Capítulo 4

MEDIDAS DE DISPERSÃO

No Capítulo 3, foi mostrado que a média e a mediana determinam um valor central de uma amostra ou variável. Enquanto a mediana localiza a posição do dado ou observação situada no centro da amostra ordenada de forma crescente, e sem considerar os valores da variável, a média determina o valor central considerando todos os valores da variável. Por exemplo, as amostras X={28, 29, 30, 31, 32} e Y={21, 25, 29, 34, 41} têm o mesmo número de dados e, também, a mesma média 30. Entretanto, os desvios são diferentes, pois os desvios da variável X são –2, –1, 0, 1 e 2, e os desvios da variável Y são –9, –5, –1, 4 e 11. A comparação dessas duas amostras aponta a variabilidade ou dispersão de seus dados com relação à média como uma medida importante para descrever uma amostra ou variável. Esse raciocínio poderia ser repetido em variáveis com medianas iguais, porém com menor aplicação do que a média.

Você deve ter em mente que, se não houver variabilidade, a maior parte das medidas estatísticas não teria utilidade. Há várias formas de medir a variabilidade dos dados de uma variável. Uma primeira tentativa é medir o intervalo ou range de variação, definido como o resultado da diferença entre os valores máximo e mínimo da amostra ou variável, como apresentado no Exemplo 2.1 do Capítulo 2.

EXEMPLO 4.1

Determine o intervalo de variação da seguinte amostra:

31	38	19	27	24	42	32	18	43	15	39

Solução. Os valores mínimo e máximo são, respectivamente, 15 e 43. O intervalo ou range de variação dos dados da amostra é 28=43–15.

O resultado do Exemplo 4.1 mostra que os dados da amostra se distribuem dentro do intervalo de variação igual a 28. O conhecimento desse intervalo não auxilia muito na tentativa de medir a dispersão dos dados da variável, pois seu cálculo envolve apenas os valores extremos, deixando de considerar os demais valores da variável que também são importantes.

Desvio absoluto médio

No Capítulo 3, vimos que os desvios dos dados de uma amostra ou variável medem sua dispersão ao redor de sua média. Portanto, a tentativa inicial de quantificar a variabilidade seria calcular a soma de todos os desvios, isto é $\sum_{i=1}^{n}(X_i - \overline{X})$. No entanto, pela primeira propriedade da média, a soma dos desvios é sempre igual a zero. Tentando manter o conceito desvio como medida de variabilidade, pode-se utilizar a média dos valores absolutos[1] dos desvios, procedimento denominado *desvio absoluto médio* ou simplesmente *DAM*.[2]

O *Desvio absoluto médio-DAM* é obtido da expressão:

$$DAM = \frac{1}{n} \times \left(|X_1 - \overline{X}| + |X_2 - \overline{X}| + \cdots + |X_n - \overline{X}|\right)$$

$$DAM = \frac{1}{n} \times \sum_{i=1}^{n} |X_i - \overline{X}|$$

onde X_i é um valor genérico e $\overline{X}$ é a média da variável ou amostra.

EXEMPLO 4.2

Calcule o desvio absoluto médio da amostra do Exemplo 4.1.

Solução. A resposta foi obtida na planilha **Exemplo 4.2**, incluída na pasta **Capítulo 4**, como mostra a figura a seguir.

- No intervalo B4:B14 foi registrada a amostra.
- Na célula G5, foi calculada a média da amostra com =MÉDIA(B4:B14), retornando o valor 29,82.
- Na célula C6 foi calculado o desvio do dado 31 da amostra registrando a fórmula =B4-$G$5, retornando o valor 1,18. Depois, essa fórmula foi copiada até a célula C14.
  - O valor de média que mostra a célula G5 é 29,82, valor arredondado com duas casas decimais. Entretanto, o valor exato e registrado na memória do Excel é 29,8181818181818. Ao mesmo tempo, no cálculo dos desvios, o Excel utiliza o valor exato da média. Portanto, você poderá encontrar diferenças entre o resultado final do *DAM* obtido manualmente com a média e os desvios arredondados e o obtido com o Excel sem arredondar nenhum resultado intermediário.
- Na célula D4, foi calculado o valor absoluto do desvio do dado 31, calculado na célula C4, registrando a fórmula =ABS(C4) que retornou o valor 1,18. Depois essa fórmula foi copiada até a célula D14.
  - Em vez de utilizar duas colunas para calcular o desvio absoluto, poderia ter sido utilizada uma única coluna registrando na célula C4; por exemplo, a fórmula combinada =ABS(B4-$G$5) que depois seria copiada.
  - A função matemática **ABS(*número*)**[3] retorna o valor absoluto do argumento *número* que pode ser qualquer número do campo real. Pode-se dizer que o valor absoluto de um número é o próprio número sem o respectivo sinal, seja positivo ou negativo.
- Na célula G6 foi registrada a fórmula =SOMA(D4:D14) que retorna o resultado da soma dos desvios absolutos igual a 92,18.

1 O valor absoluto de um número é o valor desse número considerado positivo.
2 Este procedimento é apenas um registro, pois o *DAM* não ajuda na compreensão da dispersão, nem apresenta as vantagens matemáticas da variância e do desvio padrão.
3 Em inglês, a função ABS é *ABS*.

	A	B	C	D	E	F	G	H
1	Exemplo 4.2							
2								
3		Amostra	Desvio	Desvio Absoluto		Resultados		
4		31	1,18	1,18				
5		38	8,18	8,18		Média	29,82	
6		19	-10,82	10,82		Soma dos Desvios Absolutos	92,18	
7		27	-2,82	2,82		DAM	8,38	
8		24	-5,82	5,82				
9		42	12,18	12,18		Função DESV.MÉDIO	8,38	
10		32	2,18	2,18			=DESV.MÉDIO(B4:B14)	
11		18	-11,82	11,82				
12		43	13,18	13,18				
13		15	-14,82	14,82				
14		39	9,18	9,18				
15								

Com os resultados parciais obtidos, pode-se calcular o *DAM*=8,38:

- Manualmente a fórmula $DAM = \dfrac{\sum_{i=1}^{11} |X_i - \overline{X}|}{11} = \dfrac{92,18}{11} = 8,38$
- Registrando a fórmula =G6/CONT.NÚM(D4:D14) na célula G7 da planilha.

Uma forma direta de obter o resultado desejado é utilizar a função estatística DESV.MÉDIO do Excel que retorna o desvio absoluto médio da amostra informada. Na célula G9 foi registrada a fórmula =DESV.MÉDIO(B4:B14). No Apêndice 1, você encontrará a descrição completa dessa e de outras funções que serão apresentadas neste capítulo.

Comparado com a tentativa de medir a variabilidade com o *intervalo*, o *DAM* é a média dos desvios absolutos e utiliza todos os valores da variável ou amostra. Entretanto, o valor absoluto dos desvios é um resultado difícil de compreender e não aceita tratamento matemático com as propriedades, por exemplo, do quadrado do desvio que será utilizado a seguir.

Variância

Mantendo os *desvios* para medir a variabilidade de uma variável, o procedimento recomendado é utilizar a soma dos quadrados dos desvios, pois seu resultado é um valor mínimo, como mostrou a segunda propriedade da média apresentada no Capítulo 3.

Seja a variável $X = X_1, X_2, \cdots, X_N$ uma população. Define-se variância σ_X^2 da variável X da população contendo N dados:

$$\sigma_X^2 = \frac{1}{N} \times \left((X_1 - \mu_X)^2 + (X_2 - \mu_X)^2 + \cdots + (X_n - \mu_X)^2 \right)$$

$$\sigma_X^2 = \frac{1}{N} \sum_{i=1}^{N} (X_i - \mu_X)^2$$

Seja a variável $X = X_1, X_2, \cdots, X_n$ uma amostra. Define-se a variância S_X^2 da variável X da amostra contendo n dados:

$$S_X^2 = \frac{1}{n-1} \times \left((X_1 - \overline{X})^2 + (X_2 - \overline{X})^2 + \cdots + (X_n - \overline{X})^2 \right)$$

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

EXEMPLO 4.3

Calcule a variância da amostra e da população do Exemplo 4.1 utilizando as fórmulas e as funções estatísticas do Excel.

Solução. A resposta foi obtida na planilha **Exemplo 4.3**, incluída na pasta **Capítulo 4**, como mostra a figura seguinte e tendo presente as características de arredondamento dos resultados intermediários e finais já comentadas.

- No intervalo B4:B14 foi registrada a amostra, na célula G4 foi calculada quantidade de dados da amostra e na célula G5 foi calculada a média da amostra utilizando a fórmula =MÉDIA(B4:B14).
- No intervalo C4:C14 foram calculados os desvios e no intervalo D4:D14, os quadrados dos desvios começando por registrar a fórmula =C4^2 na célula D4. Depois essa fórmula foi copiada até a célula D14.
- Na célula G8 foi calculada e registrada a soma dos quadrados dos desvios igual a 997,64 com a fórmula =SOMA(D4:D14).
  - Utilizando a função matemática SOMAQUAD não é necessário construir a coluna dos quadrados dos desvios. A fórmula =SOMAQUAD(C4:C14) registrada na célula G9 retorna a soma dos quadrados dos valores registrados no intervalo C4:C14. No Apêndice 1, você encontrará a descrição completa dessa e de outras funções que serão apresentadas.

	A	B	C	D	E	F	G	H	I
1	Exemplo 4.3								
2									
3		Amostra	Desvio	(Desvio)²		Resultados			
4		31	1,18	1,40		*n*	11	=CONT.NÚM(B4:B14)	
5		38	8,18	66,94		Média	29,82	=MÉDIA(B4:B14)	
6		19	-10,82	117,03					
7		27	-2,82	7,94		Soma (Desvios)²			
8		24	-5,82	33,85		Fórmula	997,64	=SOMA(D4:D14)	
9		42	12,18	148,40		Função SOMAQUAD	997,64	=SOMAQUAD(C4:C14)	
10		32	2,18	4,76					
11		18	-11,82	139,67		Variância amostra			
12		43	13,18	173,76		Fórmula	99,76	=G6/(G4-1)	
13		15	-14,82	219,58		Função VAR	99,76	=VAR(B4:B14)	
14		39	9,18	84,31					
15						Variância população			
16						Fórmula	90,69	=G6/G4	
17						Função VAR	90,69	=VARP(B4:B14)	
18									

Cálculo da variância da amostra. Com os resultados parciais obtidos, pode-se calcular o valor da variância da amostra$S_X^2 = 99{,}76$, utilizando:

- Manualmente a fórmula $S_X^2 = \dfrac{\sum_{i=1}^{11}(X_i - \overline{X})^2}{11-1} = \dfrac{997{,}64}{10} = 99{,}76$
- Registrando a fórmula =G8/(G4-1) na célula G12 da planilha.
- Utilizando a função estatística VAR, registrando a fórmula =VAR(B4:B14) na célula G13.

Cálculo da variância da população. Com os resultados parciais obtidos, pode-se calcular o valor da variância da amostra $\sigma_X^2 = 90{,}69$, utilizando:

- Manualmente a fórmula $\sigma_X^2 = \dfrac{\sum_{i=1}^{11}(X_i - \mu_X)^2}{11} = \dfrac{997{,}64}{11} = 90{,}69$
- Registrando a fórmula =G8/G4 na célula G16 da planilha.
- Utilizando a função estatística VARP, registrando na célula G17 a fórmula =VARP(B4:B14).

O procedimento de cálculo manual da variância é bastante trabalhoso quando comparado com a utilização das funções estatísticas do Excel; entretanto, essas funções apenas auxiliam o cálculo e podem obscurecer o conceito. O Apêndice 3 deste capítulo mostra como utilizar doze funções para banco de dados ou listas de valores, conhecidas genericamente como BDfunções(*banco_dados*; *campo*; *critérios*). Algumas dessas doze funções são equivalentes às apresentadas. Ademais, esse apêndice apresenta também as funções SUBTOTAL, CONT.SE e SOMASE úteis para realizar operações com bancos de dados ou listas de valores.

Relação entre as variâncias

A partir das definições das variâncias da amostra e da população, o Exemplo 4.3 mostra os procedimentos de cálculo, incluindo as funções estatísticas VAR e VARP. Verifique que uma das variâncias pode ser obtida da outra se o tamanho da amostra também for conhecido. Para facilitar a relação entre as variâncias da população e da amostra repetimos a seguir suas fórmulas.

$$\sum_{i=1}^{N}(X_i - \mu_X)^2 = N \times \sigma_X^2$$

$$\sum_{i=1}^{n}(X_i - \overline{X})^2 = (n-1) \times S_X^2$$

Como os dois primeiros membros dessas expressões são iguais, é possível igualar os dois segundos membros, o que nos leva à seguinte igualdade:

$$N \times \sigma_X^2 = (n-1) \times S_X^2$$

Portanto, conhecida uma das variâncias, é possível calcular a outra, sendo necessário também conhecer o tamanho da amostra.

$$\sigma_X^2 = \frac{(n-1)}{N} \times S_X^2 \text{ e } S_X^2 = \frac{N}{n-1} \times \sigma_X^2$$

EXEMPLO 4.4

Calcule a variância da população a partir da variância da amostra do Exemplo 4.3, sabendo que o tamanho da amostra é 11.

Solução. A variância da população $\sigma_X^2 = 90{,}69$ pode ser obtida com a fórmula:

$$\sigma_X^2 = \frac{n-1}{N} \times S_X^2$$

$$\sigma_X^2 = \frac{10}{11} \times 99{,}76 = 90{,}69$$

Em vez de tentar memorizar a fórmula de transformação entre as variâncias, recomenda-se ter presente a seguinte orientação:

- A variância da amostra foi obtida como resultado da divisão da soma dos quadrados dos desvios pela quantidade de valores da amostra (n–1). Para obter o valor da variância da população, será necessário multiplicar a variância da amostra por (n–1) e, em sequência, dividi-la por n.
- A variância da amostra será o resultado da multiplicação da variância da população por n e, em sequência, divida-a por (n–1).

Características da variância

O procedimento de cálculo utilizando a soma dos quadrados dos desvios é bastante trabalhoso. No Apêndice 2, mostramos um procedimento de cálculo da variância que utiliza somente os dados da amostra e os quadrados desses dados, não sendo necessário utilizar a média e os desvios. Contudo, esse procedimento de cálculo perde força quando comparado com a utilização das funções estatísticas do Excel. A fórmula e o resultado da variância têm características importantes.

