

- b. Que proporção desses estados apresenta um consumo *per capita* de energia dentro dos limites de ± 1 desvio-padrão em relação à média aritmética, dentro dos limites de ± 2 desvios-padrão em relação à média aritmética e dentro dos limites de ± 3 desvios-padrão em relação à média aritmética?
- c. Compare suas descobertas com aquilo que seria esperado com base na regra empírica. Você está surpreso com os resultados em (b)?
- d. Repita os procedimentos de (a) a (c), excluindo o Distrito de Columbia. De que modo os resultados se alteraram?
- 3.43** Trinta empresas compõem a DJIA (Dow Jones Industrial Average — Média Industrial Dow Jones). Qual o porte dessas

empresas? Um método habitual para mensurar o porte é usar o número de quotas de ações multiplicado pelo preço de uma quota de ações. Em 19 de junho de 2009, a capitalização de mercado para essas empresas variou de 20,5 bilhões de dólares para a Caterpillar até 347,4 bilhões de dólares para a ExxonMobil. A população inteira de valores para capitalização de mercado está registrada no arquivo [DowMC](#).

Fonte: Dados extraídos de [money.cnn.com](#), 19 de junho de 2009.

- a. Calcule a média aritmética e o desvio-padrão da capitalização de mercado para essa população de 30 empresas.
- b. Interprete os parâmetros calculados em (a).

3.5 A Covariância e o Coeficiente de Correlação

Na Seção 2.7, você utilizou gráficos de dispersão para examinar visualmente a relação entre duas variáveis numéricas. Esta seção apresenta duas medidas para a relação entre duas variáveis numéricas: a covariância e o coeficiente de correlação.

A Covariância

A **covariância** mede a força de uma relação linear entre duas variáveis numéricas (X e Y). A Equação (3.16) define a **covariância da amostra**, enquanto o Exemplo 3.17 ilustra o seu uso.

A COVARIÂNCIA DA AMOSTRA

$$\text{cov}(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n - 1} \quad (3.16)$$

EXEMPLO 3.17

Calculando a Covariância da Amostra

Na Figura 2.18, no início da Seção 2.7 do Capítulo 2, você construiu um gráfico de dispersão que ilustrava a relação entre a avaliação e a receita anual para os 30 times de basquete que compõem a National Basketball Association (NBA) (extraído de [www.forbes.com/lists/2008/32/nba08_NBA-Team-Valuations_MetroArea.html](#): armazenados no arquivo [ValoresNBA](#)). Agora, você deseja mensurar a associação entre o valor de uma franquia e a receita anual, calculando a covariância da amostra.

SOLUÇÃO A Tabela 3.8 fornece o valor e a receita anual para os 30 times.

A Figura 3.6 contém uma planilha que calcula a covariância para esses dados. A seção que corresponde à Área de Cálculos, na Figura 3.6, desmembra a Equação (3.16) em um conjunto de cálculos mais simples. Tomando como base a célula F9, ou utilizando diretamente a Equação (3.16) você descobre que a covariância é igual a 2.836,2069:

$$\begin{aligned} \text{cov}(X, Y) &= \frac{82.250}{30 - 1} \\ &= 2.836,2069 \end{aligned}$$

TABELA 3.8

Valores e Receitas Anuais dos 30 Times da NBA (em milhões de dólares)

Time	Valor	Receita
Atlanta	306	102
Boston	447	149
Charlotte	284	95
Chicago	504	165

TABELA 3.8

Continuação

Time	Valor	Receita
Cleveland	477	159
Dallas	466	153
Denver	329	112
Detroit	480	160
Golden State	335	112
Houston	469	156
Indiana	303	101
Los Angeles Clippers	297	99
Los Angeles Lakers	584	191
Memphis	294	95
Miami	393	131
Milwaukee	278	94
Minnesota	301	100
New Jersey	295	98
New Orleans	285	95
New York	613	208
Orlando	300	82
Philadelphia	349	100
Phoenix	360	116
Portland	452	148
Sacramento	307	114
San Antonio	350	117
Seattle	415	138
Toronto	400	138
Utah	358	119
Washington	353	118

FIGURA 3.6

Planilha do Microsoft Excel para a covariância entre o valor e as receitas anuais dos 30 times da NBA

A planilha FIGURA_3.6 da pasta de trabalho Capítulo 3 contém a Figura 3.6. Calcule a covariância utilizando as instruções na Seção GE3.5.

	A	B	C	D	E	F
1	Análise da Covariância					
2						
3	Valor	Receita	(X-XBarra)(Y-YBarra)			
4	306	102	1726,4667	Área de Cálculos		
5	447	149	1587,0333	XBarra	379,4667	
6	284	95	2911,7333	YBarra	125,5	
7	504	165	4919,0667	n - 1	29	
8	477	159	3267,3667	Soma	82250,0000	
9	466	153	2379,6667	Covariância	2836,2069	
10	329	112	681,3000			
11	480	160	3468,4000			
12	335	112	600,3000			
13	469	156	2730,7667			
14	303	101	1873,4333			
15	297	99	2185,3667			
16	584	191	13396,9333			
17	294	95	2606,7333			
18	393	131	74,4333			
19	278	94	3196,2000			
20	301	100	2000,9000			
21	295	98	2322,8333			
22	285	95	2881,2333			
23	613	208	19266,5000			
24	300	82	3456,8000			
25	349	100	776,9000			
26	360	116	184,9333			
27	452	148	1632,0000			
28	307	114	833,3667			
29	350	117	250,4667			
30	415	138	444,1667			
31	400	138	256,6667			
32	358	119	139,5333			
33	353	118	198,5000			

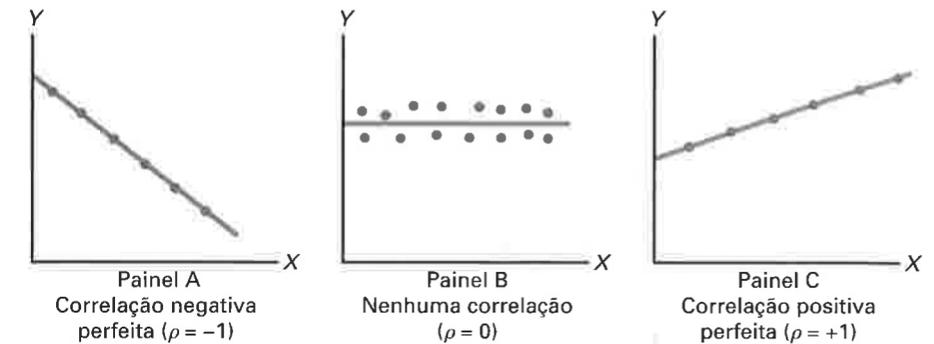
A covariância apresenta uma deficiência significativa como uma medida da relação linear entre duas variáveis numéricas. Uma vez que a covariância pode assumir qualquer valor, você não consegue utilizá-la para determinar a força relativa da relação. Em outras palavras, você não consegue afirmar se o valor 2.836,2069 é um indicativo de uma relação forte ou de uma relação fraca. Para melhor determinar a força relativa da relação, você precisa calcular o coeficiente de correlação.

O Coeficiente de Correlação

O coeficiente de correlação mede a força relativa de uma relação linear entre duas variáveis numéricas. Os valores para o coeficiente de correlação se estendem desde -1, para uma correlação negativa perfeita, até +1, para uma correlação positiva perfeita. *Perfeita*, nesse caso, significa dizer que, se os pontos fossem desenhados em um gráfico de dispersão, todos esses pontos poderiam ser interligados por meio de uma linha reta. Ao lidar com dados de populações para duas variáveis numéricas, a letra grega ρ (rô) é utilizada como o símbolo para o coeficiente de correlação. A Figura 3.7 ilustra três diferentes tipos de associação entre duas variáveis.

FIGURA 3.7

Tipos de associação entre variáveis



No Painel A da Figura 3.7, existe uma relação linear negativa perfeita entre X e Y . Por conseguinte, o coeficiente de correlação, ρ , é igual a -1 , e, quando X cresce, Y decresce de uma maneira perfeitamente previsível. O Painel B mostra uma situação na qual não existe nenhuma relação entre X e Y . Nesse caso, o coeficiente de correlação, ρ , é igual a 0 , e, à medida que X cresce, não existe nenhuma tendência para que Y cresça ou decresça. O Painel C ilustra uma relação positiva perfeita, na qual ρ é igual a $+1$. Nesse caso, Y cresce de uma maneira perfeitamente previsível quando X cresce.

A correlação, por si só, não consegue provar que existe um efeito de causalidade — ou seja, que a variação no valor de uma variável tenha causado a variação na outra variável. Uma correlação forte pode ser produzida simplesmente pelo acaso, pelo efeito de uma terceira variável não considerada no cálculo ou por uma relação do tipo causa e efeito. Você precisaria aprofundar sua análise para determinar qual dessas situações efetivamente produziu a correlação. Portanto, você pode afirmar que *causalidade implica correlação, mas correlação, por si só, não implica causalidade*.

A Equação (3.17) define o coeficiente de correlação da amostra (r).

