uma amostra de 25 observações

	Coeficiente	Desvio padrão
Constante	123,4	11,56
X_1	-12,43	11,41
X_2	0,89	0,77

- a) Teste a significância dos parâmetros.
- b) Teste a validade da regressão.
- c) Comente os resultados.

7. Com os dados da Tabela 9.31, estime o consumo em função da taxa de juros e da renda. Teste a existência de autocorrelação e, se for o caso, estime de novo o modelo, corrigindo o problema.

Tabela 9.31 – Consumo em função da taxa de juros e da renda

Ano	Juros	Renda	Consumo
1974	11	500	409,0
1975	12	550	440,9
1976	13	540	424,5
1977	9	580	494,2
1978	8	530	468,2
1979	7	500	451,0
1980	14	510	385,4
1981	16	520	366,1
1982	18	550	361,2
1983	14	570	424,2
1984	13	580	445,8
1985	11	590	471,2
1986	10	610	488,1
1987	7	620	526,5
1988	5	630	561,7
1989	8	650	549,7
1990	9	660	550,1
1991	11	650	517,5
1992	12	630	482,2
1993	11	610	482,3
1994	10	600	478,3
1995	9	620	496,6
1996	7	630	534,9
1997	9	620	514,1

8. Use o teste de White para verificar se há heterocedasticidade no Exemplo 9.3.2.1.
 9. No Exemplo 9.3.2.2, suponha que sejam dadas as populações das cidades:

Tabela 9.32 – População das cidades do Exemplo 9.3.2.2

Cidade	População (em 1.000 hab.)
A	100
B	120
C	130
D	140
E	160
F	210
G	250
H	340
I	450
J	570
K	620
L	800
M	950
N	1.020
O	1.300
P	1.400
Q	1.600

Use o teste de Goldfeld-Quandt para testar a heterocedasticidade desse modelo, usando a população como “separador”.

10. Ainda no Exemplo 9.3.2.2, faça uma estimativa corrigindo o problema da heterocedasticidade, admitindo-se que a variância (ou o desvio padrão) seja proporcional à população da cidade.

11. Suponha um modelo de oferta e demanda dado por:

$$Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 P_{t-1} + \mu_t \quad (\text{oferta})$$

$$Q_t = \beta_0 + \beta_1 P_t + \beta_2 R_t + v_t \quad (\text{demanda})$$

Sendo Q as quantidades, P o preço e R a renda, classifique cada equação em relação à identificação.

12. No Exemplo 9.4.1.1, classifique as demais equações em relação à identificação.

13. No Exemplo 9.4.2.1, suponha que a variável *salários* não tenha sido dada. Estime esse novo modelo por mínimos quadrados indiretos e mínimos quadrados de dois estágios e comente os resultados.

14. Considere o seguinte modelo de equações simultâneas:

$$Y_{1t} - \Phi_2 Y_{2t} = \gamma_{11} x_{1t} + u_{1t} \quad (\text{Equação 1})$$

$$Y_{2t} - \Phi_3 Y_{3t} = \gamma_{22} x_{2t} + u_{2t} \quad (\text{Equação 2})$$

$$Y_{3t} - \Phi_4 Y_{1t} = \gamma_{31} x_{1t} + \gamma_{32} x_{2t} + u_{3t} \quad (\text{Equação 3})$$

Em que as variáveis y são endógenas e as variáveis x são exógenas. Verifique as condições de ordem e de posto para cada equação.

15. No exemplo 9.4.1.1, foi vista a condição de posto para a equação 4. Verifique a condição de posto para as equações 2 e 3.

16. Com os dados do Exemplo 9.2.2.1, utilize o teste de Breusch-Godfrey para testar autocorrelação de segunda ordem. Corrija o modelo de novo, se indicado pelo resultado do teste.

17. No modelo a seguir:

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 Y_{t-1} + \varepsilon_t$$

$$\varepsilon_t = \rho \varepsilon_{t-1} + \mu_t$$

Indique as propriedades dos estimadores de mínimos quadrados e a validade dos testes de hipóteses nos casos em que:

- a) $\beta_2 \neq 0$ e $\rho = 0$.
- b) $\beta_2 = 0$ e $\rho \neq 0$.
- c) $\beta_2 \neq 0$ e $\rho \neq 0$.

18. Assinale verdadeiro ou falso:

- a) Quando há correlação entre as variáveis, ainda que não perfeita, embora a estimativa seja possível, devem ser feitas necessariamente as devidas correções.
- b) Como as variâncias são maiores quando há multicolinearidade, isso implica que os estimadores não são eficientes.

- c) Se os coeficientes da regressão apresentam desvios padrão muito altos, certamente há multicolinearidade.
- d) A multicolinearidade é mais um problema numérico, com os dados, do que um problema no modelo propriamente dito.
- e) Na presença de autocorrelação nos resíduos, o estimador de mínimos quadrados ordinários será sempre não viesado.
- f) Na presença de heterocedasticidade, o estimador de mínimos quadrados ordinários será viesado.
- g) Na presença de autocorrelação nos resíduos, o estimador de mínimos quadrados ordinários será eficiente.
- h) Na presença de heterocedasticidade, o estimador de mínimos quadrados ordinários será eficiente.
- i) Com o teste de Durbin-Watson, é sempre possível testar autocorrelação, desde que os erros sigam um processo do tipo AR(1).
- j) O método dos mínimos quadrados ponderados é recomendado quando há heterocedasticidade.
- k) Havendo simultaneidade, o estimador de mínimos quadrados ordinários é não viesado, porém consistente.
- l) O método dos mínimos quadrados indiretos e de dois estágios produzem estimadores não viesados.

Apêndice 9.B – O método dos mínimos quadrados generalizados

Você já viu que as hipóteses IV e V são:

$$\text{IV. } \text{var}(\varepsilon_i) = \sigma^2 \text{ (constante).}$$

$$\text{V. } E(\varepsilon_i \varepsilon_j) = 0, i \neq j \text{ (erros não são autocorrelacionados).}$$

Essas hipóteses podem ser resumidas, em notação matricial, como:

$$\text{var}(\varepsilon) = \sigma^2 \mathbf{I}$$

Um modelo que *não* siga essas hipóteses pode ter como matriz de variância e covariância do vetor de erros uma matriz qualquer, chamada Ω .

$$\text{var}(\varepsilon) = \Omega$$

Você já sabe que o estimador de mínimos quadrados, nessas condições, é ineficiente. Para encontrar um estimador eficiente para essa situação, suponha uma matriz \mathbf{T} tal que:

$$\mathbf{T} \Omega \mathbf{T}' = \mathbf{I}$$

Expressão que também pode ser escrita assim:

$$\mathbf{T}' \mathbf{T} = \Omega^{-1}$$

O modelo de regressão linear, em notação matricial, é:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

Multiplicando a equação por T:

$$\mathbf{TY} = \mathbf{TX}\boldsymbol{\beta} + \mathbf{T}\boldsymbol{\epsilon}$$

A variância dos erros desse novo modelo pode ser escrita como:

$$\text{var}(\boldsymbol{\epsilon}) = E(\mathbf{T}\boldsymbol{\epsilon}\boldsymbol{\epsilon}'\mathbf{T}') = \mathbf{T}\Omega\mathbf{T}' = \mathbf{I}$$

Que é um caso particular da hipótese usual (em que $\sigma^2 = 1$). Portanto, o modelo transformado pode ser estimado por mínimos quadrados ordinários. O estimador usual de mínimos quadrados ordinários é:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

Porém, nesse modelo transformado, não há \mathbf{X} e \mathbf{Y} , mas \mathbf{TX} e \mathbf{TY} , então:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{T}'\mathbf{TX})^{-1}\mathbf{X}'\mathbf{T}'\mathbf{TY}$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-1}\mathbf{Y}$$

Esse estimador, por levar em conta um caso mais geral em que pode haver autocorrelação e/ou heterocedasticidade, é conhecido por estimador de **mínimos quadrados generalizados**.

Não é uma grande panaceia, entretanto, pois em geral exige-se o conhecimento da estrutura da matriz Ω . Estimá-la não é uma solução viável, pois é uma matriz quadrada de ordem n , o que significa que, numa amostra com n observações, haveria n^2 elementos da matriz para serem estimados.

Nos casos vistos neste capítulo – por exemplo, uma heterocedasticidade em que a variância dos erros seja dada por $z_i\sigma^2$, com os valores de z conhecidos –, a matriz Ω é dada por:

$$\Omega = \sigma^2 \begin{bmatrix} z_1 & 0 & \dots & 0 \\ 0 & z_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & z_n \end{bmatrix}$$

Da mesma forma, se houver autocorrelação representada por um processo autorregressivo de ordem 1, com coeficiente de correlação ρ , a matriz Ω é dada por:

$$\Omega = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n-3} \\ \dots & \dots & \dots & \dots & \dots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{bmatrix}$$

Assim, conhecidos os padrões da heterocedasticidade, ou da autocorrelação, ou de ambas, pode-se montar a matriz Ω , fazer diretamente a estimativa por mínimos quadrados generalizados e obter um estimador que tenha variância mínima.



Capítulo 10

SÉRIES DE TEMPO

Este capítulo é uma introdução ao tratamento de séries temporais e, a partir delas, à previsão de valores futuros de uma variável a partir de valores passados da mesma variável.

10.1 Métodos ingênuos de previsão

O método mais simples de fazer a previsão de uma variável é o que usa justamente o último valor da variável. Por exemplo, o valor de uma ação nos últimos cinco dias foi: 23, 22, 25, 24 e 23. Então, a previsão para o valor da ação será 23, que é justamente o último valor da série.

O pressuposto desse método de previsão, na verdade, não é tão ingênuo assim. No entanto, ele só será útil se o comportamento da variável seguir o seguinte modelo:

$$y_t = y_{t-1} + \epsilon_t$$

Ou seja, o valor da variável no período t é o valor que ela possuía no período $t - 1$ mais um componente de erro. Esse processo é conhecido como **random walk** ou, traduzindo, **passeio aleatório**.

O termo de erro tem, eventualmente, as mesmas características do erro do modelo de regressão linear (homocedástico, não autocorrelacionado etc.), mas, principalmente, tem média zero. Vale dizer que não é um componente sistemático, mas aleatório, que pode subir, descer (ser positivo ou negativo), ao sabor do acaso. É um componente, por suas características, imprevisível.

Assim, a melhor forma de prever y_t é mesmo por meio do valor de y_{t-1} . De fato, se for aplicado o operador esperança na equação anterior:

$$\begin{aligned} E(y_t) &= E(y_{t-1} + \epsilon_t) \\ E(y_t) &= E(y_{t-1}) + E(\epsilon_t) \end{aligned}$$

Como y_{t-1} já é conhecido e o termo erro tem média zero:

$$\begin{aligned} E(y_t) &= y_{t-1} + 0 \\ E(y_t) &= y_{t-1} \end{aligned}$$

Portanto, a melhor previsão para y_t é realmente y_{t-1} ; isso, claro, se a variável y_t tiver o comportamento de um passeio aleatório.

10.2 Séries estacionárias e regressão espúria

Uma série que segue um comportamento como o do item anterior, isto é:

$$y_t = y_{t-1} + \epsilon_t$$

é dita **não estacionária**, porque, se em dado período ocorre um *choque*, que será dado por um valor de ϵ_t diferente de zero, esse valor fica incorporado eternamente nos valores futuros da variável y_t . Se o processo, no entanto, for dado por:

$$y_t = 0,8y_{t-1} + \epsilon_t$$

um choque que ocorra em determinado ano será amortecido nos anos seguintes. Suponha que o valor de y_t vinha sendo zero, até que, em 1990, houve um choque positivo $\epsilon_t = 20$, isto é, nesse ano, o valor de y_t foi 20. O que ocorrerá nos anos seguintes, admitindo que ϵ_t seja igual a zero para os demais anos?

$$y_{1988} = 0$$

$$y_{1989} = 0,8y_{1988} + \epsilon_{1989} = 0 + 0 = 0$$

$$y_{1990} = 0,8y_{1989} + \epsilon_{1990} = 0 + 20 = 20$$

$$y_{1991} = 0,8y_{1990} + \epsilon_{1991} = 0,8 \times 20 + 0 = 16$$

$$y_{1992} = 0,8y_{1991} + \epsilon_{1992} = 0,8 \times 16 + 0 = 12,8$$

$$y_{1993} = 0,8y_{1992} + \epsilon_{1993} = 0,8 \times 12,8 + 0 = 10,24$$

$$y_{1994} = 0,8y_{1993} + \epsilon_{1994} = 0,8 \times 10,24 + 0 = 8,192$$

$$y_{1995} = 0,8y_{1994} + \epsilon_{1995} = 0,8 \times 8,192 + 0 = 6,5536$$

$$y_{1996} = 0,8y_{1995} + \epsilon_{1996} = 0,8 \times 6,5536 + 0 = 5,24288$$

E assim sucessivamente. Verifique que y_t tende a voltar para seu valor *histórico* (zero), pois o efeito do choque é dissipado ao longo dos anos, o que não ocorre com o passeio aleatório. A série é dita, portanto, estacionária.

Mais precisamente, uma série é dita estacionária¹ se acontecer:

$$E(y_t) = \text{constante}$$

$$\text{var}(y_t) = \text{constante}$$

E a $\text{cov}(y_t, y_{t-s})$, $s \neq 0$, só depende do valor de s , isto é, só depende do tamanho da defasagem, mas não do período t . Por exemplo:

$$\text{cov}(y_{1998}, y_{1996}) = \text{cov}(y_{1997}, y_{1995}) = \text{cov}(y_{1996}, y_{1994}) = \dots$$

¹ A definição apresentada é para as chamadas séries fracamente estacionárias. A definição de séries fortemente estacionárias inclui séries que possuem média ou variância infinitas. Com isso em vista, o mais correto deveria ser $E(y_t | y_{t-1})$, ou seja, a esperança de y_t dado y_{t-1} , já que esse é conhecido.

Mais adiante você verá como testar se uma série é ou não estacionária. Para o processo apresentado:

$$y_t = 0,8y_{t-1} + \varepsilon_t$$

Tem-se que:

$$E(y_t) = E(0,8y_{t-1} + \varepsilon_t)$$

$$E(y_t) = E(0,8y_{t-1}) + E(\varepsilon_t)$$

$$E(y_t) = 0,8E(y_{t-1}) + E(\varepsilon_t)$$

Como a série é estacionária e $E(\varepsilon_t) = 0$:

$$E(y_t) = 0,8E(y_{t-1}) + 0$$

$$0,8E(y_{t-1}) = 0$$

$$E(y_t) = 0$$

A média do processo é zero. É claro que, para ser estacionária, a série não precisa ter média zero, basta ser constante. Um processo semelhante com média diferente de zero é dado por:

$$y_t = y_0 + 0,8y_{t-1} + \varepsilon_t$$

Verifique que, nesse caso, a média do processo é dada por:

$$E(y_t) = 5y_0$$

A variância é dada por:

$$\text{var}(y_t) = \text{var}(0,8y_{t-1} + \varepsilon_t)$$

$$\text{var}(y_t) = \text{var}(0,8y_{t-1}) + \text{var}(\varepsilon_t)$$

$$\text{var}(y_t) = 0,64\text{var}(y_{t-1}) + \text{var}(\varepsilon_t)$$

De novo, sendo a série estacionária e $\text{var}(\varepsilon_t) = \sigma^2$:

$$\text{var}(y_t) = 0,64\text{var}(y_{t-1}) + \sigma^2$$

$$0,36\text{var}(y_t) = \sigma^2$$

$$\text{var}(y_t) = \frac{1}{0,36} \sigma^2$$

$$\text{var}(y_t) \approx 2,77\sigma^2$$

Alguma atenção especial deve ser dada a séries que *não* são estacionárias, especialmente quando se quer fazer uma regressão entre elas, como no exemplo a seguir.

Exemplo 10.2.1

A Tabela 10.1 mostra o percentual de residências atendidas por serviços de saneamento básico na Meltávia e as exportações de trigo do Kazimenistão em milhares de toneladas. Estime a regressão com as exportações de trigo como variável dependente e o percentual de residências com esgoto como variável independente.

Tabela 10.1 – Saneamento básico de Meltávia e trigo do Kazimenistão

Ano	% de residências atendidas por saneamento básico (X)	Exportações de trigo (Y)
1971	21,15	183,6
1972	22,5	198,0
1973	24,3	234,0
1974	27,9	252,0
1975	30,6	271,8
1976	32,4	291,6
1977	35,1	316,8
1978	36,9	336,6
1979	39,6	361,8
1980	41,4	379,8
1981	43,2	394,2
1982	45,9	415,8
1983	48,6	439,2
1984	51,3	460,8
1985	54,9	500,4
1986	56,7	518,4
1987	57,6	532,8
1988	60,3	558,0
1989	63,9	577,8
1990	64,8	613,8
1991	67,5	666,0
1992	68,4	685,8
1993	69,3	709,2
1994	70,2	739,8
1995	72,0	757,8
1996	72,9	795,6
1997	74,7	820,8
1998	77,4	840,6
1999	78,3	865,8
2000	79,2	882,0

Os resultados da regressão foram:

$$Y = -93,64 + 11,59 X$$

(20,08) (0,36)

$R^2 = 0,9739$

$F = 1.043,8$

$DW = 0,1336$

Os valores entre parênteses são os desvios padrão.

O resultado da regressão foi, em princípio, excepcional. As estatísticas t foram muito altas, especialmente para o coeficiente da variável X (32,3!), mostrando que ele é altamente significante. O R^2 é próximo de 1 e o valor calculado de F também foi muito alto.

O ministro da Agricultura do Kazimenistão, ao ser informado desses resultados, deveria tomar providências para estimular a expansão do serviço de saneamento básico na Meltávia, pois isso aparentemente tem forte efeito sobre as exportações de trigo de seu país.

É claro que isso é absurdo. Apesar dos resultados aparentemente muito bons, não é possível que o número de casas atendidas por saneamento básico na Meltávia tenha algum efeito sobre as exportações do Kazimenistão, quanto mais ser tão determinante quanto indicam os resultados obtidos.

Há uma dica de que alguma coisa está errada: a estatística de Durbin-Watson encontrada foi muito próxima de zero, indicando a presença de autocorrelação positiva nos erros.

Observe o comportamento das duas variáveis nas Figuras 10.1 e 10.2.

Figura 10.1 – Evolução do percentual de residências com saneamento na Meltávia

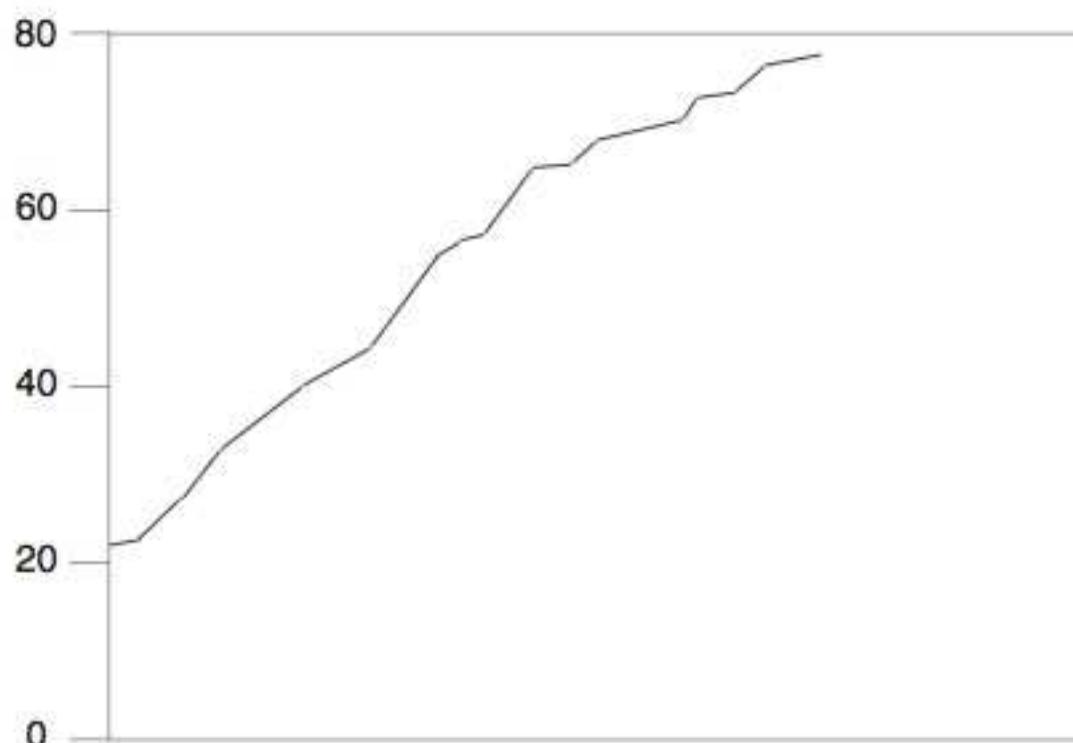
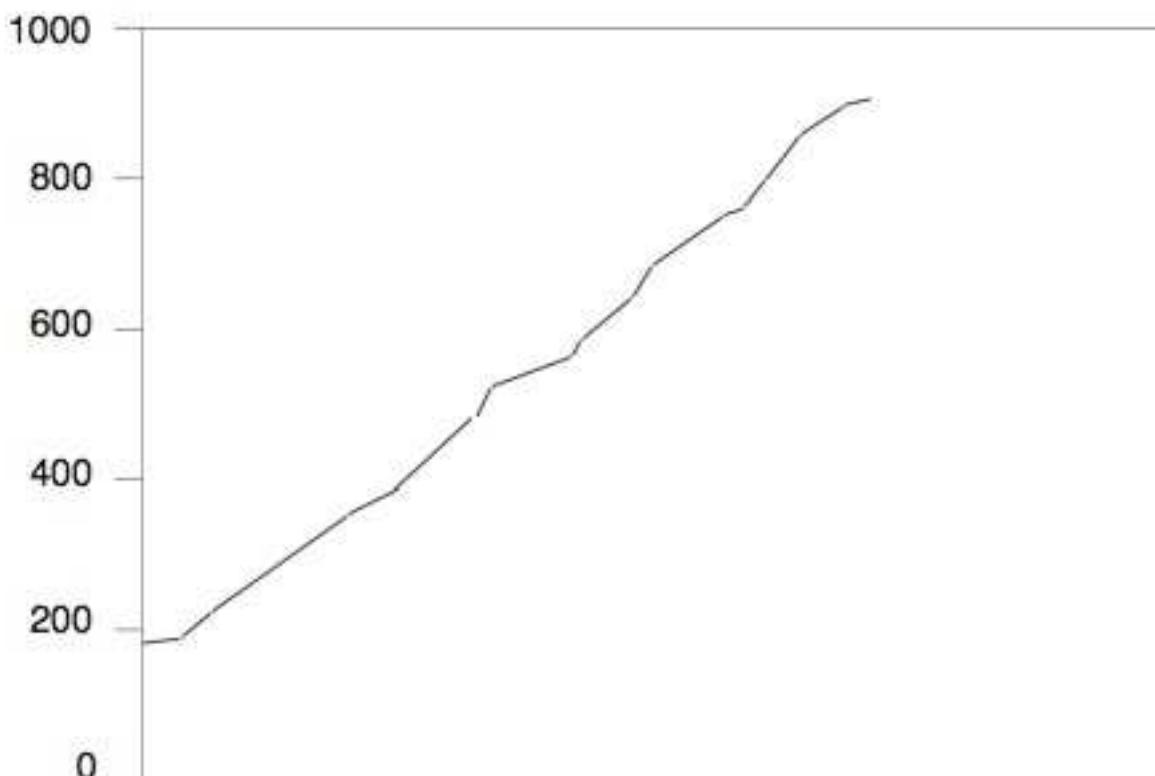


Figura 10.2 – Evolução das exportações de trigo no Kazimenistão



Pelos gráficos das Figuras 10.1 e 10.2, pode-se perceber que ambas as variáveis *não* são estacionárias, e o resultado obtido, na verdade, é típico de quando se faz uma regressão utilizando duas variáveis não estacionárias. Mesmo que uma variável não tenha nada que ver com a outra, o R^2 será muito próximo de 1, as estatísticas t e F serão muito grandes (mas, evidentemente, não terão nenhum significado; há exceções, como você verá ao longo deste capítulo) e a estatística DW será próxima de zero. Esse tipo de regressão é conhecido como **regressão espúria**.

10.3 Procedimento de Box e Jenkins (modelos ARIMA)

O procedimento de Box e Jenkins² consiste em explicar uma variável por meio de valores passados dela mesma e de valores passados de choques. Como nenhuma outra variável está explicitamente envolvida no modelo, esse é chamado de **univariado**.

10.3.1 Modelos

Em uma classe dos modelos de Box e Jenkins, a variável é explicada unicamente por valores passados dela mesma, como este:

$$y_t = \theta y_{t-1} + \varepsilon_t$$

Nesse caso, o intercepto pode ou não ser incluído, como você viu, dependendo da média do processo ser (ou não) zero.