- A variância é sempre um número positivo.
- As fórmulas para a amostra e para a população têm o mesmo numerador, a soma dos quadrados dos desvios.
- A variância de uma variável considerada como população é a média aritmética dos quadrados dos desvios.
- A variância de uma variável considerada como amostra é também um tipo de média, pois a soma dos quadrados dos desvios é dividida pela quantidade de dados da variável menos um.[4]
- Para a mesma amostra de tamanho n, a variância da amostra é sempre maior do que a da população. Na medida em que o tamanho n da amostra aumenta, para n maior do que 30, o valor da variância da amostra se aproxima do valor da variância da população.
- Da mesma forma que a média, a variância é afetada pelos valores extremos da variável, ela não é uma medida resistente.
- Uma desvantagem da variância é sua unidade de medida, o quadrado da unidade de medida dos dados da amostra ou variável; outra desvantagem é operar com os valores dos desvios ampliados, pois os desvios são elevados ao quadrado.

Regras operacionais da variância

Há propriedades operacionais muito práticas. Para evitar muitos símbolos nas fórmulas, as variâncias serão representadas como $Var(X)$. Sendo a, b e c constantes, sempre se verifica:

- Se $Y = a$, $Var(Y) = 0$
- Se $Y = aX$, $Var(Y) = a^2\ Var(X)$
- Se $Y = X + a$, $Var(Y) = Var(X)$
- Se $Y = X + Z$, $Var(Y) = Var(X) + Var(Z) + 2\ Cov(X, Z)$
- Se $Y = aX + bZ$, $Var(Y) = a^2\ Var(X) + b^2\ Var(Z) + 2\ ab\ Cov(X, Z)$

Desvio padrão

Para definir da variância nos valemos da segunda propriedade da média: *a soma dos quadrados dos desvios é sempre um valor mínimo*, como foi apresentado no Capítulo 3. Uma desvantagem da variância é sua unidade de medida, o quadrado da unidade de medida dos dados da amostra ou variável; outra desvantagem é ampliar os desvios, pois são elevados ao quadrado. Por exemplo, se a amostra do Exemplo 4.3 se refere a peças rejeitadas por lote, a unidade de medida da variância da amostra será 99,76 peças rejeitadas ao quadrado, o que não faz muito sentido. Como a unidade de medida da variância não explica nada sobre as características dos valores da amostra, é definido o *desvio padrão* que mantém a unidade de medida dos valores da variável.

4 No cálculo da variância da amostra S^2, deve-se dividir por $(n\text{-}1)$ em vez de n para corrigir a tendência de S^2 subestimar σ^2; para que S^2 seja um estimador *não viesado*.

O *desvio padrão* da variável X é a raiz quadrada positiva de sua variância. Dessa maneira:

O *desvio padrão* considerado como população é: $\sigma_X = +\sqrt{\sigma_X^2}$.

O *desvio padrão* considerado como amostra é: $S_X = +\sqrt{S_X^2}$.

Essas definições mostram que para determinar o desvio padrão é necessário conhecer o valor da variância correspondente, da amostra ou da população.

EXEMPLO 4.4

Calcular o desvio padrão da amostra e da população do Exemplo 4.1 utilizando as fórmulas e as funções estatísticas do Excel.

Solução. A resposta foi obtida na planilha **Exemplo 4.4**, incluída na pasta **Capítulo 4**, como mostra a figura a seguir e tendo presente as características de arredondamento dos resultados intermediários e finais já comentadas. O registro da amostra, os cálculos dos resultados intermediários e a obtenção dos valores das variâncias da amostra e da população foram realizados da mesma forma como foi apresentado no Exemplo 4.3. Esse procedimento é necessário para mostrar o cálculo do desvio padrão a partir de sua definição, ou a partir do conhecimento da variância correspondente, amostra ou população. No entanto, esse procedimento de cálculo perde força quando comparado com a utilização das funções estatísticas do Excel.

	A	B	C	D	E	F	G	H
1	Exemplo 4.4							
2								
3		Amostra	Desvio	(Desvio)^2		Resultados		
4		31	1,18	1,40		*n*	11	
5		38	8,18	66,94		Média	29,82	
6		19	-10,82	117,03		Soma (Desvios)^2	997,64	
7		27	-2,82	7,94		Variância amostra	99,76	
8		24	-5,82	33,85		Variância população	90,69	
9		42	12,18	148,40				
10		32	2,18	4,76		Desvio padrão amostra		
11		18	-11,82	139,67		Fórmula	9,99	=RAIZ(G7)
12		43	13,18	173,76		Função DESVPAD	9,99	=DESVPAD(B4:B14)
13		15	-14,82	219,58				
14		39	9,18	84,31		Desvio padrão população		
15						Fórmula	9,52	=RAIZ(G8)
16						Função DESVPADP	9,52	=DESVPADP(B4:B14)
17								

Cálculo do desvio padrão da amostra. O valor do desvio padrão da amostra $S_X = 9{,}99$ pode ser obtido:

- Manualmente a fórmula $S_X = +\sqrt{S_X^2} = +\sqrt{99{,}76} = 9{,}99$
- Registrando a fórmula =RAIZ(G7) na célula G11 da planilha.
  - A função matemática **RAIZ(*número*)**[5] retorna a raiz quadrada positiva do argumento *número* que deve ser qualquer número positivo.
- Utilizando a função estatística DESVPAD ao registrar na célula G12 a fórmula =DESVPAD(B4:B14).

Cálculo do desvio padrão da população. O valor da desvio padrão da população $\sigma_X = 9{,}52$ pode ser obtido:

- Manualmente pela fórmula $\sigma_X = +\sqrt{\sigma_X^2} = +\sqrt{90{,}69} = 9{,}52$
- Registrando a fórmula =RAIZ(G8) na célula G15 da planilha.
- Utilizando a função estatística DESVPADP ao registrar na célula G16 a fórmula =DESVPADP(B4:B14).

5 Em inglês, a função RAIZ é *SQRT*.

Se a amostra do Exemplo 4.4 se refere à quantidade mensal de peças rejeitadas, o desvio padrão da amostra será 9,99 peças rejeitadas, pois o desvio padrão tem a mesma unidade dos dados da amostra ou variável. Da mesma maneira, o desvio padrão da população é $\sigma_X = +\sqrt{90{,}69} = 9{,}52$ peças rejeitadas. O procedimento de cálculo manual do desvio padrão é bastante trabalhoso quando comparado com a utilização das funções estatísticas do Excel; entretanto, essas funções apenas auxiliam o cálculo e podem obscurecer o conceito.

Relação entre os desvios padrão

A partir das definições dos desvios padrão da amostra e da população, o Exemplo 4.4 mostra os procedimentos de cálculo, incluindo as funções estatísticas DESVPAD e DESVPADP. Nesse caso, também, verifica-se que um dos desvios padrão pode ser obtido do outro se o tamanho da amostra também for conhecido. Em alguns casos é necessário operar com os valores do desvio padrão da população e do desvio padrão da amostra de uma variável, tentando sempre usar uma forma prática de obter um valor do outro. Da mesma forma como foi mostrada a relação entre a variância da amostra e a variância da população, as expressões a seguir mostram a relação entre os desvios padrão da população e da amostra.

$$\sigma_X = \sqrt{\frac{n-1}{N}} \times S_X \text{ e } S_X = \sqrt{\frac{N}{n-1}} \times \sigma_X$$

O procedimento recomendado para obter o valor de um desvio padrão em função do outro é, primeiro, realizar essa operação com as variâncias equivalentes, evitando carregar uma fórmula com o símbolo de raiz quadrada. Da mesma forma que a variância, as características do desvio padrão são:

- O desvio padrão é sempre um número positivo.
- Se os dados de uma variável forem iguais, o desvio padrão será zero.
- O desvio padrão não é uma medida resistente, pois é afetada pelos valores extremos da variável.

Significado do desvio padrão

O desvio padrão depende da soma dos quadrados dos desvios dos dados da variável com relação a sua média. Portanto, quanto menor for o desvio padrão, mais os valores da variável se aproximarão de sua média. Analisando a expressão do desvio padrão, podemos chegar a conclusões importantes:

- Qualquer dado da amostra ou variável com desvio menor do que o desvio padrão da variável estará mais próximo da média do que qualquer outro valor com desvio maior.
- Quanto mais os dados se afastarem da média, maior serão os desvios e, consequentemente, maior será o desvio padrão da variável.
- Duas variáveis com médias iguais e desvios padrão diferentes têm distribuições de frequências com formas diferentes. A distribuição da variável com maior desvio padrão será mais aberta do que a da variável com menor desvio padrão.

Qual a proporção de dados incluídos em um intervalo de desvios padrão ao redor da média de uma variável ou amostra? O *Teorema de Chebyshev* dá uma resposta para uma variável com qualquer tipo de distribuição de frequências.

Teorema de Chebyshev. Para qualquer conjunto de dados de uma amostra ou população, a proporção mínima de valores que se encontram dentro de k desvios padrão ao redor da média é pelo menos igual a $1-\frac{1}{k^2}$, sendo k uma constante maior do que 1.

A próxima tabela mostra a proporção mínima de dados dentro de k desvios padrão ao redor da média. Por exemplo, 75% dos dados de uma amostra ou variável estão distribuídos no intervalo de dois desvios padrão ao redor da média; entre menos dois e mais dois desvios padrão ao redor da média.

k	1,5	2	2,5	3	3,5	4
Proporção de dados	0,56	0,75	0,84	0,89	0,92	0,94

Outro exemplo, pelo menos sete dos onze dados da amostra do Exemplo 4.1 estão distribuídos no intervalo de dois desvios padrão ao redor da média 29,8; isto é, entre menos dois desvios padrão (9,8=29,8–2×9,99) e mais dois desvios padrão (49,8=29,8+29,99) ao redor da média. Verifique que no Exemplo 4.1 todos os dados estão distribuídos no intervalo de dois desvios padrão ao redor da média.

Regra prática

Pelo teorema de *Chebyshev*, é possível determinar a proporção mínima de dados de uma variável dentro de um determinado número de desvios padrão ao redor da média. A partir da média $\overline{X}$ e o desvio padrão S_X de uma amostra ou variável X, a *Regra Prática* permite estabelecer a proporção de valores distribuídos no intervalo $\overline{X} \pm k \times S_X$, considerando a forma da distribuição de frequências da variável X.

Regra Prática

A variável X tem n dados com média $\overline{X}$ e desvio padrão S_X.

$\overline{X} \pm 1 \times S_X$. Em uma distribuição simétrica com forma de sino, a porcentagem de dados contidos no intervalo de um desvio padrão ao redor da média é 68%. Para uma distribuição assimétrica com acentuada inclinação para um lado, essa porcentagem se aproxima de 90%.

$\overline{X} \pm 2 \times S_X$. Em uma distribuição simétrica com forma de sino, a porcentagem de dados contidos no intervalo de dois desvios padrão ao redor da média é 95%. Para uma distribuição assimétrica com acentuada inclinação para um lado, a porcentagem se aproxima de 100%.

$\overline{X} \pm 3 \times S_X$. Para todas as distribuições, a porcentagem de dados contidos no intervalo de três desvios padrão ao redor da média será próxima de 100%.

A *Regra Prática* atende à maioria das distribuições; entretanto, há casos em que será necessário construir o histograma para conhecer a forma da distribuição da amostra. A partir das conclusões obtidas da aplicação da *Regra Prática*, será possível determinar a forma do histograma, da distribuição de frequências dos dados como mostra o Exemplo 4.5.

EXEMPLO 4.5

Determine a porcentagem dos dados da amostra do Exemplo 4.1 incluídos no intervalo de um, dois e três desvios padrão ao redor da média.

Solução. Na planilha **Exemplo 4.5**, incluída na pasta **Capítulo 4**, foram determinadas as quantidades de dados incluídos nos intervalos de um, dois e três desvios padrão ao redor da média, como mostra a figura seguinte.

- Na célula H5, foi calculada a quantidade de dados, na célula H6, a média e, na célula H7, o desvio padrão da amostra utilizando as funções estatísticas correspondentes.
- No intervalo H10:J11, foram calculados os valores dos limites inferiores e superiores dos intervalos de um, dois e três desvios padrão ao redor da média, acompanhando a expressão $\overline{X} \pm k \times S_X$ cujas fórmulas do primeiro intervalo são as seguintes:
  - Com a fórmula =H6-H7 registrada na célula H10, foi calculado o limite inferior do intervalo de um desvio padrão.
  - Com a fórmula =H6+H7 registrada na célula H11, foi calculado o limite superior do intervalo de um desvio padrão. Para os demais limites, procede-se da mesma forma, considerando o número de desvios padrão adequados.

	A	B	C	D	E	F	G	H	I	J
1	Exemplo 4.5									
2										
3		Amostra	1 DP	2DP	3DP		Resultados			
4		14	1	1	1					
5		16	1	1	1		*n*	11		
6		12	0	1	1		Média	15,64		
7		17	1	1	1		Desvio padrão	2,80		
8		19	0	1	1					
9		17	1	1	1			1 DP	2 DP	3 DP
10		14	1	1	1		Limite Inferior	12,83	10,03	7,23
11		18	1	1	1		Limite Superior	18,44	21,24	24,04
12		10	0	0	1		Quantidade	8	10	11
13		17	1	1	1		Porcentagem	73%	91%	100%
14		18	1	1	1					
15										

Com os limites estabelecidos, nas colunas do intervalo C4:E14, são selecionados os dados contidos em cada intervalo utilizando as seguintes fórmulas:

- **Um desvio padrão ao redor da média, coluna C**. Na célula C4 foi registrada a fórmula =SE(E(B4>=$H$10;B4<=$H$11);1;0), que depois foi copiada até a célula C14.
- **Dois desvios padrão ao redor da média, coluna D**. Na célula D4 foi registrada a fórmula =SE(E(B4>=$I$10;C4<=$I$11);1;0), que depois foi copiada até a célula D14.
- **Três desvios padrão ao redor da média, coluna E**. Na célula E4 foi registrada a fórmula =SE(E(B4>=$J$10;D4<=$J$11);1;0), que depois foi copiada até a célula E14.