COEFICIENTE DE CORRELAÇÃO DA AMOSTRA

$$r = \frac{\text{cov}(X, Y)}{S_X S_Y} \tag{3.17}$$

em que

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

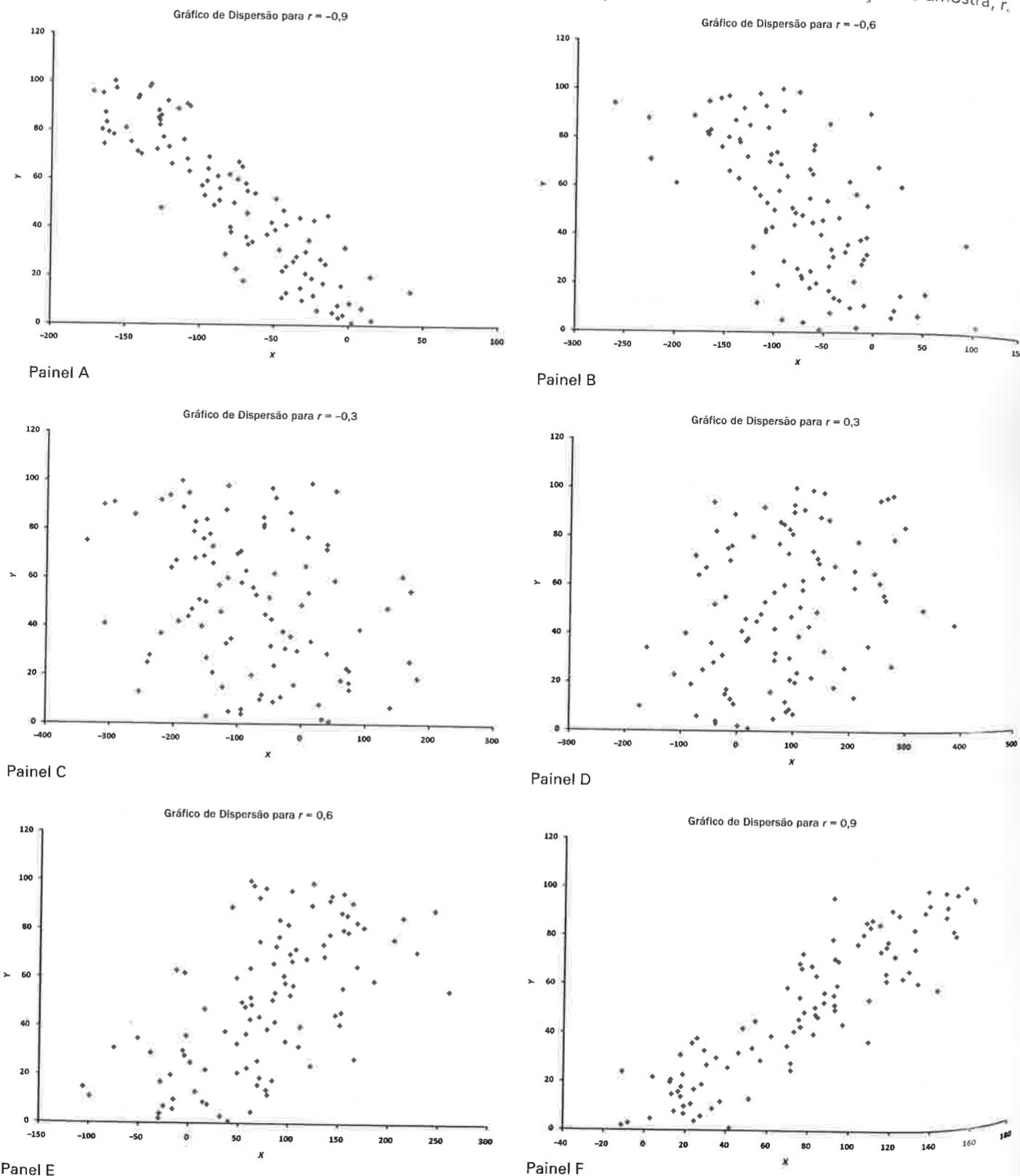
$$S_X = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}}$$

$$S_Y = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n - 1}}$$

Quando você tem em mãos dados de amostras, você pode calcular o coeficiente de correlação, r . Ao utilizar dados de amostras, é improvável que você venha a ter um coeficiente de correlação de amostra com valores exatos de +1, 0, ou -1. A Figura 3.8 apresenta gráficos de dispersão, juntamente com seus respectivos coeficientes de correlação de amostras, r , para seis conjuntos de dados, cada um dos quais contendo 100 valores para X e Y .

FIGURA 3.8

Seis gráficos de dispersão, elaborados pelo Microsoft Excel, e seus respectivos coeficientes de correlação de amostra, r .



EXEMPLO 3.18

Calculando o Coeficiente de Correlação da Amostra

No Exemplo 3.17, no início desta seção, você calculou a covariância das avaliações e receitas de 30 times de basquete da NBA. Utilizando a Figura 3.6 e a Equação (3.17), calcule o coeficiente de correlação da amostra.

SOLUÇÃO

$$r = \frac{\text{cov}(X, Y)}{S_X S_Y} = \frac{2.836,2069}{(91,9324)(31,2948)} = 0,9858$$

FIGURA 3.9

Planilha, elaborada pelo Microsoft Excel, para o coeficiente de correlação da amostra, r , entre as avaliações em termos de custo e receitas de 30 times de basquete da NBA

A FIGURA 3.9 exibe a planilha CÁLCULO da pasta de trabalho Correlação. Calcule o coeficiente de correlação utilizando as instruções na Seção GE3.5.

	A	B	C	D	E	F	G	H
1	Cálculos para o Coeficiente de Correlação							
2								
3	Valor	Receita	$(X-\bar{X})^2$	$(Y-\bar{Y})^2$	$(X-\bar{X})(Y-\bar{Y})$	Área de Cálculos		
4	306	102	5397,3511	552,2500	1726,4667	XBarra	379,4667	
5	447	149	4560,7511	552,2500	1587,0333	YBarra	125,5	
6	284	95	9113,8844	930,2500	2911,7333	n - 1	29	
7	504	165	15508,5511	1560,2500	4919,0667	Covariância	2836,2069	
8	477	159	9512,7511	1122,2500	3267,3667	S_X	91,9324	
9	466	153	7488,0178	756,2500	2379,6667	S_Y	31,2948	
10	329	112	2546,8844	182,2500	681,3000	r	0,9858	
11	480	160	10106,9511	1190,2500	3468,4000			
12	335	112	1977,2844	182,2500	600,3000			
13	469	156	8016,2178	930,2500	2730,7667			
14	303	101	5847,1511	600,2500	1873,4333			
15	297	99	6800,7511	702,2500	2185,3667			
16	584	191	41833,8844	4290,2500	13396,9333			
17	294	95	7304,5511	930,2500	2606,7333			
18	393	131	183,1511	30,2500	74,4333			
19	278	94	10295,4844	992,2500	3196,2000			
20	301	100	6157,0178	650,2500	2000,9000			
21	295	98	7134,6178	756,2500	2322,8333			
22	285	95	8923,9511	930,2500	2881,2333			
23	613	208	54537,8178	6806,2500	19266,5000			
24	300	82	6314,9511	1892,2500	3456,8000			
25	349	100	928,2178	650,2500	776,9000			
26	360	116	378,9511	90,2500	184,9333			
27	452	148	5261,0844	506,2500	1632,0000			
28	307	114	5251,4178	132,2500	833,3667			
29	350	117	868,2844	72,2500	250,4667			
30	415	138	1262,6178	156,2500	444,1667			
31	400	138	421,6178	156,2500	256,6667			
32	358	119	460,8178	42,2500	139,5333			
33	353	118	700,4844	56,2500	198,5000			
34	Somos:		245095,4667	28401,5	82250,0000			

A avaliação e as receitas dos times da NBA estão muito fortemente correlacionadas. Os times com as receitas mais baixas apresentam as avaliações mais baixas. Os times com as receitas mais altas apresentam as avaliações mais altas. Essa relação é muito forte, conforme indicado pelo coeficiente de correlação, $r = 0,9848$.

Embora, de modo geral, você não possa pressupor que simplesmente porque duas variáveis estão correlacionadas variações em uma variável tenham causado variações na outra variável, para esse exemplo faz sentido concluir que variações na receita causarão variações na avaliação de uma equipe.

Em resumo, o coeficiente de correlação indica a relação, ou associação, linear entre duas variáveis numéricas. Quando o coeficiente de correlação vai se aproximando de $+1$ ou -1 , a relação linear entre as duas variáveis numéricas vai se tornando mais forte. Quando o coeficiente de correlação se aproxima de 0 , existe pouca ou nenhuma relação linear. O sinal do coeficiente de correlação indica se os dados estão positivamente correlacionados (ou seja, os valores mais altos de X geralmente fazem par com os valores mais altos de Y), ou negativamente correlacionados (ou seja, os valores mais altos de X geralmente fazem par com os valores mais baixos de Y). A existência de uma forte correlação não implica necessariamente um efeito de causalidade. Indica tão somente as tendências presentes nos dados.

Problemas para a Seção 3.5

APRENDENDO O BÁSICO

3.44 Apresenta-se a seguir um conjunto de dados a partir de uma amostra de $n = 11$ itens:

X	7	5	8	3	6	10	12	4	9	15	18
Y	21	15	24	9	18	30	36	12	27	45	54

- Calcule a covariância.
- Calcule o coeficiente de correlação.
- Qual é a intensidade da força da relação entre X e Y ? Explique.