Esse processo é uma regressão dessa variável por ela mesma e constitui um processo **autorregressivo**. Como existe uma defasagem da variável, é um processo autorregressivo de ordem 1 ou AR(1).

O erro ε_t representa os choques que podem ocorrer sobre a variável y_t e tem todas as características das hipóteses básicas de um modelo de regressão linear, ou seja, ele mesmo é um processo estacionário com média zero, com o detalhe de não apresentar autocorrelação. Um processo desse tipo é conhecido como **ruído branco**.

É possível ter também um processo AR(2):

$$y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \varepsilon_t$$

Ou mesmo um processo autorregressivo de qualquer ordem, por exemplo, um AR(p):

$$y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \dots + \theta_p y_{t-p} + \varepsilon_t$$

Pode-se escrever esse processo de maneira mais resumida se utilizar o operador L (do inglês *last*). Alguns autores utilizam B , de *back*), definido da seguinte forma:

$$Ly_t = y_{t-1}$$

$$L^2y_t = LLy_t = Ly_{t-1} = y_{t-2}$$

$$L^n y_t = y_{t-n}$$

² Esse nome é dado a uma série de processos que foram sintetizados por Box e Jenkins (1976).

Dessa forma, o processo AR(p) pode ser escrito assim:

$$y_t - \theta_1 y_{t-1} - \theta_2 y_{t-2} - \dots - \theta_p y_{t-p} = \epsilon_t$$

$$y_t - \theta_1 L y_t - \theta_2 L^2 y_t - \dots - \theta_p L^p y_t = \epsilon_t$$

Embora o operador L não seja um número (ele, sozinho, não vale nada), pode ser tratado algebricamente como se fosse um número. Colocando-se y_t em evidência:

$$(1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_p L^p) y_t = \epsilon_t$$

Chega-se a um polinômio de ordem p na “variável” L , que pode ser chamado simplesmente de $\Theta_p(L)$. Assim:

$$\Theta_p(L) = 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_p L^p$$

Então, pode-se escrever o modelo do tipo AR(p) de maneira mais sintética, como:

$$\Theta_p(L) y_t = \epsilon_t$$

Uma forma diferente é quando o processo é uma combinação de choques passados:

$$y_t = \epsilon_t - \varphi \epsilon_{t-1}$$

Nesse caso, a variável y_t é uma combinação de um choque presente com *um* choque passado, especificamente um choque ocorrido no período imediatamente anterior. Esse processo é conhecido como de **médias móveis**, nesse caso, de ordem 1, o que é abreviado por MA(1), do inglês *moving average*.

Um processo MA(2) seria dado por:

$$y_t = \epsilon_t - \varphi_1 \epsilon_{t-1} - \varphi_2 \epsilon_{t-2}$$

E um processo de médias móveis de ordem qualquer, por exemplo, um MA(q), seria assim:

$$y_t = \epsilon_t - \varphi_1 \epsilon_{t-1} - \varphi_2 \epsilon_{t-2} - \dots - \varphi_q \epsilon_{t-q}$$

Da mesma forma que em um processo autorregressivo pode-se utilizar o operador L :

$$y_t = \epsilon_t - \varphi_1 L \epsilon_t - \varphi_2 L^2 \epsilon_t - \dots - \varphi_q L^q \epsilon_t$$

Colocando ϵ_t em evidência:

$$y_t = \epsilon_t (1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_q L^q)$$

De novo, chega-se a um polinômio em L , dessa vez de ordem q , denominado $\Phi_q(L)$:

$$\Phi_q(L) = 1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_q L^q$$

O processo MA(q) pode ser escrito como se segue:

$$y_t = \Phi_q(L) \epsilon_t$$

É possível, ainda, ter processos que são combinações de processos autorregressivos e de médias móveis, como:

$$y_t = \theta y_{t-1} + \epsilon_t - \varphi \epsilon_{t-1},$$

que é uma combinação de um processo autorregressivo de ordem 1 e de médias móveis de ordem 1, conhecido como ARMA(1,1), sendo o primeiro número a ordem do AR e o segundo a ordem do MA.

Assim, um ARMA(2,3) é dado por:

$$y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \varepsilon_t - \varphi_1 \varepsilon_{t-1} - \varphi_2 \varepsilon_{t-2} - \varphi_3 \varepsilon_{t-3}$$

Genericamente, um ARMA(p,q) seria:

$$y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \dots + \theta_p y_{t-p} + \varepsilon_t - \varphi_1 \varepsilon_{t-1} - \varphi_2 \varepsilon_{t-2} - \dots - \varphi_q \varepsilon_{t-q}$$

Ou ainda:

$$y_t - \theta_1 y_{t-1} - \theta_2 y_{t-2} - \dots - \theta_p y_{t-p} = \varepsilon_t - \varphi_1 \varepsilon_{t-1} - \varphi_2 \varepsilon_{t-2} - \dots - \varphi_q \varepsilon_{t-q}$$

$$y_t - \theta_1 L y_t - \theta_2 L^2 y_t - \dots - \theta_p L^p y_t = \varepsilon_t - \varphi_1 L \varepsilon_t - \varphi_2 L^2 \varepsilon_t - \dots - \varphi_q L^q \varepsilon_t$$

$$(1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_p L^p) y_t = \varepsilon_t (1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_q L^q)$$

Ou, simplesmente:

$$\Theta_p(L) y_t = \Phi_q(L) \varepsilon_t$$

Ainda é necessário prestar atenção a um detalhe: se, nesses processos, a variável é explicada por valores passados dela mesma (e/ou choques passados), convém que ela seja uma variável estacionária.

Quando a variável y_t não é estacionária, pode-se tentar definir uma nova variável z_t , como sendo a primeira diferença de y_t , isto é:

$$z_t = y_t - y_{t-1} = \Delta y_t$$

Se y_t não é estacionária, mas z_t é, diz-se que y_t é integrada de ordem 1, ou I(1). É uma ideia semelhante à do cálculo integral, porém em termos discretos, pois y_t é obtido a partir da soma de z_t . Às vezes, tomar a primeira diferença não é suficiente e, para obter uma variável estacionária, é preciso tomar a segunda diferença (a diferença da diferença), ou seja:

$$z_t = \Delta^2 y_t = \Delta(\Delta y_t) = \Delta y_t - \Delta y_{t-1}$$

Note que $\Delta = 1 - L$. Se só assim se obtém uma variável estacionária, então y_t é dita integrada de ordem 2, I(2).

Tome quantas diferenças forem necessárias até obter uma variável estacionária. Se forem d diferenças, então y_t é dita I(d). Evidentemente, uma variável dita I(0) é uma variável estacionária.

Se y_t não é uma variável estacionária, mas sua d -ésima diferença é, então:

$$z_t = \Delta^d y_t$$

Se essa variável z_t segue um processo ARIMA(p,q), isto é:

$$z_t = \theta_1 z_{t-1} + \theta_2 z_{t-2} + \dots + \theta_p z_{t-p} + \varepsilon_t - \varphi_1 \varepsilon_{t-1} - \varphi_2 \varepsilon_{t-2} - \dots - \varphi_q \varepsilon_{t-q},$$

então y_t segue um processo ARIMA(p,d,q), em que a letra I do meio e o número d se referem à ordem de integração. Isto é, y_t é integrada de ordem d e sua d -ésima diferença segue um processo combinado autorregressivo (de ordem p) e de médias móveis (de ordem q). O processo para y_t será dado por:

$$\Delta^d y_t = \theta_1 \Delta^d y_{t-1} + \theta_2 \Delta^d y_{t-2} + \dots + \theta_p \Delta^d y_{t-p} + \varepsilon_t - \varphi_1 \varepsilon_{t-1} - \varphi_2 \varepsilon_{t-2} - \dots - \varphi_q \varepsilon_{t-q}$$

Exemplo 10.3.1.1

Suponha que uma variável y_t segue um processo ARIMA(1,1,2). Escreva esse processo em sua forma analítica.

A variável y_t é integrada de ordem 1, ou seja, é I(1). Portanto, a variável z_t é dada por:

$$z_t = \Delta y_t$$

É estacionária e segue um processo ARIMA(1,2), ou seja:

$$z_t = \theta z_{t-1} + \epsilon_t - \varphi_1 \epsilon_{t-1} - \varphi_2 \epsilon_{t-2}$$

Portanto:

$$\Delta y_t = \theta \Delta y_{t-1} + \epsilon_t - \varphi_1 \epsilon_{t-1} - \varphi_2 \epsilon_{t-2}$$

10.3.2 Identificação dos modelos ARIMA

Antes de estimar um modelo ARIMA, é preciso descobrir (ou, pelo menos, ter uma boa ideia) qual é o processo a ser estimado. Isso é feito por meio das funções de **autocorrelação (FAC)** e **autocorrelação parcial (FACP)**.

Veja o comportamento dessas funções para um AR(1). Isto é, suponha que o processo seja do tipo:

$$y_t = \theta y_{t-1} + \epsilon_t$$

Sendo a variável estacionária, a covariância (e, portanto, o coeficiente de correlação) entre a variável e os valores defasados dela mesma é constante, se for dado o número de defasagens. Portanto, há um valor para a autocorrelação para cada número de defasagens, isto é:

$$\rho_1 = \text{corr}(y_t, y_{t-1})$$

$$\rho_2 = \text{corr}(y_t, y_{t-2})$$

...

$$\rho_k = \text{corr}(y_t, y_{t-k})$$

Como você sabe, o coeficiente de correlação é dado por:

$$\rho_k = \text{corr}(y_t, y_{t-k}) = \frac{\text{cov}(y_t, y_{t-k})}{\sqrt{\text{var}(y_t) \text{var}(y_{t-k})}} = \frac{\text{cov}(y_t, y_{t-k})}{\sqrt{\text{var}(y_t) \text{var}(y_t)}} = \frac{\text{cov}(y_t, y_{t-k})}{\text{var}(y_t)}$$

Já que, em se tratando de uma variável estacionária, a variância é constante.

Fazendo:

$$\gamma_k = \text{cov}(y_t, y_{t-k}) \quad \text{e}$$

$$\gamma_0 = \text{var}(y_t)$$

Então:

$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

A variância de y_t é dada por:

$$\begin{aligned}\text{var}(y_t) &= \text{var}(\theta y_{t-1} + \epsilon_t) \\ \text{var}(y_t) &= \text{var}(\theta y_{t-1}) + \text{var}(\epsilon_t) \\ \text{var}(y_t) &= \theta^2 \text{var}(y_{t-1}) + \text{var}(\epsilon_t) \\ \text{var}(y_t) &= \theta^2 \text{var}(y_t) + \text{var}(\epsilon_t) \\ (1-\theta^2)\text{var}(y_t) &= \sigma^2 \\ \gamma_k = \text{var}(y_t) &= \frac{\sigma^2}{1-\theta^2}\end{aligned}$$

Então, para saber como se comporta a função de autocorrelação, basta saber como se comporta a autocovariância, isto é, $\gamma_1, \gamma_2, \gamma_3$ etc.

$$\gamma_k = \text{cov}(y_t, y_{t-k}) = E(y_t y_{t-k}) - E(y_t)E(y_{t-k})$$

Como o processo tem média zero:

$$\gamma_k = E(y_t y_{t-k})$$

Portanto:

$$\gamma_1 = E(y_t y_{t-1})$$

Levando em conta que:

$$\begin{aligned}y_t &= \theta y_{t-1} + \epsilon_t \\ y_{t-1} &= \theta y_{t-2} + \epsilon_{t-1}\end{aligned}$$

Então:

$$\begin{aligned}\gamma_1 &= E(y_t y_{t-1}) = E[(\theta y_{t-1} + \epsilon_t) y_{t-1}] \\ \gamma_1 &= E[\theta y_{t-1}^2 + \epsilon_t y_{t-1}] \\ \gamma_1 &= E(\theta y_{t-1}^2) + E(\epsilon_t y_{t-1}) \\ \gamma_1 &= \theta E(y_{t-1}^2) + 0 \\ \gamma_1 &= \theta \text{var}(y_t) = \theta \gamma_0\end{aligned}$$

Assim sendo:

$$\rho_1 = \theta$$

O mesmo procedimento será feito para γ_2 :

$$\begin{aligned}\gamma_2 &= E(y_t y_{t-2}) \\ \gamma_2 &= E[(\theta y_{t-1} + \epsilon_t) y_{t-2}] \\ \gamma_2 &= E[(\theta(\theta y_{t-2} + \epsilon_{t-1}) + \epsilon_t) y_{t-2}] \\ \gamma_2 &= E[\theta^2 y_{t-2}^2 + \theta \epsilon_{t-1} y_{t-2} + \epsilon_t y_{t-2}] \\ \gamma_2 &= E(\theta^2 y_{t-2}^2) + E(\theta \epsilon_{t-1} y_{t-2}) + E(\epsilon_t y_{t-2}) \\ \gamma_2 &= \theta^2 E(y_{t-2}^2) + \theta E(\epsilon_{t-1} y_{t-2}) + E(\epsilon_t y_{t-2}) \\ \gamma_2 &= \theta^2 \text{var}(y_t) + 0 + 0 \\ \gamma_2 &= \theta^2 \gamma_0\end{aligned}$$

Portanto:

$$\rho_2 = \theta^2$$

Como θ é menor do que 1, em módulo (porque caso contrário a série não seria estacionária), θ^2 é menor do que θ (em módulo). É fácil ver que os valores seguintes para a função de autocorrelação serão θ^3 , θ^4 etc., de modo que a função de autocorrelação de um processo AR(1) será declinante. Isso, entretanto, não é suficiente para identificar o processo como AR(1).

O conceito de correlação parcial se refere à correlação entre duas variáveis eliminando o efeito de outras variáveis, o que é feito por meio de uma regressão. De fato, a função de autocorrelação parcial é dada pelos coeficientes ϕ_1, ϕ_2, ϕ_3 etc., que são encontrados como segue.

O coeficiente ϕ_1 é encontrado na regressão a seguir:

$$y_t = \alpha + \phi_1 y_{t-1} + v_t$$

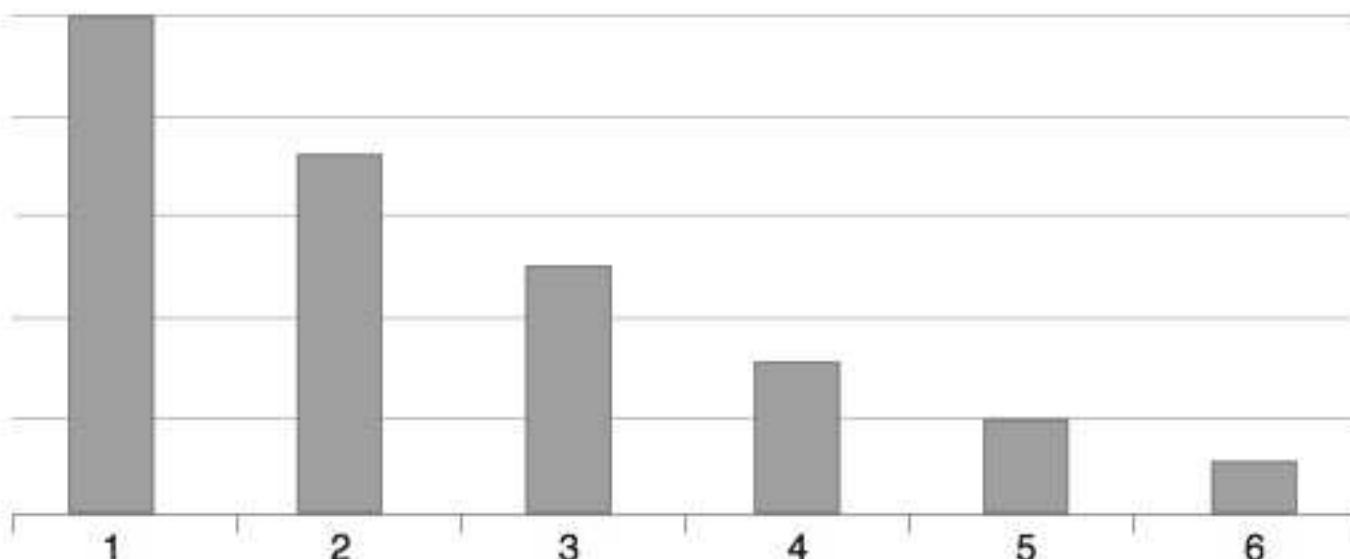
Enquanto o coeficiente ϕ_2 é o correspondente estimado pela seguinte regressão:

$$y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + v_t$$

E assim sucessivamente. É fácil ver que, se o processo é AR(1), o coeficiente ϕ_2 não existe (não será significante numa regressão). De modo geral, num AR(p) $\phi_k \neq 0$ para k menor ou igual a p e $\phi_k = 0$ para valores maiores do que k .

Portanto, um processo autorregressivo apresenta função de autocorrelação declinante³ e a função de autocorrelação parcial truncada exatamente na ordem do processo.

Figura 10.3 – Função de autocorrelação de um AR(p) – qualquer p



³ Aqui só foi feito para AR(1), mas o resultado pode ser generalizado.

Figura 10.4 – Função de autocorrelação parcial de um AR(1)

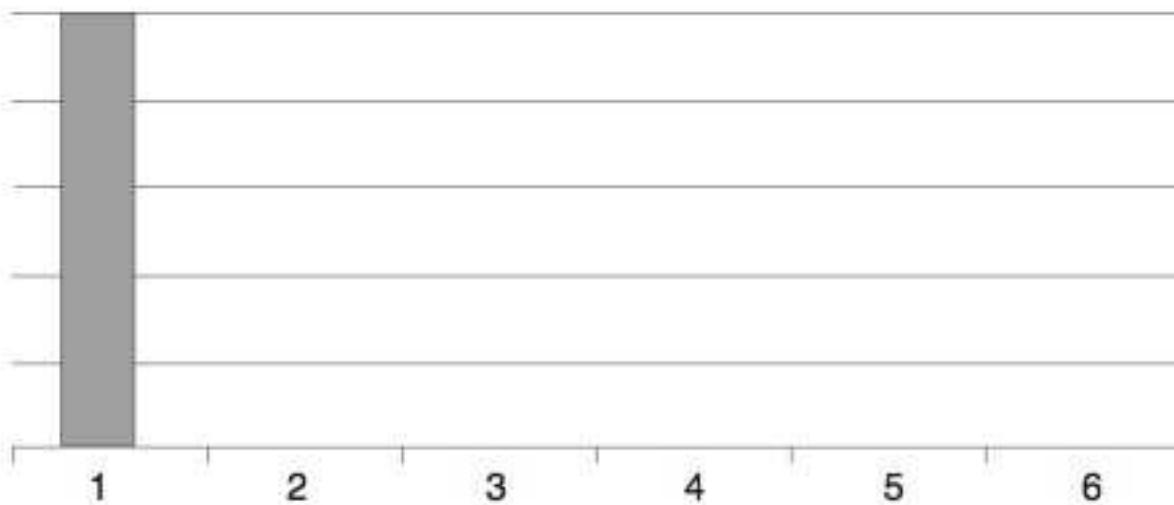
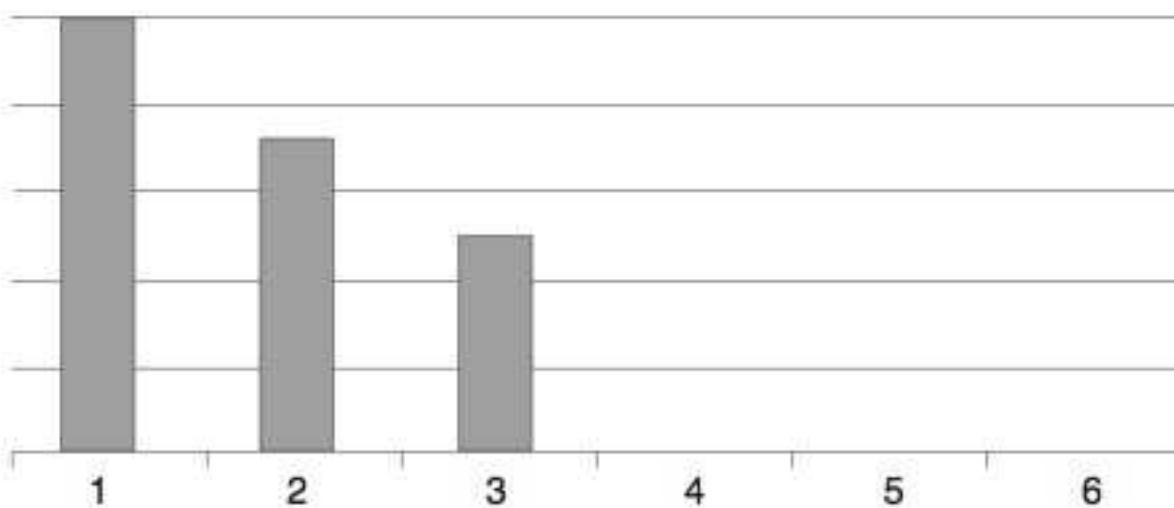


Figura 10.5 – Função de autocorrelação parcial de um AR(3)



Veja o comportamento dessas duas funções para um MA(1).

$$y_t = \varepsilon_t - \varphi \varepsilon_{t-1}$$

A variância é dada por:

$$\begin{aligned} \text{var}(y_t) &= \text{var}(\varepsilon_t - \varphi \varepsilon_{t-1}) \\ \text{var}(y_t) &= \text{var}(\varepsilon_t) + \text{var}(\varphi \varepsilon_{t-1}) \\ \text{var}(y_t) &= \text{var}(\varepsilon_t) + \varphi^2 \text{var}(\varepsilon_{t-1}) \\ \text{var}(y_t) &= \text{var}(\varepsilon_t) + \varphi^2 \text{var}(\varepsilon_t) \\ \text{var}(y_t) &= (1 + \varphi^2)\text{var}(\varepsilon_t) \\ \text{var}(y_t) &= (1 + \varphi^2) \sigma^2 \end{aligned}$$

Determine a autocovariância de ordem 1:

$$\begin{aligned} \gamma_1 &= E(y_t y_{t-1}) \\ \gamma_1 &= E[(\varepsilon_t - \varphi \varepsilon_{t-1})(\varepsilon_{t-1} - \varphi \varepsilon_{t-2})] \\ \gamma_1 &= E(\varepsilon_t \varepsilon_{t-1} - \varphi \varepsilon_{t-1}^2 - \varphi \varepsilon_t \varepsilon_{t-2} + \varphi^2 \varepsilon_{t-1} \varepsilon_{t-2}) \\ \gamma_1 &= E(\varepsilon_t \varepsilon_{t-1}) - E(\varphi \varepsilon_{t-1}^2) - E(\varphi \varepsilon_t \varepsilon_{t-2}) + E(\varphi^2 \varepsilon_{t-1} \varepsilon_{t-2}) \end{aligned}$$

$$\gamma_1 = E(\varepsilon_t \varepsilon_{t-1}) - \varphi E(\varepsilon_{t-1}^2) - \varphi E(\varepsilon_t \varepsilon_{t-2}) + \varphi^2 E(\varepsilon_{t-1} \varepsilon_{t-2})$$

$$\gamma_1 = 0 - \varphi E(\varepsilon_{t-1}^2) - 0 + 0$$

$$\gamma_1 = -\varphi \text{ var}(\varepsilon_t)$$

$$\gamma_1 = -\varphi \sigma^2$$

Portanto:

$$\rho_1 = -\varphi / (1 + \varphi^2)$$

Para ordem 2:

$$\gamma_2 = E(y_t y_{t-2})$$

$$\gamma_2 = E[(\varepsilon_t - \varphi \varepsilon_{t-1})(\varepsilon_{t-2} - \varphi \varepsilon_{t-3})]$$

$$\gamma_2 = E(\varepsilon_t \varepsilon_{t-2} - \varphi \varepsilon_{t-1} \varepsilon_{t-2} - \varphi \varepsilon_t \varepsilon_{t-3} + \varphi^2 \varepsilon_{t-1} \varepsilon_{t-3})$$

$$\gamma_2 = E(\varepsilon_t \varepsilon_{t-2}) - E(\varphi \varepsilon_{t-1} \varepsilon_{t-2}) - E(\varphi \varepsilon_t \varepsilon_{t-3}) + E(\varphi^2 \varepsilon_{t-1} \varepsilon_{t-3})$$

$$\gamma_2 = E(\varepsilon_t \varepsilon_{t-1}) - \varphi E(\varepsilon_{t-1} \varepsilon_{t-2}) - \varphi E(\varepsilon_t \varepsilon_{t-3}) + \varphi^2 E(\varepsilon_{t-1} \varepsilon_{t-3})$$

$$\gamma_2 = 0 - 0 - 0 + 0 = 0$$

A função de autocorrelação só é diferente de zero para $k = 1$ quando se trata de um MA(1). Generalizando, a função de autocorrelação de um MA(q) é diferente de zero para valores de k menores ou iguais a q , e é zero para k maior do que q . O ponto em que a função de autocorrelação é truncada determina a ordem do processo MA.