Para terminar, no intervalo H12:J12 são contados os dados contidos no intervalo de um, dois e três desvios padrão ao redor da média, e no intervalo H13:J13 são calculadas as respectivas porcentagens, obtendo os seguintes resultados $\overline{X} \pm 1 \times S_X$=73%, $\overline{X} \pm 2 \times S_X$=91% e $\overline{X} \pm 3 \times S_X$=100%. Portanto, 73%, 91% e 100% dos dados ou observações se distribuem, respectivamente, no intervalo de um, dois e três desvios padrão ao redor da média.

Medida relativa de dispersão

O desvio padrão tem duas características importantes:[6]

6 A variância também tem essas duas características.

- Considera que os desvios se distribuem de forma homogênea ao redor da média.
- É uma medida absoluta.

A comparação da dispersão de duas ou mais distribuições pelo simples confronto de seus desvios padrão nem sempre é suficiente, pois as amostras ou populações podem ter unidades diferentes ou, tendo a mesma unidade, seus valores de média podem estar bastante afastados.

O *coeficiente de variação CV* é o resultado de dividir o desvio padrão da variável pela sua média:

$$CV_{pop} = \frac{\sigma_X}{\mu_X} \qquad CV_{amo} = \frac{S_X}{\overline{X}}$$

A medida relativa de dispersão coeficiente de variação *CV* permite a comparação de distribuições, pois seu resultado é o desvio padrão por unidade de média. Em alguns casos, o resultado do *CV* é apresentado multiplicado por 100, em porcentagem. Comparando duas variáveis, a variável que tiver menor *CV* tem menor dispersão ou variabilidade.

EXEMPLO 4.6

A tabela a seguir registra os retornos mensais dos investimentos *A* e *B* durante os últimos seis meses. Interessa conhecer qual dos dois investimentos apresentou maior dispersão.

Solução. Na planilha **Exemplo 4.6**, incluída na pasta **Capítulo 4**, foi resolvido o exemplo, começando pelo cálculo das médias e dos desvios padrão dos retornos dos dois investimentos e terminando pelo cálculo do coeficiente de variação de cada investimento.

	A	B	C	D	E	F	G
1	Exemplo 4.6						
2							
3		Retornos			Resultados		
4		A	B				
5		5%	6%			A	B
6		9%	7%		Média	9,33%	7,17%
7		15%	9%		Desvio padrão	3,72%	1,17%
8		12%	7%		CV amostra	39,9%	16,3%
9		9%	6%				
10		6%	8%				
11							

Como o *CV* do investimento *A* é maior do que o *CV* do investimento *B*, a variabilidade[7] do investimento *A* foi maior do que a do investimento *B*.

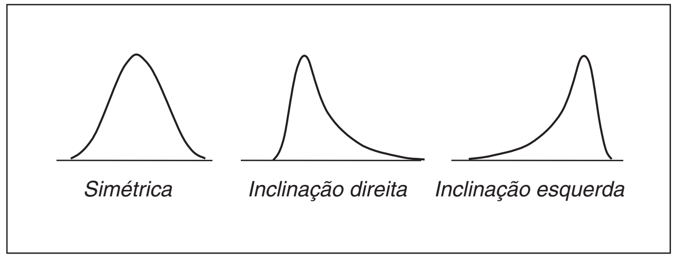
Análise da forma da distribuição de uma amostra

Como nem todas as amostras ou variáveis têm moda ou um único valor de moda, a mediana e a média são as medidas mais usuais de tendência central. Em uma distribuição simétrica de frequências, a média e a mediana têm o mesmo valor. Se os valores da média e da mediana forem diferentes, a distribuição será assimétrica e quanto mais os valores da média e da mediana se afastarem, maior será a inclina-

7 Em finanças, a variabilidade é o *risco* do investimento; o investimento A apresentou mais risco do que o investimento *B*.

ção da distribuição na direção de uma das caudas. Por exemplo, se um ou mais dados da amostra forem valores maiores do que a maioria dos demais dados, então a média será maior do que a mediana e a distribuição de frequências terá inclinação direita ou positiva, conforme mostra a Figura 4.1. Da mesma forma, é possível analisar para o lado esquerdo. A inclinação de uma distribuição é medida pelo *coeficiente de inclinação* da distribuição.

FIGURA 4.1
Distribuições de frequências, simétrica e inclinada.



Apesar de duas amostras ou variáveis apresentarem a mesma dispersão e inclinação, essas características não serão suficientes para supor que as duas distribuições tenham a mesma forma, atributo denominado *curtose*. A curtose é medida pelo *coeficiente de curtose* que compara a distribuição de frequências da amostra com a distribuição normal.

EXEMPLO 4.7

A tabela a seguir registra uma amostra ordenada de 28 retornos de diversos investimentos no mesmo período. Calcule e analise a forma da distribuição dessa amostra.

–2,1%	10,1%	10,6%	16,3%	16,3%	20,4%	21,0%
23,6%	24,7%	24,8%	26,2%	26,6%	27,0%	27,8%
28,6%	30,2%	30,3%	30,7%	32,0%	32,5%	32,6%
34,3%	35,5%	36,7%	52,9%	59,5%	76,2%	114,7%

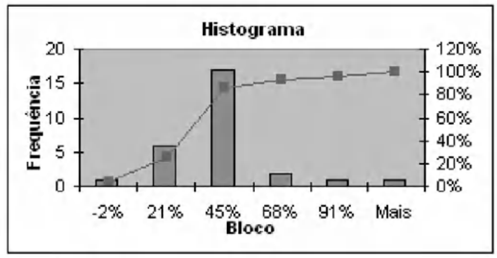
Solução. Na planilha **Exemplo 4.7**, incluída na pasta **Capítulo 4**, foi analisada a forma da distribuição da amostra anterior registrada no intervalo B4:C17, como mostra a figura seguinte. Analisemos os resultados registrados nessa planilha.

- O histograma foi construído utilizando a ferramenta de análise *Histograma*, depois de ajustar a formatação do gráfico, os títulos e as escalas. O histograma mostra que a distribuição apresenta inclinação para a direita.
- No intervalo F14:F15, foram calculadas a média e a mediana, respectivamente iguais a 32,1% e 28,2%. Como a média é maior do que a mediana, a distribuição tem inclinação para a direita.
- O *coeficiente de inclinação* igual a 2,24 foi calculado com a função estatística DISTORÇÃO do Excel, registrando a fórmula =DISTORÇÃO(B4:C17) na célula F16. O resultado positivo mostra que a distribuição tem inclinação para a direita. Se o resultado fosse negativo, a inclinação seria negativa, e se fosse igual a zero, a distribuição seria simétrica.
  - **DISTORÇÃO (*núm1; núm2; ... ; núm30*)**

    A função estatística DISTORÇÃO (*núm1; núm2; ... ; núm30*) retorna o coeficiente de inclinação dos valores numéricos *núm1; núm2; ... ; núm30*. Cada um desses *núm* pode ser um intervalo de células de uma planilha contendo valores numéricos ou assemelhados. Nesse exemplo, a amostra do intervalo B4:C17 foi registrado no primeiro argumento *núm1*. Mais informações sobre essa função e outras formas de utilizá-la estão disponíveis no Apêndice 1deste capítulo.

Exemplo 4.7

Retorno	
-2,1%	28,6%
10,1%	30,2%
10,6%	30,3%
16,3%	30,7%
16,3%	32,0%
20,4%	32,5%
21,0%	32,6%
23,6%	34,3%
24,7%	35,5%
24,8%	36,7%
26,2%	52,9%
26,6%	59,5%
27,0%	76,2%
27,8%	114,7%



Resultados		
Média	32,1%	=MÉDIA(B4:C17)
Mediana	28,2%	=MED(B4:C17)
Inclinação	2,24	=DISTORÇÃO(B4:C17)
Curtose	6,94	=CURT(B4:C17)

- O *coeficiente de curtose* igual a 6,94 foi calculado com a função estatística CURT do Excel registrando a fórmula =CURT(B4:C17) na célula F17. O resultado positivo mostra que a distribuição de frequências será concentrada ao redor da média, distribuição com pico. Se o resultado fosse negativo, a distribuição seria achatada, plana, e se fosse igual a zero, a distribuição de frequências seria a própria distribuição normal.
  - **CURT (*núm1; núm2; ... ; núm30*)**
    A função estatística CURT (*núm1; núm2; ... ; núm30*) retorna o *coeficiente de curtose* dos valores numéricos *núm1; núm2; ... ; núm30*. Cada um desses *núm* pode ser um intervalo de células de uma planilha contendo valores numéricos ou assemelhados. Nesse exemplo, a amostra do intervalo B4:C17 foi registrada no primeiro argumento *núm1*. Mais informações sobre essa função e outras formas de utilizá-la estão disponíveis no Apêndice 1 deste capítulo.

Modelo análise numérica

A determinação isolada de medidas estatísticas numéricas leva à obtenção de respostas parciais. O conjunto dessas medidas melhora a compreensão e a visualização das medidas numéricas em um gráfico complementa a análise da amostra. Realizar esse processo de medição de forma manual é muito trabalhoso; entretanto, utilizando a planilha Excel, consegue-se diminuir um pouco esse trabalho. O conjunto desses resultados é apresentado no *Modelo Análise Numérica* construído pelo autor na pasta **Modelo Análise Numérica** que está disponível na página do livro, no site da Editora. A Figura 4.2 mostra esse modelo para uma amostra de tamanho n=77, incluindo o gráfico que destaca as medidas numéricas mais importantes.

O *Modelo Análise Numérica* calcula as medidas mais importantes e constrói um gráfico com os dados da amostra, os intervalos de um, dois e três desvios padrão ao redor da média, a identificação de uma linha (no eixo de abscissas) com as cinco medidas estatísticas que ajudam a descrever a forma da distribuição de frequências, e a identificação de linhas verticais da média e do primeiro, segundo e terceiro quartis. Para operar o *modelo*:

- Recomenda-se zerar os dados da amostra diretamente na planilha.
- Informar a série de valores numéricos a partir da célula B4. Não há limite de tamanho da amostra, apenas os limites impostos pela planilha Excel e a memória do microcomputador utilizada.
- Depois de informar a amostra, pressione o botão **Calcular**. O *modelo* fornecerá os resultados do intervalo F4:F24 e construirá ou atualizará o gráfico.
- No intervalo E9:F11, é possível obter respostas específicas para um dado da amostra quanto à sua posição no intervalo E9:F9, ao seu percentil no intervalo E10:F10 e ao dado referente a um determinado percentil no intervalo E11:F11.
- Preste atenção ao aviso de recálculo que o *modelo* apresenta na célula mesclada H2.

	A	B	C	D	E	F
1	Análise numérica de uma amostra					
2						
3		Amostra		Análise da amostra		
4		699		Quantidade de dados		77
5		668		Valor Mínimo		651,0
6		655		Valor Máximo		795,0
7		730				
8		684		Medidas de Posicionamento		
9		733		Ordem de	684,0	32
10		688		Percentil de	734,0	92,1%
11		675		Percentil	53%	700,3
12		729		Primeiro quartil		678,0
13		706		Segundo quartil		699,0
14		719		Terceiro quartil		719,0
15		733				
16		795		Medidas de Tendência Central		
17		674		Média		697,1
18		678		Mediana		699,0
19		661		Moda		678,0
20		669				
21		657		Medidas de Dispersão		
22		732		Variância		850,9
23		701		Desvio padrão		29,2
24		680		CV		4,18%
25		704				

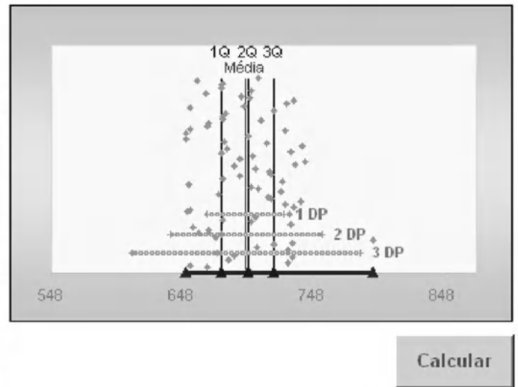


FIGURA 4.2 Modelo Análise Numérica.

Ainda, no intervalo D26:F29, não mostrado na Figura 4.2, o modelo apresenta a contagem e a proporção de dados dentro de um, dois e três desvios padrão ao redor da média.

Ferramenta de análise *Estatística Descritiva*

A partir de uma amostra quantitativa discreta registrada em uma planilha Excel, a ferramenta de análise *Estatística descritiva* retornará uma tabela com um grupo de resultados estatísticos. Para utilizar a ferramenta de análise *Estatística descritiva*,[8] a amostra que será analisada deve estar registrada em uma planilha como a **Ferram. Estatística Descritiva**, incluída na pasta **Capítulo 4**, com a amostra do Exemplo 4.1, onde:

- No intervalo B3:B14 foram registrados os valores numéricos da amostra, incluindo o nome Amostra na célula B3. Os valores da amostra devem ser registrados em uma coluna identificados com um único intervalo. Essa ferramenta de análise pode gerar tabelas para mais de uma amostra simultaneamente com a condição de terem o mesmo tamanho e serem registradas em intervalos contíguos.
- Depois de selecionar **Análise de dados** dentro do menu **Ferramentas**, o Excel apresentará a caixa de diálogo **Análise de dados** com todas as ferramentas de análise disponíveis, como mostrado na Figura 1.7 do Capítulo 1 deste livro.
- Escolhendo a ferramenta **Estatística descritiva** e depois pressionando o botão **OK**, você receberá a caixa de diálogo **Estatística descritiva** mostrada na Figura 4.3 depois de selecionadas algumas opções.
  - Pressionando o botão **Ajuda** dessa caixa de diálogo, o Excel apresentará a página *Sobre a caixa de diálogo Estatística descritiva* pertencente à *Ajuda do Excel*.

As informações que devem ser registradas no quadro **Entrada** da caixa de diálogo da ferramenta *Estatística descritiva* são, conforme apresentado na Figura 4.3:

- **Intervalo de entrada:** Informe o intervalo de células da planilha, no qual os dados estão registrados, nesse caso, o intervalo B3:B14 que inclui a célula onde foi registrado o título *Amostra*, ou rótulo no Excel.

8 Em inglês, a ferramenta ESTATÍSTICA DESCRITIVA é *DESCRIPTIVE STATISTICS*.

- **Agrupado por:** Selecionamos **Colunas**, pois a amostra foi registrada em uma coluna. Em geral, o Excel selecionará automaticamente depois de ter informado intervalo da amostra.
- **Rótulos na primeira linha**. Tendo escolhido **Colunas** no item anterior, necessariamente selecionaremos **Rótulos na primeira linha**, pois na primeira célula da série foi incluído o nome *Amostra*.