APLICANDO OS CONCEITOS

3.45 Um estudo realizado junto a 218 alunos da Ohio State University sugere uma ligação entre o tempo gasto no portal eletrônico de relacionamento social Facebook e a média geral de notas do aluno. Alunos que raramente ou jamais utilizaram o Facebook apresentaram médias gerais mais altas do que alunos que fazem uso do Facebook.

Fonte: Dados extraídos de M. B. Marklein, "Facebook Use Linked to Less Textbook Time", www.usatoday.com, 14 de abril de 2009.

- O estudo sugere que o tempo gasto no Facebook e a média geral de notas do aluno estão positivamente relacionados ou negativamente correlacionados?
- Você acredita que possa existir uma relação de causa e efeito entre o tempo gasto no Facebook e a média geral de notas do aluno? Explique.

3.46 O arquivo **BebidaCafé** contém a quantidade de calorias e a gordura, em gramas, relativas a bebidas geladas à base de café com conteúdo de 16 onças na Dunkin' Donuts e na Starbucks:

Produto	Calorias	Gordura
Iced Mocha Swirl Latte (leite integral), da Dunkin' Donuts	240	8,0
Coffee Frappuccino blended, da Starbucks	260	3,5
Coffee Coolatta (creme), da Dunkin' Donuts	350	22,0
Mocha Espresso Café Gelado (leite integral e creme chantili), da Starbucks	350	20,0
Mocha Frappuccino blended coffee (creme chantili), da Starbucks	420	16,0
Chocolate Brownie Frappuccino blended coffee (creme chantili), da Starbucks	510	22,0
Chocolate Frappuccino blended crème (creme chantili), da Starbucks	530	19,0

Fonte: Dados extraídos de "Coffee as Candy at Dunkin' Donuts and Starbucks", *Consumer Reports*, junho de 2004, p. 9.

- Calcule a covariância.
- Calcule o coeficiente de correlação.
- Qual deles você acredita que tenha mais valor no que diz respeito a expressar a relação entre calorias e gordura — a covariância ou o coeficiente de correlação? Explique.
- Com base em (a) e (b), a que conclusões você chega quanto à relação entre calorias e gordura?

3.47 Existem diversos métodos para calcular a eficiência em termos de economia no consumo de combustível. A tabela a seguir (contida no arquivo **Milhagem**) indica a milhagem, calculada por proprietários de automóveis e tendo como base os padrões correntes do governo americano:

Automóvel	Proprietário	Governo
2005 Ford F-150	14,3	16,8
2005 Chevrolet Silverado	15,0	17,8
2002 Honda Accord LX	27,8	26,2
2002 Honda Civic	27,9	34,2
2004 Honda Civic Hybrid	48,8	47,6
2002 Ford Explorer	16,8	18,3
2005 Toyota Camry	23,7	28,5
2003 Toyota Corolla	32,8	33,1
2005 Toyota Prius	37,3	56,0

Fonte: Dados extraídos de J. Healey "Fuel Economy Calculations to Be Altered", *USA Today*, 11 de janeiro de 2006, p. 1B.

- Calcule a covariância.
- Calcule o coeficiente de correlação.
- O que você acredita que tenha mais valor no sentido de expressar a relação entre a milhagem calculada pelos proprietários dos carros e a milhagem calculada com base nos padrões correntes do governo americano — a covariância ou o coeficiente de correlação? Explique.

d. Com base em (a) e (b), a que conclusões você consegue chegar no que diz respeito à relação entre a milhagem calculada pelos proprietários dos carros e a milhagem calculada com base nos padrões correntes do governo americano?

3.48 O basquete nas faculdades é um grande negócio, com salários de treinadores, receitas e despesas na casa dos milhões de dólares. O arquivo **Basquete.Faculdades** contém os salários de treinadores e a receita do basquete de faculdades em escolas selecionadas, em um ano recente

Fonte: Dados extraídos de R. Adams, "Pay for Playoffs", *The Wall Street Journal*, 11-12 de março de 2006, pp. P1.P8.

- Calcule a covariância.
- Calcule o coeficiente de correlação.
- Com base em (a) e (b), a que conclusões você consegue chegar quanto à relação entre salários de treinadores e receitas?

3.49 Jogadores universitários de futebol americano que estão tentando se habilitar à NFL são submetidos ao teste de inteligência padronizado Wonderlic. Os arquivos **Wonderlic** contém a média dos resultados do Wonderlic de jogadores de futebol que estão tentando se habilitar à NFL e a avaliação durante a graduação de jogadores de futebol americano em escolas selecionadas.

Fonte: Dados extraídos de S. Walker, "The NFL's Smartest Team", *The Wall Street Journal*, 30 de setembro de 2005, pp. W1, W10.

- Calcule a covariância.
- Calcule o coeficiente de correlação.
- Com base em (a) e (b), a que conclusões você consegue chegar quanto à relação entre a média do resultado do Wonderlic e a avaliação durante a graduação?

3.6 Estatísticas Descritivas: Armadilhas e Questões Éticas

Este capítulo descreve o modo como um conjunto de dados numéricos pode ser caracterizado por meio das estatísticas que medem as propriedades de tendência central, variação e formato. No mundo dos negócios, estatísticas descritivas tais como aquelas que você passou a conhecer estão frequentemente incluídas nos relatórios analíticos que são preparados periodicamente.

O volume de informações disponíveis na Internet, nos jornais e nas revistas tem produzido grande ceticismo quanto à objetividade dos dados. Quando estiver lendo informações que contêm estatísticas descritivas, você deve ter em mente o dito popular frequentemente atribuído ao famoso estadista britânico do século XIX Benjamin Disraeli: "Existem três tipos de mentiras: as mentiras, as grandes mentiras e a estatística."

Por exemplo, ao examinar estatísticas, você precisa comparar a média aritmética e a mediana. Elas são semelhantes ou são significativamente diferentes? Ou, ainda, somente a média aritmética está sendo apresentada? As respostas para essas perguntas farão com que você seja capaz de determinar se os dados são assimétricos ou simétricos e se a mediana poderia ser uma medida mais eficaz de tendência central do que a média aritmética. Além disso, você deve tentar verificar se o desvio-padrão foi incluído nas estatísticas apresentadas. Sem o desvio-padrão, é difícil determinar o montante de variação que existe nos dados.

Considerações éticas surgem quando você está decidindo sobre quais resultados deve incluir em um relatório. Você deve documentar tanto os resultados favoráveis quanto os desfavoráveis. Além disso, ao fazer apresentações orais e ao apresentar relatórios por escrito, você precisa apresentar os resultados de maneira correta, objetiva e neutra. Um comportamento fora dos padrões da ética ocorre quando você deliberadamente deixa de relatar descobertas pertinentes que sejam prejudiciais à defesa de um determinado ponto de vista.

UTILIZANDO A ESTATÍSTICA



@ Choice Is Yours (A Escolha É Sua) Parte II Revisitada

Na parte II do cenário Choice Is Yours (A Escolha É Sua), você foi contratado por esta empresa prestadora de serviços de investimentos para aconselhar investidores interessados em fundos mútuos compostos por títulos. Uma amostra constituída por 180 fundos mútuos compostos por títulos incluía 90 fundos de títulos do governo de médio prazo e 90 fundos de títulos corporativos de curto prazo. Ao comparar essas duas categorias, você conseguiu proporcionar valiosas linhas de raciocínio aos investidores.

Os retornos de 2008 para os fundos de títulos do governo de médio prazo e os fundos de títulos corporativos de curto prazo eram, ambos, assimétricos à esquerda. Os retornos para os fundos de títulos corporativos de curto prazo eram especialmente assimétricos à esquerda, conforme indicado pelos box-plots (veja a Figura 3.4, Exemplo 3.14, Seção 3.3). As estatísticas descritivas (veja a Figura 3.2, Exemplo 3.10, Seção 3.2) permitiu que você comparasse a tendência central e a variabilidade dos retornos para os fundos de títulos do governo de médio prazo e os fundos de títulos corporativos de curto prazo. A média aritmética indicou que os fundos de títulos do governo de médio prazo apresentaram um retorno médio de 5,96, enquanto a mediana indicou que metade dos fundos apresentou retornos de 7,0 ou mais. As tendências centrais dos fundos de títulos corporativos de curto prazo se mostraram bem mais baixas do que aquelas para os fundos de títulos do governo de médio prazo — elas apresentaram uma média de -3,32, e metade dos fundos apresentou retornos inferiores a -1,45. Além disso, os fundos de títulos do governo de médio prazo mostraram menor variabilidade em seus retornos do que os fundos de títulos corporativos de curto prazo — eles apresentaram uma amplitude de 23,62, comparada a 39,3; uma amplitude interquartil de 5,00, comparada a 7,30; e um desvio-padrão de 4,9, comparado a 7,09. Assim, um fundo típico composto de títulos do governo de médio prazo não somente apresentou um retorno significativamente mais alto em 2008 do que um fundo composto por títulos corporativos de curto prazo; a amostra de 90 fundos de títulos do governo de médio prazo apresentou retornos que estavam bem mais estreitamente concentrados próximo ao centro. Uma linha de raciocínio interessante é o fato de que, enquanto 75% dos fundos do governo de médio prazo apresentaram retornos correspondentes a 3,70 ou mais ($Q_1 = 3,70$), somente 25% fundos corporativos de curto prazo apresentaram retornos correspondentes a 1,8 ou mais ($Q_1 = 1,80$). Embora o desempenho no passado não constitua garantia de desempenho no futuro, em 2008 os fundos de títulos do governo de médio prazo tiveram desempenho significativamente melhor do que os fundos corporativos de curto prazo.