Agora, passe à função de autocorrelação parcial. Antes, faça uma transformação no modelo:

$$y_t = \varepsilon_t - \varphi \varepsilon_{t-1}$$

$$\varepsilon_t = y_t + \varphi \varepsilon_{t-1}$$

Mas:

$$\varepsilon_{t-1} = y_{t-1} + \varphi \varepsilon_{t-2}$$

Substituindo, vem:

$$\varepsilon_t = y_t + \varphi(y_{t-1} + \varphi \varepsilon_{t-2})$$

$$\varepsilon_t = y_t + \varphi y_{t-1} + \varphi^2 \varepsilon_{t-2}$$

De novo:

$$\varepsilon_{t-2} = y_{t-2} + \varphi \varepsilon_{t-3}$$

Substituindo mais uma vez:

$$\varepsilon_t = y_t + \varphi y_{t-1} + \varphi^2 (y_{t-2} + \varphi \varepsilon_{t-3})$$

$$\varepsilon_t = y_t + \varphi y_{t-1} + \varphi^2 y_{t-2} + \varphi^3 \varepsilon_{t-3}$$

Se o processo for repetido indefinidamente, chega-se a:

$$\varepsilon_t = y_t + \varphi y_{t-1} + \varphi^2 y_{t-2} + \varphi^3 y_{t-3} + \varphi^4 y_{t-4} + \varphi^5 y_{t-5} + \dots,$$

que é uma representação de um processo autorregressivo de ordem infinita. Portanto, um processo MA pode ser escrito como um AR infinito. Como o coeficiente φ tem de ser menor do que 1, em módulo (caso contrário, essa *inversão* não seria possível, pois o valor

de ε_t , não convergiria na expressão acima), os coeficientes são declinantes. Assim, a função de autocorrelação parcial de um MA(1) seria equivalente à desse processo AR infinito, isto é, apresentaria coeficientes declinantes.

Figura 10.6 – Função de autocorrelação de um MA(1)

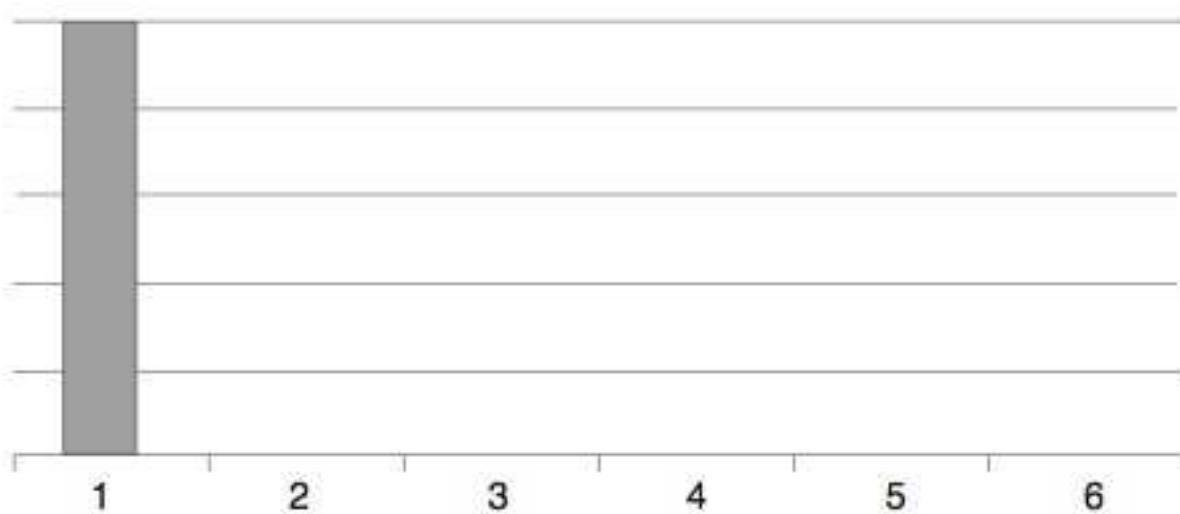
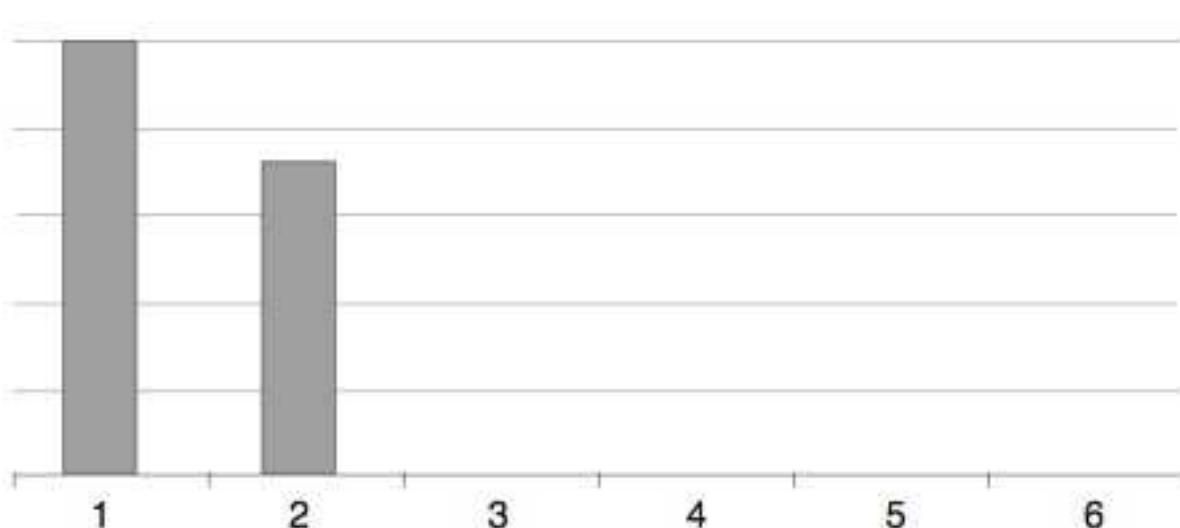
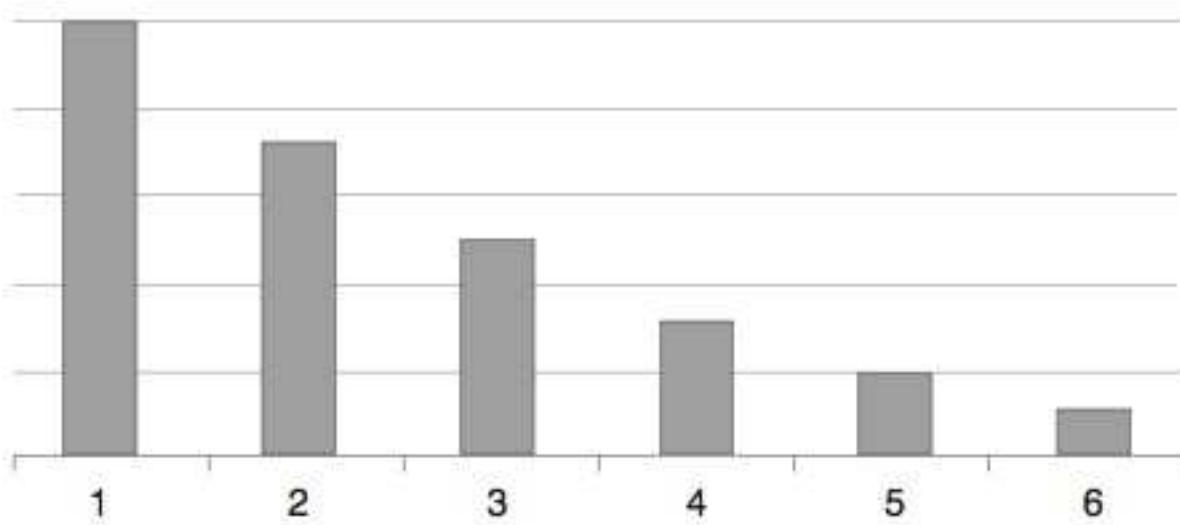


Figura 10.7 – Função de autocorrelação de um MA(2)

Figura 10.8 – Função de autocorrelação parcial de um MA(q) – qualquer q 

Finalmente, se o processo for um ARMA(p,q), ele terá as funções de autocorrelação e autocorrelação parcial combinadas dos dois processos. Dessa forma, um processo desse tipo apresentará as duas funções indefinidamente declinantes. A Tabela 10.2 resume a identificação de processos ARMA.

Tabela 10.2 – Identificação de processos ARMA

Tipo de processo	Função de autocorrelação	Função de autocorrelação parcial
AR(p)	declinante	truncada em p
MA(q)	truncada em q	declinante
ARMA(p,q)	declinante	declinante

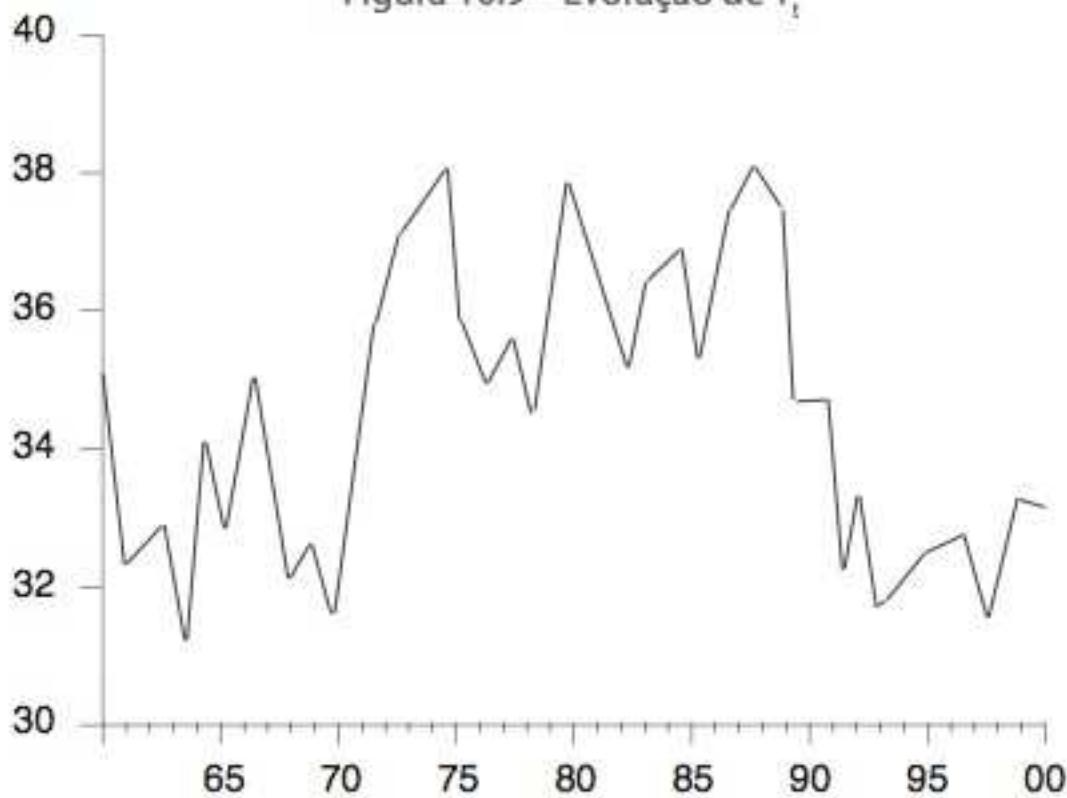
Exemplo 10.3.2.1

Identifique o processo da variável dada na Tabela 10.3.

Tabela 10.3 – Y_t de 1961 a 2000

Ano	Y _t	Ano	Y _t
1961	32,2	1981	36,5
1962	32,7	1982	35,1
1963	31,4	1983	36,4
1964	34,2	1984	36,8
1965	32,8	1985	35,3
1966	35,1	1986	37,1
1967	33,5	1987	38,2
1968	32,1	1988	37,5
1969	32,7	1989	34,7
1970	31,7	1990	34,8
1971	34,4	1991	32,4
1972	36,3	1992	33,5
1973	37,5	1993	31,8
1974	38,0	1994	32,0
1975	35,9	1995	32,4
1976	35,0	1996	32,6
1977	35,5	1997	31,6
1978	34,7	1998	33,0
1979	36,4	1999	32,9
1980	37,9	2000	33,3

Observe, na Figura 10.9, o gráfico de Y_t .

Figura 10.9 – Evolução de Y_t 

Aparentemente, é uma variável estacionária. Então, calcule as autocorrelações e autocorrelações parciais. A Tabela 10.4 mostra os valores de Y_t e suas defasagens:

Tabela 10.4 – Y_t e suas defasagens

Ano	Y_t	Y_{t-1}	Y_{t-2}	Y_{t-3}	Y_{t-4}	Y_{t-5}
1961	32,2					
1962	32,7	32,2				
1963	31,4	32,7	32,2			
1964	34,2	31,4	32,7	32,2		
1965	32,8	34,2	31,4	32,7	32,2	
1966	35,1	32,8	34,2	31,4	32,7	32,2
1967	33,5	35,1	32,8	34,2	31,4	32,7
1968	32,1	33,5	35,1	32,8	34,2	31,4
1969	32,7	32,1	33,5	35,1	32,8	34,2
1970	31,7	32,7	32,1	33,5	35,1	32,8
1971	34,4	31,7	32,7	32,1	33,5	35,1
1972	36,3	34,4	31,7	32,7	32,1	33,5
1973	37,5	36,3	34,4	31,7	32,7	32,1
1974	38,0	37,5	36,3	34,4	31,7	32,7
1975	35,9	38,0	37,5	36,3	34,4	31,7
1976	35,0	35,9	38,0	37,5	36,3	34,4
1977	35,5	35,0	35,9	38,0	37,5	36,3

Ano	Y_t	Y_{t-1}	Y_{t-2}	Y_{t-3}	Y_{t-4}	Y_{t-5}
1978	34,7	35,5	35,0	35,9	38,0	37,5
1979	36,4	34,7	35,5	35,0	35,9	38,0
1980	37,9	36,4	34,7	35,5	35,0	35,9
1981	36,5	37,9	36,4	34,7	35,5	35,0
1982	35,1	36,5	37,9	36,4	34,7	35,5
1983	36,4	35,1	36,5	37,9	36,4	34,7
1984	36,8	36,4	35,1	36,5	37,9	36,4
1985	35,3	36,8	36,4	35,1	36,5	37,9
1986	37,1	35,3	36,8	36,4	35,1	36,5
1987	38,2	37,1	35,3	36,8	36,4	35,1
1988	37,5	38,2	37,1	35,3	36,8	36,4
1989	34,7	37,5	38,2	37,1	35,3	36,8
1990	34,8	34,7	37,5	38,2	37,1	35,3
1991	32,4	34,8	34,7	37,5	38,2	37,1
1992	33,5	32,4	34,8	34,7	37,5	38,2
1993	31,8	33,5	32,4	34,8	34,7	37,5
1994	32,0	31,8	33,5	32,4	34,8	34,7
1995	32,4	32,0	31,8	33,5	32,4	34,8
1996	32,6	32,4	32,0	31,8	33,5	32,4
1997	31,6	32,6	32,4	32,0	31,8	33,5
1998	33,0	31,6	32,6	32,4	32,0	31,8
1999	32,9	33,0	31,6	32,6	32,4	32,0
2000	33,3	32,9	33,0	31,6	32,6	32,4

Usando os dados da Tabela 10.4, é possível encontrar os valores da função de autocorrelação:⁴

$$\rho_1 = \text{corr}(Y_t, Y_{t-1}) = 0,7538$$

$$\rho_2 = \text{corr}(Y_t, Y_{t-2}) = 0,6015$$

$$\rho_3 = \text{corr}(Y_t, Y_{t-3}) = 0,3928$$

$$\rho_4 = \text{corr}(Y_t, Y_{t-4}) = 0,2645$$

$$\rho_5 = \text{corr}(Y_t, Y_{t-5}) = 0,1927$$

O que indica uma função de autocorrelação declinante, típica de um processo AR ou ARMA. De fato, pode-se mostrar que o intervalo de 95% de confiança é dado por:

$$IC_{95\%} \cong \pm \frac{2}{\sqrt{n}} = \pm \frac{2}{\sqrt{40}} \cong 0,3162$$

⁴ Os valores amostrais das FAC e FACP são conhecidos por correlograma.

Portanto, os valores de ρ_1 , ρ_2 e ρ_3 são significantes, então, a função de autocorrelação é declinante (possivelmente, tendo em vista os demais valores) ou truncada em 3. Lembre-se de que, como em qualquer processo de estimativa (a identificação seria o primeiro passo), você está lidando com valores amostrais.

Para encontrar os valores da função de autocorrelação parcial, estimam-se as regressões com os valores defasados. Os resultados foram:

$$Y_t = 9,03 + 0,7370Y_{t-1}$$

$$Y_t = 8,05 + 0,6861Y_{t-1} + 0,0811Y_{t-2}$$

$$Y_t = 10,12 + 0,6905Y_{t-1} + 0,2242Y_{t-2} - 0,2066Y_{t-3}$$

$$Y_t = 9,92 + 0,6959Y_{t-1} + 0,1666Y_{t-2} - 0,2007Y_{t-3} + 0,0527Y_{t-4}$$

$$Y_t = 8,24 + 0,7299Y_{t-1} + 0,1622Y_{t-2} - 0,1580Y_{t-3} - 0,0340Y_{t-4} + 0,0617Y_{t-5}$$

Os valores da função de autocorrelação parcial, então, são:

$$\phi_1 = 0,7370$$

$$\phi_2 = 0,0811$$

$$\phi_3 = -0,2066$$

$$\phi_4 = 0,0527$$

$$\phi_5 = 0,0617$$

Nesse caso, fica claro que a função é truncada em 1, pois não só a queda de ϕ_1 para ϕ_2 é abrupta, como todos os valores de ϕ_2 em diante ficam bem abaixo do valor crítico de 0,3162 (em módulo).

Portanto, há uma função de autocorrelação declinante e uma função de autocorrelação parcial truncada em 1, o que nos indica que o processo é um AR(1).

10.3.3 Estimação de modelos ARIMA

A estimação de um modelo AR pode ser feita por mínimos quadrados ordinários. Para um modelo MA ou ARMA, a estimação deve ser feita por um processo recursivo, já que os erros (choques) passados, que atuam como variáveis independentes no modelo, não são diretamente observáveis.

Exemplo 10.3.3.1

Estime um modelo ARIMA para a variável apresentada no Exemplo 10.3.2.1.

A identificação sugere um modelo AR(1), que pode ser estimado por mínimos quadrados ordinários, o que, aliás, já foi feito quando você estimou a função de autocorrelação parcial. O resultado foi:

$$Y_t = 9,03 + 0,7370Y_{t-1}$$

Exemplo 10.3.3.2

Dada a série da Tabela 10.5, suponha que ela é um MA(1) e estime o modelo.

Tabela 10.5 – Z de 1961 a 2000			
Ano	Z _t	Ano	Z _t
1961	3,8	1981	2,0
1962	2,9	1982	1,3
1963	3,3	1983	2,8
1964	0,4	1984	4,6
1965	0,4	1985	4,0
1966	3,1	1986	2,5
1967	5,4	1987	1,9
1968	0,8	1988	1,2
1969	-0,7	1989	-1,1
1970	-0,6	1990	-1,6
1971	-1,3	1991	3,3
1972	-1,1	1992	0,7
1973	0,8	1993	0,2
1974	4,3	1994	4,3
1975	4,1	1995	4,7
1976	-0,5	1996	3,8
1977	-0,1	1997	4,6
1978	1,1	1998	3,3
1979	-1,0	1999	4,5
1980	2,5	2000	3,0

Se é um MA(1), é do tipo:

$$Z_t = \alpha + \varepsilon_t - \varphi \varepsilon_{t-1}$$

Como ε_{t-1} não é observável, uma forma de estimar é dar um *chute* inicial para α e φ . O *chute* inicial para α é fácil, pois:

$$E(Z_t) = E(\alpha) + E(\varepsilon_t) - \varphi E(\varepsilon_{t-1}) = \alpha$$

Como α é a própria média do processo, chuta-se o valor inicial para α como sendo a média amostral dos Z_t , que é dada por 1,9.

Para o chute inicial do coeficiente φ , usa-se o fato de que um MA(1) pode ser escrito como um AR infinito, isto é:

$$\varepsilon_t = y_t + \varphi y_{t-1} + \varphi^2 y_{t-2} + \varphi^3 y_{t-3} + \varphi^4 y_{t-4} + \varphi^5 y_{t-5} + \dots$$

Ou:

$$y_t = -\varphi y_{t-1} - \varphi^2 y_{t-2} - \varphi^3 y_{t-3} - \varphi^4 y_{t-4} - \varphi^5 y_{t-5} + \dots + \varepsilon_t$$

Evidentemente, não é possível estimar um AR infinito, mas se pode ter uma boa ideia do coeficiente φ se um processo AR for estimado com várias defasagens. Estimou-se um AR(5) e o resultado foi:

$$y_t = 1,34 + 0,67 y_{t-1} - 0,42 y_{t-2} + 0,35 y_{t-3} - 0,27 y_{t-4} - 0,04 y_{t-5}$$

O *chute* inicial é $\varphi = -0,67$.

Então, o modelo *inicial* é dado por:

$$Z_t = 1,9 + \varepsilon_t + 0,67 \varepsilon_{t-1}$$

Considerando⁵ $\varepsilon_{1960} = 0$, computa-se ε_t a partir de:

$$\hat{\varepsilon}_t = Z_t - 1,9 - 0,67 \hat{\varepsilon}_{t-1}$$

O que é feito na Tabela 10.6.

Tabela 10.6 – Variável Z_t e resíduos

Ano	Z_t	$\hat{\varepsilon}_t$	$\hat{\varepsilon}_{t-1}$
1961	3,8	1,9	0
1962	2,9	-0,273	1,9
1963	3,3	1,58291	-0,273
1964	0,4	-2,56055	1,58291
1965	0,4	0,215568	-2,56055
1966	3,1	1,055569	0,215568
1967	5,4	2,792769	1,055569
1968	0,8	-2,97115	2,792769
1969	-0,7	-0,60933	-2,97115
1970	-0,6	-2,09175	-0,60933
1971	-1,3	-1,79853	-2,09175
1972	-1,1	-1,79499	-1,79853
1973	0,8	0,102641	-1,79499
1974	4,3	2,33123	0,102641
1975	4,1	0,638076	2,33123
1976	-0,5	-2,82751	0,638076
1977	-0,1	-0,10557	-2,82751
1978	1,1	-0,72927	-0,10557
1979	-1,0	-2,41139	-0,72927
1980	2,5	2,215631	-2,41139
1981	2,0	-1,38447	2,215631
1982	1,3	0,327597	-1,38447

⁵ Poderia ser outro critério. Note que a estimação feita usando outro critério poderá dar resultados diferentes.

Ano	Z_t	$\hat{\epsilon}_t$	$\hat{\epsilon}_{t-1}$
1983	2,8	0,68051	0,327597
1984	4,6	2,244058	0,68051
1985	4,0	0,596481	2,244058
1986	2,5	0,200358	0,596481
1987	1,9	-0,13424	0,200358
1988	1,2	-0,61006	-0,13424
1989	-1,1	-2,59126	-0,61006
1990	-1,6	-1,76386	-2,59126
1991	3,3	2,581783	-1,76386
1992	0,7	-2,92979	2,581783
1993	0,2	0,262963	-2,92979
1994	4,3	2,223815	0,262963
1995	4,7	1,310044	2,223815
1996	3,8	1,022271	1,310044
1997	4,6	2,015079	1,022271
1998	3,3	0,049897	2,015079
1999	4,5	2,566569	0,049897
2000	3,0	-0,6196	2,566569

E, então, usa-se $\hat{\epsilon}_{t-1}$ computado como uma variável numa nova estimação. O resultado obtido foi:

$$Z_t = 1,9193 + \epsilon_t + 0,6232\epsilon_{t-1}$$

Repete-se o procedimento com esses novos valores. Computam-se novamente $\hat{\epsilon}_t$ e $\hat{\epsilon}_{t-1}$ e refaz-se a estimativa, cujo resultado foi:

$$Z_t = 1,9273 + \epsilon_t + 0,6297\epsilon_{t-1}$$

Repetindo:

$$Z_t = 1,9302 + \epsilon_t + 0,6295\epsilon_{t-1}$$

E novamente:

$$Z_t = 1,9313 + \epsilon_t + 0,6296\epsilon_{t-1}$$

Repete-se o procedimento quantas vezes forem necessárias, até que a diferença entre os coeficientes seja suficientemente pequena dentro de um critério estabelecido. Note que a diferença está na terceira casa decimal, isto é, o erro já é menor do que 0,01. Portanto, o resultado da estimativa é:

$$Z_t = 1,93 + \epsilon_t + 0,63\epsilon_{t-1}$$

10.3.4 Diagnóstico de modelos ARIMA

Você já sabe que, quando se faz a identificação do modelo, as funções de autocorrelação e autocorrelação parcial não são populacionais, mas amostrais. Assim, a identificação, na maioria dos casos, não dá uma resposta definitiva de qual o modelo a ser estimado. Após a estimação, um diagnóstico do modelo deve ser feito para se ter certeza de que o modelo escolhido foi adequado. Note que é possível que mais de um modelo ARIMA se mostre adequado para estimar uma série. A definição do modelo recai, então, nos critérios de escolha, como os critérios de informação de Schwarz e de Akaike.