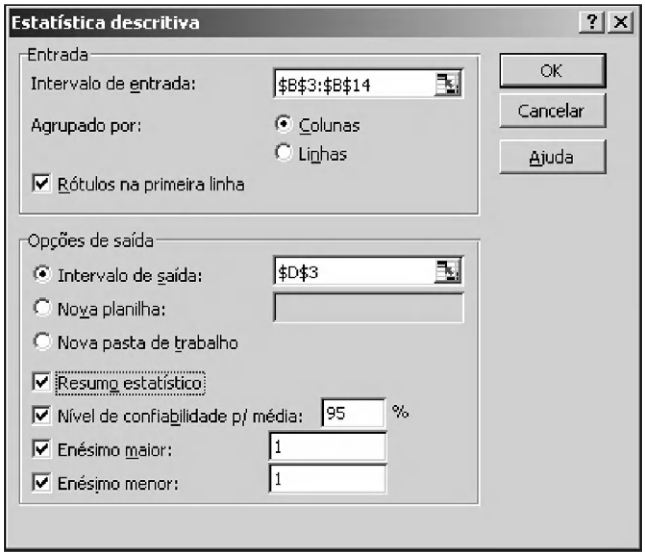


FIGURA 4.3 Caixa de diálogo da ferramenta **Estatística descritiva.**

Na primeira parte do quadro **Opções de saída**, deve ser obrigatoriamente informado um endereço a partir do qual a ferramenta *Estatística descritiva* registrará os resultados. Há três alternativas nas quais não é necessário informar esse endereço, identificadas por três *botões de opção* que aceitam a escolha de uma única alternativa:

- **Intervalo de saída:** Os resultados serão apresentados na mesma planilha a partir da célula informada, nesse caso D3. Depois de clicar com o botão esquerdo do mouse dentro da caixa correspondente, o endereço pode ser registrado digitando D3, ou *clicando* com o botão esquerdo do *mouse* na célula D3; nesse caso, será registrado o endereço com os dois cifrões, $D$3. Esse endereço é o da célula superior esquerda da tabela de respostas que a ferramenta construirá. Também, o Excel automaticamente definirá o tamanho da área dos resultados e exibirá uma mensagem se a tabela de saída estiver prestes a substituir dados existentes.
- **Nova planilha:** Os resultados serão apresentados a partir da célula A1 de uma nova planilha da mesma pasta.
  - Se não for informado nenhum endereço, a ferramenta inserirá uma nova planilha com o nome **Plan** seguido de um número sequencial; por exemplo, escolhendo essa alternativa na pasta **Capítulo 4**, a ferramenta inserirá a planilha **Plan1**.
  - Há a alternativa de informar o nome da planilha na caixa dessa alternativa; por exemplo, registrando o nome *Teste*, a ferramenta inserirá na mesma pasta uma nova planilha com o nome **Teste**.
- **Nova pasta de trabalho**. Os resultados serão apresentados em uma nova pasta e a partir da célula A1 da planilha **Plan1**.

Em continuação, no quadro **Opções de saída**, há quatro alternativas não excludentes de resultados possíveis. Nelas é possível selecionar qualquer combinação marcando nas quatro *caixas de seleção*, com a condição de selecionar pelo menos uma delas.

FIGURA 4.4
Resumo estatístico da ferramenta Estatística Descritiva.

	A	B	C	D	E	F	G
1	Ferramenta de Análise ESTATÍSTICA DESCRITIVA						
2							
3		Amostra		*Amostra*			Resultados
4		31					com fórmulas
5		38		Média	29,81818		29,81818
6		19		Erro padrão	3,011548		3,011548
7		27		Mediana	31		31
8		24		Modo	#N/D		#N/D
9		42		Desvio padrão	9,988175		9,988175
10		32		Variância da amostra	99,76364		99,76364
11		18		Curtose	-1,46891		-1,46891
12		43		Assimetria	-0,11748		-0,11748
13		15		Intervalo	28		28
14		39		Mínimo	15		15
15				Máximo	43		43
16				Soma	328		328
17				Contagem	11		11
18				Maior(1)	43		43
19				Menor(1)	15		15
20				Nível de confiança(95,0%)	6,710148		6,710148
21							

- **Resumo estatístico**: Marcando este item, a ferramenta de análise apresentará o resumo estatístico completo, conforme apresentada na Figura 4.4.
- **Nível de confiabilidade p/a média:** A resposta dessa seleção será compreendida ao estudar *Estimação* no Capítulo 11 deste livro. Neste caso, registramos 95, que representa 95% de *intervalo de confiança*.
- **Enésimo maior:** escolhendo este item e informando o valor 1, a ferramenta fornecerá o maior valor da *Amostra* ordenada de forma crescente. Se for informado o valor 2, então a ferramenta apresentará o penúltimo valor da amostra, e assim sucessivamente. **Enésimo maior** retorna o mesmo resultado da função estatística MAIOR, apresentada no Capítulo 3.
- **Enésimo menor:** Escolhendo este item e informando o valor 1, a ferramenta fornecerá o menor valor da *Amostra* ordenada de forma crescente. Se for informado o valor 2, então a ferramenta apresentará o segundo elemento da série, e assim sucessivamente. **Enésimo menor** retorna o resultado da função estatística MENOR, apresentada no Capítulo 3.

Depois de realizar as escolhas e pressionar o botão **OK**, a ferramenta registra os resultados a partir da célula D3, Figura 4.4.

Análise dos resultados

No intervalo G5:G20, foram registrados os mesmos resultados do intervalo E5:E20 da ferramenta de análise, porém calculados com fórmulas e funções estatísticas, algumas delas já conhecidas. Nem todos os resultados registrados na tabela da Figura 4.4 foram apresentados até o momento no livro; por exemplo, *Erro padrão* e *Nível de confiança (95%)*, mas que a seguir é feita uma introdução.

- **Erro padrão**. O *Erro padrão* é o *erro amostral* S_e estudado na Distribuição Amostral, no Capítulo 10 deste livro. O valor de S_e é calculado com a expressão, também registrado na célula G6 da planilha **Ferram. Estatística Descritiva**:

$$S_e = \frac{S_X}{\sqrt{n}} = \frac{9{,}988175}{\sqrt{11}} = 3{,}0011548$$

- **Nível de confiança(95%).** O *Nível de confiança* é estudado no Capítulo 11, sendo 95% o percentual de acerto da estimativa da média da população. O resultado 6,710148 da célula E20 é o *erro de estimação* com distribuição t, $t_{(1-0{,}95)/2} \times S_e = 2{,}228139 \times 3{,}011548 = 6{,}710148$. Esse resultado foi calculado na célula G20 com a função estatística INVT e a fórmula =INVT(0,05;10)*G7, que será apresentada no Capítulo 11.

EXEMPLO 4.8

Analise as distribuições de frequências das amostras *A* e *B* registradas na tabela seguinte utilizando a ferramenta de análise *Estatística descritiva.*

A	100	120	120	120	120	120	120	140	140	140	140	160	160	180
B	88,6	108,5	108,6	128,5	128,6	128,5	128,6	148,6	148,5	148,6	148,6	148,6	148,6	168,6

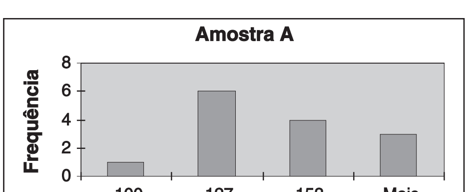
Solução. As amostras *A* e *B* e os resultados da ferramenta de análise foram registrados na planilha **Exemplo 4.8**, incluída na pasta **Capítulo** 4, como mostra a figura seguinte depois de ajustar as larguras das colunas e a formatação dos resultados. Analisando os resultados das medidas estatísticas, verificamos que as amostras *A* e *B* têm o mesmo valor de média igual a 134,29, medianas diferentes, respectivamente, 130 e 138,55, e desvio padrão praticamente iguais, respectivamente, 21,38 e 21,39. Comparando somente as médias e os desvios padrão, aparentemente, parece que as amostras têm a mesma forma de distribuição. Entretanto, a diferença de medianas mostra que não é assim.

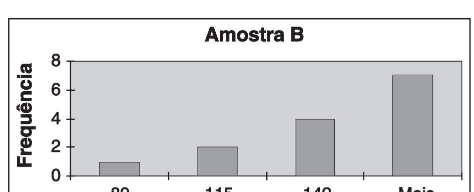
- Como a média da amostra *A* é maior do que a mediana, pode-se deduzir que a distribuição de frequências da amostra *A* tem inclinação positiva. Essa inclinação também é confirmada pelo resultado *Assimetria* igual a 0,67 que, por ser positivo, indica a inclinação positiva da distribuição.

No caso da amostra *B*, ocorre o contrário: ela tem inclinação para a esquerda, como confirmado também pelo resultado *Assimetria* igual a –0,66 que, por ser negativo, indica a inclinação negativa da distribuição.

	A	B	C	D	E	F	G	H
1	**Exemplo 4.8**							
2								
3		**Amostra A**	**Amostra B**		*Amostra A*		*Amostra B*	
4		100	88,6					
5		120	108,5		Média	134,29	Média	134,29
6		120	108,6		Erro padrão	5,71	Erro padrão	5,72
7		120	128,5		Mediana	130,00	Mediana	138,55
8		120	128,6		Modo	120,00	Modo	148,60
9		120	128,5		Desvio padrão	21,38	Desvio padrão	21,39
10		120	128,6		Variância da amostra	457,14	Variância da amostra	457,50
11		140	148,6		Curtose	0,19	Curtose	0,18
12		140	148,5		Assimetria	0,67	Assimetria	-0,66
13		140	148,6		Intervalo	80,00	Intervalo	80,00
14		140	148,6		Mínimo	100,00	Mínimo	88,60
15		160	148,6		Máximo	180,00	Máximo	168,60
16		160	148,6		Soma	1880,00	Soma	1880,00
17		180	168,6		Contagem	14,00	Contagem	14,00
18								

Para facilitar a confirmação da análise anterior, com a ferramenta de análise *Histograma*, foram construídos os histogramas a partir da linha 20 da planilha **Exemplo 4.8**. Analisando os histograma, verifica-se que as distribuições são diferentes, pois enquanto a distribuição de frequências da amostra *A* tem inclinação para a direita, a da amostra *B* é para a esquerda.





EXEMPLO 4.9

Continuando com o Exemplo 4.8. Analise as distribuições das amostras *A* e *B* considerando as seguintes cinco medidas de posição, mínimo, primeiro quartil, mediana, terceiro quartil e máximo.

Solução. As amostras *A* e *B* e os resultados da ferramenta de análise foram registrados na planilha **Exemplo 4.9**, incluída na pasta **Capítulo** 4, como mostra a próxima figura. No intervalo F6:G10, estão registrados os resultados: *Mínimo*, Q_1, *Mediana*, Q_3 e *Máximo* de cada amostra. Note que essas cinco medidas estão registradas em ordem crescente dos resultados. Analisando esses resultados, obtemos:

- As duas amostras têm o mesmo *intervalo* igual a 80=180–100=168,8–88,6.
- A diferença entre o terceiro quartil e o primeiro quartil das duas amostras é o mesmo valor e igual a 20. Esse resultado mostra que 50% dos dados em cada amostra se distribuem entre os dois quartis.
- A *mediana* de cada amostra está situada no centro de Q_1 e Q_3.
- A diferença entre o Q_1 e o *Mínimo* da amostra *A* é 20, enquanto a da amostra *B* é 39,9.
- Da mesma maneira, a diferença entre o *Máximo* e o Q_3 da amostra *A* é 40, e a da amostra *B* é 20.

Exemplo 4.9

Amostra A	Amostra B
100	88,6
120	108,5
120	108,6
120	128,5
120	128,6
120	128,5
120	128,6
140	148,6
140	148,5
140	148,6
140	148,6
160	148,6
160	148,6
180	168,6

Resultados

	Amostra A	Amostra B
Mínimo	100	88,6
Q_1	120	128,5
Mediana	130,0	138,6
Q_3	140	148,6
Máximo	180	168,6

Intervalo entre Q_1 e Q_3

Os resultados do Exemplo 4.9 ajudarão a compreender o intervalo entre o primeiro quartil e o terceiro quartil, denominado *IEQ*,[9] e as vantagens do diagrama *Boxplot* que será apresentado em sequência. O primeiro quartil, a mediana e o terceiro quartil avaliam a forma da parte central e a variabilidade da distribuição de frequências da amostra. O *IEQ* é o resultado da diferença entre o terceiro quartil Q_3 e o primeiro quartil Q_1:

$$IEQ = Q_3 - Q_1$$

As características importantes do *IEQ* são:

- É uma medida simples, fácil de ser calculada e automatizada.
- Mede a distribuição da metade dos dados da amostra situados ao redor da mediana.
- É uma medida resistente, pois não é afetado pelos dados extremos da amostra ou variável.
- É parecido com o *intervalo*; entretanto, essas três medidas Q_1, *mediana* e Q_3 dão mais informações.