RESUMO

Neste capítulo e no capítulo anterior, você estudou estatísticas descritivas — como você consegue visualizar dados por meio de tabelas e gráficos e, a partir de então, como você consegue utilizar diferentes estatísticas para ajudar a analisar os dados e chegar a conclusões. No Capítulo 2, você foi capaz de visualizar dados por meio do uso de gráficos de barra e de pizza, histogramas e outros métodos gráficos. Neste capítulo, você aprendeu como estatísticas descritivas tais como a média aritmética, a mediana,

quartis, amplitude, desvio-padrão são utilizadas para descrever as características de tendência central, variabilidade e formato. A Tabela 3.9 apresenta uma lista com as medidas numéricas descritivas abordadas neste capítulo.

No próximo capítulo, serão apresentados os princípios básicos da probabilidade, com o objetivo de estabelecer uma ponte entre o objeto da estatística descritiva e o objeto da estatística inferencial.

TABELA 3.9
Resumo de Estatísticas Descritivas

Tipo de Análise	Dados Numéricos
Descrever tendência central, variação e formato de uma variável numérica	Média aritmética, mediana, moda, quartis, amplitude, amplitude interquartil, variância, desvio-padrão, coeficiente de variação, escores Z, box-plot (Seções 3.1 a 3.4)
Descrever a relação entre duas variáveis numéricas	Covariância, coeficiente de correlação (Seção 3.5)

EQUAÇÕES-CHAVE

Média Aritmética da Amostra

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \quad (3.1)$$

Mediana

$$\text{Mediana} = \frac{n+1}{2} \text{ valor ordenado} \quad (3.2)$$

Média Geométrica

$$\bar{X}_G = (X_1 \times X_2 \times \dots \times X_n)^{1/n} \quad (3.3)$$

Média Geométrica da Taxa de Retorno

$$\bar{R}_G = [(1 + R_1) \times (1 + R_2) \times \dots \times (1 + R_n)]^{1/n} - 1 \quad (3.4)$$

Amplitude

$$\text{Amplitude} = X_{\text{maior}} - X_{\text{menor}} \quad (3.5)$$

Variância da Amostra

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} \quad (3.6)$$

Desvio-Padrão da Amostra

$$S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}} \quad (3.7)$$

Coefficiente de Variação

$$CV = \left(\frac{S}{\bar{X}}\right) 100\% \quad (3.8)$$

Escore Z

$$Z = \frac{X - \bar{X}}{S} \quad (3.9)$$

Primeiro Quartil, Q_1

$$Q_1 = \frac{n+1}{4} \text{ valor ordenado} \quad (3.10)$$

Terceiro Quartil, Q_3

$$Q_3 = \frac{3(n+1)}{4} \text{ valor ordenado} \quad (3.11)$$

Amplitude Interquartil

$$\text{Amplitude interquartil} = Q_3 - Q_1 \quad (3.12)$$

Média Aritmética da População

$$\mu = \frac{\sum_{i=1}^N X_i}{N} \quad (3.13)$$

Variância da População

$$\sigma^2 = \frac{\sum_{i=1}^N (X_i - \mu)^2}{N} \quad (3.14)$$

Desvio-Padrão da População

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (X_i - \mu)^2}{N}} \quad (3.15)$$

Covariância da Amostra

$$\text{cov}(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n-1} \quad (3.16)$$

Coefficiente de Correlação da Amostra

$$r = \frac{\text{cov}(X, Y)}{S_X S_Y} \quad (3.17)$$

TERMOS-CHAVE

- | | | | |
|--------------------------------------|-----------------------------|-------------------------------|-----------------------------|
| amplitude | covariância da amostra | média aritmética da população | regra de Chebyshev |
| amplitude interquartil | curtose | média geométrica | regra empírica |
| assimetria | desvio-padrão | média geométrica do retorno | resumo de cinco números |
| assimetria à direita | desvio-padrão da amostra | mediana | simetria |
| assimetria à esquerda | desvio-padrão da população | medida resistente | soma dos quadrados (SQ) |
| assimétrico | dispersão | moda | <i>spread</i> |
| box-plot | dispersão média | <i>outlier</i> | tendência central |
| coeficiente de correlação | escore Z | Q_1 : primeiro quartil | valor extremo |
| coeficiente de correlação da amostra | formato | Q_2 : segundo quartil | variação |
| coeficiente de variação | média | Q_3 : terceiro quartil | variância |
| covariância | média aritmética | quartil | variância da amostra |
| | média aritmética da amostra | | variância da população |

PROBLEMAS DE REVISÃO DO CAPÍTULO 3

AVALIANDO O SEU ENTENDIMENTO

3.50 Quais são as propriedades de um conjunto de dados numéricos?

3.51 Qual é o significado da propriedade de tendência central?

3.52 Quais são as diferenças entre a média aritmética, a mediana e a moda, e quais são as vantagens e desvantagens de cada uma delas?

3.53 Como você interpreta o primeiro quartil, a mediana e o terceiro quartil?

3.54 Qual é o significado da propriedade de variação?

3.55 Qual é o significado da medida *escore Z*?

3.56 Quais são as diferenças entre as diversas medidas de variação, tais como a amplitude, a amplitude interquartil, a variância, o desvio-padrão e o coeficiente de variação, e quais são as vantagens e desvantagens de cada uma delas?

3.57 De que modo a regra empírica ajuda a explicar os meios pelos quais se concentram e se distribuem os valores em um conjunto de dados numéricos?

3.58 De que modo a regra empírica e a regra de Chebyshev diferem?

3.59 Qual é o significado da propriedade do formato?

3.60 De que modo a covariância e o coeficiente de correlação diferem?

APLICANDO OS CONCEITOS

3.61 A American Society for Quality (ASQ) conduziu uma pesquisa salarial junto a todos os seus membros. Os membros da ASQ atuam em todas as áreas de instituições de manufatura e de serviços, com um tema comum que diz respeito a um interesse na qualidade. Para a pesquisa nos Estados Unidos, foram encaminhadas mensagens de correio eletrônico a 70.227 membros, e foram recebidas 9.449 respostas válidas. Os dois títulos mais comuns para funções foram gerente e engenheiros de qualidade. Um outro título é o Mestre Faixa-Preta (Master Black Belt), que é a pessoa que assume uma função de liderança como o guardião do processo Seis Sigma (veja a Seção 17.8). Um título adicional é o de Inspetor. Estatísticas descritivas relativas a salários para essas quatro funções são apresentadas a seguir. Compare os salários dos inspetores, dos gerentes, dos engenheiros de qualidade e dos Mestres Faixa-Preta.

Título	Tamanho da Amostra	Mínimo	Máximo	Desvio-padrão	Média Aritmética	Mediana
Inspetor	154	21.000	110.000	17.197	45.949	41.800
Gerente	2.228	27.000	202.516	24.109	85.551	83.500
Engenheiro de qualidade	1.387	20.000	174.000	18.796	72.824	70.000
Mestre Faixa-Preta	134	33.000	185.250	24.738	113.385	114.100

Fonte: Dados extraídos de I. E. Allen, "Salary Survey: Seeing Green", Quality Progress, dezembro de 2008, pp. 20-53.

3.62 No estado de Nova York, instituições de poupança e investimentos podem comercializar um tipo de seguro de vida co-

nhecido como seguro de vida das instituições de poupança e investimentos (*savings bank life insurance* — SBLI). O processo de aprovação consiste na subscrição, que inclui a revisão da proposta; verificação das informações médicas; possíveis pedidos de outras informações médicas ou exames adicionais; e o estágio de compilação da apólice, durante o qual as páginas da apólice são geradas e enviadas ao banco para que sejam então remetidas. A capacidade de entregar as apólices aprovadas ao cliente em tempo hábil é crítica para a instituição no que diz respeito à eficácia desse serviço. Durante um período correspondente a um mês, foi selecionada uma amostra aleatória de 27 apólices aprovadas, e foram registrados os seguintes dados sobre o total de tempo, em dias, para o processamento (contidos no arquivo **Seguro**):

73 19 16 64 28 28 31 90 60 56 31 56 22 18
45 48 17 17 17 91 92 63 50 51 69 16 17

- Calcule a média aritmética, a mediana, o primeiro quartil e o terceiro quartil.
- Calcule a amplitude, a amplitude interquartil, a variância, o desvio-padrão e o coeficiente de variação.
- Construa um box-plot. Os dados são assimétricos? Em caso afirmativo, qual a direção da assimetria?
- O que você diria a um cliente que entrasse na instituição para adquirir esse tipo de apólice de seguro e perguntasse quanto tempo leva o processo de aprovação?