E o que é um modelo adequado? É aquele que explica todas as interações entre a variável e valores passados dela ou de choques passados. Isso significa que os resíduos devem ser desprovidos de qualquer tipo de autocorrelação, portanto, devem ter características de um ruído branco.

Para tanto, calcule a função de autocorrelação dos resíduos. Para se testar a hipótese nula de que todas as autocovariâncias são nulas, usa-se a estatística de Box e Pierce:

$$Q = n \sum_{k=1}^m \hat{\rho}_k^2$$

Que segue uma distribuição de χ^2 com m graus de liberdade, ou, ainda, a estatística de Ljung e Box (que costuma apresentar melhor desempenho em amostras pequenas):

$$Q = n(n+2) \sum_{k=1}^m \frac{\hat{\rho}_k^2}{n-k}$$

Que segue uma distribuição de χ^2 com os mesmos m graus de liberdade.

Exemplo 10.3.4.1

Faça o diagnóstico do modelo estimado no Exemplo 10.3.3.1.

Os resíduos estão na Tabela 10.7.

Tabela 10.7 – Resíduos do modelo do Exemplo 10.3.3.1

-0,05911	1,91947	-0,82825	0,59349
-1,72761	1,719151	1,503563	-1,91722
2,0305	1,334739	0,94545	-0,4643
-1,43313	-1,13377	-0,84935	-0,21171
1,898686	-0,48605	2,056161	-0,30651
-1,39644	0,677264	1,829543	-1,45391
-1,61722	-0,49124	0,318832	0,683098
0,014593	1,798367	-1,96526	-0,44872
-1,42761	2,04545	0,198367	0,024985
2,009397	-0,46007	-2,27533	

A partir de tais resíduos, calcule os ρ_k e as estatísticas Q, mostradas na Tabela 10.8.

Tabela 10.8 – ρ_k e estatísticas Q

k	ρ_k	Ljung-Box	Box-Pierce	$\chi^2(k, 90\%)$
1	-0,0609	0,1562	0,1448	2,71
2	0,1421	1,0289	0,9323	4,61
3	-0,0462	1,1239	1,0157	6,25
4	-0,1029	1,6075	1,4285	7,78
5	-0,0883	1,9738	1,7323	9,24
6	0,0751	2,2470	1,9522	10,64
7	0,2058	4,3632	3,6039	12,02
8	0,1854	6,1358	4,9441	13,36
9	-0,1772	7,8085	6,1681	14,68
10	-0,1881	9,7595	7,5480	15,99
11	-0,2229	12,5960	9,4852	17,28
12	-0,0860	13,0344	9,7739	18,55
13	0,1073	13,7418	10,2225	19,81
14	0,0145	13,7553	10,2307	21,06
15	0,2171	16,8968	12,0696	22,31

Primeiro, volte sua atenção para a coluna dos ρ_k . O valor limite é dado por:

$$\pm \frac{2}{\sqrt{39}} \cong \pm 0,32$$

Todos os valores individuais de ρ_k estão dentro do limite, o que já é alentador, pois, pelo menos tomadas uma a uma, as autocorrelações são não significantes. O teste conjunto é feito pelas estatísticas Q, e tanto a de Ljung e Box como a de Box e Pierce estão abaixo do valor limite da distribuição χ^2 com os respectivos graus de liberdade. Portanto, aceita-se a hipótese nula de que *todas* as autocorrelações são nulas e, assim, os resíduos se comportam como um ruído branco, donde se conclui que o modelo estimado foi adequado.

10.3.5 Condições de estacionariedade e invertibilidade de um modelo ARIMA

Tome um modelo AR(1):

$$y_t = \theta y_{t-1} + \varepsilon_t$$

Você sabe que a série y_t só será estacionária se θ , em módulo, for menor do que 1, isto é:

$$|\theta| < 1$$

O que vale dizer, se escrever o modelo como se segue:

$$\Theta_1(L) y_t = \varepsilon_t$$

Sendo:

$$\Theta_1(L) \equiv 1 - \theta L$$

É um polinômio em L , cuja raiz é dada por (substituindo L por λ):

$$1 - \theta\lambda = 0$$

$$\lambda = \frac{1}{\theta}$$

Se θ for menor do que 1, em módulo, λ será maior do que 1 (também em módulo). A raiz do polinômio deve, então, ser maior do que 1 em valores absolutos, o que se diz, de maneira um tanto sofisticada, que a raiz cai *fora do círculo unitário*. Isso é dito porque, para o caso de processos de ordens maiores do que 1, é possível que as raízes não sejam todas reais, de modo que pode haver raízes não reais cujo módulo é maior do que 1.

Para um modelo AR(p) qualquer, isto é:

$$y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \dots + \theta_p y_{t-p} + \varepsilon_t$$

Que pode ser escrito como:

$$\Theta_p(L) y_t = \varepsilon_t$$

Sendo:

$$\Theta_p(L) \equiv 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_p L^p$$

A condição de estacionariedade desse processo é a de que *todas* as raízes de $\Theta_p(L)$ caiam fora do círculo unitário.

A contrapartida da condição de estacionariedade do modelo autorregressivo é a condição de invertibilidade do modelo de médias móveis.

Dado um modelo MA(1):

$$y_t = \varepsilon_t - \varphi \varepsilon_{t-1}$$

Vimos que esse modelo pode ser escrito (invertido) como um AR infinito, mas, para isso, é necessário que o coeficiente φ seja menor do que 1, em módulo.

Vale dizer que a raiz do polinômio $\Phi_1(L)$, dado por:

$$\Phi_1(L) \equiv 1 - \varphi L$$

Caia fora do círculo unitário.

Da mesma forma, um modelo MA(q) dado por:

$$y_t = \Phi_q(L) \varepsilon_t$$

Sendo:

$$\Phi_q(L) \equiv 1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_q L^q$$

Para que esse modelo possa ser invertido para um AR infinito, é necessário que todas as raízes de $\Phi_q(L)$ caiam fora do círculo unitário.

10.4 Testes de raízes unitárias

Fica clara a importância, pelo que foi visto até agora, de testar, para uma série y_t , num modelo do tipo AR(1):

$$y_t = \rho y_{t-1} + \varepsilon_t$$

se o coeficiente ρ é igual a 1. Se isso ocorrer, y_t não é estacionário e diz-se que apresenta uma raiz unitária, isto é, a raiz do polinômio autorregressivo é igual a 1.

Se ρ for mesmo igual a 1, a variância de y_t vai para infinito à medida que t aumenta. Dessa forma, os testes usuais (usando a distribuição de Student, por exemplo) não são válidos.

Por meio de simulações, Dickey e Fuller chegaram a valores limite que são válidos para quando se testa a hipótese de que ρ é igual a 1.

Na verdade, o que se testa é um pouco diferente: subtrai-se y_{t-1} do modelo anterior:

$$y_t - y_{t-1} = \rho y_{t-1} - y_{t-1} + \varepsilon_t$$

$$\Delta y_t = (\rho - 1) y_{t-1} + \varepsilon_t$$

$$\Delta y_t = \delta y_{t-1} + \varepsilon_t$$

Testar ρ igual a 1 equivale a testar $\delta = 0$. O teste é feito computando-se a estatística t como se fosse um teste comum numa regressão qualquer, mas, como os limites não são dados pela distribuição de Student, a estatística é denominada τ , e o teste é conhecido como teste de Dickey e Fuller (DF), cujos valores limite são dados ao final do livro.

Usualmente, são testadas também as seguintes formas:

$$\Delta y_t = \alpha + \delta y_{t-1} + \varepsilon_t \quad (\text{com intercepto})$$

$$\Delta y_t = \alpha + \beta_t + \delta y_{t-1} + \varepsilon_t \quad (\text{com intercepto e tendência determinística})$$

Vale uma lembrança: um modelo do tipo $y_t = \alpha + \beta_t + \varepsilon_t$, isto é, com tendência determinística, não é um modelo estacionário da maneira como foi definida, uma vez que a média não é constante. Porém, se se subtrai a tendência, chega-se a $y_t - \beta_t$, que é variável estacionária. Diz-se que y_t é **estacionária em torno da tendência**.

Cada um dos limites com hipótese alternativa e valores críticos próprios. Para o teste sem intercepto ou tendência, a hipótese alternativa é um processo estacionário com média zero. Para o teste com intercepto, mas sem tendência, a hipótese alternativa é um processo estacionário com média qualquer, enquanto a rejeição do teste com intercepto e a tendência indicam um processo estacionário em torno da tendência. Os valores críticos podem ser consultados nas tabelas ao final do livro.

Exemplo 10.4.1

Teste a presença de raiz unitária na variável *percentual de residências atendidas por saneamento básico na Meltávia*.

Os valores são repetidos na Tabela 10.9.

Tabela 10.9 – Percentual de residências com saneamento básico na Meltávia

Ano	y_t	y_{t-1}	Δy_t
1971	21,15		
1972	22,5	21,15	1,35
1973	24,3	22,5	1,8
1974	27,9	24,3	3,6
1975	30,6	27,9	2,7
1976	32,4	30,6	1,8
1977	35,1	32,4	2,7
1978	36,9	35,1	1,8
1979	39,6	36,9	2,7
1980	41,4	39,6	1,8
1981	43,2	41,4	1,8
1982	45,9	43,2	2,7
1983	48,6	45,9	2,7
1984	51,3	48,6	2,7
1985	54,9	51,3	3,6
1986	56,7	54,9	1,8
1987	57,6	56,7	0,9
1988	60,3	57,6	2,7
1989	63,9	60,3	3,6
1990	64,8	63,9	0,9
1991	67,5	64,8	2,7
1992	68,4	67,5	0,9
1993	69,3	68,4	0,9
1994	70,2	69,3	0,9
1995	72,0	70,2	1,8
1996	72,9	72,0	0,9
1997	74,7	72,9	1,8
1998	77,4	74,7	2,7
1999	78,3	77,4	0,9
2000	79,2	78,3	0,9

$$\Delta y_t = 3,35 - 0,0195 y_{t-1} \quad \tau_\mu = -2,22 \\ (0,0088)$$

$$\Delta y_t = 3,32 - 0,0034 t - 0,0180 y_{t-1} \quad \tau_t = -0,20 \\ (0,0886)$$

Não é necessário realizar o teste sem intercepto, pois as variáveis, se estacionárias, não possuem média zero.

Os valores críticos da tabela, para $n = 25$ (o valor mais próximo, já que se utiliza uma regressão com 29 observações) e 10% de significância, são: -1,60 (sem intercepto),

-2,62 (com intercepto) e -3,24 (com intercepto e tendência). Portanto, aceita-se a hipótese nula de que $\delta = 0$ e, portanto, $\rho = 1$. Conclui-se, então, que a variável apresenta raiz unitária e, assim, *não* é estacionária.

O teste de Dickey e Fuller assim formulado testa apenas a raiz unitária num processo do tipo AR(1). Para um processo AR(p), deve-se utilizar o teste de **Dickey e Fuller Aumentado (ADF)**, fazendo as seguintes regressões:

$$\Delta y_t = \delta y_{t-1} + \sum_{i=1}^{p-1} \omega_i \Delta y_{t-i} + \epsilon_t \quad (\text{sem intercepto})$$

$$\Delta y_t = \alpha + \delta y_{t-1} + \sum_{i=1}^{p-1} \omega_i \Delta y_{t-i} + \epsilon_t \quad (\text{com intercepto})$$

$$\Delta y_t = \alpha + \beta t + \delta y_{t-1} + \sum_{i=1}^{p-1} \omega_i \Delta y_{t-i} + \epsilon_t \quad (\text{com intercepto e tendência})$$

Uma variável pode apresentar mais de uma raiz unitária, que é o caso já discutido sobre as variáveis que, para se tornarem estacionárias, precisam de duas ou mais diferenças. Uma variável I(2) (estacionária na segunda diferença), por exemplo, apresenta duas raízes unitárias.

10.5 Cointegração

Como você já viu, uma regressão entre duas variáveis não estacionárias *pode* ser espúria; nesse caso, os testes usuais não têm validade. Portanto, na regressão:

$$Y_t = \alpha + \beta X_t + \epsilon_t$$

Se X e Y apresentam raiz unitária, há boa chance de que a regressão seja espúria. Entretanto, se X e Y são integradas de mesma ordem (são ambas I(1), por exemplo), é possível que elas *caminhem juntas* e, assim, o resultado da regressão entre as variáveis (bem como os testes) passa a fazer sentido.

Quando duas séries são integradas de mesma ordem e *caminham juntas*, elas são ditas **cointegradas**. Como testar se duas variáveis são cointegradas? Imagine os resíduos da regressão de Y por X : se elas não *caminham juntas*, o resíduo dessa regressão tenderá a aumentar, em valor absoluto. O resíduo de uma regressão espúria não é estacionário (o que é consistente com o fato de que os resíduos apresentam autocorrelação positiva), portanto, a maneira mais simples⁶ de verificar se duas séries são cointegradas é testar a existência de uma raiz unitária nos resíduos. Como os resíduos terão média zero, vale o teste:

$$\hat{\Delta \epsilon}_t = \hat{\delta \epsilon}_{t-1} + \mu_t$$

* Para testes mais complexos de cointegração, ou mesmo de raízes unitárias, procure textos mais avançados sobre o tema, como HAMILTON (1994) ou ENDERS (1994).

Que nada mais é que uma aplicação do teste Dickey-Fuller para os resíduos da regressão. Como se trata de um valor estimado, isso acrescenta uma incerteza ao teste e, portanto, valores críticos modificados, que podem ser consultados ao final do livro (na tabela “Valores Críticos para Testes de Cointegração Baseados em Resíduos de Regressões”). Esse teste é conhecido como **teste de Engle-Granger**.

Exemplo 10.5.1

Sejam Y_t e X_t duas variáveis aleatórias. Considere os resultados dos seguintes modelos de regressão estimados por mínimos quadrados ordinários (MQO) numa amostra de 100 elementos (entre parênteses estão os desvios padrão):

$$\begin{aligned}\Delta \hat{Y}_t &= 4,88 - 0,15 Y_{t-1} & \Delta \hat{X}_t &= 0,11 - 0,18 X_{t-1} \\ (2,87) &(0,08) & (0,09) &(0,08)\end{aligned}$$

Considere também os resultados da regressão de Y_t por X_t ,

$$Y_t = 23,39 + 14,40 X_t + \hat{\epsilon}_t \\ (13,76) (7,30)$$

em que $\hat{\epsilon}_t$ é o resíduo. Finalmente, considere a seguinte regressão:

$$\Delta \hat{\epsilon}_t = -0,38 \hat{\epsilon}_{t-1} \\ (0,11)$$

A regressão de Y_t por X_t é espúria? Ou seriam essas duas variáveis cointegradas?

A variável Y_t não é estacionária, pois a estatística do teste de Dickey-Fuller para Y_t é dada por:

$$T_y = -\frac{0,15}{0,08} = -1,865.$$

Como o valor crítico para o teste de Dickey-Fuller para uma amostra de 100 elementos é $-2,58$ a 10% de significância, aceita-se a hipótese nula, que é a não estacionariedade de Y_t . O mesmo ocorre para X_t , pois:

$$T_x = -\frac{0,18}{0,08} = -2,25.$$

A estatística do teste foi maior, mas não o suficiente para atingir o valor crítico. Então, as duas variáveis são não estacionárias e, dessa forma, é possível que a regressão entre elas seja espúria, o que deve ser testado pelos resíduos da regressão. A estatística do teste Engle-Granger é dada por:

$$\tau_\epsilon = -\frac{0,38}{0,11} = -3,45.$$

O valor ultrapassa os $-2,58$ do teste Dickey-Fuller, mas isso não é suficiente para se rejeitar a hipótese nula, pois os valores críticos devem incorporar a incerteza resultante do fato de que os resíduos são valores amostrais. O valor crítico deve ser consultado na tabela “Valores Críticos para Testes de Cointegração Baseados em Resíduos de Regressões”, em Tabelas Estatísticas, e, para uma amostra de 100, a 10% de significância é 3,03; a 5% de significância, 3,37. Agora, sim, pode-se dizer que a hipótese nula é rejeitada, e os resíduos são estacionários.

Ao rejeitar a hipótese de que os resíduos são não estacionários, rejeita-se também, em consequência, a hipótese de não cointegração. Portanto, as variáveis X_t e Y_t são cointegradas e a regressão não é espúria.

Exercícios

1. Dê a forma analítica dos seguintes processos:
 - a) ARMA(3,1);
 - b) ARIMA(2,2,1);
 - c) IMA(1,4);
 - d) ARI (1,2);
2. Teste a existência de uma raiz unitária na variável “exportações de trigo do Kazimenistão”, apresentada no Exemplo 10.2.1.
3. Faça a identificação da variável apresentada no Exemplo 10.3.3.2.
4. Com base no Exercício 3, é possível encontrar algum outro modelo, que não um MA(1), para Z_t ? Se sim, estime-o.
5. Faça o diagnóstico do modelo MA(1) e do modelo estimado (se houver) no Exercício 4 para a variável Z_t . Se ambos forem adequados, escolha o melhor modelo usando um dos critérios de informação vistos no Capítulo 8.
6. Usando o teste de Dickey-Fuller para os resíduos, verifique se as duas variáveis do Exemplo 10.2.1 são cointegradas.
7. Dado o modelo:

$$Y_t = 10 + 0,7Y_{t-1} + \varepsilon_t$$

- a) Determine a média do processo, isto é $E(Y_t)$.
- b) Se $Y_t = 7$, qual o valor previsto para Y_{t+2} ? (Isto é, $E(Y_{t+2} | Y_t)$?)
- c) Determine a variância do processo, considerando $\text{var}(\varepsilon_t) = 2$.

8. Dado o modelo:

$$Y_t = 6 + \varepsilon_t + 0,2 \varepsilon_{t-1}$$

- a) Determine a média do processo, isto é, $E(Y_t)$.
- b) Se $Y_t = 3,5$, qual o valor previsto para Y_{t+1} ? (Isto é, $E(Y_{t+1} | Y_t)$?)
- c) Determine a variância do processo, considerando $\text{var}(\varepsilon_t) = 1$.

9. Dado o modelo:

$$Y_t = 2 + 0,5Y_{t-1} + \varepsilon_t - 0,1\varepsilon_{t-1}$$

- a) Determine a média do processo, isto é $E(Y_t)$.
- b) Se $Y_t = 2$, qual o valor previsto para Y_{t+1} ? (Isto é, $E(Y_{t+1} | Y_t)$?)
- c) Determine a variância do processo, considerando $\text{var}(\varepsilon_t) = 1$.

10. Assinale verdadeiro ou falso:

- a) Se $z_t = w_1 z_{t-1} + w_2 z_{t-2} + w_3 z_{t-3} + \varepsilon_t$, se $w_1 + w_2 + w_3 = 1$, então z_t não é estacionário.
- b) No modelo de regressão $Y_t = \alpha + \beta X_t + \varepsilon_t$, se Y_t e X_t apresentam raiz unitária, a regressão é espúria.
- c) Na regressão $Y_t = \alpha + \beta Y_{t-1} + \varepsilon_t$, é possível testar a hipótese de que $\beta = 1$ pela distribuição t , de Student.

11. Considerando os operadores defasagem (L) e diferença (Δ), mostre que:

- a) $\Delta^2 = 1 - 2L + L^2$
- b) $\frac{1}{1-L} = 1 + L + L^2 + L^3 + L^4 + \dots$

12. Sejam Y_t e X_t duas variáveis aleatórias. Considere os resultados dos seguintes modelos de regressão estimados por mínimos quadrados ordinários (MQO) numa amostra de 50 elementos (entre parênteses estão os desvios padrão):

$$\begin{aligned}\hat{\Delta}Y_t &= 2,34 - 0,23\hat{Y}_{t-1} & \hat{\Delta}X_t &= 4,08 - 0,55\hat{X}_{t-1} \\ (1,26) & (0,12)(2,13) & (0,36) &\end{aligned}$$

Considere também os resultados da regressão de Y_t por X_t

$$\begin{aligned}Y_t &= 11,82 + 27,15X_t + \hat{\varepsilon}_t \\ (2,66) & (5,41)\end{aligned}$$

em que $\hat{\varepsilon}_t$ é o resíduo. Finalmente, considere a seguinte regressão:

$$\begin{aligned}\hat{\Delta} \hat{\varepsilon}_t &= -0,41 \hat{\varepsilon}_{t-1} \\ (0,15) &\end{aligned}$$

A regressão de Y_t por X_t é espúria? Ou seriam estas duas variáveis cointegradas?



Capítulo 11

NÚMEROS-ÍNDICES

11.1 Construindo números-índices

Suponha que se esteja fazendo um estudo das exportações da Xenodávia, medidas em moeda local, o xenodávio. As exportações da Xenodávia na década de 1990 estão dadas na Tabela 11.1.

Tabela 11.1 – Exportações da Xenodávia	
Ano	Valor das exportações em X\$
1991	1.234.321
1992	2.345.678
1993	3.456.809
1994	3.312.090
1995	3.211.601
1996	4.567.011
1997	5.299.181
1998	6.450.222
1999	5.878.477
2000	4.990.670

O objetivo da apresentação da Tabela 11.1 é mostrar a evolução das exportações desse país ao longo da década, uma vez que não se sabe o que significa 1 milhão de xenodávios. Assim, a apresentação dos valores em si não é tão importante.

Daí a utilidade do **número-índice**: é uma sequência que apresenta a mesma evolução da sequência original (isto é, os números mantêm a mesma proporção entre si), mas, como o valor propriamente dito não é importante, seus números são mais “amigáveis” e, supostamente, de leitura mais fácil.

Para a construção do número-índice, escolhe-se um valor qualquer da tabela, por exemplo, o valor correspondente a 1995 (porque a partir desse ano as exportações passam a crescer muito, mas poderia ser por outro motivo qualquer ou mesmo sem nenhuma

razão em particular). Atribui-se a esse ano o valor 100, o que, diga-se de passagem, é bem mais “amigável” do que 3.211.601.

Parte-se do valor de 1995 (que será então o **ano-base**) para encontrar os valores dos demais anos, o que pode ser feito por meio de uma regra de três simples. Por exemplo, para o ano de 1991:

$$3.211.601 - 100$$

$$1.234.321 - x$$

Portanto, o valor correspondente a 1991 é:

$$x = \frac{1.234.321 \times 100}{3.211.601} = 308,43$$

Dessa forma, pode-se estabelecer uma regra prática para calcular os valores do número-índice para os demais anos: multiplicar por 100 e dividir pelo valor da base. Assim:

$$1992: \quad 2.345.768 \times \frac{100}{3.211.601} = 73,04$$

$$1993: \quad 3.456.809 \times \frac{100}{3.211.601} = 107,64$$

$$1994: \quad 3.312.090 \times \frac{100}{3.211.601} = 103,13$$

$$1995: \quad 3.211.604 \times \frac{100}{3.211.601} = 100$$

$$1996: \quad 4.567.011 \times \frac{100}{3.211.601} = 142,20$$

$$1997: \quad 5.299.181 \times \frac{100}{3.211.601} = 165,00$$

$$1998: \quad 6.450.222 \times \frac{100}{3.211.601} = 200,84$$

$$1999: \quad 5.878.477 \times \frac{100}{3.211.601} = 183,04$$

$$2000: \quad 4.990.670 \times \frac{100}{3.211.601} = 155,40$$

Repare que a conta referente ao ano 1995 é desnecessária, já que o valor de 1995 foi definido *a priori* como 100.

Então, o número-índice referente aos valores das exportações da Xenodávia está na Tabela 11.2.

Tabela 11.2 – Número-índice das exportações da Xenodávia	
Ano	Índice de valor das exportações (base: 1995 = 100)
1991	38,43
1992	73,04
1993	107,64
1994	103,13
1995	100,00
1996	142,20
1997	165,00
1998	200,84
1999	183,04
2000	155,40

Repare que é fundamental que apareça na tabela qual o ano tomado como base, porque não necessariamente ele será indicado na tabela apresentada (seria possível, por exemplo, apresentar os valores a partir de 1997 usando a mesma base). Nesse caso específico das exportações da Xenodávia, o *ano* foi a variável escolhida, mas poderia ser outra.

Analizando a Tabela 11.2, constata-se que, entre 1995 e 1997, houve crescimento de 65% no valor das exportações; e, em 1992, o valor das exportações foi cerca de 27% menor em relação às exportações de 1995.

Exemplo 11.1.1

A partir da Tabela 11.2, construa um novo número-índice de tal modo que o ano-base seja 1991.