9 Em inglês, IEQ é *IQR – InterQuartile Range*.

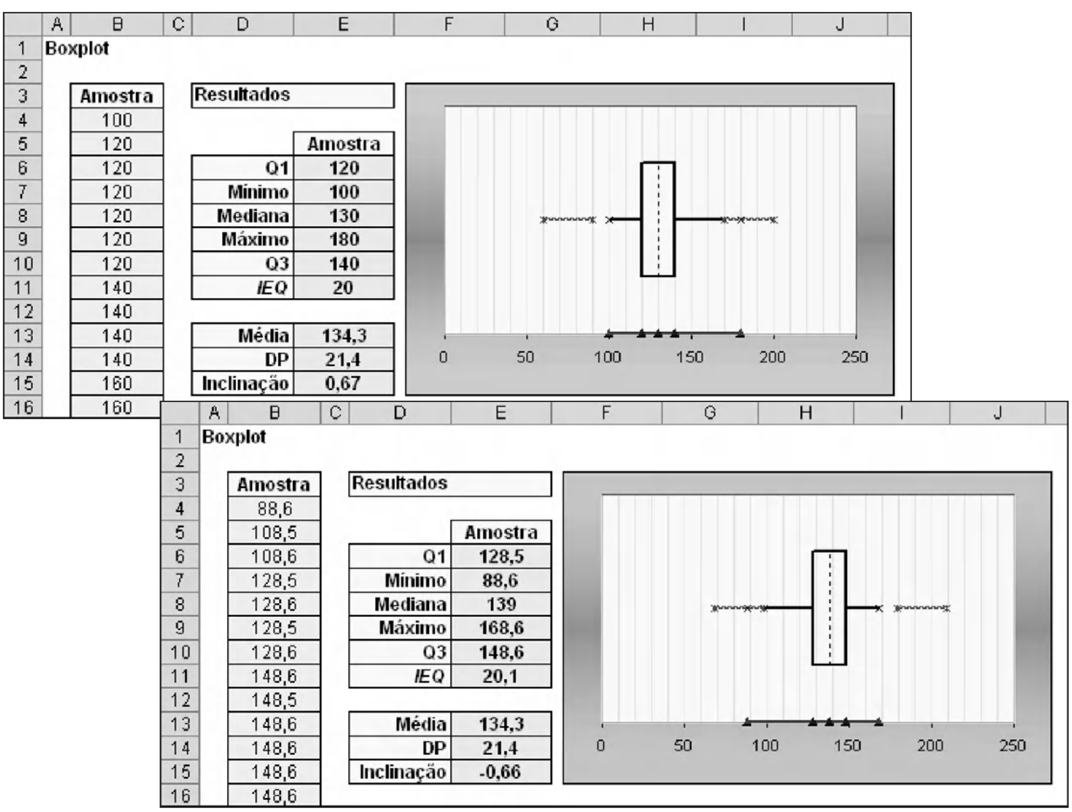
- Contudo, essa medida não é suficiente para avaliar a variabilidade de uma amostra ou variável, pois envolve apenas os valores centrais, deixando de considerar os valores extremos que também são importantes, os restantes 50% dos dados.

Boxplot

Embora os três resultados Q_1, mediana e Q_3 mostrem a forma da distribuição de 50% dos valores ao redor da mediana de uma amostra ou variável, o conjunto formado por esses cinco resultados:[10] *mínimo*, Q_1, *mediana*, Q_3 e *máximo* permitirão obter muitas informações sobre a forma da distribuição de frequências.

O *boxplot*[11] é a forma gráfica para mostrar o conjunto dos cinco resultados estatísticos e obter informações diretas sobre a forma da distribuição de frequências da amostra ou variável. O *boxplot* da Figura 4.5, planilha à esquerda, mostra que a inclinação da amostra *A* é positiva ou para a direita, confirmando o resultado obtido no Exemplo 4.9. O *boxplot* da Figura 4.5, planilha à direita, mostra que a inclinação da amostra *B* é negativa ou para a esquerda, confirmando também o resultado obtido no Exemplo 4.9. No gráfico do *boxplot*, foi incluída uma linha (no eixo de abscissas) com as cinco medidas estatísticas que ajudam a descrever a forma da distribuição de frequências, como mostrado no **Modelo Análise Numérica**. Observe que cada amostra tem um *boxplot* diferente que registra:

- Uma medida de tendência central, a *mediana*.
- Duas medidas de variabilidade ou dispersão, o *intervalo* e o *IEQ*.
- O tipo de *inclinação* por comparação da *mediana* com relação aos valores extremos.
- Os possíveis dados suspeitos.


Boxplot

Amostra
100
120
120
120
120
120
120
140
140
140
140
160
160

Resultados

	Amostra
Q1	120
Mínimo	100
Mediana	130
Máximo	180
Q3	140
IEQ	20
Média	134,3
DP	21,4
Inclinação	0,67

Boxplot

Amostra
88,6
108,5
108,6
128,5
128,6
128,5
128,6
148,6
148,5
148,6
148,6
148,6
148,6

Resultados

	Amostra
Q1	128,5
Mínimo	88,6
Mediana	139
Máximo	168,6
Q3	148,6
IEQ	20,1
Média	134,3
DP	21,4
Inclinação	-0,66

FIGURA 4.5 *Boxplot* das amostras *A* e *B*, Exemplo 4.9.

10 Em inglês, conhecido como *Five-number Summary*.
11 Preferimos manter o nome *boxplot* em inglês.

Dado suspeito

É denominado *dado suspeito*[12] o dado de uma amostra extremamente diferente da maioria dos dados da amostra. Como qualquer amostra pode conter *dados suspeitos*, é importante estar preparado para detectá-lo e analisar sua causa.

- Se o dado suspeito tiver sua origem em um erro de registro; por exemplo, o valor medido 135 foi registrado como 2.135. Nesse caso, o erro pode ser corrigido e a característica suspeita pode ser eliminada do dado amostrado.
- O que fazer se o dado suspeito foi corretamente amostrado e registrado? Se a população está sendo amostrada através de uma pesquisa de indivíduos de uma determinada população, um dado suspeito poderá ser originado por um individuo que não pertence à população definida. O dado suspeito também pode ser evidência de um acontecimento extraordinário ou uma variabilidade não esperada da variável. Em qualquer caso, os dados suspeitos sem causa aparente associada à população devem ser retirados da amostra, registrando esse evento.

O valor X de uma variável é considerado *possível suspeito* se estiver no intervalo $Q_1 - 3\times IEQ < X < Q_1 - 1{,}5\times IEQ$ ou no intervalo $Q_3 + 1{,}5\times IEQ < X < Q_3 + 3\times IEQ$.
O valor X de uma variável é considerado *suspeito* se $X < Q_1 - 3\times IEQ$ ou $X > Q_3 + 3\times IEQ$.

Uma estratégia para tratar dados suspeitos e outras irregularidades é utilizar métodos numéricos resistentes que pouco são afetados pelos dados suspeitos. Uma das aplicações do *IEQ* é a detecção de valores suspeitos de uma variável. Embora o *IEQ* ajude a retirar um dado da amostra por considerá-lo suspeito, essa decisão deve ser acompanhada de um criterioso julgamento.

EXEMPLO 4.10

Calcule o *IEQ* das amostras *A* e *B* do Exemplo 4.9 e verifique a existência de dados suspeitos.

Solução. A figura a seguir mostra a resolução deste exemplo na planilha **Exemplo 4.10**, incluída na pasta **Capítulo 4**. A primeira parte dos resultados é igual ado Exemplo 4.9, adicionando o intervalo F11:G11 para o cálculo do *IEQ* de cada amostra. Depois, no intervalo E13:I15, foram calculados os limites dos dados suspeitos indicados nesta tabela.

Exemplo 4.10

Amostra A	Amostra B
100	88,6
120	108,5
120	108,6
120	128,5
120	128,6
120	128,5
120	128,6
140	148,6
140	148,5
140	148,6
140	148,6
160	148,6
160	148,6
180	168,6

Resultados

	Amostra A	Amostra B
Mínimo	100	88,6
Q_1	120	128,5
Mediana	130,0	138,6
Q_3	140	148,6
Máximo	180	168,6
IEQ	20	20,1

	Q_1-3x*IEQ*	Q_1-1,5x*IEQ*	Q_3+1,5x*IEQ*	Q_3+3x*IEQ*
Amostra *A*	60	90	170	200
Amostra *B*	68,2	98,4	178,8	208,9

12 Em inglês, *dados suspeitos* são *outliers*.

- Amostra *A*.
  - Na cauda inferior da distribuição, são suspeitos os valores menores do que 60, e os valores entre 60 e 90 são possíveis suspeitos. Como o valor mínimo é 100, essa amostra não tem valores suspeitos nessa região.
  - Na cauda superior da distribuição, são suspeitos os valores maiores do que 200, e os valores entre 170 e 200 são possíveis suspeitos. O único valor possível de suspeita é o valor máximo 180.
- Amostra *B*.
  - Na cauda inferior da distribuição, são suspeitos os valores menores do que 68,2. Os valores entre 68,2 e 98,4 são possíveis suspeitos como o valor mínimo é 88,6.
  - Na cauda superior da distribuição, são suspeitos os valores maiores do que 208,9, e os valores entre 178,8 e 208,9 são possíveis suspeitos. Nenhum valor deve ser considerado suspeito.

Os intervalos de detecção de valores suspeitos foram adicionados ao *boxplot* da planilha **Boxplot**, como mostram as planilhas da Figura 4.5 referentes às amostras *A* e *B*. Verifique que:

- Nos extremos da distribuição, são representados os segmentos dos valores potencialmente suspeitos, linhas de cor vermelha.
- A amostra *A* não tem valores suspeitos na cauda inferior; entretanto, pode ter valores suspeitos na cauda superior da distribuição.
- A amostra *B* tem valores suspeitos na cauda inferior; entretanto, pode não ter valores suspeitos na cauda superior da distribuição.

Boxplot com Excel

O *boxplot* de uma amostra também pode ser construído utilizando os recursos gráficos do Excel. Na planilha **Boxplot com Excel**, incluída na pasta **Capítulo 4**, foram repetidos os dados e os resultados da planilha **Exemplo 4.9**, fazendo uma cópia dessa planilha. Depois, as posições dos resultados dos cinco números, *mínimo*, Q_1, *mediana*, Q_3 e *máximo* foram mudadas para a nova sequência dos cinco resultados, Q_1, *mínimo*, *mediana*, *máximo* e Q_3.

Construção de um Boxplot

Depois de ter mudado as posições dos cinco resultados na planilha **Boxplot com Excel** proceda assim:[13]

- Selecione o intervalo E5:F10 da planilha **Boxplot com Excel**.
- Clique no ícone assistente de gráfico e, na página **Tipos padrão** de gráficos, selecione o tipo de gráfico **Linha** e o subtipo de gráfico **Linhas com marcadores exibidos a cada valor de dado**.
- Depois, clique no botão **Avançar**. Na guia **Intervalo de dados** você deverá selecionar **Linhas** apesar de os dados estarem registrados em colunas, como mostra a Figura 4.6, à esquerda. Depois clique no botão **Concluir**.

Agora temos um gráfico como o mostrado na Figura 4.6, à esquerda. Para construir a forma do *boxplot* proceda desta forma:

- Clique duas vezes seguidas com o botão esquerdo do mouse em cima de um dos pontos do gráfico construído. Aparecerá a caixa de diálogo **Formatar sequência de dados**.
- Na caixa de diálogo **Formatar sequência de dados**, selecione a guia **Opções**. Nessa página, marque as caixas **Linhas de máximo/mínimo** e **Barras superiores/inferiores** como mostrado na Figura 4.6, à direita.

13 Adptado de Hunt N. – *Boxplots in Excel* em *http://www.mis.coventry.ac.uk/~nhunt/boxplot.htm*.

- Para terminar, ajuste a formatação do gráfico da forma que achar mais conveniente, mudando a posição da legenda, a cor do fundo do gráfico, a identificação dos cinco pontos etc.

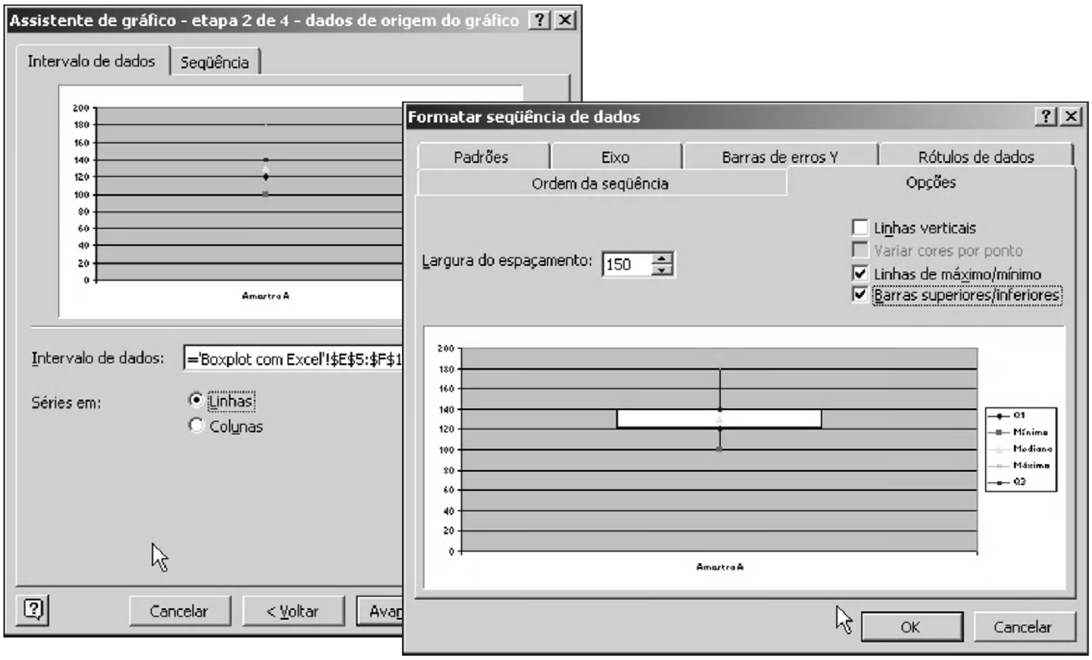


FIGURA 4.6 Construção de um *boxplot* com Excel.

Construção de dois ou mais Boxplot

O procedimento é parecido com o de um *boxplot* anterior e pode ser utilizado para mais de dois *boxplot*:

- Selecione o intervalo E5:G10 da planilha **Boxplot com Excel**.
- Clique no ícone assistente de gráfico e, na página **Tipos padrão** de gráficos, selecione o tipo de gráfico **Linha** e o subtipo de gráfico **Linhas com marcadores exibidos a cada valor de dado**.
- Depois clique no botão **Avançar**. Na guia **Intervalo de dados**, deverá selecionar **Linhas**, apesar de os dados estarem registrados em colunas, como mostrado na Figura 4.7, à esquerda. Depois clique no botão **Concluir**.

Agora temos um gráfico como o mostrado na Figura 4.7, à esquerda. Para construir a forma de dois *boxplot*, siga este procedimento:

- Clique duas vezes seguidas com o botão esquerdo do mouse na primeira linha do gráfico construído. Será exibida a caixa de diálogo **Formatar sequência de dados**.
- Na caixa de diálogo **Formatar sequência de dados**, selecione a guia **Padrões** e, no quadro **Linha**, marque **Nenhuma** e depois pressione **OK**. Verifique se, com essa instrução, a linha que ligava os dois pontos foi removida.
- Repita o procedimento anterior com as quatro linhas restantes.
- Na caixa de diálogo **Formatar sequência de dados**, selecione a guia **Opções**. Nessa página, marque as caixas **Linhas de máximo/mínimo** e **Barras superiores/inferiores** como mostra a Figura 4.7, à direita.
- Para terminar, ajuste a formatação do gráfico da forma que achar mais conveniente, mudando a posição da legenda, a cor do fundo do gráfico, a identificação dos cinco pontos etc.