3.63 Um dos mais importantes indicadores da qualidade de serviços prestados por qualquer organização é a velocidade com que ela responde às reclamações dos clientes. Uma grande loja de departamentos, de propriedade familiar, que vende mobiliário e pisos, incluindo carpetes, passou por uma grande expansão ao longo dos últimos anos. Particularmente, o departamento de pisos se expandiu, partindo de apenas dois especialistas em instalação para um supervisor de instalações, um medidor e 15 especialistas em instalação. O objetivo estratégico da empresa era reduzir o intervalo de tempo entre o momento em que a reclamação é recebida e o momento em que ela é solucionada. Durante um ano recente, a empresa obteve 50 reclamações referentes à instalação de carpetes. Os dados oriundos das 50 reclamações, organizados no arquivo **Mobiliário**, representam o número de dias entre o recebimento de uma reclamação e a solução do problema relacionado à reclamação:

54 5 35 137 31 27 152 2 123 81 74 27 11
19 126 110 110 29 61 35 94 31 26 5 12 4
165 32 29 28 29 26 25 1 14 13 13 10 5
27 4 52 30 22 36 26 20 23 33 68

- Calcule a média aritmética, a mediana, o primeiro quartil e o terceiro quartil.
- Calcule a amplitude, a amplitude interquartil, a variância, o desvio-padrão e o coeficiente de variação.
- Construa um box-plot. Os dados são assimétricos? Em caso afirmativo, qual a direção da assimetria?
- Com base nos resultados de (a) até (c), se você tivesse que informar ao presidente da empresa quanto tempo o cliente deve supostamente esperar para ter uma reclamação solucionada, o que você diria? Explique.

3.64 Uma indústria produz suportes de aço para equipamentos elétricos. O principal componente do suporte é uma placa de aço, em baixo-relevo, feita de uma bobina de aço de calibre 14. Ela é produzida por punção progressiva de uma prensa de 250 toneladas, com uma operação de limpeza que coloca duas formas de 90 graus no aço plano no intuito de fabricar o baixo-relevo. A distância de um lado da forma até o outro é crítica em razão da necessidade de o suporte ser à prova d'água quando utilizado em ambientes externos. A empresa exige que a largura do baixo-relevo esteja entre 8,31 polegadas e 8,61 polegadas. Os dados são coletados de uma amostra de 49 a baixos-relevos e armazenados em **Baixo-Relevo**, que contém as amplitudes dos baixos-relevos, em polegadas, apresentadas a seguir:

8,312 8,343 8,317 8,383 8,348 8,410 8,351 8,373 8,481 8,422
8,476 8,382 8,484 8,403 8,414 8,419 8,385 8,465 8,498 8,447
8,436 8,413 8,489 8,414 8,481 8,415 8,479 8,429 8,458 8,462
8,460 8,444 8,429 8,460 8,412 8,420 8,410 8,405 8,323 8,420
8,396 8,447 8,405 8,439 8,411 8,427 8,420 8,498 8,409

- Calcule a média aritmética, a mediana, a amplitude e o desvio-padrão para a amplitude. Interprete essas medidas de tendência central e variabilidade.
- Faça a lista com o resumo de cinco números.
- Construa um box-plot e descreva o respectivo formato.
- O que você consegue concluir em relação ao número de baixos-relevos que atenderão aos pré-requisitos da empresa para que esses baixos-relevos tenham uma largura entre 8,31 e 8,61 polegadas?

3.65 A indústria citada no Problema 3.64 também produz isolantes elétricos. Se os isolantes se rompem durante sua utilização, existe a possibilidade de um curto-circuito. Para testar a resistência desses isolantes, são realizados testes de destruição com o objetivo de determinar o volume de força necessário para que os isolantes venham a se romper. A força é medida por meio da observação da quantidade de libras que podem ser aplicadas no isolante antes que ele venha a se romper. São coletados dados oriundos de uma amostra de 30 isolantes. O arquivo **Força** apresenta as respectivas forças, como segue:

1.870 1.728 1.656 1.610 1.634 1.784 1.522 1.696 1.592 1.662
1.866 1.764 1.734 1.662 1.734 1.774 1.550 1.756 1.762 1.866
1.820 1.744 1.788 1.688 1.810 1.752 1.680 1.810 1.652 1.736

- Calcule a média aritmética, a mediana, a amplitude e o desvio-padrão para a força necessária para romper o isolante.
- Interprete as medidas de tendência central e variabilidade em (a).
- Construa um box-plot e descreva o respectivo formato.
- O que você pode concluir sobre a resistência dos isolantes se a empresa requer uma medida de força correspondente a pelo menos 1.500 libras, antes que o isolante venha a se romper?

3.66 O arquivo **VeggieBurger** contém dados sobre as calorias e o total de gordura (em gramas por porção) de uma amostra de 12 hambúrgueres vegetarianos.

Fonte: Dados extraídos de "Healthful Burgers That Taste Good", Consumer Reports, junho de 2008, p. 8.

- Para cada uma das variáveis, calcule a média aritmética, a mediana, o primeiro quartil e o terceiro quartil.

b. Para cada uma das variáveis, calcule a amplitude, a amplitude interquartil, a variância, o desvio-padrão e o coeficiente de variação.

c. Para cada uma das variáveis, construa um box-plot. Os dados são assimétricos? Em caso afirmativo, qual a direção da assimetria?

d. Calcule o coeficiente de correlação entre calorias e o total de gordura.

e. A que conclusões você consegue chegar no que diz respeito a calorias e total de gordura?

3.67 Uma característica de interesse relativa à qualidade em um processo de abastecimento de saquinhos de chá é o peso do chá em cada um dos saquinhos. Se os saquinhos estiverem subabastecidos, surgem dois problemas. Em primeiro lugar, os clientes podem não conseguir fazer com que o chá fique tão forte quanto desejam. Em segundo lugar, a empresa pode estar violando a legislação pertinente aos pesos e medidas apresentados nas embalagens. Para esse produto, o peso descrito na embalagem indica que, em média, existem 5,5 gramas de chá em um saquinho. Se a média aritmética da quantidade de chá em cada saquinho exceder a descrição na embalagem, a empresa estará desperdiçando produto. Conseguir uma medida exata de chá em cada saquinho é problemático, em razão de variação de temperatura e umidade no interior da fábrica, diferenças na densidade do chá e operação de abastecimento extremamente rápida da máquina (aproximadamente 170 saquinhos por minuto). O arquivo **SaquinhosChá**, apresentado a seguir, contém os pesos, em gramas, correspondentes a uma amostra de 50 saquinhos de chá, produzidos no intervalo de uma hora por uma única máquina.

5,65 5,44 5,42 5,40 5,53 5,34 5,54 5,45 5,52 5,41
5,57 5,40 5,53 5,54 5,55 5,62 5,56 5,46 5,44 5,51
5,47 5,40 5,47 5,61 5,53 5,32 5,67 5,29 5,49 5,55
5,77 5,57 5,42 5,58 5,58 5,50 5,32 5,50 5,53 5,58
5,61 5,45 5,44 5,25 5,56 5,63 5,50 5,57 5,67 5,36

- Calcule a média aritmética, a mediana, o primeiro quartil e o terceiro quartil.
- Calcule a amplitude, a amplitude interquartil, a variância, o desvio-padrão e o coeficiente de variação.
- Interprete as medidas de tendência central e a variação no contexto deste problema. Por que razão a empresa que produz os saquinhos de chá deve estar atenta para a tendência central e a variação?
- Construa um box-plot. Os dados são assimétricos? Em caso afirmativo, qual a direção da assimetria?
- A empresa está atendendo às exigências estabelecidas na embalagem no sentido de que, em média, existem 5,5 gramas de chá em um saquinho? Caso você fosse o encarregado desse processo, que modificações, se é que alguma, você tentaria fazer no que concerne à distribuição de pesos em cada um dos saquinhos individuais?

3.68 O produtor das placas de asfalto das marcas Boston e Vermont fornece a seus clientes uma garantia de 20 anos para a maior parte de seus produtos. Para determinar se uma placa durará todo o período estabelecido na garantia, são conduzidos, na área de produção, testes de aceleração de vida útil. O teste de aceleração de vida útil expõe a placa, em um laboratório, às condições de desgaste às quais ela estaria sujeita ao longo de uma vida útil de

utilização normal, por meio de um experimento que precisa apenas de alguns poucos minutos para ser conduzido. Nesse teste, uma placa é repetidamente esfregada com uma escova por um curto período de tempo e os grãos da placa removidos por meio da escovação são pesados (em gramas). Espera-se que as placas que sofrem a perda de pequenas quantidades de grãos durem mais sob condições de utilização normal do que as placas que sofrem a perda de grandes quantidades de grãos. Nesse tipo de situação, uma placa deve sofrer a perda de uma quantidade de grãos não superior a 0,8 grama, caso seja esperado que ela dure toda a extensão do período da garantia. O arquivo **Grão** contém uma amostra com 170 medições realizadas nas placas Boston e 140 medições realizadas nas placas Vermont da empresa.

- Faça uma lista com o resumo de cinco números para as placas Boston e para as placas Vermont.
- Construa box-plots paralelos para as duas marcas de placas e descreva os formatos das distribuições.
- Comente sobre a capacidade de cada um dos tipos de placas no que concerne a alcançar uma perda de grãos de 0,8 grama ou menos.