Supõe-se, nessa situação, que a tabela original não seja conhecida; parte-se, então, da tabela com o número-índice cuja base é 1995. Trata-se, pois, de simplesmente construir um número-índice da mesma forma que foi feita anteriormente; a única diferença é que se parte de uma sequência de dados que já está na forma de número-índice. Cada ano, então, é multiplicado por 100 e dividido pelo valor do ano-base, que agora é 38,43 (1991).

$$1992: \quad 73,04 \times \frac{100}{38,43} = 190,04$$

$$1993: \quad 107,64 \times \frac{100}{38,43} = 280,06$$

$$1994: \quad 103,13 \times \frac{100}{38,43} = 268,33$$

$$1995: \quad 100 \times \frac{100}{38,43} = 260,19$$

$$1996: \quad 142,20 \times \frac{100}{38,43} = 370,00$$

$$1997: \quad 165,00 \times \frac{100}{38,43} = 429,32$$

$$1998: \quad 200,84 \times \frac{100}{38,43} = 522,57$$

$$1999: \quad 183,04 \times \frac{100}{38,43} = 476,25$$

$$2000: \quad 155,40 \times \frac{100}{38,43} = 404,33$$

Repare que se chegaria aos mesmos valores se o índice fosse construído a partir dos dados originais.

11.2 Índices de preços

Uma variável que é uma candidata natural a ser representada por um número-índice é o preço, especialmente quando se refere ao nível geral de preços, em vez do preço de um bem específico.

Quando se diz que “a taxa de inflação foi de 10%”, o que é algo comprehensível para a maioria das pessoas, o que se quer dizer exatamente? Que o nível geral de preços subiu de 1.000.000.000.000 reais para 1.100.000.000.000 reais? Bom, isso já não é muito comprehensível.

Porém, na verdade, é algo parecido. A tal taxa de inflação aumentou 10% ou, o que talvez seja mais preciso, o nível de preços aumentou 10% significa que o preço de uma cesta de bens, que representaria o consumo da sociedade, aumentou 10%.

Como medir essa variação? Bom, como os preços não variam todos na mesma proporção ao mesmo tempo, essa resposta não é óbvia. Há, como você verá nas seções seguintes, mais de uma resposta possível.

11.2.1 Índice agregativo simples

A ideia desse índice é simplesmente comparar os preços entre um período e outro.

$$\text{IAS} = \frac{\sum_{i=1}^n p_i^1}{\sum_{i=1}^n p_i^0}$$

O subscrito representa o bem e o sobreescrito, o período. Assim, p_i^0 representa o preço do bem i no período zero.

Exemplo 11.2.1.1

Suponha que existam apenas três bens: arroz, feijão e televisão, cujos preços em 1999 e 2000 são mostrados na Tabela 11.3. Determine a variação de preços pelo IAS.

Tabela 11.3 – Preços de arroz, feijão e televisão

Bem	Preço 1999 (R\$)	Preço 2000 (R\$)
Arroz (kg)	1,00	2,00
Feijão (kg)	0,50	1,20
Televisão	400,00	440,00

$$\text{IAS} = \frac{\sum_{i=1}^n p_i^1}{\sum_{i=1}^n p_i^0} = \frac{2 + 1,2 + 440}{1 + 0,5 + 400} = \frac{443,2}{401,5} \cong 1,1039$$

Portanto, a variação do nível de preços medida pelo IAS é de 10,39%. Pode ser obtida facilmente por $(\text{IAS} - 1) \times 100\%$ ou, ainda, pode-se manter a representação utilizada para os números-índices de modo geral: se 1999 for o ano-base (valor do índice igual a 100), o índice em 2000 é 110,39.

Fica fácil perceber que essa não é uma boa forma de medir a variação de preços, pois como é possível que o arroz dobre de preço, o feijão mais que dobre, e a variação total seja apenas cerca de 10%, não por coincidência, muito próxima da variação do preço da televisão? É que, calculando dessa forma, o bem que tem preço maior terá, ainda que involuntariamente, maior peso na medição, já que uma variação de 70 centavos no preço do feijão acaba sendo comparada com o preço de 400 reais da televisão.

11.2.2 Índice de Sauerbeck

O **índice de Sauerbeck** apresenta uma mudança importante em relação ao IAS. É calculado da seguinte forma:

$$S = \frac{1}{n} \sum_{i=1}^n \frac{p_i^1}{p_i^0}$$

Ou seja, é uma média aritmética simples da razão entre os preços dos bens nos dois períodos. Essa razão é conhecida como **relativo de preços** ou, mais comumente, **preço relativo**.

Exemplo 11.2.2.1

Suponha que existam apenas três bens: arroz, feijão e caviar, cujos preços em 1999 e 2000 são mostrados na Tabela 11.4. Determine a variação de preços pelo índice de Sauerbeck.

Tabela 11.4 – Preços de arroz, feijão e caviar

Bem	Preço 1999 (R\$)	Preço 2000 (R\$)
Arroz (kg)	1,00	1,00
Feijão (kg)	0,90	1,00
Caviar(kg)	200,00	400,00

$$S = \frac{1}{3} \times \frac{1}{1} + \frac{1}{0,9} + \frac{400}{200} \cong 1,3704$$

Portanto, a variação de preços medida pelo índice de Sauerbeck é 37,04%.

Claramente, esse resultado também não é dos mais adequados. O arroz ficou estável, o feijão aumentou 11%, e esses dois bens (entre os três existentes) devem ter um peso muito maior no gasto dos consumidores que o caviar, que “puxou” o índice para cima, certamente bem mais do que deveria. É necessário levar-se em conta quanto cada bem é consumido. Não é possível fazer uma medida que represente a variação dos preços sem que se considerem também as *quantidades* consumidas.

11.2.3 Índices de Laspeyres e Paasche

Quando, ao comparar preços em dois períodos, levam-se em conta as quantidades consumidas, um problema que é preciso ter em mente é o de que as quantidades também podem mudar de um período para outro. Fica a questão de quais devem ser as quantidades escolhidas, o que é respondido no Exemplo 11.2.3.1.

Exemplo 11.2.3.1

Em uma sociedade em que há apenas três bens (denominados A, B e C), os preços e as quantidades consumidas em dois anos são mostrados na Tabela 11.5. Determine a variação de preços no período.

Tabela 11.5 – Preços dos bens A, B e C

	1999		2000	
	Preço	Quantidade	Preço	Quantidade
Bem A	\$1	1.000	\$2	500
Bem B	\$3	1.500	\$4	1.200
Bem C	\$4	1.000	\$3	1.200

Num primeiro momento, pode-se imaginar que a ponderação dos preços pelas quantidades se daria período a período, isto é, os preços de 2000 seriam ponderados pelas quantidades daquele ano e o mesmo ocorreria com os preços de 1999.

Entretanto, se o objetivo é a comparação dos preços, o uso de quantidades diferentes em diversos períodos *contaminaria* a comparação. É preciso escolher o período cujas quantidades serão utilizadas ou, o que também é possível, como você verá adiante, tomar a média das quantidades.

Essa escolha é arbitrária: não há, em princípio, nenhum motivo pelo qual se possa dizer que as quantidades de um período sejam mais adequadas do que outras. É possível escolher o período inicial, nesse caso 1999. Então, cada preço será multiplicado pela respectiva quantidade consumida em 1999.

$$L = \frac{1.000 \times 2 + 1.500 \times 4 + 1.000 \times 3}{1.000 \times 1 + 1.500 \times 3 + 1.000 \times 4} = \frac{11.000}{9.500} \cong 1,1579$$

A variação de preços, calculada dessa forma, é de 15,79%. A letra *L* colocada nesse cálculo se deve ao fato de que, quando se utilizam as quantidades iniciais, o índice é chamado **índice de Laspeyres**. Se escolher as quantidades do período final, o que é feito a seguir, então ele é chamado **índice de Paasche**.

$$P = \frac{500 \times 2 + 1.200 \times 4 + 1.200 \times 3}{500 \times 1 + 1.200 \times 3 + 1.200 \times 4} = \frac{9.400}{8.900} \cong 1,0562$$

Portanto, pelo índice de Paasche, a variação foi de 5,62%. O resultado foi um tanto assustador à primeira vista, já que a diferença foi substancial. Entretanto, é preciso lembrar que, em geral, índices de preços são calculados para períodos mais curtos (um mês, por exemplo), em que as mudanças nas quantidades não são tão grandes. Mesmo em períodos longos, é pouco provável que se observe

uma mudança tão radical no consumo de todos os bens de uma economia como nos três bens do exemplo anterior.

Independentemente dessas questões, o fato é que qualquer dos critérios é válido. Existem, portanto, duas formas de calcular o índice de preços: pelo índice de Laspeyres e pelo de Paasche:

$$L = \frac{\sum_{i=1}^n p_i^1 q_i^0}{\sum_{i=1}^n p_i^0 q_i^0} \quad P = \frac{\sum_{i=1}^n p_i^1 q_i^1}{\sum_{i=1}^n p_i^0 q_i^1}$$

Exemplo 11.2.3.2

Dada a Tabela 11.6, determine a variação de preços pelos índices de Laspeyres e Paasche.

Tabela 11.6 – Preços dos bens A, B e C

	1999		2000	
	Preço	Quantidade	Preço	Quantidade
Bem A	\$2	1.000	\$4	800
Bem B	\$6	1.000	\$6	900
Bem C	\$4	1.800	\$3	2.200

$$L = \frac{1.000 \times 4 + 1.000 \times 6 + 1.800 \times 3}{1.000 \times 2 + 1.000 \times 6 + 1.800 \times 4} = \frac{15.400}{15.200} \cong 1,0132$$

$$P = \frac{800 \times 4 + 900 \times 6 + 2.200 \times 3}{800 \times 2 + 900 \times 6 + 2.200 \times 4} = \frac{15.200}{15.800} \cong 0,962$$

Constatam-se um aumento de 1,32% no nível de preços por Laspeyres e uma queda de 3,8% por Paasche.

Note que, de novo, o valor para Laspeyres é maior do que para Paasche, isto é, $L > P$ nos dois exemplos. Isso vale sempre? Veja o Exemplo 11.2.3.3.

Exemplo 11.2.3.3

Dada a Tabela 11.7, determine novamente a variação de preços pelos índices de Laspeyres e de Paasche.

Tabela 11.7 – Preços dos bens A, B e C				
	1999		2000	
	Preço	Quantidade	Preço	Quantidade
Bem A	\$1	1.000	\$2	1.400
Bem B	\$2	1.000	\$3	1.200
Bem C	\$3	1.000	\$2	900

$$L = \frac{1.000 \times 2 + 1.000 \times 3 + 1.000 \times 2}{1.000 \times 1 + 1.000 \times 2 + 1.000 \times 3} = \frac{7.000}{6.000} \cong 1,1667$$

$$P = \frac{1.400 \times 2 + 1.200 \times 3 + 900 \times 2}{1.400 \times 1 + 1.200 \times 2 + 900 \times 3} = \frac{8.200}{6.500} \cong 1,2615$$

Dessa vez, houve aumento de 16,67%, calculado por Laspeyres, e de 26,15% por Paasche, isto é, $P > L$.

Agora que você já sabe que nem sempre $L > P$, resta verificar o que há de diferente neste exemplo em relação aos dois anteriores. É imediato que, neste último, a queda nos preços foi acompanhada da queda das quantidades, aumento dos preços, de aumento das quantidades. Foi o contrário do que ocorreu nos exemplos anteriores.

É possível ainda que o índice de Paasche seja maior do que o de Laspeyres, mesmo que a direção dos preços absolutos seja a mesma. Veja o Exemplo 11.2.3.4.

Exemplo 11.2.3.4

Mais uma vez, calcule os índices de Laspeyres e Paasche com base nos dados da Tabela 11.8.

Tabela 11.8 – Preços dos bens A, B e C				
	1999		2000	
	Preço	Quantidade	Preço	Quantidade
Bem A	\$1	1.000	\$2	900
Bem B	\$2	1.000	\$3	800
Bem C	\$3	1.000	\$4	500

$$L = \frac{1.000 \times 2 + 1.000 \times 3 + 1.000 \times 4}{1.000 \times 1 + 1.000 \times 2 + 1.000 \times 3} = \frac{9.000}{6.000} = 1,5$$

$$P = \frac{900 \times 2 + 800 \times 3 + 500 \times 4}{900 \times 1 + 800 \times 2 + 500 \times 3} = \frac{6.200}{4.000} \cong 1,55$$

O índice de Paasche foi maior do que o índice de Laspeyres, todos os preços aumentaram e todas as quantidades caíram. O movimento dos preços e quantidades absolutas foi o mesmo dos Exemplos 11.2.3.1 e 11.2.3.2, mas note que aqueles bens cujos preços aumentaram mais, relativamente, tiveram queda menor.

Nos dois últimos exemplos, preços e quantidades relativas se moveram *na mesma direção*, enquanto nos dois primeiros o movimento se deu *em direções opostas*.

No primeiro exemplo, os preços relativos foram, respectivamente, 2 ($=2/1$), 1,33 ($=4/3$) e 0,75 ($=3/4$). Já as quantidades relativas foram 0,5 ($=500/1.000$), 0,8 ($=1.200/1.500$) e 1,2 ($=1.200/1.000$). O coeficiente de correlação entre preços e quantidades relativas é -0,9927.

Quanto ao segundo exemplo, os preços relativos foram 2, 1 e 0,75, respectivamente, enquanto as quantidades relativas foram 0,8, 0,9 e 1,22. O coeficiente de correlação encontrado agora foi -0,8089.

Já no terceiro, os preços relativos são iguais a 2, 1,5 e 0,67; as quantidades relativas, 1,4, 1,2 e 0,9, o que dá um coeficiente de correlação de 0,9996.

Finalmente, no quarto exemplo, os preços relativos são 2, 1,5 e 1,33, enquanto as quantidades relativas são 0,9, 0,8 e 0,5 e o coeficiente de correlação entre eles é 0,8462. Nos dois últimos exemplos, o *coeficiente de correlação* é *positivo* entre preços e quantidades relativas, enquanto nos dois primeiros o *coeficiente de correlação* é *negativo*¹ entre essas duas variáveis.

Portanto:

$$\rho_{pq} < 0 \Rightarrow L > P$$

$$\rho_{pq} > 0 \Rightarrow P > L$$

Onde ρ_{pq} é o coeficiente de correlação entre preços e quantidades relativas.

Vale dizer que, num caso pouco provável, se o coeficiente de correlação for nulo, teremos $L = P$.

Os índices de Laspeyres e Paasche podem ser calculados de forma alternativa, por meio de transformações algébricas da fórmula original. Veja como isso é feito para o índice de Laspeyres:

$$L = \frac{\sum_{i=1}^n p_i^1 q_i^0}{\sum_{i=1}^n p_i^0 q_i^0}$$

¹ Muitas vezes se diz que o índice de Laspeyres é, em geral, maior que o de Paasche, o que é verdade. Entretanto, pela teoria econômica, as duas situações são possíveis, dependendo da origem da variação de preços; se resulta de uma variação da curva de oferta, a correlação é negativa, e é positiva se é originária de um deslocamento da curva de demanda.

Desmembrando, vem:

$$L = \frac{P_1^1 q_1^0 + P_2^1 q_2^0 + \dots + P_n^1 q_n^0}{\sum_{i=1}^n P_i^0 q_i^0}$$

Ou, ainda:

$$L = \frac{P_1^1 q_1^0}{\sum_{i=1}^n P_i^0 q_i^0} + \frac{P_2^1 q_2^0}{\sum_{i=1}^n P_i^0 q_i^0} + \dots + \frac{P_n^1 q_n^0}{\sum_{i=1}^n P_i^0 q_i^0}$$

Se cada termo dessa equação for multiplicado e dividido por P_i^0 , chega-se a:

$$L = \frac{P_1^1}{P_1^0} \times \frac{P_1^1 q_1^0}{\sum_{i=1}^n P_i^0 q_i^0} + \frac{P_2^1}{P_2^0} \times \frac{P_2^1 q_2^0}{\sum_{i=1}^n P_i^0 q_i^0} + \dots + \frac{P_n^1}{P_n^0} \times \frac{P_n^1 q_n^0}{\sum_{i=1}^n P_i^0 q_i^0}$$

Dessa forma, a exemplo do que ocorria com o índice de Sauerbeck, calcula-se uma média dos preços relativos de cada bem, só que dessa vez é uma média *aritmética ponderada*, cujos pesos são dados por:

$$w_i^0 = \frac{P_i^0 q_i^0}{\sum_{i=1}^n P_i^0 q_i^0}$$

Esses pesos têm significado muito claro, pois a expressão $P_i^0 q_i^0$ (preço vezes a quantidade do bem i no período zero) significa o gasto no bem i no período zero, enquanto a expressão $\sum_{i=1}^n P_i^0 q_i^0$ significa o gasto total (em todos os bens) no mesmo período.

Portanto, w_i^0 significa a participação relativa (percentual) no gasto do bem i no período zero, isto é, cada um dos bens será ponderado por sua participação no orçamento das famílias no período zero. Assim:

$$L = \frac{P_1^1}{P_1^0} \times w_1^0 + \frac{P_2^1}{P_2^0} \times w_2^0 + \dots + \frac{P_n^1}{P_n^0} \times w_n^0$$

Ou, resumidamente:

$$L = \sum_{i=1}^n \frac{P_i^1}{P_i^0} \times w_i^0$$

Portanto, o índice de Laspeyres pode ser interpretado como uma média aritmética (ponderada) dos preços relativos, na qual os pesos são o percentual que cada bem representa no orçamento, considerando-se o período inicial (zero).

Já foi mencionada uma *forma alternativa* de calcular o índice. Na verdade, é essa a forma mais comum, uma vez que uma pesquisa de quantidades é muito mais trabalhosa do que uma pesquisa de preços (é muito mais fácil ir ao supermercado ou à feira e verificar qual o preço de determinado bem do que saber quanto as pessoas compram desse bem). Normalmente, os institutos que calculam os índices de preços fazem pesquisas sobre as quantidades (na verdade, sobre os orçamentos) apenas uma vez em certo número de anos e, então, são estabelecidos os pesos que serão utilizados para as pesquisas de preços.

Transformação semelhante pode ser feita com o índice de Paasche:

$$P = \frac{\sum_{i=1}^n p_i^1 q_i^1}{\sum_{i=1}^n p_i^0 q_i^1}$$

Que pode ser escrito assim:

$$P = \frac{1}{\frac{\sum_{i=1}^n p_i^0 q_i^1}{\sum_{i=1}^n p_i^1 q_i^1}}$$

Desmembrando:

$$P = \frac{1}{\frac{p_1^0 q_1^1}{\sum_{i=1}^n p_i^1 q_i^1} + \frac{p_2^0 q_2^1}{\sum_{i=1}^n p_i^1 q_i^1} + \dots + \frac{p_n^0 q_n^1}{\sum_{i=1}^n p_i^1 q_i^1}}$$

Multiplicando e dividindo cada termo do denominador por p_i^1 :

$$P = \frac{1}{\frac{p_1^0}{p_1^1} \times \frac{p_1^1 q_1^1}{\sum_{i=1}^n p_i^1 q_i^1} + \frac{p_2^0}{p_2^1} \times \frac{p_2^1 q_2^1}{\sum_{i=1}^n p_i^1 q_i^1} + \dots + \frac{p_n^0}{p_n^1} \times \frac{p_n^1 q_n^1}{\sum_{i=1}^n p_i^1 q_i^1}}$$

Chega-se de novo aos preços relativos, só que invertidos e, no denominador, multiplicados por um peso que agora é definido por:

$$w_i^1 = \frac{p_i^1 q_i^1}{\sum_{i=1}^n p_i^1 q_i^1}$$

Que é a participação relativa no gasto do bem i no período 1. Assim, o índice de Paasche pode ser escrito da seguinte forma:

$$P = \frac{1}{\frac{p_1^0 \times w_1^1 + p_2^0 \times w_2^1 + \dots + p_n^0 \times w_n^1}{p_1^1 \times w_1^1 + p_2^1 \times w_2^1 + \dots + p_n^1 \times w_n^1}}$$

Que é uma média *harmônica*² (e ponderada) dos preços relativos, e pode ser escrita resumidamente como se segue:

$$P = \frac{1}{\sum_{i=1}^n \frac{p_i^0}{p_i^1} \times w_i^1}$$

É preciso fazer duas observações importantes: a primeira é que o peso utilizado no cálculo do índice de Paasche é obtido por meio das quantidades consumidas finais (atuais). Portanto, é necessário pesquisar quantidades com a mesma periodicidade que se pesquisam preços, o que torna a pesquisa muito trabalhosa e cara. Não é surpreendente, portanto, que os institutos que pesquisam preços sistematicamente prefiram o índice de Laspeyres.

A outra é que o fato de o índice de Laspeyres ser uma média aritmética dos preços relativos, enquanto Paasche é uma média harmônica, induz à noção (errada, como você já viu) de que o primeiro é *sempre* maior, porque a média aritmética é sempre maior ou, no mínimo, igual à média harmônica, desde que, obviamente, os pesos sejam os mesmos, o que não é o caso.

Exemplo 11.2.3.5

Calcule a variação do nível de preços pelos índices de Laspeyres e de Paasche usando os dados da Tabela 11.9.

² Se você precisar recordar a média harmônica, veja o Capítulo 2.

Tabela 11.9 – Preços dos bens A, B e C

	1999		2000	
	Preço	% do gasto	Preço	% do gasto
Bem A	\$ 11	25%	\$ 12	40%
Bem B	\$ 15	35%	\$ 18	20%
Bem C	\$ 22	40%	\$ 23	40%

Agora os dados não são as quantidades, mas as participações relativas no gasto em cada período. Devem ser calculados os dois índices como médias (aritmética e harmônica, respectivamente) dos preços relativos.

$$L = \frac{12}{11} \times 0,25 + \frac{18}{15} \times 0,35 + \frac{23}{22} \times 0,4 = 1,0509$$

$$P = \frac{\frac{1}{\frac{12}{11} \times 0,4 + \frac{18}{15} \times 0,2 + \frac{23}{22} \times 0,4}}{1} = 1,0918$$

Portanto, verificou-se um aumento de 5,09% no nível de preços pelo índice de Laspeyres e de 9,18% pelo índice de Paasche.

11.2.4 Critérios e índice de Fisher

Você viu que há diferentes maneiras de calcular índices de preços, mas como dizer se um tipo de índice de preços é *bom* ou *ruim*? Uma tentativa de responder a essa questão foi o estabelecimento de determinados critérios por Fisher³. São eles⁴:

- I. *Critério de identidade*: se o período para o qual o índice é calculado é o mesmo do período-base, então o valor do índice tem de ser igual a 1. Isto é:

$$I_{00} = 1$$

Esse critério é atendido por Laspeyres e Paasche. Senão, veja:

$$L_{00} = P_{00} = \frac{\sum_{i=1}^n p_i^0 q_i^0}{\sum_{i=1}^n p_i^0 q_i^0} = 1$$

Já que os dois períodos coincidem.

- II. *Critério da homogeneidade*: o valor do índice não deve ser modificado por alterações nas unidades de medida.

³ Irving Fisher, economista americano (1867-1947).

⁴ Será usada agora a seguinte notação: I_{01} é o índice do período um em relação ao período zero.

É fácil ver que tanto Laspeyres como Paasche atendem a esse critério, já que, se forem trocados os pesos de quilogramas por libras⁵¹⁰, ou os preços de reais para UFIR, essa alteração se dará tanto no numerador como no denominador, deixando inalterado o resultado.

III. Critério da proporcionalidade: se os preços relativos forem todos iguais a certo valor, o índice também o será.

Basta lembrar que Laspeyres e Paasche podem ser escritos como médias de preços relativos, e a média de valores iguais tem de ser o mesmo valor, caso contrário não seria média.

IV. Critério da determinação: o índice não pode ser nulo, infinito ou indeterminado se um único preço ou quantidade for nulo.

Seria nulo se o numerador fosse zero; infinito se o denominador se anulasse; e indeterminado no caso de ambos. Enfim, isso não ocorreria nem em Laspeyres, nem em Paasche, já que tanto o numerador como o denominador são somatórios e, portanto, uma única parcela sendo zero não tornaria a soma total zero.

V. Critério da reversibilidade: se o índice de março for calculado em relação a fevereiro, por exemplo, e se constatar aumento nos preços, quando se calculasse o índice de fevereiro em relação a março (invertendo a ordem), deveria ser encontrada uma queda que *cancelaria* o aumento constatado anteriormente. Isto é:

$$I_{01} \times I_{10} = 1$$

Isso *não vale* para Laspeyres e Paasche. Veja:

$$L_{01} \times L_{10} = \frac{\sum_{i=1}^n p_i^1 q_i^0}{\sum_{i=1}^n p_i^0 q_i^0} \times \frac{\sum_{i=1}^n p_i^0 q_i^1}{\sum_{i=1}^n p_i^1 q_i^1} \neq 1$$

$$P_{01} \times P_{10} = \frac{\sum_{i=1}^n p_i^1 q_i^1}{\sum_{i=1}^n p_i^0 q_i^1} \times \frac{\sum_{i=1}^n p_i^0 q_i^0}{\sum_{i=1}^n p_i^1 q_i^0} \neq 1$$

VI. Critério da circularidade: se, por exemplo, calcula-se o índice de fevereiro em relação a janeiro e o de março em relação a fevereiro, o *acumulado* dos dois deveria ser igual ao cálculo feito diretamente entre março e janeiro, ou seja:

$$I_{01} \times I_{12} = I_{02}$$

⁵¹⁰ Nesse caso, seria preciso alterar os preços também, já que eles são dados em R\$/kg ou RS/libra, o que manteria o total do gasto no bem inalterado.