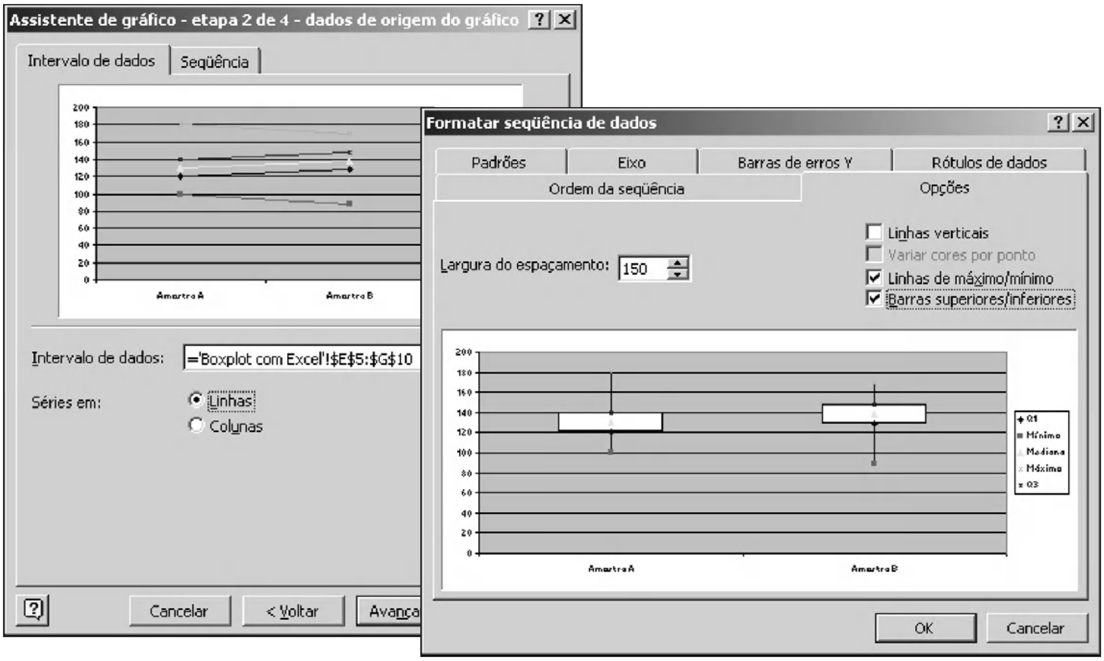


FIGURA 4.7 Construção de dois *boxplot* com Excel.

Problemas

Problema 1

Calcule a variância e o desvio padrão da amostra registrada na tabela seguinte:

10	15	14	23	21	18	11	12	14	15	23	12	15

R: S^2=19,09 e S=4,37

Problema 2

Calcule a variância e o desvio do Problema 1, considerando, como população.

R: σ^2=17,62 e σ=4,20

Problema 3

Repita o Problema 2, calculando a variância e o desvio padrão da população a partir da variância e do desvio padrão da amostra e utilizando as fórmulas.

Problema 4

A tabela a seguir registra uma amostra do número de gerentes operacionais que respondem diretamente a um diretor em empresas do ramo químico. Calcule a média e o desvio padrão do número de gerentes por empresa:

7	7	9	8	7	13	10	14	8	9	8	6
9	9	10	11	7	8	9	6	8	11	12	10

R: $\overline{X}$=9 e S=2,09

Problema 5

Calcule a variância e o desvio padrão da amostra registrada na tabela:

10	15	14	23	21	18	11	12	14	15	23	12	18	16	15

R: S^2=16,74 e S=4,09

Problema 6

A tabela seguinte registra as notas finais de um grupo de alunos da disciplina Estatística. Calcule a variância e o desvio padrão da amostra.

89,5	74,7	99,4	84,9	96,5	82,1	77,7	92,7	59,1	74,7	91,0	100	77,6	98,5	2,2	60,8
83,1	20,1	84,2	70,1	90,8	97,5	78,2	31,7	98,1	99,0	94,3	73,4	85,7	94,1	61,0	77,8

R: $\overline{X}$=78,14 e S=23,15

Problema 7

Continuando com Problema 6. Calcule a mediana da amostra e analise a inclinação da distribuição.

R: *Md* =83,65. A distribuição tem inclinação para a esquerda, pois $\overline{X} < Md$, como mostra o *coeficiente de inclinação* igual a 1,87.

Problema 8

Continuando com o Problema 6, determine a porcentagem das notas finais do grupo de alunos que estão incluídos em um, dois e três desvios padrão.

R: $\overline{X} \pm 1 \times S$=91%; $\overline{X} \pm 2 \times S$=91% e $\overline{X} \pm 3 \times S$=97%.

Problema 9

Repita o Problema 8, excluindo as observações 2,2; 20,1; e 31,7.

R: $\overline{X} \pm 1 \times S$=66%; $\overline{X} \pm 2 \times S$=97% e $\overline{X} \pm 3 \times S$=100%.

Problema 10

Calcule a variância e o desvio padrão dos retornos da tabela seguinte.

Aplicação	Retorno mensal %
Ouro	–1,74%
Curto prazo	0,52%
Dólar paralelo	0,87%
CDB para <$5.000	1,15%
Caderneta de poupança	1,16%
FRF 30 dias	1,30%
FRF 60 dias	1,49%
CDB para >$100.000	1,58%
Bolsa RJ	2,12%
Bolsa SP	2,99%

R: S^2=0,00015 e S=1,22%

Problema 11

Continuando com o Problema 10, determine a porcentagem dos retornos incluídos em um, dois e três desvios padrão.

R: $\overline{X} \pm 1 \times S$=80%; $\overline{X} \pm 2 \times S$=90% e $\overline{X} \pm 3 \times S$=100%.

Problema 12

Calcule o coeficiente de variação dos retornos do Problema 10.

R: CV=1,07

Problema 13

Os retornos anuais das ações *X* e *Y* durante os últimos cinco anos estão registrados na tabela seguinte. Qual dos dois retornos tem maior dispersão?

X	*Y*
12%	12%
15%	16%
12%	15%
11%	9%
14%	13%

R: A dispersão do retorno da ação *Y* é maior do que a dispersão da ação *X*.

Problema 14

Continuando com o Problema 13. Calcule os coeficientes de variação de *X* e *Y*. Qual é a ação com maior risco?

R: CV_X=0,13 e CV_Y=0,21

Problema 15

As taxas de juros cobradas nos empréstimos para compra de eletrodomésticos em oito das maiores lojas da cidade estão registradas na tabela seguinte. Calcule a média, a variância e o desvio padrão das taxas de juros.

6,00%	4,80%	5,30%	4,75%	4,10%	5,40%	3,90%	5,20%

R: $\overline{X}$=4,93% S^2=0,0000482 e S=0,69%

Problema 16

A tabela seguinte registra uma amostra do tempo que os *caixas* do banco gastam para realizar as transações dos clientes. Calcule a média, a variância e o desvio padrão da amostra.

2,5	8,0	4,5	7,5	2,0	11,0	4,0	5,0	8,0	6,5	3,5

R: $\overline{X}$=5,68 minutos, S^2=7,61 e S=2,76 minutos

Problema 17

Para conhecer o número de horas por semana que os principais executivos das maiores empresas do país trabalham, a empresa de consultoria realizou uma pesquisa com doze executivos escolhidos aleatoriamente dentre as 500 maiores empresas. Calcule a média, a variância e o desvio padrão da amostra registrada na tabela a seguir.

60	66	64	62	58	62	62	60	62	60	64	66

R: $\overline{X}$=62,17 hs/sem. S^2=6,15 e S=2,48 hs/sem.

Problema 18

Ao comparar os retornos de duas ações, a ação que apresentar maior *coeficiente de variação* terá maior risco. A tabela seguinte registra os retornos da *Ação A* e da *Ação B* durante cinco anos. Determine a ação com maior risco.

Ação A	*Ação B*
9,00%	12,00%
10,00%	10,50%
12,00%	9,50%
10,50%	11,00%
9,50%	12,50%

R: A *Ação A* teve maior *coeficiente de variação* e, portanto, maior risco.

Problema 19

Calcule a variância e o desvio padrão da amostra das notas finais da *Turma C* da disciplina Estatística registradas no Problema 6.

R: *Var*=535,90 e *S*=23,15

Problema 20

Determine os cinco números: *mínimo*, Q_1, *mediana*, Q_3 e *máximo* da amostra do Problema 19.

R: *Min*=2,20; Q_1=74,38; *Med*=83,65; Q_3=94,15; *Max*=100

Problema 21

Construa o *boxplot* do Problema 19.

Problema 22

Com os resultados do Problema 21, analise a distribuição de frequências dessa amostra.

Problema 23

Repita o Problema 22 utilizando o **Modelo Análise Numérica.**

Problema 24

Verifique a existência de dados suspeitos na amostra do Problema 19.

Problema 25

Construa o *boxplot* da amostra do Problema 10, analise a distribuição e verifique a existência de dados suspeitos.

Problema 26

A rede de restaurantes AQUIeAGORA, especializada em almoços pelo sistema *refeição por quilo*, tem 30 lojas distribuídas em diversos bairros de São Paulo, todas com o mesmo padrão e capacidade de atendimento. A tabela seguinte apresenta o número de refeições servidas pelas 30 lojas em um dia típico.

290	243	295	275	216	253
266	232	256	224	252	298
316	247	234	278	270	280
226	233	298	278	266	278
252	269	239	325	240	295

Construa o *boxplot*, analise a distribuição e verifique a existência de dados suspeitos.

Problema 27

Repita o Problema 26 utilizando as vendas das 50 primeiras empresas por vendas em 2002, cujos dados estão registrados na planilha **Problemas** deste capítulo.

Apêndice 1

Funções de medida de dispersão do Excel

O cálculo das medidas de dispersão utilizando o Excel pode ser realizado utilizando expressões matemáticas e procedimentos combinados com os recursos da planilha e funções estatísticas. Na planilha **Funções de Dispersão**, incluída na pasta **Capítulo 4**, está registrada a utilização de cada função utilizando a amostra do Exemplo 4.1, como se pode ver na Figura 4.8. Uma característica comum das funções a seguir são os 30 argumentos (*núm1; núm2; ... ; núm30*) utilizados para registrar os valores de intervalos. Na apresentação da primeira função DESV.MÉDIO, será mostrado como utilizar esses argumentos, procedimentos que se repetem com as demais funções com o mesmo tipo de argumentos. As sintaxes dessas funções estatísticas são apresentadas a seguir.

DESV.MÉDIO(*núm1; núm2; ... ; núm30*)

A função estatística DESV.MÉDIO[14] retorna o *desvio absoluto médio* dos valores numéricos *núm1; núm2; ... ; núm30*. Cada um desses *núm* pode ser um intervalo de células de uma planilha contendo valores numéricos ou assemelhados.[15] Por exemplo, a função DESV.MÉDIO aplicada aos valores do Exemplo 4.1 retorna o resultado 8,38. Para obter esse resultado, a função DESV.MÉDIO pode ser utilizada das seguintes maneiras:

- Registrando os valores da amostra em um intervalo de células da planilha.
  - Se os valores da variável estiverem registrados em um único intervalo, ou intervalos contíguos, apenas será necessário informar um único intervalo no argumento *num1*. Por exemplo, na célula E6 foi registrada a fórmula =DESV.MÉDIO(B4:B14), conforme apresenta a Figura 4.8.
  - Se os valores da variável estiverem registrados em intervalos não adjacentes, será necessário informar o endereço de cada intervalo no lugar de cada *núm* de *núm1; núm2; ... ; núm30*, até um máximo de 30; por exemplo, na célula E7 foi registrada a seguinte fórmula =DESV.MÉDIO(B4:B7;B8:B12;B13:B14).
- Registrando os valores da amostra como *matriz* na própria fórmula da função, evitando registrar os valores da amostra em um intervalo de células da planilha.
  - Na célula F6, os valores foram registrados em uma única matriz:
    =DESV.MÉDIO({31;38;19;27;24;42;32;18;43;15;39})
  - Na célula F7, os valores foram registrados em três matrizes:
    =DESV.MÉDIO({31;38;19};{27;24;42;32;18;43};{15;39}) correspondentes aos três primeiros argumentos da função DESV.MÉDIO *núm1; núm2; núm3*.

DESVQ(*núm1; núm2; ... ; núm30*)

A função estatística DESVQ[16] retorna a *soma dos quadrados dos desvios* dos valores numéricos *núm1; núm2; ... ; núm30* com relação à média. Cada um desses *núm* pode ser um intervalo de células de uma

14 Em inglês, a função DESV.MÉDIO é *AVEDEV*.
15 Assemelhados são os intervalos definidos por nomes, valores lógicos, representações em forma de texto de números, por exemplo, com a função de texto VALOR("10")=10.
16 Em inglês, a função DESVQ é *DEVSQ*.

planilha contendo valores numéricos ou assemelhados. A função DESVQ pode ser registrada de diversas formas equivalentes às descritas na função DESV.MÉDIO, mencionada anteriormente, conforme mostrado na Figura 4.8.

VARP(*núm1; núm2; ... ; núm30*)

A função estatística VARP[17] retorna a *variância da população* dos valores numéricos *núm1; núm2; ... ; núm30*. Cada um desses *núm* pode ser um intervalo de células de uma planilha contendo valores numéricos ou assemelhados. A função VARP pode ser registrada de diversas formas equivalentes às descritas na função DESV.MÉDIO citada anteriormente.

VAR(*núm1; núm2; ... ; núm30*)

A função estatística VAR[18] retorna a *variância da amostra* dos valores numéricos *núm1; núm2; ... ; núm30*. Cada um desses *núm* pode ser um intervalo de células de uma planilha que contém valores numéricos ou *assemelhados*. A função VAR pode ser registrada de diversas formas equivalentes às descritas na função DESV.MÉDIO definida anteriormente.