3.69 Um estudo conduzido pela Zagat Survey concluiu que muitos restaurantes de primeira classe estão localizados em hotéis em todos os Estados Unidos. Os turistas podem encontrar ambientes requintadamente decorados, refeições e serviços de qualidade sem ter que deixar o hotel onde estão hospedados. O restaurante de hotel com melhor avaliação é o The French Room, localizado no The Adolphus Hotel, em Dallas, Texas. O preço estimado para um jantar, incluindo uma bebida e gorjeta, no The French Room, é de \$80. O preço mais alto relatado é \$179, no Alain Ducasse, localizado no Jumeirah Essex House, na cidade de Nova York. O arquivo **MelhoresRest** contém os 100 melhores restaurantes em hotéis nos Estados Unidos e as variáveis estado, cidade, restaurante, hotel, custo (preço estimado de um jantar, incluindo uma bebida e gorjeta) e a classificação (1 a 100, sendo 1 o restaurante com melhor avaliação).

Fonte: Dados extraídos de Gary Stoller, "Top Restaurants Check into Luxury Hotels", USA Today, 11 de abril de 2006, p. 5B.

- Construa o resumo de cinco números para o preço de jantar.
- Construa um box-plots para preço de jantar e interprete a distribuição para os preços de jantar.
- Calcule e interprete o coeficiente de correlação para a classificação e o preço de jantar.

3.70 O arquivo **Proteína** contém as calorias, proteínas e colesterol para alimentos populares ricos em proteínas (carnes vermelhas frescas, aves e peixes).

Fonte: U.S. Department of Agriculture.

- Calcule o coeficiente de correlação entre calorias e proteínas.
- Calcule o coeficiente de correlação entre calorias e colesterol.
- Calcule o coeficiente de correlação entre proteínas e colesterol.
- Com base nos resultados de (a) a (c), a que conclusões você consegue chegar no que se refere a calorias, proteínas e colesterol?

3.71 O arquivo **Custo de Vida** contém o índice geral do custo de vida, o aluguel mensal para um apartamento com dois quartos, o custo de uma xícara de café incluindo a taxa de serviço, o custo de uma refeição rápida à base de hambúrguer, o custo de lavagem a seco de um paletó masculino, o custo de uma pasta de dentes e o custo de ingressos de cinema em 10 diferentes cidades americanas.

- Calcule o coeficiente de correlação entre o índice geral do custo de vida, o aluguel mensal para um apartamento com dois quartos, o custo de uma xícara de café incluindo a taxa de serviço, o custo de uma refeição rápida à base de hambúrguer, o custo de lavagem a seco de um paletó masculino, o custo de uma pasta de dentes e o custo de ingressos de cinema.
- A que conclusões você consegue chegar no que se refere à relação entre o índice geral do custo de vida e cada uma dessas seis variáveis?

3.72 O arquivo **ImpostosPropriedade** contém os impostos cobrados sobre propriedades, per capita, para os 50 estados dos EUA e o Distrito de Columbia.

- Calcule a média aritmética, a mediana, o primeiro quartil e o terceiro quartil.
- Calcule a amplitude, a amplitude interquartil, a variância, o desvio-padrão e o coeficiente de correlação.
- Construa um box-plots. Os dados são assimétricos? Em caso afirmativo, qual a direção da assimetria?
- Com base nos resultados de (a) até (c), a que conclusões você consegue chegar no que se refere aos impostos cobrados sobre propriedades, per capita, em milhares de dólares, para cada um dos 50 estados dos EUA e o Distrito de Columbia?

3.73 Os dados no arquivo **Remuneração** representam a remuneração total (em dólares) dos executivos-chefes (CEO) de grandes empresas públicas nos EUA, em 2008.

Fonte: Dados extraídos de D. Jones e B. Hansen "CEO Pay Dives in a Rough 2008", www.usatoday.com, 1º de maio de 2009.

- Calcule a média aritmética, a mediana, o primeiro quartil e o terceiro quartil.
- Calcule a amplitude, a amplitude interquartil, a variância, o desvio-padrão e o coeficiente de variação.
- Construa um box-plot. Os dados são assimétricos? Em caso afirmativo, qual a direção da assimetria?
- Com base nos resultados de (a) a (c), a que conclusões você consegue chegar no que concerne ao total da remuneração (em milhões de dólares) dos executivos (CEO)?
- Calcule o coeficiente de correlação entre a remuneração e o montante relativo a bonificações.
- Calcule o coeficiente de correlação entre a remuneração e a variação nos preços de ações em bolsa, em 2008.
- A que conclusões você consegue chegar a partir dos resultados de (e) e (f)?

3.74 Você está planejando estudar para sua prova de estatística com um grupo de colegas de classe, um dos quais você particularmente deseja impressionar. Essa pessoa se ofereceu como voluntária para utilizar o Microsoft Excel no intuito de obter as informações resumidas, tabelas e gráficos para um conjunto de dados contendo diversas variáveis numéricas e categóricas atribuídas pelo professor para propósitos de estudo. Essa pessoa se dirige a você com o material impresso e exclama: "Já tenho tudo — as médias aritméticas, as medianas, os desvios-padrão, os box-plots e os gráficos de pizza — para todas as nossas variáveis. O problema é que parte do resultado parece estranha — como é o caso dos box-plots para gênero, para disciplina de especialização e os gráficos de pizza para o índice médio de avaliação e estatutura. Do mesmo modo, não consigo entender a razão pela qual o Professor Krehbiel afirmou que não podemos obter as estatísticas descritivas para algumas das variáveis: consegui essas estatísticas para todas elas! Veja: a média aritmética para a estatura é 68,23; a média aritmética para o índice médio de avaliação é 2,76; a

média aritmética para o gênero é 1,50; a média aritmética para disciplina de especialização é 4,33." O que você responde?

EXERCÍCIOS DE REDAÇÃO DE RELATÓRIOS

3.75 O arquivo **Cerveja Artesanal** contém a porcentagem de álcool, a quantidade de calorias para 12 onças e a quantidade de carboidratos (em gramas) para 12 onças correspondentes a 128 dentre as cervejas artesanais com maior índice de vendas nos Estados Unidos.

Sua tarefa é redigir um relatório baseado em uma completa avaliação descritiva de cada uma das variáveis numéricas — porcentagem de álcool, quantidade de calorias para 12 onças e quantidade de carboidratos (em gramas) para 12 onças. Em anexo a seu relatório devem estar todas as tabelas, gráficos e as medidas numéricas descritivas.

Fonte: Dados extraídos de www.Beer100.com, 15 de junho de 2009.

PROJETOS DE GRUPO

O arquivo **Fundos de Títulos** contém informações relativas a nove variáveis extraídas de uma amostra de 180 fundos mútuos:

Numeração do Fundo — Número de identificação para cada um dos fundos de títulos

Tipo — Tipos de títulos que compõem o fundo mútuo (títulos do governo de médio prazo e títulos corporativos de curto prazo)

Ativos — Em milhões de dólares

Comissões — Tarifas cobradas sobre vendas (sim ou não)

Proporção de despesas — Proporção entre despesas incorridas e ativos líquidos, em percentagens

Retorno de 2008 — Retorno relativo a doze meses, em 2008

Retorno de três anos — Retorno a cada ano, 2006-2008

Retorno de cinco anos — Retorno a cada ano, 2004-2008

Risco — Fator de risco de perda para o fundo mútuo (baixo, médio ou alto)

3.76 Para a proporção de despesas, em termos percentuais, retorno de três anos e retorno de cinco anos,

- Calcule a média aritmética, a mediana, o primeiro quartil e o terceiro quartil.
- Calcule a amplitude, a amplitude interquartil, a variância, o desvio-padrão e o coeficiente de variação.
- Construa um box-plot. Os dados são assimétricos? Em caso afirmativo, qual a direção da assimetria?
- Com base nos resultados de (a) a (c), a que conclusões você consegue chegar quanto a essas variáveis?

3.77 Você deseja comparar fundos de títulos que cobram tarifas com fundos de títulos que não cobram tarifas. Para cada um desses dois grupos, no que diz respeito às variáveis proporção de despesas, em termos percentuais, retorno de 2008, retorno de três anos e retorno de cinco anos,

- Calcule a média aritmética, a mediana, o primeiro quartil e o terceiro quartil.
- Calcule a amplitude, a amplitude interquartil, a variância, o desvio-padrão e o coeficiente de variação.
- Construa um box-plot. Os dados são assimétricos? Em caso afirmativo, qual a direção da assimetria?
- Com base nos resultados de (a) a (c), a que conclusões você consegue chegar quanto às diferenças entre fundos mútuos que cobram tarifas e fundos mútuos que não cobram tarifas?

3.78 Você deseja comparar fundos de títulos do governo de médio prazo com fundos de títulos corporativos de curto prazo. Para cada um desses dois grupos, no que diz respeito às variáveis proporção de despesas em termos percentuais, retorno de três anos e retorno de cinco anos,

- Calcule a média aritmética, a mediana, o primeiro quartil e o terceiro quartil.
- Calcule a amplitude, a amplitude interquartil, a variância, o desvio-padrão e o coeficiente de variação.
- Construa um box-plot. Os dados são assimétricos? Em caso afirmativo, qual a direção da assimetria?
- Com base nos resultados de (a) a (c), a que conclusões você consegue chegar quanto às diferenças entre fundos de títulos do governo de médio prazo e fundos de títulos corporativos de curto prazo?