De novo, esse critério *não vale* para Laspeyres e Paasche, como é verificado a seguir:

$$L_{01} \times L_{12} = \frac{\sum_{i=1}^n p_i^1 q_i^0}{\sum_{i=1}^n p_i^0 q_i^0} \times \frac{\sum_{i=1}^n p_i^2 q_i^1}{\sum_{i=1}^n p_i^1 q_i^1} \neq \frac{\sum_{i=1}^n p_i^2 q_i^0}{\sum_{i=1}^n p_i^0 q_i^0} = L_{02}$$

$$P_{01} \times P_{12} = \frac{\sum_{i=1}^n p_i^1 q_i^1}{\sum_{i=1}^n p_i^0 q_i^1} \times \frac{\sum_{i=1}^n p_i^2 q_i^2}{\sum_{i=1}^n p_i^2 q_i^1} \neq \frac{\sum_{i=1}^n p_i^2 q_i^2}{\sum_{i=1}^n p_i^0 q_i^2} = P_{02}$$

O fato de Laspeyres e Paasche não atenderem aos dois últimos critérios pode trazer certo incômodo. Por isso, Fisher propôs um novo índice, chamado, de maneira talvez um pouco pretensiosa, **índice ideal de Fisher**, que nada mais é do que a média geométrica dos índices de Laspeyres e de Paasche.

$$F = \sqrt{L \times P}$$

É fácil verificar que o índice de Fisher atende ao critério da reversibilidade, mas, como os de Laspeyres e Paasche, não atende ao da circularidade⁶.

Exemplo 11.2.4.1

No Exemplo 11.2.3.1, consulte a Tabela 11.5 e determine a variação de preços pelo índice de Fisher.

Como você já calculou o índice de Laspeyres e o de Paasche, o cálculo do índice de Fisher é imediato.

$$F = \sqrt{L \times P} = \sqrt{1,1579 \times 1,0562} = 1,1059$$

Portanto, pelo índice de Fisher, constata-se um aumento de 10,59%.

É claro que, independentemente de qual seja o maior entre Laspeyres e Paasche, Fisher será sempre um valor intermediário entre os dois, uma vez que é uma média geométrica de ambos.

Quanto à utilidade prática do índice de Fisher, ele tem, no mínimo, os mesmos inconvenientes do índice de Paasche, já que as quantidades (ou a proporção do gasto) devem ser atualizadas como os preços. No mínimo, porque as quantidades iniciais também precisam ser conhecidas.

⁶ O que, por si só, torna bastante discutível o termo ideal.

11.2.5 Índice de Marshall-Edgeworth

Na dúvida entre escolher as quantidades iniciais (Laspeyres) ou as atuais (Paasche), é possível ficar “em cima do muro”, escolhendo a média das duas. Quando se faz isso, você está calculando o **índice de Marshall-Edgeworth**.

O índice de Marshall-Edgeworth é, portanto, calculado da seguinte forma:

$$ME = \frac{\sum_{i=1}^n p_i^1 \frac{(q_i^0 + q_i^1)}{2}}{\sum_{i=1}^n p_i^0 \frac{(q_i^0 + q_i^1)}{2}}$$

Que, simplificando, fica:

$$ME = \frac{\sum_{i=1}^n p_i^1 (q_i^0 + q_i^1)}{\sum_{i=1}^n p_i^0 (q_i^0 + q_i^1)}$$

Do ponto de vista prático, entretanto, o índice de Marshall-Edgeworth apresenta os mesmos inconvenientes do índice de Fisher, pois são necessárias as quantidades dos dois períodos para calcular o índice.

11.3 Índices de quantidades e de valor

Da mesma forma que você calcula índices de preços, o que vale dizer, compara preços de períodos diferentes, é possível também comparar quantidades.

Se antes você usou as quantidades para ponderar os preços, agora utilizará os preços para ponderar as quantidades. Dessa forma, obterá, por exemplo, índice de Laspeyres de quantidades e índice de Paasche de quantidades:

$$L_q = \frac{\sum_{i=1}^n p_i^0 q_i^1}{\sum_{i=1}^n p_i^0 q_i^0}$$

$$P_q = \frac{\sum_{i=1}^n p_i^1 q_i^1}{\sum_{i=1}^n p_i^1 q_i^0}$$

Repare que, em ambos os casos citados (e ao contrário do que ocorre com os índices de preços), os preços estão fixos e as quantidades é que variam.

E se ambos variam? Nesse caso, você não está comparando preços nem quantidades, mas gasto ou, mais genericamente, valor. De fato, quando faz isso, você está calculando o chamado **índice de valor**:

$$V = \frac{\sum_{i=1}^n p_i^1 q_i^1}{\sum_{i=1}^n p_i^0 q_i^0}$$

Uma propriedade interessante para os índices (que poderia ser um sétimo critério) é a de que o índice de preços multiplicado pelo índice de quantidades seja igual ao índice de valor. Essa propriedade *não* é atendida pelos índices de Laspeyres e Paasche, como verificado a seguir:

$$L_p \times L_q = \frac{\sum_{i=1}^n p_i^1 q_i^0}{\sum_{i=1}^n p_i^0 q_i^0} \times \frac{\sum_{i=1}^n p_i^0 q_i^1}{\sum_{i=1}^n p_i^0 q_i^0} \neq \frac{\sum_{i=1}^n p_i^1 q_i^1}{\sum_{i=1}^n p_i^0 q_i^0} = V$$

$$P_p \times P_q = \frac{\sum_{i=1}^n p_i^1 q_i^1}{\sum_{i=1}^n p_i^0 q_i^1} \times \frac{\sum_{i=1}^n p_i^1 q_i^0}{\sum_{i=1}^n p_i^1 q_i^0} \neq \frac{\sum_{i=1}^n p_i^1 q_i^1}{\sum_{i=1}^n p_i^0 q_i^0} = V$$

Mas é fácil verificar que o índice de Fisher tem essa propriedade.

Exemplo 11.3.1

No Exemplo 11.2.3.1, consulte de novo a Tabela 11.5 e determine o índice de quantidades de Laspeyres e Paasche e o índice de valor.

$$L_q = \frac{1 \times 500 + 3 \times 1.200 + 4 \times 1.200}{1 \times 1.000 + 3 \times 1.500 + 4 \times 1.000} = 0,9368$$

$$P_q = \frac{2 \times 500 + 4 \times 1.200 + 3 \times 1.200}{2 \times 1.000 + 4 \times 1.500 + 3 \times 1.000} = 0,8545$$

$$V = \frac{2 \times 500 + 4 \times 1.200 + 3 \times 1.200}{1 \times 1.000 + 3 \times 1.500 + 4 \times 1.000} = 0,9895$$

Todos os índices apresentaram queda: o índice de quantidades teve queda de 6,32%, medido por Laspeyres, e 4,55%, medido por Paasche. Já o índice de valor apresentou queda de 1,05%.

11.4 Valores nominais e reais – deflacionamento de séries

Observe a Tabela 11.10, que mostra os salários de uma categoria profissional em um período inflacionário.

Tabela 11.10 – Salários em período inflacionário

Mês	Salários a preços correntes	Índice de preços (base: jan./YY = 100)
jan./XX	R\$ 1.000	300
fev./XX	R\$ 1.100	320
mar./XX	R\$ 1.200	340
abr./XX	R\$ 1.300	360
mai./XX	R\$ 1.400	400
jun./XX	R\$ 1.500	410
jul./XX	R\$ 1.600	430

Repare que essa categoria teve um aumento (alguns preferem falar reajuste), em fevereiro, de 10%. O valor, em moeda, foi 10% maior. Isso significa que o trabalhador pertencente a essa categoria pode comprar 10% a mais em bens e serviços? A resposta é claramente não, bastando para isso dar uma rápida olhada na coluna referente ao índice de preços.

Quando se observa essa coluna, verifica-se que os preços aumentaram de janeiro para fevereiro. De fato, é possível inclusive quantificar esse aumento no nível de preços:

$$\frac{320}{300} = 1,0667$$

Ou seja, houve aumento de preços (inflação) de 6,67%. O aumento dos salários é superior a essa taxa, o que vale dizer que ocorreu, sim, aumento do poder aquisitivo, mas não de 10%. Aliás, pela matemática financeira,⁷ é possível facilmente calcular esse aumento de poder aquisitivo, que foi de 3,12%.

Esse aumento de poder aquisitivo significa aumento de salário *real*, isto é, não expresso simplesmente no valor monetário, mas em termos de bens e serviços que podem ser adquiridos.

⁷ Basta fazer a conta $1,1/1,0667$, que nada mais é que o aumento dos salários (mais 1) dividido pela taxa de inflação (mais 1).

Ora, se o aumento de 10% em moeda não significa aumento equivalente em bens e serviços, isso significa que a moeda perdeu valor. Reais em fevereiro valem menos do que reais em janeiro.

Seria útil que a unidade de medida tivesse um valor constante, de tal modo que fosse possível identificar diretamente quando o poder aquisitivo aumentou ou caiu. Isso seria possível se todos os valores da tabela estivessem no mesmo *real*, isto é, fosse estabelecido o valor da moeda em um mês específico e, então, todos os valores seriam calculados com base nessa *moeda*. Isso equivale a encontrar uma série de **valores reais**, ou seja, retirar os efeitos da desvalorização da moeda (inflação), o que é conhecido como **deflacionamento** de uma série.

Exemplo 11.4.1

Com base na Tabela 11.10, construa uma série de salários reais medidos em reais constantes de abril.

A questão é: qual seria o valor equivalente ao salário de cada mês se os preços de abril fossem válidos em todos os meses? Ou, melhor dizendo, qual o valor do salário de cada mês a preços constantes de abril?

Esse cálculo pode ser feito a partir de uma simples regra de três. O valor de maio, por exemplo, a preços de maio (índice = 400) é R\$ 1.400. Então, pode-se encontrar o valor de maio a preços de abril (índice = 360) por:

$$1.400 : 400$$

$$x : 360$$

$$\text{Salário real de maio (preços de abril)} = 1.400 \times 360/400 = \text{R\$ } 1.260,00$$

Portanto, nota-se que o salário real em maio sofreu *queda* (diminuição de poder aquisitivo) de aproximadamente 3%.

Para os outros meses, o cálculo é feito da mesma forma: multiplica-se pelo índice de abril e divide-se pelo índice do mês em questão:

$$\text{Salário real de janeiro (preços de abril)} =$$

$$1.000 \times 360/300 = \text{R\$ } 1.200,00$$

$$\text{Salário real de fevereiro (preços de abril)} =$$

$$1.100 \times 360/320 = \text{R\$ } 1.237,50$$

$$\text{Salário real de março (preços de abril)} =$$

$$1.200 \times 360/340 = \text{R\$ } 1.270,59$$

$$\text{Salário real de abril (preços de abril)} =$$

$$1.300 \times 360/360 = \text{R\$ } 1.300,00$$

$$\text{Salário real de junho (preços de abril)} =$$

$$1.500 \times 360/410 = \text{R\$ } 1.317,07$$

$$\text{Salário real de julho (preços de abril)} =$$

$$1.600 \times 360/430 = \text{R\$ } 1.339,53$$

Seria possível então montar a Tabela 11.11 da seguinte maneira:

Tabela 11.11 – Salários reais em período inflacionário			
Mês	Salários a preços correntes	Índice de preços (base: Jan./YY = 100)	Salários reais (preços constantes de abr./XX)
jan./XX	R\$ 1.000	300	R\$ 1.200,00
fev./XX	R\$ 1.100	320	R\$ 1.237,50
mar./XX	R\$ 1.200	340	R\$ 1.270,59
abr./XX	R\$ 1.300	360	R\$ 1.300,00
mai./XX	R\$ 1.400	400	R\$ 1.260,00
jun./XX	R\$ 1.500	410	R\$ 1.317,07
jul./XX	R\$ 1.600	430	R\$ 1.339,53

Houve queda no poder aquisitivo do salário apenas em maio; nos demais meses, o salário real aumentou.

Repare que, de janeiro a fevereiro, a variação no salário real foi de 3,12%, como você já havia calculado.

Outra coisa importante é que o mês tomado como base para os valores reais não tem de ser o mesmo mês-base utilizado para o índice. De fato, o mês-base do índice nem sequer aparece na tabela (é janeiro de outro ano).

11.5 Tipos de índices de preços

Alguns assuntos bombardeiam os leitores com uma infinidade de índices que, com frequência, apresentam valores diferentes, muitas vezes de maneira significativa. Na verdade, são diferentes porque medem coisas diferentes.

Os índices são calculados por vários institutos (no Brasil, por exemplo, há índices calculados pelo IBGE, FIPE e Fundação Getúlio Vargas, DIEESE, entre outros), mas essa não é a única diferença.

Os índices podem ser especificamente de preços finais ao consumidor. Recebem abreviações do tipo IPC (índice de preços ao consumidor) e ICV (índice de custo de vida). Esses índices ainda variam segundo a faixa de renda da população que abrangem (isto é, da faixa de renda das famílias de cujos orçamentos são extraídos os pesos para o cálculo do índice).

Os índices podem ser também de preços no atacado, normalmente conhecidos como IPA, ou podem se referir especificamente a um setor da economia, como o da construção civil.

Há, ainda, índices gerais de preços (IGP), que, como o próprio nome diz, são uma média de índices como o de preços ao consumidor, atacado e construção civil.

Exercícios

1. Na Tabela 11.12 são dados os valores das exportações de um país em moeda local.

Tabela 11.12 – Exportações	
Ano	Exportações (X\$)
1994	1.234.567
1995	1.345.234
1996	1.027.123
1997	1.825.621
1998	1.975.454
1999	1.754.141

- a) Construa um índice tomando como base o ano 1997.
 b) Transforme a base do índice para 1994.
2. Na Tabela 11.13 é dada uma série de números-índices.

Tabela 11.13 – Números-índices	
Mês	Índice (base: jan./96 = 100)
jan./99	410
fev./99	430
mar./99	427
abr./99	450
mai./99	478
jun./99	490
jul./99	465
ago./99	481

- a) Calcule a variação percentual em cada mês.
 b) Transforme a base do índice para agosto de 1999.

3. Use os dados das Tabelas 11.14, 11.15 e 11.16 para calcular as variações de preços pelos índices de Laspeyres, Paasche, Fisher e Marshall–Edgeworth.

a)

Tabela 11.14 – Preços e quantidades dos bens A, B e C				
	1997		1998	
	Preço	Quantidade	Preço	Quantidade
Bem A	\$ 1	1.000	\$ 2	500
Bem B	\$ 3	1.500	\$ 4	1.200
Bem C	\$ 4	2.000	\$ 3	2.500

b)

Tabela 11.15 – Preços e quantidades dos bens 1, 2, 3 e 4				
	1999		2000	
	Preço	Quantidade	Preço	Quantidade
Bem 1	\$ 10	1.000	\$ 12	800
Bem 2	\$ 3	2.000	\$ 5	1.500
Bem 3	\$ 2	3.000	\$ 3	2.500
Bem 4	\$ 5	500	\$ 4	700

c)

Tabela 11.16 – Preços e quantidades dos bens X, Y e Z				
	2000		2001	
	Preço	Quantidade	Preço	Quantidade
Bem X	\$ 5	1.500	\$ 7	1.800
Bem Y	\$ 8	1.500	\$ 6	1.200
Bem Z	\$ 4	1.000	\$ 4	800

4. Calcule as variações de preço pelos índices de Laspeyres e Paasche usando os dados da Tabela 11.17.

Tabela 11.17 – Preços e participação no gasto dos bens A, B e C				
	1998		1999	
	Preço	% do gasto	Preço	% do gasto
Bem A	\$ 10	30%	\$ 14	20%
Bem B	\$ 20	40%	\$ 18	60%
Bem C	\$ 22	30%	\$ 25	20%

5. Usando os dados da Tabela 11.18, calcule a participação percentual de cada bem no gasto total para o ano 1997.

Tabela 11.18 – Preços e quantidade dos bens 1, 2, 3 e 4

	1997	
	Preço	Quantidade
Bem 1	\$ 15	1.000
Bem 2	\$ 20	1.200
Bem 3	\$ 25	800
Bem 4	\$ 22	600

6. Utilizando os dados da Tabela 11.19, calcule o índice de Laspeyres em 1998, 1999 e 2000.

Tabela 11.19 – Preços dos bens 1, 2, 3 e 4

	Preços		
	1998	1999	2000
Bem 1	\$ 16	\$ 18	\$ 20
Bem 2	\$ 22	\$ 25	\$ 26
Bem 3	\$ 24	\$ 23	\$ 22
Bem 4	\$ 22	\$ 23	\$ 25

7. Verifique se o índice de Fisher atende aos critérios de reversibilidade e circularidade e se tem a propriedade de que o índice de preços multiplicado pelo de quantidades é igual ao índice de valor.
8. Verifique se o índice de Marshall-Edgeworth atende aos critérios de Fisher e se tem a propriedade de que o índice de preços multiplicado pelo de quantidades é igual ao índice de valor.
9. O **índice geométrico simples** é uma média geométrica (simples, não ponderada) dos preços relativos. Verifique se esse índice atende aos critérios de Fisher.
10. Na Tabela 11.20 são dados os salários nominais de uma categoria profissional e o índice de preços.

Tabela 11.20 – Salários nominais

Mês	Salário nominal (R\$)	Índice de preços (base: jan. = 100)
janeiro	1.000,00	100
fevereiro	1.100,00	120
março	1.300,00	140
abril	1.650,00	170
maio	1.700,00	190
junho	2.000,00	220

- a) Determine a variação percentual dos salários nominais.
- b) Determine a variação percentual dos preços (taxa de inflação).
- c) Determine a variação percentual dos salários reais (a preços de janeiro).
11. Na Tabela 11.21 são dados os valores das importações de um país em moeda corrente local e o índice de preços desse país.

Tabela 11.21 – Valores das importações

Ano	Importações (X\$)	Índice de preços (base: 1999 = 100)
1996	978.503	127
1997	1.130.544	150
1998	1.475.612	171
1999	1.121.300	187

- a) Construa um índice para as importações tomando como base o ano 1997.
- b) Calcule a taxa de inflação (variação no nível de preços) em cada ano.
- c) Construa uma série com os valores reais das importações (utilize os preços de 1999).
12. São dados:

índice de valor = 120

índice de quantidades de Laspeyres = 80

Determine a variação de preços medida pelo índice de Paasche.

13. Um produto teve aumento de 20%. Se isso representou aumento de 0,5% no custo de vida, qual é o percentual do orçamento representado por esse produto na época do período-base?
14. Qual o significado do produto de um índice de preços de Laspeyres por um índice de quantidades de Paasche? E o produto de um índice de preços de Paasche por um índice de quantidades de Laspeyres?

15. Assinale verdadeiro ou falso:

- a) Se há inflação, o salário real sempre cai.
- b) O índice de preços de Laspeyres compara o custo de aquisição de uma cesta de bens em certo período com o custo de aquisição dessa mesma cesta no período-base.
- c) O índice de preços de Paasche compara o custo de aquisição de uma cesta de bens em certo período com o custo de aquisição dessa mesma cesta no período-base.
- d) O índice de preços de Laspeyres é sempre maior (ou igual) do que o índice de preços de Paasche.

- e) O índice de Fisher é sempre maior do que os índices de Laspeyres e de Paasche.
- f) A diferença entre o índice de preços de Laspeyres e o índice de preços de Paasche é que, para o primeiro, a ponderação é fixa na época-base e para o segundo é variável na época atual.
- g) Sob condições normais de demanda, um índice de valor é sempre menor do que o produto de um índice de Paasche-preço por um índice de Paasche-quantidade e sempre maior do que o produto de um índice de Laspeyres-preço por um índice de Laspeyres-quantidade.

Tabelas estatísticas

Tabela de probabilidades – distribuição normal padronizada											
z	0	1	2	3	4	5	6	7	8	9	
0,0	0,000000	0,003989	0,007978	0,011966	0,015953	0,019939	0,023922	0,027903	0,031881	0,035856	
0,1	0,039828	0,043795	0,047758	0,051717	0,055670	0,059618	0,063559	0,067495	0,071424	0,075345	
0,2	0,079260	0,083166	0,087064	0,090954	0,094835	0,098706	0,102568	0,106420	0,110261	0,114092	
0,3	0,117911	0,121720	0,125516	0,129300	0,133072	0,136831	0,140576	0,144309	0,148027	0,151732	
0,4	0,155422	0,159097	0,162757	0,166402	0,170031	0,173645	0,177242	0,180822	0,184386	0,187933	
0,5	0,191462	0,194974	0,198468	0,201944	0,205401	0,208840	0,212260	0,215661	0,219043	0,222405	
0,6	0,225747	0,229069	0,232371	0,235653	0,238914	0,242154	0,245373	0,248571	0,251748	0,254903	
0,7	0,258036	0,261148	0,264238	0,267305	0,270350	0,273373	0,276373	0,279350	0,282305	0,285236	
0,8	0,288145	0,291030	0,293892	0,296731	0,299546	0,302337	0,305105	0,307850	0,310570	0,313267	
0,9	0,315940	0,318589	0,321214	0,323814	0,326391	0,328944	0,331472	0,333977	0,336457	0,338913	
1,0	0,341345	0,343752	0,346136	0,348495	0,350830	0,353141	0,355428	0,357690	0,359929	0,362143	
1,1	0,364334	0,366500	0,368643	0,370762	0,372857	0,374928	0,376976	0,379000	0,381000	0,382977	
1,2	0,384930	0,386861	0,388768	0,390651	0,392512	0,394350	0,396165	0,397958	0,399727	0,401475	
1,3	0,403200	0,404902	0,406582	0,408241	0,409877	0,411492	0,413085	0,414657	0,416207	0,417736	
1,4	0,419243	0,420730	0,422196	0,423641	0,425066	0,426471	0,427855	0,429219	0,430563	0,431888	
1,5	0,433193	0,434478	0,435745	0,436992	0,438220	0,439429	0,440620	0,441792	0,442947	0,444083	
1,6	0,445201	0,446301	0,447384	0,448449	0,449497	0,450529	0,451543	0,452540	0,453521	0,454486	
1,7	0,455435	0,456367	0,457284	0,458185	0,459070	0,459941	0,460796	0,461636	0,462462	0,463273	
1,8	0,464070	0,464852	0,465620	0,466375	0,467116	0,467843	0,468557	0,469258	0,469946	0,470621	
1,9	0,471283	0,471933	0,472571	0,473197	0,473810	0,474412	0,475002	0,475581	0,476148	0,476705	
2,0	0,477250	0,477784	0,478308	0,478822	0,479325	0,479818	0,480301	0,480774	0,481237	0,481691	
2,1	0,482136	0,482571	0,482997	0,483414	0,483823	0,484222	0,484614	0,484997	0,485371	0,485738	
2,2	0,486097	0,486447	0,486791	0,487126	0,487455	0,487776	0,488089	0,488396	0,488696	0,488989	
2,3	0,489276	0,489556	0,489830	0,490097	0,490358	0,490613	0,490863	0,491106	0,491344	0,491576	
2,4	0,491802	0,492024	0,492240	0,492451	0,492656	0,492857	0,493053	0,493244	0,493431	0,493613	
2,5	0,493790	0,493963	0,494132	0,494297	0,494457	0,494614	0,494766	0,494915	0,495060	0,495201	
2,6	0,495339	0,495473	0,495604	0,495731	0,495855	0,495975	0,496093	0,496207	0,496319	0,496427	
2,7	0,496533	0,496636	0,496736	0,496833	0,496928	0,497020	0,497110	0,497197	0,497282	0,497365	
2,8	0,497445	0,497523	0,497599	0,497673	0,497744	0,497814	0,497882	0,497948	0,498012	0,498074	
2,9	0,498134	0,498193	0,498250	0,498305	0,498359	0,498411	0,498462	0,498511	0,498559	0,498605	
3,0	0,498650	0,498694	0,498736	0,498777	0,498817	0,498856	0,498893	0,498930	0,498965	0,498999	
3,1	0,499032	0,499065	0,499096	0,499126	0,499155	0,499184	0,499211	0,499238	0,499264	0,499289	
3,2	0,499313	0,499336	0,499359	0,499381	0,499402	0,499423	0,499443	0,499462	0,499481	0,499499	
3,3	0,499517	0,499534	0,499550	0,499566	0,499581	0,499596	0,499610	0,499624	0,499638	0,499651	
3,4	0,499663	0,499675	0,499687	0,499698	0,499709	0,499720	0,499730	0,499740	0,499749	0,499758	
3,5	0,499767	0,499776	0,499784	0,499792	0,499800	0,499807	0,499815	0,499822	0,499828	0,499835	
3,6	0,499841	0,499847	0,499853	0,499858	0,499864	0,499869	0,499874	0,499879	0,499883	0,499888	
3,7	0,499892	0,499896	0,499900	0,499904	0,499908	0,499912	0,499915	0,499918	0,499922	0,499925	
3,8	0,499928	0,499931	0,499933	0,499936	0,499938	0,499941	0,499943	0,499946	0,499948	0,499950	
3,9	0,499952	0,499954	0,499956	0,499958	0,499959	0,499961	0,499963	0,499964	0,499966	0,499967	
4,0	0,499968	0,499970	0,499971	0,499972	0,499973	0,499974	0,499975	0,499976	0,499977	0,499978	
4,1	0,499979	0,499980	0,499981	0,499982	0,499983	0,499983	0,499984	0,499985	0,499985	0,499986	
4,2	0,499987	0,499987	0,499988	0,499988	0,499989	0,499989	0,499990	0,499990	0,499991	0,499991	
4,3	0,499991	0,499992	0,499992	0,499993	0,499993	0,499993	0,499993	0,499994	0,499994	0,499994	
4,4	0,499995	0,499995	0,499995	0,499995	0,499996	0,499996	0,499996	0,499996	0,499996	0,499996	