DESVPADP(*núm1; núm2; ... ; núm30*)

A função estatística DESVPADP[19] retorna o *desvio padrão da população* dos valores numéricos *núm1; núm2; ... ; núm30*. Cada um desses *núm* pode ser um intervalo de células de uma planilha que contém valores numéricos ou assemelhados. A função DESVPADP pode ser registrada de diversas formas equivalentes às descritas na função DESV.MÉDIO mencionada anteriormente.

DESVPAD(*núm1; núm2; ... ; núm30*)

A função estatística DESVPAD[20] retorna o *desvio padrão da amostra* dos valores numéricos *núm1; núm2; ... ; núm30*. Cada um desses *núm* pode ser um intervalo de células de uma planilha que contém valores numéricos ou assemelhados. A função DESVPADP pode ser registrada de diversas formas equivalentes às descritas na função DESV.MÉDIO detalhada anteriormente.

VARPA(*núm1; núm2; ... ; núm30)*

A função estatística VARPA[21] é equivalente à função anterior VARP. A diferença está relacionada com os valores registrados nos argumentos *núm1; núm2; ... ; núm30* que, nesta função, além de números, podem ser valores lógicos e de texto, como VERDADEIRO e FALSO.

VARA(*núm1; núm2; ... ; núm30)*

A função estatística VARPA[22] é equivalente à função anterior VAR. A diferença está relacionada com os valores registrados nos argumentos *núm1; núm2; ... ; núm30* que, nesta função, além de números, podem ser valores lógicos e de texto, como VERDADEIRO e FALSO.

17 Em inglês, a função VARP é *VARP*.
18 Em inglês, a função VAR é *VAR*.
19 Em inglês, a função DESVPADP é *STDEVP*.
20 Em inglês, a função DESVPAD é *STDEV*.
21 Em inglês, a função VARPA é *VARPA*.
22 Em inglês, a função VARA é *VARA*.

	A	B	C	D	E	F
1	**Funções de medição de dispersão**					
2						
3		**Amostra**			**Dados informados como**	
4		31		**Funções Estatísticas**	**Intervalo**	**Matriz**
5		38				
6		19		**DESV.MÉDIO**	**8,38**	**8,38**
7		27			**8,38**	**8,38**
8		24				
9		42		**DESVQ**	**997,64**	**997,64**
10		32			**997,64**	**997,64**
11		18				
12		43		**VARP**	**90,69**	**90,69**
13		15			**90,69**	**90,69**
14		39				
15				**VAR**	**99,76**	**99,76**
16					**99,76**	**99,76**
17						
18				**DESVPADP**	**9,52**	**9,52**
19					**9,52**	**9,52**
20						
21				**DESVPAD**	**9,99**	**9,99**
22					**9,99**	**9,99**
23						
24				**DISTORÇÃO**	(0,12)	(0,12)
25					(0,12)	(0,12)
26						
27				**CURT**	(1,47)	(1,47)
28					(1,47)	(1,47)
29						

FIGURA 4.8 Aplicando as funções de medidas de dispersão no Exemplo 4.1.

DESVPADPA(*núm1; núm2; ... ; núm30)*

A função estatística DESVPADPA[23] é equivalente à função anterior DESVPADP. A diferença está relacionada com os valores registrados nos argumentos *núm1; núm2; ... ; núm30* que, nesta função, além de números, podem ser valores lógicos e de texto, como VERDADEIRO e FALSO.

DESVPADA(*núm1; núm2; ... ; núm30)*

A função estatística DESVPADA[24] é equivalente à função anterior DESVPAD. A diferença está relacionada com os valores registrados nos argumentos *núm1; núm2; ... ; núm30* que, nesta função, além de números, podem ser valores lógicos e de texto, como VERDADEIRO e FALSO.

DISTORÇÃO(*núm1; núm2; ... ; núm30*)

A função estatística DISTORÇÃO[25] retorna o *coeficiente de inclinação* dos valores numéricos *núm1; núm2; ... ; núm30*. Cada um desses *núm* pode ser um intervalo de células de uma planilha que contém valores numéricos ou *assemelhados*. A fórmula utilizada pela função DISTORÇÃO para calcular o coeficiente de inclinação é:

$$\textit{Coeficiente de Inclinação} = \frac{n}{(n-1)\times(n-2)}\sum_{i=1}^{n}\left(\frac{X_i - \overline{X}}{S_X}\right)^3$$

23 Em inglês, a função DESVPADPA é *STDEVPA*.
24 Em inglês, a função DESVPADA é *STDEVA*.
25 Em inglês, a função DISTORÇÃO é *SKEW*.

O coeficiente de inclinação é o resultado da comparação da distribuição de frequências dos valores informados com a distribuição normal, apresentada no Capítulo 8, e seu resultado deve ser interpretado como segue. Se o coeficiente de inclinação for igual a zero, então a distribuição de frequências é simétrica, se for negativo, a distribuição de frequências terá inclinação para a esquerda ou negativa, e se for positivo, a distribuição de frequências terá inclinação para a direita ou positiva. A função DISTORÇÃO pode ser registrada de diversas formas equivalentes às descritas na função DESV.MÉDIO, definida anteriormente.

CURT(*núm1; núm2; ... ; núm30*)

A função estatística CURT[26] retorna o *coeficiente de curtose* dos valores numéricos *núm1; núm2; ... ; núm30*. Cada um desses *núm* pode ser um intervalo de células de uma planilha que contém valores numéricos ou *assemelhados*. A fórmula utilizada pela função CURT para calcular o coeficiente de curtose é a seguinte:

$$\textit{Coeficiente de Curtose} = \left\{ \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^{n} \left(\frac{X_i - \overline{X}}{S_X} \right)^4 \right\} - \frac{3(n-1)^2}{(n-2)(n-3)}$$

O coeficiente de curtose é o resultado da comparação da distribuição de frequências dos valores informados com a distribuição normal apresentada no Capítulo 8, e seu resultado deve ser interpretado como segue. Se o coeficiente de curtose for igual a zero, então a distribuição de frequências será a própria distribuição normal; se for negativo, a distribuição será achatada, plana; e se for positivo, a distribuição de frequências será concentrada ao redor da média, distribuição com pico. A função CURT pode ser registrada de diversas formas equivalentes às descritas na função DESV.MÉDIO mencionada anteriormente.

Apêndice 2

Outra forma de calcular a variância

O cálculo da variância da variável *X* pode ser realizado utilizando apenas os valores da variável, sem necessidade de calcular a média e os desvios da variável. Se na fórmula da soma dos quadrados dos desvios desenvolvemos o quadrado do binômio indicado, obtemos a seguinte igualdade:

$$\sum_{i=1}^{N} (X_i - \mu_X)^2 = \sum_{i=1}^{N} (X_i^2 - 2X\mu_X + \mu_X^2)$$

Continuando com o desenvolvimento algébrico, obtemos:

$$\sum_{i=1}^{N} (X_i - \mu_X)^2 = \sum_{i=1}^{N} X_i^2 - 2\mu_X \sum_{i=1}^{N} X_i + \sum_{i=1}^{N} \mu_X^2$$

26 Em inglês, a função CURT é *KURT*.

No segundo membro dessa expressão reconhecemos que $\sum_{i=1}^{N} X_i = N\mu_X$ e $\sum_{i=1}^{N} \mu_X^2 = N\mu_X^2$. Dessa maneira, o segundo membro pode ser reescrito da seguinte forma: $\sum_{i=1}^{N} X_i^2 - 2\mu_X N\mu_X + \mu_X^2 = \sum_{i=1}^{N} X_i^2 - N\mu_X^2$.

Voltando à primeira fórmula, formamos a igualdade que nos interessa:

$$\sum_{i=1}^{N} (X_i - \mu_X)^2 = \sum_{i=1}^{N} X_i^2 - N\mu_X^2$$

Ainda, pela definição de média da população $\mu_X^2 = \frac{\left(\sum_{i=1}^{N} X_i\right)^2}{N^2}$. Substituindo essa relação na expressão da soma dos quadrados dos desvios, teremos:

$$\sum_{i=1}^{N} (X_i - \mu_X)^2 = \sum_{i=1}^{N} X_i^2 - N = \frac{\left(\sum_{i=1}^{N} X_i\right)^2}{N^2}$$

Agora, o cálculo da soma dos quadrados dos desvios depende somente dos dados da amostra e dos quadrados desses dados. Dessa maneira, as expressões das variâncias são:

- Da população: $\sigma_X^2 = \frac{1}{N}\left\{\sum_{i=1}^{N} X_i^2 - \frac{\left(\sum_{i=1}^{N} X_i\right)^2}{N}\right\}$
- Da amostra: $S_X^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}\right\}$

Para calcular a variância, será necessário gerar a série dos quadrados dos valores da variável, não sendo necessário calcular a média nem os desvios. Na realidade, esse procedimento de cálculo perde sua força quando comparado com a utilização das funções estatísticas do Excel, como mostra a planilha **Apêndice 2**, incluída na pasta **Capítulo 4**. Essa expressão da variância será utilizada no Apêndice 1 do Capítulo 9.

Apêndice 3

Funções para banco de dados do Excel

As funções estatísticas apresentadas até este momento foram utilizadas para obter alguma medida estatística de uma amostra ou variável, atendendo a algumas especificações dessas funções:

- Os dados foram registrados em um intervalo de células da planilha e a fórmula com a função em outra célula fora daquele intervalo.
- Os dados da amostra foram registrados como matriz na própria fórmula da função em uma única célula da planilha.

Há situações em que os dados ou variáveis para análise fazem parte de uma tabela contendo outras variáveis. Por exemplo, os resultados mensais significativos de uma empresa durante um ano estão registrados na planilha **Funções Banco de Dados**, incluída na pasta **Capítulo 4**, conforme apresenta a Figura 4.9. Os resultados estão registrados em uma tabela com as colunas identificadas com os nomes *Mês*, *Vendas*, *Custos*, *Lucro Bruto* e *Lucro Líquido*. A tabela com os resultados da empresa é denominada *banco de dados*, e cada uma de suas colunas é denominada *campo*; em termos técnicos, cada linha da tabela é uma unidade elementar de informação que contém quatro variáveis. Para essas situações, o Excel dispõe de funções denominadas genericamente BDfunções e equivalentes a algumas das funções apresentadas neste capítulo.

	A	B	C	D	E	F
1	**Funções estatísticas para banco de dados**					
2						
3		**Mês**	**Vendas**	**Custos**	**Lucro Bruto**	**Lucro Líquido**
4		jan/2004	$ 6.423	$ 3.270	$ 3.153	$ 2.193
5		fev/2004	$ 5.467	$ 3.649	$ 1.818	$ 1.198
6		mar/2004	$ 5.191	$ 3.381	$ 1.810	$ 1.156
7		abr/2004	$ 6.315	$ 3.513	$ 2.802	$ 1.844
8		mai/2004	$ 6.080	$ 3.316	$ 2.764	$ 1.754
9		jun/2004	$ 6.195	$ 3.768	$ 2.427	$ 1.664
10		jul/2004	$ 6.131	$ 3.564	$ 2.567	$ 1.746
11		ago/2004	$ 6.386	$ 3.258	$ 3.128	$ 2.072
12		set/2004	$ 6.014	$ 3.122	$ 2.892	$ 1.931
13		out/2004	$ 5.993	$ 3.706	$ 2.287	$ 1.522
14		nov/2004	$ 5.237	$ 3.533	$ 1.704	$ 1.129
15		dez/2004	$ 5.374	$ 3.115	$ 2.259	$ 1.457
16						

FIGURA 4.9 Resultados mensais da empresa.

EXEMPLO 4.11

Calcule a média, o desvio padrão e o valor máximo das *Vendas* da empresa durante o primeiro mês dos quatro trimestres do ano 2004, e cujos resultados estão registrados na tabela da Figura 4.9.

Solução. Os resultados foram obtidos de diversas formas, a partir da célula H1 da planilha **Funções Banco de Dados**, incluída na pasta **Capítulo 4**, como mostra a figura seguinte. A média das vendas da empresa nos primeiros meses dos quatro trimestres do ano 2004 é igual a $6.215,50, resultado obtido:

- Calculando com a função estatística MÉDIA, registrando na célula K4 a fórmula =MÉDIA(C4;C7;C10;C13).
- Calculando com a função estatística SUBTOTAL, registrando na célula K5 a fórmula =SUBTOTAL (1;C4;C7;C10;C13). Com a função SUBTOTAL, é possível obter 11 resultados diferentes informando um número de 1 a 11 no primeiro argumento da função, como será apresentado mais adiante neste apêndice.
- Calculando com a função estatística para banco de dados BDMÉDIA, registrando na célula K6 a fórmula =BDMÉDIA(B3:F15;C3;I3:I7). A função BDMÉDIA é uma das doze funções para listas ou banco de dados disponíveis no Excel e denominadas genericamente BDfunções, pois todas elas utilizam a mesma sintaxe, BDFunção(*banco_dados; campo; critérios*).
  - No argumento *banco_dados*, deve ser informado o intervalo do banco de dados incluindo a primeira linha com os títulos, neste exemplo B3:F15.
  - No argumento *campo*, deve ser informado o nome da coluna do banco de dados onde será aplicada a função. Neste exemplo, pode ser informado o texto "Vendas", entre aspas duplas, ou o endereço da célula C3.
  - No argumento *critérios*, deve ser registrada a especificação da escolha dos dados. Neste exemplo, no intervalo I3:I7 foi construída a tabela de meses, ou linhas, que identificam os valores correspondentes da coluna *Vendas* o argumento *campus* da função. Como alternativa, pode-se utilizar a fórmula =BDMÉDIA(B3:F15; ;"Vendas";I3:I7) para obter o mesmo resultado Ou, como um número que represente a posição da coluna dentro da lista, começando com 1 para a primeira coluna, 2 para a segunda coluna e assim sucessivamente, até esgotar as colunas do banco de dados.