3.79 Você deseja comparar fundos de títulos com base no risco. Para cada um desses três níveis de risco (abaixo da média, médio e acima da média), no que diz respeito às variáveis proporção de despesas em termos percentuais, retorno de três anos e retorno de cinco anos,

- Calcule a média aritmética, a mediana, o primeiro quartil e o terceiro quartil.
- Calcule a amplitude, a amplitude interquartil, a variância, o desvio-padrão e o coeficiente de variação.
- Construa um box-plot. Os dados são assimétricos? Em caso afirmativo, qual a direção da assimetria?
- Com base nos resultados de (a) a (c), a que conclusões você consegue chegar quanto às diferenças entre fundos de títulos, tomando como base o risco?

BANCO DE DADOS DA PESQUISA REALIZADA JUNTO A ALUNOS

3.80 O Problema 2.117, nos Problemas de Revisão do Capítulo 2, descreve uma pesquisa realizada junto a 50 alunos da graduação (contida em **PesquisaGrad**). Para esses dados, no que diz respeito a cada variável numérica,

- Calcule a média aritmética, a mediana, o primeiro quartil e o terceiro quartil.
- Calcule a amplitude, a amplitude interquartil, a variância, o desvio-padrão e o coeficiente de variação.
- Construa um box-plot. Os dados são assimétricos? Em caso afirmativo, qual a direção da assimetria?
- Escreva um relatório resumindo suas conclusões.

3.81 O Problema 2.117, nos Problemas de Revisão do Capítulo 2, descreve uma pesquisa realizada junto a 50 alunos da graduação (contida em **PesquisaGrad**).

- Selecione uma amostra de 50 alunos da graduação em sua faculdade e conduza uma pesquisa semelhante junto a esses alunos.
- Para os dados coletados em (a), repita os itens de (a) a (d) do Problema 3.80.
- Compare os resultados em (b) com os resultados do Problema 3.80.

3.82 O Problema 2.119, nos Problemas de Revisão do Capítulo 2, descreve uma pesquisa realizada junto a 40 alunos de MBA (contida em **PesquisaPosGrad**). Para esses dados, no que diz respeito a cada variável numérica,

- Calcule a média aritmética, a mediana, o primeiro quartil e o terceiro quartil.

- b. Calcule a amplitude, a amplitude interquartil, a variância, o desvio-padrão e o coeficiente de variação.
 c. Construa um box-plot. Os dados são assimétricos? Em caso afirmativo, qual a direção da assimetria?
 d. Escreva um relatório resumindo suas conclusões.

3.83 O Problema 2.119, nos Problemas de Revisão do Capítulo 2, descreve uma pesquisa realizada junto a 40 alunos de MBA (contida em **PesquisaPosGrad**).

- a. Selecione uma amostra de 40 alunos de MBA em sua faculdade e conduza uma pesquisa semelhante junto a esses alunos.
 b. Para os dados coletados em (a), repita os itens de (a) a (d) do Problema 3.82.
 c. Compare os resultados em (b) com os resultados do Problema 3.82.

ADMINISTRANDO O SPRINGVILLE HERALD

Para qual variável no caso “Administrando o *Springville Herald*”, do Capítulo 2 (veja o final do Capítulo 2), são necessárias medidas numéricas descritivas?

1. Para a variável que você identifica, calcule as medidas numéricas descritivas apropriadas e construa um box-plot.
2. Para a variável que você identifica, aponte uma outra apresentação gráfica que possa vir a ser útil e construa essa apre-

sentação. Que conclusões você pode formar a partir desse outro gráfico que não podem ser formadas a partir do box-plot?

3. Faça um resumo de suas descobertas em um relatório que possa ser incluído junto ao estudo da força-tarefa.

CASO DE INTERNET

Aplique os seus conhecimentos sobre o uso apropriado de medidas numéricas descritivas nesta continuação do Caso de Internet do Capítulo 2.

Utilizando o CD-ROM que acompanha este livro, abra o arquivo **EndRun.htm** na pasta Caso de Internet do Capítulo 3 uma segunda vez, com o objetivo de visitar a prestadora de serviços financeiros EndRun Financial Services. Reexamine os dados que dão suporte à EndRun e, depois, responda ao seguinte:

1. Medidas descritivas podem ser calculadas para qualquer tipo de variável? De que modo essas estatísticas resumidas dão respaldo para as declarações da EndRun? De que modo essas estatísticas resumidas afetam a sua percepção referente aos registros da EndRun?

2. Avalie os métodos que a EndRun utilizou para sintetizar os resultados da pesquisa de satisfação do consumidor por ela realizada (clique no vínculo **Mostre-me os resultados da pesquisa** no arquivo **ER_Survey.htm** contido na pasta Caso de Internet do Capítulo 2, disponível no site da LTC Editora). Existe alguma coisa que você faria de modo diferente no intuito de sintetizar esses resultados?
3. Observe que a última pergunta da pesquisa apresenta um menor número de respostas do que as outras perguntas. Que fatores podem ter limitado o número de respostas para essa pergunta?

REFERÊNCIAS

1. Kendall, M. G., A. Stuart, and J. K. Ord, *Kendall's Advanced Theory of Statistics, Volume 1: Distribution Theory*, 6th ed. (New York: Oxford University Press, 1994).
2. *Microsoft Excel 2007* (Redmond, WA: Microsoft Corporation, 2007).

GUIA DO EXCEL PARA O CAPÍTULO 3

GE3.1 MEDIDAS de TENDÊNCIA CENTRAL

Média Aritmética, Mediana e Moda

Excel Avançado. Utilize as funções **MÉDIA** (para a média aritmética), **MEDIANA** ou **MODO** em fórmulas de planilhas para calcular medidas de tendência central. Insira as funções no formato **FUNÇÃO(intervalo de células dos valores dos dados)**. Veja a Seção GE3.2 para um exemplo do uso dessas funções.

Ferramentas de Análise Utilize o procedimento **Estatística Descritiva** para criar uma lista que inclui medidas de tendência central. (A Seção GE3.2 explica integralmente esse procedimento.)

GE3.2 VARIAÇÃO e FORMATO

Amplitude, Variância, Desvio-Padrão, Coeficiente de Variação e Formato

Excel Avançado. Utilize as funções **VAR** (variância da amostra), **DESVPAD** (desvio-padrão da amostra), **CURT** (curtose) **DISTORÇÃO** (assimetria), **MÍNIMO** (valor mínimo) e **MÁXIMO** (valor máximo) para calcular medidas de variação. Insira essas funções no formato **FUNÇÃO(intervalo de células dos valores dos dados)**.

Utilize a diferença entre **MÁXIMO** e **MÍNIMO** para calcular a amplitude. Utilize a função **CONT.NÚM** para determinar o tamanho da amostra e divida o desvio-padrão da amostra pela raiz quadrada (**RAIZ**) do tamanho da amostra, de modo a calcular o erro-padrão.

Por exemplo, para criar a planilha da Figura 3.2, que apresenta estatísticas descritivas (incluindo medidas de tendência central) relativas ao retorno de 2008 para fundos de títulos do governo de médio prazo e de fundos de títulos corporativos de curto prazo (veja o Exemplo 3.10, Seção 3.2), abra a planilha **RETORNO2008** da pasta de trabalho **Fundos De Títulos**, e:

Clique à direita na guia correspondente à planilha **RETORNO2008** e clique em **Inserir** no menu de atalhos. Na caixa de diálogo **Inserir**, clique no ícone **Planilha** e, depois, clique em **OK**. Na nova planilha:

1. Insira os títulos e cabeçalhos de coluna para a linha 1 e a linha 2, bem como as legendas de linha para a coluna A, utilizando a Figura 3.2 como orientação.
2. Insira as fórmulas no intervalo de células **B3:B15**, conforme ilustrado na Figura GE3.1.
3. Selecione o intervalo de células **B3:B15** e copie o intervalo de células para a célula C3. (O Excel ajusta os intervalos de células em todas as funções para **RETORNO2008!B2:B91**, o intervalo de células para os valores relativos ao retorno de 2008 dos fundos corporativos de curto prazo.)

Embora não faça parte da planilha da Figura 3.2 para acrescentar o coeficiente de variação insira inicialmente **Coeficiente de variação** na célula **A16**. Em seguida, insira a fórmula $= B7/B3$ na célula **B16** e, depois, copie para a célula **C16**. Por fim, formate as células **B16** e **C16** para exibir o estilo de porcentagem.

A	B
1	Estatísticas Descritivas para o Retorno de 2008
2	Governo de Médio Prazo
3	Média Aritmética =MÉDIA(RETORNO2008!A2:A91)
4	Erro-padrão =B7/RAIZ(B15)
5	Mediana =MEDIANA(RETORNO2008!A2:A91)
6	Moda =MODA(RETORNO2008!A2:A91)
7	Desvio-padrão =DEVPAD(RETORNO2008!A2:A91)
8	Variância da Amostra =VAR(RETORNO2008!A2:A91)
9	Curtose =CURT(RETORNO2008!A2:A91)
10	Assimetria =DISTORÇÃO(RETORNO2008!A2:A91)
11	Amplitude =B13-B12
12	Mínimo =MÍNIMO(RETORNO2008!A2:A91)
13	Máximo =MÁXIMO(RETORNO2008!A2:A91)
14	Soma =SOMA(RETORNO2008!A2:A91)
15	Contagem =CONT.NÚM(RETORNO2008!A2:A91)

FIGURA GE3.1

Fórmulas da Coluna B para a planilha DESCRITIVA.