Valores da distribuição qui-quadrado							
G.L.	0,5%	1%	2,5%	5%	10%	90%	95%
1	0,000039	0,00016	0,00098	0,0039	0,0158	2,71	3,84
2	0,0100	0,0201	0,0506	0,103	0,211	4,61	5,99
3	0,072	0,115	0,216	0,352	0,584	6,25	7,81
4	0,21	0,30	0,48	0,71	1,06	7,78	9,49
5	0,41	0,55	0,83	1,15	1,61	9,24	11,14
6	0,68	0,87	1,24	1,64	2,20	10,64	12,59
7	0,99	1,24	1,69	2,17	2,83	12,02	14,07
8	1,34	1,65	2,18	2,73	3,49	13,36	15,51
9	1,73	2,09	2,70	3,33	4,17	14,68	16,92
10	2,16	2,56	3,25	3,94	4,87	15,99	18,31
11	2,60	3,05	3,82	4,57	5,58	17,28	19,68
12	3,07	3,57	4,40	5,23	6,30	18,55	21,03
13	3,57	4,11	5,01	5,89	7,04	19,81	22,36
14	4,07	4,66	5,63	6,57	7,79	21,06	23,68
15	4,60	5,23	6,26	7,26	8,55	22,31	25,00
16	5,14	5,81	6,91	7,96	9,31	23,54	26,30
17	5,70	6,41	7,56	8,67	10,09	24,77	27,59
18	6,26	7,01	8,23	9,39	10,86	25,99	28,87
19	6,84	7,63	8,91	10,12	11,65	27,20	30,14
20	7,43	8,26	9,59	10,85	12,44	28,41	31,41
25	10,52	11,52	13,12	14,61	16,47	34,38	37,65
30	13,79	14,95	16,79	18,49	20,60	40,26	43,77
40	20,71	22,16	24,43	26,51	29,05	51,81	55,76
50	27,99	29,71	32,36	34,76	37,69	63,17	67,50
60	35,53	37,48	40,48	43,19	46,46	74,40	79,08
70	43,28	45,44	48,76	51,74	55,33	85,53	90,53
80	51,17	53,54	57,15	60,39	64,28	96,58	101,88
90	59,20	61,75	65,65	69,13	73,29	107,57	113,15
100	67,33	70,06	74,22	77,93	82,36	118,50	124,34

Distribuição de Student – Valores críticos de t , tais que $P(-tc < t < tc) = 1 - p$ em itálico: teste monocaudal

G.L.	90%	80%	70%	60%	50%	40%	30%	25%	20%	15%	10%	5%	2,5%	2%	1%	0,5%
	45%	40%	35%	30%	25%	20%	15%	10%	8%	6%	4%	3%	2%	1%	0,5%	
1	0,1584	0,3249	0,5095	0,7265	1,0000	1,3764	1,9626	3,0777	6,3138	12,7062	31,8205	63,6567				
2	0,1421	0,2887	0,4447	0,6172	0,8165	1,0607	1,3862	1,8856	2,9200	4,3027	6,9646	9,9248				
3	0,1366	0,2767	0,4242	0,5844	0,7649	0,9785	1,2498	1,6377	2,3534	3,1824	4,5407	5,8409				
4	0,1338	0,2707	0,4142	0,5686	0,7407	0,9410	1,1896	1,5332	2,1318	2,7764	3,7469	4,6041				
5	0,1322	0,2672	0,4082	0,5594	0,7267	0,9195	1,1558	1,4759	2,0150	2,5706	3,3649	4,0321				
6	0,1311	0,2648	0,4043	0,5534	0,7176	0,9057	1,1342	1,4398	1,9432	2,4469	3,1427	3,7074				
7	0,1303	0,2632	0,4015	0,5491	0,7111	0,8960	1,1192	1,4149	1,8946	2,3646	2,9980	3,4995				
8	0,1297	0,2619	0,3995	0,5459	0,7064	0,8889	1,1081	1,3968	1,8595	2,3060	2,8965	3,3554				
9	0,1293	0,2610	0,3979	0,5435	0,7027	0,8834	1,0997	1,3830	1,8331	2,2622	2,8214	3,2498				
10	0,1289	0,2602	0,3966	0,5415	0,6998	0,8791	1,0931	1,3722	1,8125	2,2281	2,7638	3,1693				
11	0,1286	0,2596	0,3956	0,5399	0,6974	0,8755	1,0877	1,3634	1,7959	2,2010	2,7181	3,1058				
12	0,1283	0,2590	0,3947	0,5386	0,6955	0,8726	1,0832	1,3562	1,7823	2,1788	2,6810	3,0545				
13	0,1281	0,2586	0,3940	0,5375	0,6938	0,8702	1,0795	1,3502	1,7709	2,1604	2,6503	3,0123				
14	0,1280	0,2582	0,3933	0,5366	0,6924	0,8681	1,0763	1,3450	1,7613	2,1448	2,6245	2,9768				
15	0,1278	0,2579	0,3928	0,5357	0,6912	0,8662	1,0735	1,3406	1,7531	2,1314	2,6025	2,9467				
16	0,1277	0,2576	0,3923	0,5350	0,6901	0,8647	1,0711	1,3368	1,7459	2,1199	2,5835	2,9208				
17	0,1276	0,2573	0,3919	0,5344	0,6892	0,8633	1,0690	1,3334	1,7396	2,1098	2,5669	2,8982				
18	0,1274	0,2571	0,3915	0,5338	0,6884	0,8620	1,0672	1,3304	1,7341	2,1009	2,5524	2,8784				
19	0,1274	0,2569	0,3912	0,5333	0,6876	0,8610	1,0655	1,3277	1,7291	2,0930	2,5395	2,8609				
20	0,1273	0,2567	0,3909	0,5329	0,6870	0,8600	1,0640	1,3253	1,7247	2,0860	2,5280	2,8453				
21	0,1272	0,2566	0,3906	0,5325	0,6864	0,8591	1,0627	1,3232	1,7207	2,0796	2,5176	2,8314				
22	0,1271	0,2564	0,3904	0,5321	0,6858	0,8583	1,0614	1,3212	1,7171	2,0739	2,5083	2,8188				
23	0,1271	0,2563	0,3902	0,5317	0,6853	0,8575	1,0603	1,3195	1,7139	2,0687	2,4999	2,8073				
24	0,1270	0,2562	0,3900	0,5314	0,6848	0,8569	1,0593	1,3178	1,7109	2,0639	2,4922	2,7969				
25	0,1269	0,2561	0,3898	0,5312	0,6844	0,8562	1,0584	1,3163	1,7081	2,0595	2,4851	2,7874				
26	0,1269	0,2560	0,3896	0,5309	0,6840	0,8557	1,0575	1,3150	1,7056	2,0555	2,4786	2,7787				
27	0,1268	0,2559	0,3894	0,5306	0,6837	0,8551	1,0567	1,3137	1,7033	2,0518	2,4727	2,7707				
28	0,1268	0,2558	0,3893	0,5304	0,6834	0,8546	1,0560	1,3125	1,7011	2,0484	2,4671	2,7633				
29	0,1268	0,2557	0,3892	0,5302	0,6830	0,8542	1,0553	1,3114	1,6991	2,0452	2,4620	2,7564				
30	0,1267	0,2556	0,3890	0,5300	0,6828	0,8538	1,0547	1,3104	1,6973	2,0423	2,4570	2,7500				
40	0,1265	0,2550	0,3881	0,5286	0,6807	0,8507	1,0500	1,3031	1,6839	2,0211	2,4233	2,7045				
50	0,1263	0,2547	0,3875	0,5278	0,6794	0,8489	1,0473	1,2987	1,6759	2,0086	2,4033	2,6778				
100	0,1260	0,2540	0,3864	0,5261	0,6770	0,8452	1,0418	1,2901	1,6602	1,9840	2,3642	2,6259				
Inf.	0,1257	0,2533	0,3853	0,5244	0,6745	0,8416	1,0364	1,2816	1,6449	1,9600	2,3263	2,5758				

Distribuição F ($P = 10\% \text{ monoaural}, P = 20\% \text{ biaural}$)

G.I. do denomin.	Graus de liberdade do numerador																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	39,86	49,50	53,59	55,83	57,24	58,20	58,91	59,44	59,86	60,19	60,47	60,71	60,90	61,07	61,35	61,46	61,57	61,66	61,74	
2	8,53	9,00	9,16	9,24	9,29	9,33	9,35	9,37	9,38	9,39	9,40	9,41	9,41	9,42	9,42	9,43	9,43	9,44	9,44	9,44
3	5,54	5,46	5,39	5,34	5,31	5,28	5,27	5,25	5,24	5,23	5,22	5,21	5,20	5,20	5,20	5,20	5,19	5,19	5,19	5,18
4	4,54	4,32	4,19	4,11	4,05	4,01	3,98	3,95	3,94	3,92	3,91	3,90	3,89	3,88	3,87	3,86	3,85	3,85	3,84	3,84
5	4,06	3,78	3,62	3,52	3,45	3,40	3,37	3,34	3,32	3,30	3,28	3,27	3,26	3,25	3,24	3,23	3,22	3,22	3,21	3,21
6	3,78	3,46	3,29	3,18	3,11	3,05	3,01	2,98	2,96	2,94	2,92	2,90	2,89	2,88	2,87	2,86	2,85	2,84	2,84	2,84
7	3,59	3,26	3,07	2,96	2,88	2,83	2,78	2,75	2,72	2,70	2,68	2,67	2,65	2,64	2,63	2,62	2,61	2,60	2,59	2,59
8	3,46	3,11	2,92	2,81	2,73	2,67	2,62	2,59	2,56	2,54	2,52	2,50	2,49	2,48	2,46	2,45	2,45	2,44	2,43	2,42
9	3,36	3,01	2,81	2,69	2,61	2,55	2,51	2,47	2,44	2,42	2,40	2,38	2,36	2,35	2,34	2,33	2,32	2,31	2,30	2,30
10	3,29	2,92	2,73	2,61	2,52	2,46	2,41	2,38	2,35	2,32	2,30	2,28	2,27	2,26	2,24	2,23	2,22	2,21	2,20	2,20
11	3,23	2,86	2,66	2,54	2,45	2,39	2,34	2,30	2,27	2,25	2,23	2,21	2,19	2,18	2,17	2,16	2,15	2,14	2,13	2,12
12	3,18	2,81	2,61	2,48	2,39	2,33	2,28	2,24	2,21	2,19	2,17	2,15	2,13	2,12	2,10	2,09	2,08	2,07	2,06	2,06
13	3,14	2,76	2,56	2,43	2,35	2,28	2,23	2,20	2,16	2,14	2,12	2,10	2,08	2,07	2,05	2,04	2,03	2,02	2,01	2,01
14	3,10	2,73	2,52	2,39	2,31	2,24	2,19	2,15	2,12	2,10	2,07	2,05	2,04	2,02	2,01	2,00	1,99	1,98	1,97	1,96
15	3,07	2,70	2,49	2,36	2,27	2,21	2,16	2,12	2,09	2,06	2,04	2,02	2,00	1,99	1,97	1,96	1,95	1,94	1,93	1,92
16	3,05	2,67	2,46	2,33	2,24	2,18	2,13	2,09	2,06	2,03	2,01	1,99	1,97	1,95	1,94	1,93	1,92	1,91	1,90	1,89
17	3,03	2,64	2,44	2,31	2,22	2,15	2,10	2,06	2,03	2,00	1,98	1,96	1,94	1,93	1,91	1,90	1,89	1,88	1,87	1,86
18	3,01	2,62	2,42	2,29	2,20	2,13	2,08	2,04	2,00	1,98	1,95	1,93	1,92	1,90	1,89	1,87	1,86	1,85	1,84	1,84
19	2,99	2,61	2,40	2,27	2,18	2,11	2,06	2,02	1,98	1,96	1,93	1,91	1,89	1,88	1,86	1,85	1,84	1,83	1,82	1,81
20	2,97	2,59	2,38	2,25	2,16	2,09	2,04	2,00	1,96	1,94	1,91	1,89	1,87	1,86	1,84	1,83	1,82	1,81	1,80	1,79
25	2,92	2,53	2,32	2,18	2,09	2,02	1,97	1,93	1,89	1,87	1,84	1,82	1,80	1,79	1,77	1,76	1,75	1,74	1,73	1,72
30	2,88	2,49	2,28	2,14	2,05	1,98	1,93	1,88	1,85	1,82	1,79	1,77	1,75	1,74	1,72	1,71	1,70	1,69	1,68	1,67
40	2,84	2,44	2,23	2,09	2,00	1,93	1,87	1,83	1,79	1,76	1,74	1,71	1,70	1,68	1,66	1,65	1,64	1,62	1,61	1,61
50	2,81	2,41	2,20	2,06	1,97	1,90	1,84	1,80	1,76	1,73	1,70	1,68	1,66	1,64	1,63	1,61	1,60	1,59	1,58	1,57
60	2,79	2,39	2,18	2,04	1,95	1,87	1,82	1,77	1,74	1,71	1,68	1,66	1,64	1,62	1,60	1,59	1,58	1,56	1,55	1,54
70	2,78	2,38	2,16	2,03	1,93	1,86	1,80	1,76	1,72	1,69	1,66	1,64	1,62	1,60	1,59	1,57	1,56	1,55	1,54	1,53
80	2,77	2,37	2,15	2,02	1,92	1,85	1,79	1,75	1,71	1,68	1,65	1,63	1,61	1,59	1,57	1,56	1,55	1,53	1,52	1,51
90	2,76	2,36	2,15	2,01	1,91	1,84	1,78	1,74	1,70	1,67	1,64	1,62	1,60	1,58	1,56	1,55	1,54	1,52	1,51	1,50
100	2,76	2,36	2,14	2,00	1,91	1,83	1,78	1,73	1,69	1,66	1,64	1,61	1,59	1,57	1,56	1,54	1,53	1,52	1,50	1,49
1.000	2,71	2,31	2,09	1,95	1,85	1,78	1,72	1,68	1,64	1,61	1,58	1,55	1,53	1,51	1,49	1,46	1,45	1,44	1,43	1,43

Distribuição F ($P = 5\%$ monocaudal, $\alpha = 10\%$ bicaudal)

G.L do denomin.	Graus de liberdade do numerador																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9	243,0	243,9	244,7	245,4	245,9	246,5	246,9	247,3	247,7	248,0
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,42	19,43	19,44	19,44	19,44	19,44	19,44	19,44	19,45	
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,76	8,74	8,73	8,71	8,70	8,69	8,68	8,67	8,67	8,66
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,94	5,91	5,89	5,87	5,86	5,84	5,83	5,82	5,81	5,80
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,70	4,68	4,66	4,64	4,62	4,60	4,59	4,58	4,57	4,56
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00	3,98	3,96	3,94	3,92	3,91	3,90	3,88	3,87
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,60	3,57	3,55	3,53	3,51	3,49	3,48	3,47	3,46	3,44
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,31	3,28	3,26	3,24	3,22	3,20	3,19	3,17	3,16	3,15
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,10	3,07	3,05	3,03	3,01	2,99	2,97	2,96	2,95	2,94
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,94	2,91	2,89	2,86	2,85	2,83	2,81	2,80	2,79	2,77
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,82	2,79	2,76	2,74	2,72	2,70	2,69	2,67	2,66	2,65
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,72	2,69	2,66	2,64	2,62	2,60	2,58	2,57	2,56	2,54
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,63	2,60	2,58	2,55	2,53	2,51	2,50	2,48	2,47	2,46
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,57	2,53	2,51	2,48	2,46	2,44	2,43	2,41	2,40	2,39
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,51	2,48	2,45	2,42	2,40	2,38	2,37	2,35	2,34	2,33
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,46	2,42	2,40	2,37	2,35	2,33	2,32	2,30	2,29	2,28
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,41	2,38	2,35	2,33	2,31	2,29	2,27	2,26	2,24	2,23
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,37	2,34	2,31	2,29	2,27	2,25	2,23	2,22	2,20	2,19
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,34	2,31	2,28	2,26	2,23	2,21	2,20	2,18	2,17	2,16
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,31	2,28	2,25	2,22	2,20	2,18	2,17	2,15	2,14	2,12
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,20	2,16	2,14	2,11	2,09	2,07	2,05	2,04	2,02	2,01
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,13	2,09	2,06	2,04	2,01	1,99	1,98	1,96	1,95	1,93
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,04	2,00	1,97	1,95	1,92	1,90	1,89	1,87	1,85	1,84
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03	1,99	1,95	1,92	1,89	1,86	1,84	1,81	1,80	1,78	
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,95	1,92	1,89	1,86	1,84	1,82	1,80	1,78	1,76	1,75
70	3,98	3,13	2,74	2,50	2,35	2,23	2,14	2,07	2,02	1,97	1,93	1,89	1,86	1,84	1,81	1,79	1,77	1,75	1,74	1,72
80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00	1,95	1,91	1,88	1,84	1,82	1,79	1,75	1,73	1,72	1,70	
90	3,95	3,10	2,71	2,47	2,32	2,20	2,11	2,04	1,99	1,94	1,90	1,86	1,83	1,80	1,78	1,76	1,74	1,72	1,70	1,69
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93	1,89	1,85	1,82	1,79	1,77	1,75	1,73	1,71	1,69	1,68
1.000	3,85	3,00	2,61	2,38	2,22	2,11	2,02	1,95	1,89	1,84	1,80	1,76	1,73	1,70	1,66	1,63	1,61	1,60	1,58	

Distribuição F ($P = 1\%$ mono-caudal, $P = 2\%$ bicaudal)

G.L. do denomin.	Graus de liberdade do numerador																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	4052,2	4999,5	5403,4	5624,6	5763,6	5859,0	5928,4	5981,1	6022,5	6055,8	6083,3	6106,3	6125,9	6142,7	6157,3	6170,1	6181,4	6191,5	6200,6	6208,7
2	98,50	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,39	99,40	99,41	99,42	99,43	99,44	99,44	99,44	99,44	99,44	99,45	99,45
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,35	27,23	27,13	27,05	26,98	26,92	26,87	26,83	26,79	26,75	26,72	26,69
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,55	14,45	14,37	14,31	14,25	14,20	14,15	14,11	14,08	14,05	14,02
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05	9,96	9,89	9,82	9,77	9,72	9,68	9,64	9,61	9,58	9,55
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72	7,66	7,60	7,56	7,52	7,48	7,45	7,42	7,40
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62	6,54	6,47	6,41	6,36	6,31	6,28	6,24	6,21	6,18	6,16
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,73	5,67	5,61	5,56	5,52	5,48	5,44	5,41	5,38	5,36
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,18	5,11	5,05	5,01	4,96	4,92	4,89	4,86	4,83	4,81
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,77	4,71	4,65	4,60	4,56	4,52	4,49	4,46	4,43	4,41
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,46	4,40	4,34	4,29	4,25	4,21	4,18	4,15	4,12	4,10
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,22	4,16	4,10	4,05	4,01	3,97	3,94	3,91	3,88	3,86
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96	3,91	3,86	3,82	3,78	3,75	3,72	3,69	3,66
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80	3,75	3,70	3,66	3,62	3,59	3,56	3,53	3,51
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67	3,61	3,56	3,52	3,49	3,45	3,42	3,40	3,37
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,62	3,55	3,50	3,45	3,41	3,37	3,34	3,31	3,28	3,26
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,52	3,46	3,40	3,35	3,31	3,27	3,24	3,21	3,19	3,16
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51	3,43	3,37	3,32	3,27	3,23	3,19	3,16	3,13	3,10	3,08
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,36	3,30	3,24	3,19	3,15	3,12	3,08	3,05	3,03	3,00
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37	3,29	3,23	3,18	3,13	3,09	3,05	3,02	2,99	2,96	2,94
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22	3,13	3,06	2,99	2,94	2,89	2,85	2,81	2,78	2,75	2,72	2,70
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98	2,91	2,84	2,79	2,74	2,70	2,66	2,63	2,60	2,57	2,55
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80	2,73	2,66	2,61	2,56	2,52	2,48	2,45	2,42	2,39	2,37
50	7,17	5,06	4,20	3,72	3,41	3,19	3,02	2,89	2,78	2,70	2,63	2,56	2,51	2,46	2,42	2,38	2,35	2,32	2,29	2,27
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63	2,56	2,50	2,44	2,39	2,35	2,31	2,28	2,25	2,22	2,20
70	7,01	4,92	4,07	3,60	3,29	3,07	2,91	2,78	2,67	2,59	2,51	2,45	2,40	2,35	2,31	2,27	2,23	2,20	2,18	2,15
80	6,96	4,88	4,04	3,56	3,26	3,04	2,87	2,74	2,64	2,55	2,48	2,42	2,36	2,31	2,27	2,23	2,20	2,17	2,14	2,12
90	6,93	4,85	4,01	3,53	3,23	3,01	2,84	2,72	2,61	2,52	2,45	2,39	2,33	2,29	2,24	2,21	2,17	2,14	2,11	2,09
100	6,90	4,82	3,98	3,51	3,21	2,99	2,82	2,69	2,59	2,50	2,43	2,37	2,31	2,27	2,22	2,19	2,15	2,12	2,09	2,07
1.000	6,66	4,63	3,80	3,34	3,04	2,82	2,66	2,53	2,43	2,34	2,27	2,20	2,15	2,10	2,06	2,02	1,98	1,95	1,92	1,90

Valores-limite da estatística de Durbin-Watson – (5% de significância)										
n	k' = 1		k' = 2		k' = 3		k' = 4		k' = 5	
	d _l	d _u								
15	1,08	1,36	0,95	1,54	0,82	1,75	0,69	1,97	0,56	2,21
16	1,10	1,37	0,98	1,54	0,86	1,73	0,74	1,93	0,62	2,15
17	1,13	1,38	1,02	1,54	0,90	1,71	0,78	1,90	0,67	2,10
18	1,16	1,39	1,05	1,53	0,93	1,69	0,82	1,87	0,71	2,06
19	1,18	1,40	1,08	1,53	0,97	1,68	0,86	1,85	0,75	2,02
20	1,20	1,41	1,10	1,54	1,00	1,68	0,90	1,83	0,79	1,99
21	1,22	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,96
22	1,24	1,43	1,15	1,54	1,05	1,66	0,96	1,80	0,86	1,94
23	1,26	1,44	1,17	1,54	1,08	1,66	0,99	1,79	0,90	1,92
24	1,27	1,45	1,19	1,55	1,10	1,66	1,01	1,78	0,93	1,90
25	1,29	1,45	1,21	1,55	1,12	1,66	1,04	1,77	0,95	1,89
26	1,30	1,46	1,22	1,55	1,14	1,65	1,06	1,76	0,98	1,88
27	1,32	1,47	1,24	1,56	1,16	1,65	1,08	1,76	1,01	1,86
28	1,33	1,48	1,26	1,56	1,18	1,65	1,10	1,75	1,03	1,85
29	1,34	1,48	1,27	1,56	1,20	1,65	1,12	1,74	1,05	1,84
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83
31	1,36	1,50	1,30	1,57	1,23	1,65	1,16	1,74	1,09	1,83
32	1,37	1,50	1,31	1,57	1,24	1,65	1,18	1,73	1,11	1,82
33	1,38	1,51	1,32	1,58	1,26	1,65	1,19	1,73	1,13	1,81
34	1,39	1,51	1,33	1,58	1,27	1,65	1,21	1,73	1,15	1,81
35	1,40	1,52	1,34	1,58	1,28	1,65	1,22	1,73	1,16	1,80
36	1,41	1,52	1,35	1,59	1,29	1,65	1,24	1,73	1,18	1,80
37	1,42	1,53	1,36	1,59	1,31	1,66	1,25	1,72	1,19	1,80
38	1,43	1,54	1,37	1,59	1,32	1,66	1,26	1,72	1,21	1,79
39	1,43	1,54	1,38	1,60	1,33	1,66	1,27	1,72	1,22	1,79
40	1,44	1,54	1,39	1,60	1,34	1,66	1,29	1,72	1,23	1,79
45	1,48	1,57	1,43	1,62	1,38	1,67	1,34	1,72	1,29	1,78
50	1,50	1,59	1,46	1,63	1,42	1,67	1,38	1,72	1,34	1,77
55	1,53	1,60	1,49	1,64	1,45	1,68	1,41	1,72	1,38	1,77
60	1,55	1,62	1,51	1,65	1,48	1,69	1,44	1,73	1,41	1,77
65	1,57	1,63	1,54	1,66	1,50	1,70	1,47	1,73	1,44	1,77
70	1,58	1,64	1,55	1,67	1,52	1,70	1,49	1,74	1,46	1,77
75	1,60	1,65	1,57	1,68	1,54	1,71	1,51	1,74	1,49	1,77
80	1,61	1,66	1,59	1,69	1,56	1,72	1,53	1,74	1,51	1,77
85	1,62	1,67	1,60	1,70	1,57	1,72	1,55	1,75	1,52	1,77
90	1,63	1,68	1,61	1,70	1,59	1,73	1,57	1,75	1,54	1,78
95	1,64	1,69	1,62	1,71	1,60	1,73	1,58	1,75	1,56	1,78
100	1,65	1,69	1,63	1,72	1,61	1,74	1,59	1,76	1,57	1,78

n = número de observações

k' = número de variáveis explicativas (excluindo o intercepto)

Fonte: Biometrika, v. 41, p. 173, 1951.