	H	I	J	K	L	M
1	Exemplo 4.11					
2						
3		Mês		Cálculo da média		
4		jan/2004		$ 6.215,50	=MÉDIA(C4;C7;C10;C13)	
5		abr/2004		$ 6.215,50	=SUBTOTAL(1;C4;C7;C10;C13)	
6		jul/2004		$ 6.215,50	=BDMÉDIA(B3:F15;C3;I3:I7)	
7		out/2004				
8				Cálculo do desvio padrão, como população		
9				$ 165,53	=DESVPADP(C4;C7;C10;C13)	
10				$ 165,53	=SUBTOTAL(8;C4;C7;C10;C13)	
11				$ 165,53	=BDDESVPA(B3:F15;C3;I3:I7)	
12						
13				Determinação do valor máximo		
14				$ 6.423,00	=MÁXIMO(C4;C7;C10;C13)	
15				$ 6.423,00	=SUBTOTAL(4;C4;C7;C10;C13)	
16				$ 6.423,00	=BDMÁX(B3:F15;C3;I3:I7)	
17						
18				$ 6.423,00	=BDMÁX(B3:F15;"Vendas";I3:I7)	
19				$ 6.423,00	=BDMÁX(B3:F15;2;I3:I7)	
20						

A partir das linhas 8 e 13 da planilha **Funções Banco de Dados**, foram calculados, respectivamente, o desvio padrão e o valor máximo das *Vendas* da empresa durante o primeiro mês dos quatro trimestres do ano 2004, utilizando as três funções apresentadas e adequadas para esses cálculos.

Incluindo outros critérios

Com as funções para banco de dados operamos a distância sem necessidade de definir intervalos dentro do banco de dados. A tabela de *critérios* pode incluir condições lógicas nos *campos* do *banco de dados*. Sem esgotar este assunto, a seguir mostraremos outra forma de incluir critérios.

EXEMPLO 4.12

Calcule a média das vendas da empresa durante o primeiro mês dos quatro trimestres do ano 2004, considerando somente os meses com lucro líquido maior ou igual a $1.600.

Solução. Para calcular a média das vendas dos primeiros meses dos quatro trimestres do ano 2004, considerando apenas as vendas dos meses com *Lucro Líquido* igual ou maior do que $1.600, no intervalo O3:P7, foi construída a tabela com os campos *Mês* e *Lucro Líquido*, sendo que, neste último cálculo, foram registradas as restrições de seleção de cada mês, a fórmula >=1600. A média das vendas da empresa durante o primeiro mês dos quatro trimestres do ano 2004, considerando somente os meses com lucro líquido maior ou igual a $1.600 é igual a $6.289,67, resultado obtido com =BDMÉDIA(B3:F15;C3;O3:P7), fórmula registrada na célula R4.

	N	O	P	Q	R	S	T
1	Exemplo 4.12						
2							
3		Mês	Lucro Líquido		Cálculo da média, com restrições		
4		jan/2004	>=1600		$ 6.289,67	=BDMÉDIA(B3:F15;C3;O3:P7)	
5		abr/2004	>=1600				
6		jul/2004	>=1600				
7		out/2004	>=1600				
8							

Resumo das funções de banco de dados do Excel

O Excel dispõe de doze funções orientadas para banco de dados, denominadas genericamente **BDfunções**, pois cada uma dessas funções tem os mesmos três argumentos: *banco de dados*, *campo* e *critérios*. Sua sintaxe geral é:

BDfunção(*banco_dados*; *campo*; *critérios*)

- O argumento *banco_dados* é o intervalo de células que delimita a tabela com as informações, que pode ser uma lista ou um banco de dados. Um banco de dados é uma lista de dados na qual cada linha é um registro formado por um ou mais campos identificados por um nome na primeira linha de cada coluna. O argumento *banco_dados* pode ser informado como um intervalo de células ou como um nome representando o intervalo.
- O argumento *campo* define o nome da coluna do banco de dados que será utilizada para realizar um cálculo ou uma seleção, podendo ser informado:
  - Como texto, por exemplo, "Vendas" ou "Lucro Líquido".
  - Como endereço da célula onde está registrado nome do campo.
  - Como um número que represente a posição da coluna dentro da lista, começando com 1 para a primeira coluna, 2 para a segunda coluna e assim sucessivamente, até esgotar as colunas do banco de dados.
- O argumento *critérios* é o intervalo de células que especifica a forma de seleção. Pode ser informado qualquer intervalo, sempre que ele incluir pelo menos um título de coluna e ao menos uma célula abaixo desse título que especifique alguma condição para seleção nessa coluna.

A seguir, são apresentadas as sintaxes das doze funções para bancos de dados disponíveis no Excel. As primeiras onze funções foram registradas com o mesmo argumento (B3:F15;C3;I3:I7) no intervalo K23:L34 da planilha **Funções Banco de Dados**, incluída na pasta **Capítulo 4**, cujos resultados são mostrados na Figura 4.10.

BDMÉDIA(*banco_dados*; *campo*; *critérios*)

A função BDMÉDIA[27] retorna a média dos valores da coluna *campo* do *banco_dados* que coincide com os *critérios* especificados.

BDCONTAR(*banco_dados*; *campo*; *critérios*)

A função BDCONTAR[28] retorna a quantidade de células contendo números da coluna *campo* do *banco_dados* que coincide com os *critérios* especificados.

BDCONTARA(*banco_dados*; *campo*; *critérios*)

A função BDCONTARA[29] retorna a quantidade de células não vazias da coluna *campo* do *banco_dados* que coincide com os *critérios* especificados

BDMÁX(*banco_dados*; *campo*; *critérios*)

A função BDMÁX[30] retorna o valor máximo da coluna *campo* do *banco_dados* que coincide com os *critérios* especificados.

BDMÍN(*banco_dados*; *campo*; *critérios*)

A função BDMÍN[31] retorna o valor mínimo da coluna *campo* do *banco_dados* que coincide com os *critérios* especificados.

BDMULTIPL(*banco_dados*; *campo*; *critérios*)

A função BDMULTIPL[32] retorna o resultado da multiplicação dos valores da coluna *campo* do *banco_dados* que coincide com os *critérios* especificados.

BDEST(*banco_dados*; *campo*; *critérios*)

A função BDEST[33] retorna o desvio padrão da amostra dos valores da coluna *campo* do *banco_dados* que coincide com os *critérios* especificados.

BDDESVPA(*banco_dados*; *campo*; *critérios*)

A função BDDESVPA[34] retorna o desvio padrão da população dos valores da coluna *campo* do *banco_dados* que coincide com os *critérios* especificados.

BDSOMA(*banco_dados*; *campo*; *critérios*)

A função BDSOMA[35] retorna a soma dos valores da coluna *campo* do *banco_dados* que coincide com os *critérios* especificados.

27 Em inglês, a função BDMÉDIA é *DAVERAGE*.
28 Em inglês, a função BDCONTAR é *DCOUNT*.
29 Em inglês, a função BDCONTARA é *DCOUNTA*.
30 Em inglês, a função BDMÁX é *DMAX*.
31 Em inglês, a função BDMÍN é *DMIN*.
32 Em inglês, a função BDMULTIPL é *DPRODUCT*.
33 Em inglês, a função BDEST é *DSTDEV*.
34 Em inglês, a função BDDESVP é *DSTDEVP*.
35 Em inglês, a função BDSOMA é *DSUM*.

	H	I	J	K	L	M	N	O
21	BDfunções							
22								
23		Mês		Função	Resultado			
24		jan/2004		BDMÉDIA	$ 6.215,50	=BDMÉDIA(B3:F15;C3;I23:I27)		
25		abr/2004		BDCONTAR	4	=BDCONTAR(B3:F15;C3;I23:I27)		
26		jul/2004		BDCONTARA	4	=BDCONTARA(B3:F15;C3;I23:I27)		
27		out/2004		BDMÁX	$ 6.423	=BDMÁX(B3:F15;C3;I23:I27)		
28				BDMÍN	$ 5.993	=BDMÍN(B3:F15;C3;I23:I27)		
29				BDMULTIPL	1,49035E+15	=BDMULTIPL(B3:F15;C3;I23:I27)		
30				BDEST	$ 191,14	=BDEST(B3:F15;C3;I23:I27)		
31				BDDESVPA	$ 165,53	=BDDESVPA(B3:F15;C3;I23:I27)		
32				BDSOMA	$ 24.862	=BDSOMA(B3:F15;C3;I23:I27)		
33				BDVAREST	36.534,33	=BDVAREST(B3:F15;C3;I23:I27)		
34				BDVARP	27.400,75	=BDVARP(B3:F15;C3;I23:I27)		
35								
36				BDEXTRAIR	#NÚM!	=BDEXTRAIR(B3:F15;C3;I23:I27)		
37				BDEXTRAIR	$ 6.423	=BDEXTRAIR(B3:F15;C3;I23:I24)		
38								

FIGURA 4.10 Aplicação das BDfunções.

BDVAREST(*banco_dados*; *campo*; *critérios*)

A função BDVAREST[36] retorna a variância da amostra dos valores da coluna *campo* do *banco_dados* que coincide com os *critérios* especificados.

BDVARP(*banco_dados*; *campo*; *critérios*)

A função BDVARP[37] retorna a variância da população dos valores da coluna *campo* do *banco_dados* que coincide com os *critérios* especificados.

BDEXTRAIR(*banco_dados*; *campo*; *critérios*)

A função BDEXTRAIR[38] extrai do *banco_dados* um único registro da coluna *campo* que coincide com os *critérios* especificados. A seguir, apresentamos como se deve utilizar essa função:

- A fórmula =BDEXTRAIR(B3:F15;C3;I23:I27) registrada na célula L36 retorna como resultado o valor de erro #NUM!, conforme mostrado na Figura 4.10. Isso ocorre porque a função BDEXTRAIR não consegue identificar um valor único no intervalo I23:I27, no qual há quatro datas possíveis.
- A fórmula =BDEXTRAIR(B3:F15;C3;I23:I24) registrada na célula L37 retorna o resultado $6.423, pois no intervalo I23:I24 há apenas uma única data Jan/2004.

Outras funções do Excel

O Excel dispõe também das funções matemáticas SUBTOTAL, CONT.SE e SOMASE que realizam operações equivalentes às apresentadas para banco de dados.

SUBTOTAL(*número_função; ref1; ref2; ...; ref29*)

A função SUBTOTAL[39] retorna o resultado das primeiras onze funções do grupo de BDfunções. O argumento *número_função* é um número de 1 a 11 que identifica a função que deverá ser utilizada no cálculo de subtotais do banco de dados, de uma lista ou grupo de valores, como mostra a Figura 4.11. Os argumentos *ref1; ref2; ...; ref29* são intervalos de células de uma planilha, ou referências, sobre os quais será calculado o subtotal.

36 Em inglês, a função BDVAREST é *DVAR*.
37 Em inglês, a função BDVARP é *DVARP*.
38 Em inglês, BDEXTRAIR é *DGET*.
39 Em inglês, SUBTOTAL é *SUBTOTAL*.

número_função	Função equivalente
1	MÉDIA
2	CONT.NÚM
3	CONT.VALORES
4	MÁXIMO
5	MÍNIMO
6	MULT
7	DESVPAD
8	DESVPADP
9	SOMA
10	VAR
11	VARP

FIGURA 4.11 Significado do argumento *número_ função.*

A Figura 4.12 mostra os onze resultados possíveis da função SUBTOTAL, registrados a partir da célula J39 da planilha **Funções Banco de Dados**, incluída na pasta **Capítulo 4**. Por exemplo, para calcular a média das vendas da empresa do Exemplo 4.11 referentes aos primeiros meses dos quatro trimestres do ano 2004, na célula L42 foi registrada a fórmula =SUBTOTAL(1;C4;C7;C10;C13), cujo resultado é $6.215,50.

O leitor atento deve ter percebido que a função SUBTOTAL pode ser utilizada como substituta de algumas das funções básicas apresentadas nos Capítulos 3 e 4 do livro. Como ajuda, a partir da célula H10 da planilha **Funções de Dispersão**, incluída na pasta **Capítulo 4**, foram registradas fórmulas utilizando a função SUBTOTAL ao lado da função equivalente original. Uma vantagem da utilização da função SUBTOTAL é que com um único nome de função poderíamos agrupar onze funções, com a desvantagem de ter de lembrar a tabela de equivalência da Figura 4.11, que também não é muito amigável.

	J	K	L	M
39	Função SUBTOTAL			
40				Função
41		Número	Resultado	Equivalente
42		1	$ 6.215,50	BDMÉDIA
43		2	4	BDCONTAR
44		3	4	BDCONTARA
45		4	$ 6.423	BDMÁX
46		5	$ 5.993	BDMÍN
47		6	1,49035E+15	BDMULTIPL
48		7	$ 191,14	BDEST
49		8	$ 165,53	BDDESVPA
50		9	$ 24.862,00	BDSOMA
51		10	36.534,33	BDVAREST
52		11	27.400,75	BDVARP
53				

FIGURA 4.12
Resultados com a função SUBTOTAL.

CONT.SE(*intervalo; critérios*)

A função CONT.SE[40] retorna o número de células não vazias da série de dados definida no argumento *intervalo* e que atendem a *critérios* definidos em forma de texto. Por exemplo, gostaríamos de conhecer, na tabela de resultados da Figura 4.9, em quantos meses do ano 2004 o lucro líquido da empresa foi igual ou maior do que $1.500. O resultado foi obtido com a função CONT.SE, registrando a fórmula

40 Em inglês, CONT.SE é *COUNTIF*.

=CONT.SE(F4:F15;">=1500") na célula K57 da planilha **Funções Banco de Dados**. Portanto, em oito meses do ano 2004, a empresa registrou lucro líquido igual ou maior do que $1.500.

SOMASE(*intervalo; critérios; intervalo_soma*)

A função SOMASE[41] retorna a soma de valores das células que atendem a um determinado critério.

- No argumento *intervalo* é registrado o intervalo de células utilizado para aplicar o critério de seleção.
- No argumento *critérios* é registrado um número, expressão ou texto, que define como as células serão selecionadas.
- No argumento *intervalo_soma* é registrado o intervalo das células que poderão ser somadas, sendo somadas somente as células correspondentes ao argumento *intervalo* que atendam ao argumento *critérios*. Se *intervalo_soma* for omitido, serão somadas as células do argumento *intervalo*.

Por exemplo, gostaríamos de conhecer, da empresa cujos resultados estão registrados na tabela de resultados da Figura 4.9, o total das vendas com lucro líquido igual ou maior do que $2.000 durante o ano 2004. O resultado foi obtido com a fórmula =SOMASE(F4:F15;">=2000";C4:C15) registrada na célula K62 da planilha **Funções Banco de Dados**. Portanto, o total das vendas com lucro líquido igual ou maior do que $2.000 durante o ano 2004 foi $12.809.

41 Em inglês, SOMASE é *SUMIF*.