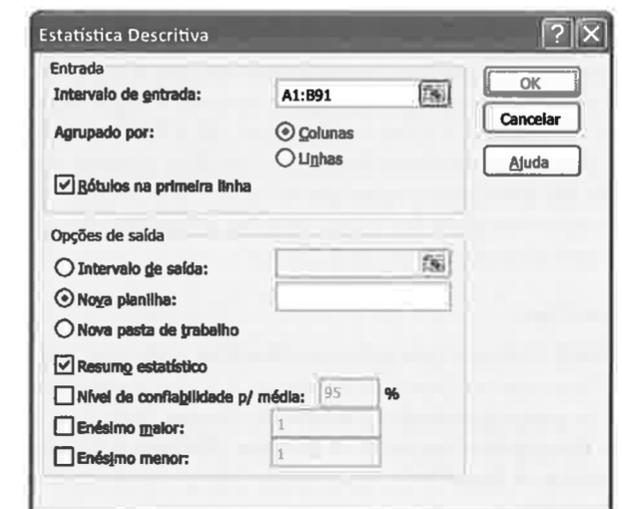
Ferramentas de Análise Utilize o procedimento **Estatística Descritiva** de modo a criar uma lista que contém medidas de variação e formato, juntamente com medidas de tendência central.

Por exemplo, para criar a planilha da Figura 3.2, que apresenta estatísticas descritivas para o retorno de 2008 relativo aos fundos de títulos do governo de médio prazo e aos fundos corporativos de curto prazo (veja o Exemplo 3.10, na Seção 3.2), abra a planilha **RETORNO2008** da pasta de trabalho **Fundos de Títulos** e:

1. Selecione **Dados → Análise de dados** (Excel 2007) ou **Ferramentas → Análise de dados** (Excel 2003).
2. Na caixa de diálogo **Análise de dados**, selecione **Estatística Descritiva** a partir da lista **Ferramentas de Análise** e, em seguida, clique em **OK**.

Na caixa de diálogo **Estatística Descritiva** (apresentada a seguir):

3. Insira **A1:B91** como o **Intervalo de entrada**. Clique na opção **Colunas** e marque a opção **Rótulos na primeira linha**.
4. Clique em **Nova planilha**, marque a opção **Resumo estatístico**, a caixa de diálogo a seguir precisa aparecer aqui depois da etapa 4, e, depois, clique em **OK**.



Na nova planilha:

5. Selecione a coluna C, clique à direita e clique em **Excluir** no menu de atalho (para eliminar as legendas de linhas em duplicata).
6. Edite a entrada de célula da linha 1 como **Retorno 2008**. Insira **Governo de Médio Prazo** na célula B2 e **Corporativo de Curto Prazo** na célula C2.
7. Ajuste os cabeçalhos de colunas e ajuste a formatação de células utilizando a Figura 3.2 como orientação. (Veja a Seção C5 do Apêndice para obter ajuda com esses ajustes.)

O procedimento cria uma planilha que é semelhante à planilha **DESCRITIVAS_FA** da pasta de trabalho **Descritivas**. Embora visualmente idêntica à planilha **DESCRITIVAS, DESCRITIVAS_FA** não contém nenhum tipo de fórmula em suas células.

Para acrescentar o coeficiente de variação a essa planilha, insira, inicialmente, **Coefficiente de variação** na célula A16. Depois, insira a fórmula $=B7/B3$ na célula B16 e, então, copie essa fórmula para a célula C16. Por fim, formate as células B16 e C16 para exibição no estilo de porcentagem.

Calculando Escores Z

Excel Avançado Utilize a função **PADRONIZAR** para calcular escores Z. Insira a função no formato **PADRONIZAR(valor, média aritmética, desvio-padrão)**, em que *valor* corresponde a um valor de *X*. Abra a planilha **TABELA_3.3** da pasta de trabalho **Descritivas** para verificar um exemplo do uso da função **PADRONIZAR** para calcular os escores Z ilustrados na Tabela 3.3, Seção 3.2.

GE3.3 EXPLORANDO DADOS NUMÉRICOS

Quartis

Excel Avançado Conforme observado anteriormente ao se tratar de quartis, a função **QUARTIL**, do Excel, inserida no formato **QUARTIL(intervalo de células dos dados a serem resumidos, número do quartil)** utiliza regras que diferem das regras apresentadas na Seção 3.3 para calcular quartis. Para calcular quartis utilizando as regras da Seção 3.3, abra a planilha **CÁLCULO** da pasta de trabalho **QUARTIS** e insira na coluna A um conjunto de valores de dados. (A planilha contém os valores para o Exemplo 3.11. Sobrescreva esses valores quando inserir algum outro conjunto de valores de dados.)

Os resultados para os quartis utilizando-se as regras da Seção 3.3 estão ilustrados na coluna D (a coluna **Regras do Livro**). Os resultados da coluna D se fundamentam em uma série de fórmulas avançadas nas colunas de G até I, de modo a implementar as regras da Seção 3.3. Abra a planilha **CÁLCULO_FÓRMULAS** para examinar essas fórmulas e verificar o modo como a função **SE** seleciona a regra que se aplica aos cálculos. (Uma explanação completa das outras funções utilizadas nas colunas G a I está além do escopo deste livro.)

O Box-Plot

PHStat2 Utilize o procedimento **BoxPlot** para criar um box-plot. Para criar um box-plot da Figura 3.4, abra a planilha **DA DOS** da pasta de trabalho **Fundos de Títulos**. Selecione **PHStat** → **Descriptive Statistics** → **Boxplot** (**PHStat** → **Estatística Descritiva** → **Box-Plot**). Na caixa de diálogo do procedimento (apresentada a seguir):

1. Insira **F1:F181** na legenda **Raw Data Cell Range (Intervalo de Células de Dados Brutos)** e marque a opção **First cell contains label (Primeira célula contém rótulo)**.
2. Clique em **Multiple Groups-Stacked (Múltiplos Grupos-Empilhados)** e insira **B1:B181** como **Grouping Variable Cell Range (Intervalo de Células de Agrupamento da Variável)**.
3. Insira um título em **Title**, marque a opção **Five-Number Summary (Resumo de Cinco Números)** e clique em **OK**.



O box-plot aparece em sua própria planilha de gráfico, em separado da planilha que contém o resumo de cinco números.

Excel Avançado Utilize a planilha **Plot** da pasta de trabalho **Boxplot** como um modelo para criar um box-plot. Como o Excel não inclui um box-plot como um de seus tipos de gráfico, criar um box-plot requer a aplicação de um “subterfúgio” criativo e avançado nas funções de elaboração de gráficos do Excel. Abra a planilha **PLOT_FORMULAS** para examinar esse “subterfúgio” (fora do escopo deste livro para que possa ser integralmente explicitado). Explore o gráfico contendo o box-plot para descobrir os 8(!) gráficos de linha em separado aos quais é aplicado o “subterfúgio” para que seja criada a imagem do box-plot. (Utilize a planilha **PLOT_DATA**, na mesma pasta de trabalho, como um modelo para calcular um resumo de cinco números e um box-plot em uma única planilha.)

GE3.4 MEDIDAS NUMÉRICAS DESCRITIVAS para uma POPULAÇÃO

A Variância e o Desvio-Padrão da População

Excel Avançado Utilize as funções **VARP** e **DESVPADP** para calcular a variância e o desvio-padrão da população, respectivamente. Insira essas funções no formato **VARP(intervalo de células da população)** e **DESVPADP(intervalo de células da população)**.

GE3.5 A COVARIÂNCIA e o COEFICIENTE de CORRELAÇÃO

A Covariância

Excel Avançado Utilize a planilha **CÁLCULO** da pasta de trabalho **Covariância** como um modelo para análise da covariância. (A planilha contém o conjunto com os 30 valores do Exemplo 3.17. Sobrescreva esses valores quando inserir o conjunto de valores de dados.) Siga as instruções na planilha para modificá-la quando você tiver menos ou mais do que 30 valores.

Abra a planilha **CÁLCULO_FÓRMULAS** para examinar as fórmulas utilizadas na planilha. Observe com especial atenção a **Área de Cálculos** (veja a Figura GE3.2 a seguir) que calcula os valores para \bar{X} e \bar{Y} , bem como os valores utilizados para calcular a covariância.

	E	F
4	Área de Cálculos	
5	XBarra	=MÉDIA(A4:A33)
6	YBarra	=MÉDIA(B4:B33)
7	n - 1	=CONT.NÚM(A4:A33)-1
8	Soma	=SOMA(C4:C33)
9	Covariância	=FB/F7

FIGURA GE3.2 Área de Cálculos para a covariância

O Coeficiente de Correlação

Excel Avançado Utilize a função **CORREL** para calcular o coeficiente de correlação. Insira essa função no formato **CORREL(intervalo de células dos valores de X, intervalo de células dos valores de Y)**.

Utilize a planilha **CÁLCULO** da pasta de trabalho **Correlação** como um modelo para análise da correlação ilustrado na Figura 3.9, no Exemplo 3.18, Seção 3.5. Nessa planilha, a fórmula da célula H10, **=CORREL(A4:A33,B4:B33)** calcula o coeficiente de correlação.