Valores-limite da estatística de Durbin-Watson – (1% de significância)										
n	$k^2 = 1$		$k^2 = 2$		$k^2 = 3$		$k^2 = 4$		$k^2 = 5$	
	d_l	d_s								
15	0,81	1,07	0,70	1,25	0,59	1,46	0,49	1,70	0,39	1,96
16	0,84	1,09	0,74	1,25	0,63	1,44	0,53	1,66	0,44	1,90
17	0,87	1,10	0,77	1,25	0,67	1,43	0,57	1,63	0,48	1,85
18	0,90	1,12	0,80	1,26	0,71	1,42	0,61	1,60	0,52	1,80
19	0,93	1,13	0,83	1,26	0,74	1,41	0,65	1,58	0,56	1,77
20	0,95	1,15	0,86	1,27	0,77	1,41	0,68	1,57	0,60	1,74
21	0,97	1,16	0,89	1,27	0,80	1,41	0,72	1,55	0,63	1,71
22	1,00	1,17	0,91	1,28	0,83	1,40	0,75	1,54	0,66	1,69
23	1,02	1,19	0,94	1,29	0,86	1,40	0,77	1,53	0,70	1,67
24	1,04	1,20	0,96	1,30	0,88	1,41	0,80	1,53	0,72	1,66
25	1,05	1,21	0,98	1,30	0,90	1,41	0,83	1,52	0,75	1,65
26	1,07	1,22	1,00	1,31	0,93	1,41	0,85	1,52	0,78	1,64
27	1,09	1,23	1,02	1,32	0,95	1,41	0,88	1,51	0,81	1,63
28	1,10	1,24	1,04	1,32	0,97	1,41	0,90	1,51	0,83	1,62
29	1,12	1,25	1,05	1,33	0,99	1,42	0,92	1,51	0,85	1,61
30	1,13	1,26	1,07	1,34	1,01	1,42	0,94	1,51	0,88	1,61
31	1,15	1,27	1,08	1,34	1,02	1,42	0,96	1,51	0,90	1,60
32	1,16	1,28	1,10	1,35	1,04	1,43	0,98	1,51	0,92	1,60
33	1,17	1,29	1,11	1,36	1,05	1,43	1,00	1,51	0,94	1,59
34	1,18	1,30	1,13	1,36	1,07	1,43	1,01	1,51	0,95	1,59
35	1,19	1,31	1,14	1,37	1,08	1,44	1,03	1,51	0,97	1,59
36	1,21	1,32	1,15	1,38	1,10	1,44	1,04	1,51	0,99	1,59
37	1,22	1,32	1,16	1,38	1,11	1,45	1,06	1,51	1,00	1,59
38	1,23	1,33	1,18	1,39	1,12	1,45	1,07	1,52	1,02	1,58
39	1,24	1,34	1,19	1,39	1,14	1,45	1,09	1,52	1,03	1,58
40	1,25	1,34	1,20	1,40	1,15	1,46	1,10	1,52	1,05	1,58
45	1,29	1,38	1,24	1,42	1,20	1,48	1,16	1,53	1,11	1,58
50	1,32	1,40	1,28	1,45	1,24	1,49	1,20	1,54	1,16	1,59
55	1,36	1,43	1,32	1,47	1,28	1,51	1,25	1,55	1,21	1,59
60	1,38	1,45	1,35	1,48	1,32	1,52	1,28	1,56	1,25	1,60
65	1,41	1,47	1,38	1,50	1,35	1,53	1,31	1,57	1,28	1,61
70	1,43	1,49	1,40	1,52	1,37	1,55	1,34	1,58	1,31	1,61
75	1,45	1,50	1,42	1,53	1,39	1,56	1,37	1,59	1,34	1,62
80	1,47	1,52	1,44	1,54	1,42	1,57	1,39	1,60	1,36	1,62
85	1,48	1,53	1,46	1,55	1,43	1,58	1,41	1,60	1,39	1,63
90	1,50	1,54	1,47	1,56	1,45	1,59	1,43	1,61	1,41	1,64
95	1,51	1,55	1,49	1,57	1,47	1,60	1,45	1,62	1,42	1,64
100	1,52	1,56	1,50	1,58	1,48	1,60	1,46	1,63	1,44	1,65

n = número de observações

k^2 = número de variáveis explicativas (excluindo o intercepto)

Fonte: *Biometrika*, v. 41, p. 173, 1951.

**Valores-limite da estatística τ
(teste de raízes unitárias de Dickey-Fuller):**

n	Sem intercepto ou tendência ($a = 0$ e $b = 0$) τ							
	1%	2,5%	5%	10%	90%	95%	97,5%	99%
25	-2,66	-2,26	-1,95	-1,60	0,92	1,33	1,70	2,16
50	-2,62	-2,25	-1,95	-1,61	0,91	1,31	1,66	2,08
100	-2,60	-2,24	-1,95	-1,61	0,90	1,29	1,64	2,03
250	-2,58	-2,23	-1,95	-1,62	0,89	1,29	1,63	2,01
500	-2,58	-2,23	-1,95	-1,62	0,89	1,28	1,62	2,00
∞	-2,58	-2,23	-1,95	-1,62	0,89	1,28	1,62	2,00

n	Com intercepto e sem tendência ($\alpha \neq 0$ e $\beta = 0$) τ_{μ}							
	1%	2,5%	5%	10%	90%	95%	97,5%	99%
25	-3,75	-3,33	-3,00	-2,62	-0,37	0,00	0,34	0,72
50	-3,58	-3,22	-2,93	-2,60	-0,40	-0,03	0,29	0,66
100	-3,51	-3,17	-2,89	-2,58	-0,42	-0,05	0,26	0,63
250	-3,46	-3,14	-2,88	-2,57	-0,42	-0,06	0,24	0,62
500	-3,44	-3,13	-2,87	-2,57	-0,43	-0,07	0,24	0,61
∞	-3,43	-3,12	-2,86	-2,57	-0,44	-0,07	0,23	0,60

n	Com intercepto e tendência ($\alpha \neq 0$ e $\beta \neq 0$) τ_{ϵ}							
	1%	2,5%	5%	10%	90%	95%	97,5%	99%
25	-4,38	-3,95	-3,60	-3,24	-1,14	-0,80	-0,50	-0,15
50	-4,15	-3,80	-3,50	-3,18	-1,19	-0,87	-0,58	-0,24
100	-4,04	-3,73	-3,45	-3,15	-1,22	-0,90	-0,62	-0,28
250	-3,99	-3,69	-3,43	-3,13	-1,23	-0,92	-0,64	-0,31
500	-3,98	-3,68	-3,42	-3,13	-1,24	-0,93	-0,65	-0,32
∞	-3,96	-3,66	-3,41	-3,12	-1,25	-0,94	-0,66	-0,33

Fonte: FULLER, Wayne. *Introduction to Statistical Time Series*. Nova York: John Wiley, 1976.

Valores críticos para testes de cointegração baseados em resíduos de regressões

Número de variáveis (n + 1)	Tamanho da amostra	Valores críticos		
		10%	5%	1%
2	50	3,28	3,67	4,32
	100	3,03	3,37	4,07
	200	3,02	3,37	4,00
3	50	3,73	4,11	4,84
	100	3,59	3,93	4,45
	200	3,47	3,78	4,35
4	50	4,02	4,35	4,94
	100	3,89	4,22	4,75
	200	3,89	4,18	4,70
5	50	4,42	4,76	5,41
	100	4,26	4,58	5,18
	200	4,18	4,48	5,02

Fonte: ENGLE, Robert F.; YOO, Byung Sam. Forecasting and Testing in Co-Integrated Systems.
Journal of Econometrics, v. 35, p. 157, 1987.

Referências Bibliográficas

- BUSSAB, W.; MORETTIN, P. *Estatística básica*. São Paulo: Saraiva, 2002.
- DAVIDSON, R.; MACKINNON, J. *Estimation and Inference in Econometrics*. New York: Oxford University Press, 1993.
- ENDERS, W. *Applied Econometric Time Series*. New York: John Wiley & Sons, 1994.
- ENDO, S. *Números-índices*. São Paulo: Atual, 1986.
- FISHER, R. *Statistical Methods and Scientific Inference*. Edinburgh: Oliver & Boyd, 1959.
- GOLDBERG, A. *A Course in Econometrics*. Cambridge: Harvard University Press, 1991.
- GREENE, W. *Econometric Analysis*. Upper Saddle River: Prentice Hall, 1997.
- GUJARATI, D. *Econometria básica*. São Paulo: Makron, 2000.
- HAMILTON, J. *Time Series Analysis*. Princeton: Princeton University Press, 1994.
- HILL, C.; GRIFFITHS, W.; JUDGE, G. *Econometria*. São Paulo: Saraiva, 2000.
- HOFFMANN, R. *Estatística para economistas*. Rio de Janeiro: Pioneira, 1998.
- KMENTA, J. *Elementos de econometria*. São Paulo: Atual, 1988.
- MADDALA, G. *Introduction to Econometrics*. New York: MacMillan, 1988.
- MEYER, P. *Probabilidade: aplicações à estatística*. São Paulo: Livros Técnicos e Científicos, 1983.
- MITTLEHAMMER, R. *Mathematical Statistics for Economics and Business*. New York: Springer-Verlag, 1996.
- PINDYCK, R.; RUBINFELD, D. *Econometric Models and Economic Forecasts*. New York: McGraw-Hill, 1998.
- RAMANATHAN, R. *Introductory Econometrics*. Fort Worth: The Dryden Press, 1989.
- _____. *Statistical Methods in Econometrics*. San Diego: Academic Press, 1993.
- ROUSSAS, G. *A Course in Mathematical Statistics*. San Diego: Academic Press, 1997.
- SPIEGEL, M. R. *Estatística*. São Paulo: Makron, 1994.
- SVESHNIKOV, A. *Problems in Probability Theory, Mathematical Statistics and Theory of Random Functions*. New York: Dover, 1968.
- VANDAELE, W. *Applied Time Series and Box-Jenkins Models*. New York: Academic Press, 1983.
- VASCONCELLOS, M.; ALVES, D. *Manual de econometria*. São Paulo: Atlas, 2000.
- VINIOTIS, Y. *Probability and Random Processes*. New York: McGraw-Hill, 1998.

Índice Remissivo

A

Alta correlação, 283
Amostra, 156
retirada com
reposição, 180
sem reposição, 181

Análise
combinatória, 39-42
de variância, 237

ARIMA, 330, 332, 333, 342, 346, 347

ARMA, 331, 332, 339, 341, 342

Arranjos, 41

Assintóticas
propriedades, 172, 175, 192

Atribuir pesos, 46

Atributo qualitativo, 258

Autocorrelação dos erros,
conceito de, 289
consequências da, 290
de ordem 290
identificação, 291
providências, 296

Autorregressivo
AR(1), 290, 330, 333, 335, 336, 342,
347, 349, 351, 396
AR(2), 290, 330
AR(p), 330, 331, 335, 348, 351
Processo, 290, 324, 330, 331, 335, 337

Axiomas, 42-44

B

Bimodal, 51, 52

C

Cálculo diferencial e integral, 118

Cauda, 200
Classe dos estimadores lineares, 172

Classes, 49

Coeficiente de
assimetria, 113
correlação, 59-61
negativo, 364
positivo, 364
propriedades, 61, 62
determinação, 238
variação, 55

Cointegração, 351-353

Combinações, 41

Condição invertida, 33

Condições de
estacionariedade, 47, 48
invertibilidade, 47, 48

Conjunto

contínuo, 19, 20, 22
discreto, 19, 69

Convergência
em distribuição, 178
em probabilidade, 191
quase certa, 191

Covariância, 57, 58, 61, 62, 64, 72,
132-134, 257, 333

Critério de informação de
Akaike, 262, 263, 279, 346
Schwarz, 262, 263, 279, 346

Curtose, 113

D

Dados agrupados, 50, 52
variância a partir de, 41

Deflacionamento de séries, 373

Derivadas, 118, 120
 Desvio padrão
 conceito de, 54, 55
 Diagrama de Venn, 22
 Disjuntos, 24, 27
 Distribuição
 achatamento de uma, 113
 log-normal, 121
 normal bivariada, 145, 154
 Distribuição de probabilidade, 45, 69
 binomial, 75, 89
 com três variáveis, 153
 conceito de, 69
 conjunta
 contínua, 134, 150
 discreta, 124
 de Bernoulli, 74
 média e variância
 de Poisson, 82, 83, 90
 de variáveis aleatórias
 contínuas, 100-103
 discretas, 69
 exponencial, 98
 F, de Fisher-Snedecor, 218, 229
 gaussiana, 106
 geométrica, 90
 hipergeométrica, 90
 leptocúrtica, 113
 log-normal, 121
 marginal, 126
 mesocúrtica, 113
 normal, 103-110, 209
 padronizada, 106
 platicúrtica, 113
 qui-quadrado, 209, 296
 t, de Student, 214
 uniforme
 contínua, 91
 discreta, 73
 Distribuições contínuas, 91-123
 conjuntas, 126, 134
 independência em uma, 152

E
 Elasticidade, 250
 Empate técnico, 193
 Equação
 da demanda, 313
 de oferta, 311, 313
 identificação da, 313, 314
 condição de posto, 316
 exatamente identificada, 313, 314
 subidentificada, 314
 superidentificada, 314
 Equações simultâneas, 309, 316
 condição de
 ordem, 314
 posto, 316
 estimação de, 316
 exatamente identificadas, 313, 314
 forma
 estrutural, 311
 reduzida, 311
 identificação, 311
 problema de, 309
 subidentificadas, 313, 314
 superidentificadas, 314
 Erro
 do tipo I, 220
 do tipo II, 220
 Erro quadrático
 médio, 165
 menor, 165
 Espaço
 amostral, 17, 21, 43
 de eventos, 43
 de probabilidade, 43
 Esperança
 conceito de, 69-70
 condicionada, 129, 153
 de matrizes, 275
 de variáveis contínuas, 91
 matemática, 69-72
 Estimação, 156-192
 por máxima verossimilhança, 183

E

- Estimador
 assintoticamente não viesado, 173, 175
 conceito de, 156
 consistente, 173-175
 da média, 174, 182
 da variância, 168, 171, 182
 de máxima verossimilhança, 183, 277
 eficiente, 162, 165
 assintoticamente, 192
 MELiNV, 171, 172, 249
 não viesado, 158-162, 171
 relativamente mais eficiente, 165, 166
 viesado, 160

E

- Eventos
 dependentes, 27
 independentes, 27
 mutuamente exclusivos, 27
- Expressões
 demonstrações, 66-68

F

- Fatorial, 39
- Frequência
 absoluta, 46
 relativa, 46
- Função de autocorrelação, 334, 335, 337
 parcial, 333, 335, 336
- Função de distribuição,
 acumulada, 70, 99
 conceito de, 70
 de variáveis contínuas, 99
- Função de probabilidade, 70
- Função densidade de probabilidade
 conceito de, 93
 condicional, 137
 conjunta, 134
 de uma distribuição normal, 103
 de variáveis transformadas, 110-112
 marginal, 136-139
- Funções
 máximos e mínimos de, 119
 primitivas, 119

G

- Graus de liberdade, 209

H

- Heterocedasticidade
 conceito de, 299
 consequências da, 299
 identificação, 299
 providências, 306

H

- Hipótese
 alternativa, 200
 básica, 281, 282
 violando a, 282-324
 de normalidade, 246
 nula, 200
 aceita-se a, 215
 rejeita-se a, 217, 220
- III, 309
 problemas da simultaneidade, 309

H

- Histograma, 49, 50

H**I**

- Importância, 58
- Índices de
 preços, 358
 no atacado, 375
 agregativo simples, 359
 ao consumidor, 375
 critérios de Fisher para, 368, 378
 da circularidade, 369
 da construção civil, 375
 da determinação, 369
 da homogeneidade, 368
 da proporcionalidade, 369
 da reversibilidade, 369
 de custo de vida, 375
 de Fisher, 370
 de identidade, 368
 de Laspeyres, 360, 361, 369

- de Marshall-Edgeworth, 371
- de Paasche, 360, 361, 364, 369
- de Sauerbeck, 378
- quantidades, 371
- valor, 371, 372
- Inferência estatística, 156
 - amostra, 156
 - população, 156
 - infinita, 157
- Integral, 119
- Intersecção, 23, 24
- Intervalo, 19, 20
 - aberto, 20
 - fechado, 20
- Intervalo de confiança
 - conceito de, 193
 - para a média, 194
 - para proporções, 198
 - para variância, 217
- L**
- Lei dos grandes números, 176, 191
- Leptocúrtica, 113
- Linha *inf*, 217
- M**
- Margens de erro, 75, 193, 195
- Matriz
 - conceito de, 272
 - derivada de, 275
 - determinante de uma, 271, 286
 - forma quadrática de uma, 274, 279
 - identidade, 254, 257, 273
 - inversa, 253, 273
 - operações com, 253, 272, 273
 - ordem dos fatores, 273
 - quadrado especial, 273
- Média
 - amostral, 158, 159, 160
 - aritmética, 45, 46, 48, 50
 - da média amostral, 195
 - binomial, 75, 89
 - da distribuição
 - de Bernoulli, 89
 - de Poisson, 82
 - F, 218
 - geométrica, 80
 - hipergeométrica, 81
 - normal, 103, 106
 - qui-quadrado, 209, 296
 - t, de Student, 214, 215
 - de dados agrupados, 50, 52
 - geométrica, 47
 - harmônica, 47, 367
 - ponderada, 46-48, 160
 - populacional, 158-160, 163
 - testando a, 213
- Média e variância
 - das principais distribuições discretas, 90
 - de uma distribuição de Bernoulli, 89
- Mediana, 51
- Médias móveis (MA)
 - MA(1), 331
 - MA(2), 331
 - MA(q), 331, 337
 - Processo de, 331
- Medidas de
 - dispersão, 52
 - posição central, 45-51
 - posição e dispersão, 45-68
- Mesocúrtica, 113
- Métodos ingênuos de previsão, 325
- Micronumerosidade, 289
- Mínimos quadrados
 - estimador de, 290, 324
 - consistência do, 280
 - métodos dos, 323
 - de dois estágios, 316
 - generalizados, 323, 324
 - indiretos, 316
 - ordinários, 290, 291, 299, 310, 342
 - ponderados, 306
 - propriedades dos estimadores de, 254

- Moda, 51
- Modelo
- ARIMA, 333-348
 - exponencial, 250
 - log-linear, 251
 - log-log, 250
 - multiplicativo, 250
 - não linear, 249
- Momentos de uma distribuição, 113-117
- Multicolinearidade
- conceito de, 283
 - consequências da, 285
 - exata, 283
 - identificação, 286
 - perfeita, 283
 - providências, 287
- Mutuamente exclusivos, 24
-
- N**
- Números-índice
- ano-base, 356
 - construção de, 355
 - mudança de base, 360
-
- O**
- Octis, 51
-
- P**
- Padronização, 109
- Parâmetro populacional, 157
- Parte integrante, 296
- Passeio aleatório, 325, 326
- Percentis, 51
- Permutações, 40
- Pesquisa eleitoral, 157, 180, 198
- Platicúrtica, 113
- Poder explicativo, 238
- População,
- conceito de, 156
 - finita, 180
 - infinita, 157
- Probabilidade, 17-44
- axiomas de, 42-43
 - conceito de, 17-20
 - condicional, 25-32, 127
 - da intersecção, 23
 - da união, 23
 - de distribuição normal, 103
 - de eventos
 - disjuntos, 25
 - independentes, 27
 - mutuamente exclusivos, 25, 27 - de variáveis contínuas, 99
 - definição
 - axiomática de, 42-44
 - frequentista de, 19
 - distribuição de, 45, 69-90
 - do e, 21-25
 - do ou, 21-25
 - espaço de, 43
 - função de, 70
 - subjetiva, 20-21
- Procedimento de Box e Jenkins
- conceito geral, 330
 - diagnóstico, 346
 - estimação, 342
 - identificação, 333
 - modelo univariado, 330
- Progressão geométrica, 87
- razão da, 87
- Propriedades assintóticas, 172, 175, 192
-
- Q**
- Quadrado da soma, 168
- Quartis, 51
-
- R**
- R^2 , 238
- ajustado, 262
- Random walk*, 325
- Região de aceitação, 203
- Regra
- de Bayes, 32-33
 - do ou, 25
- Regras de derivação, 118

- Regressão linear, 230-281
 conceito de, 230
 espúria, 330, 351
 hipóteses básicas, 282-324
 modelo
 exponencial (log-linear), 251
 multiplicativo (log-log), 250
 múltipla, 252, 379
 pelo método da máxima verossimilhança, 183, 277
 pelo método dos mínimos quadrados
 em dois estágios, 316
 generalizados, 323, 324
 indiretos, 316
 ordinários, 290, 291, 299, 310, 342
 ponderados, 306
 resíduos da, 237
 simples, 238, 352, 375
- Ruído branco, 330
-
- S**
- Séries de tempo, 325-353
 estacionárias, 326
 não estacionárias, 327, 330
- Soma dos quadrados
 dos resíduos, 237
 dos totais, 237
 explicados, 237
-
- T**
- Teorema
 de Gauss-Markov, 249, 276, 290, 299
 de Tchebichev, 91, 112, 121
 do limite central, 176, 178, 179
- Teoremas, 43-44
 demonstração dos, 120, 121
- Teste de
 Dickey-Fuller, 352, 389
 aumentado, 351
- Durbin-Watson, 291
- F, 239
- para selecionar modelos, 263
 não restrito, 263
 restrito, 263
- Goldfeld-Quandt, 321
- hipóteses, 199-207
 bicaudal, 200
 caso geral dos, 227
 conceito de, 200
 erros em um, 220
 inadmissível, 277
 mais poderoso, 277
 monocaudal, 200
 não viesado, 227
 para comparação de médias, 228
 quando a variância é desconhecida, 213
 para comparação de variâncias, 217
 para média, 201
 com variância desconhecida, 213
 para proporções, 206
 para variância, 208
 poder de um, 220
 propriedades desejáveis de, 227
 significância dos, 200
 Jarque-Bera, 246
 raízes unitárias, 349
 White, 303
- Transações possíveis, 157
- Triângulo de Pascal, 42
-
- U**
- União, 23
- Unimodal, 51
- Último valor, 70
-
- V**
- Valor
 médio esperado, 70
 verdadeiro da média populacional, 195
- Valores
 nominais, 373
 reais, 373, 374

- Variância, 52
amostral, 168
para população finita, 180
assintótica, 192
cálculo, 60
comparação de, 217
conceito de, 54
da distribuição
binomial, 75, 89
de Bernoulli, 89
de Poisson, 82, 83
 F , 218
geométrica, 80, 81
hipergeométrica, 81
normal, 103, 106, 113, 122, 145
qui-quadrado, 209, 296
 t , de Student, 214, 215, 349
da média amostral, 163
de dados agrupados, 50, 52
de estimadores, 162
de matrizes, 272
de variáveis contínuas, 99, 134
desconhecida, 213
testando a, 208-213
- Variáveis
aleatórias, 45
conceito de, 45
contínuas, 100-103
esperança e variância de, 100-103
independentes, 124, 132
dependentes, 139
dummy, 258
endógenas e exógenas, 310
forma reduzida e estrutural, 311
identificação, 311
independentes, 132
correlação entre, 59-60
transformações de, 110-112
- Verossimilhança
estimação por máxima, 183, 277
função de, 183
- Vetor aleatório, 124

