Capítulo 3

MEDIDAS DE TENDÊNCIA CENTRAL

Para tentar conhecer uma ou mais características de uma população, extraímos uma amostra dessa população, conforme descrito no Capítulo 1. Em geral, quando o tamanho da amostra é grande, maior do que quinze dados, a simples inspeção das observações não será suficiente para obter as características relevantes desses valores. Para facilitar a análise e a interpretação, esses dados devem ser organizados ou resumidos, por exemplo, em tabelas de frequências e histogramas, como foi apresentado no Capítulo 2. As *medidas de ordenamento* e as *medidas de posição* são os métodos numéricos para resumir e analisar os valores de uma série de dados numéricos, seja uma amostra ou a própria população, denominados como *medidas de tendência central*. No Capítulo 4, serão apresentadas as *medidas de dispersão*.

Ordenamento de dados

Em algumas situações, o objetivo é conhecer a *posição* de um determinado valor numérico em relação aos demais valores da amostra; por exemplo, qual a posição de um determinado candidato a *trainee* comparando seu *QI* com os *QIs* dos outros candidatos que concorrem? O *QI* desse candidato é baixo ou alto? Quantos candidatos têm *QI* maior do que o candidato sob análise? Ou, quão maior é o *QI* do candidato? Outro exemplo, o retorno de 15% ao ano é baixo ou alto quando comparado com as rentabilidades das aplicações do mercado financeiro durante o mesmo período? Quantos retornos do mercado financeiro são maiores do que 15%?

Para responder a perguntas desse tipo, primeiro, os valores da série de dados devem estar ordenados em ordem crescente ou decrescente. Depois, deve-se estabelecer um critério que permita definir a posição de um determinado valor da série dentro da própria série de valores numéricos.

EXEMPLO 3.1

Ordene de forma crescente os valores da amostra registrada na tabela a seguir:

31	38	19	27	24	42	32	18	43	15	39

Solução. Depois de ordenar de forma crescente os onze valores numéricos da amostra, a seguir são associados os números 1, 2, ..., 11 aos valores ordenados como mostra esta tabela:

Amostra	15	18	19	24	27	31	32	38	39	42	43
Ordem	1	2	3	4	5	6	7	8	9	10	11

Agora, o valor 15 tem a posição 1, o 19 a posição 3 e o 43 a posição final 11.

De forma geral, o Exemplo 3.1 mostra que os *n* valores numéricos de uma amostra ordenada de forma crescente foram associados à série dos números naturais 1,2, 3, ... até *n*. Foi estabelecida uma relação de ordem entre os valores numéricos da amostra.

EXEMPLO 3.2

Determine a *ordem* de cada valor da amostra seguinte:

27	32	64	65	58	62	59	54	29	30	26	48	47
46	43	38	29	32	35	37	31	43	45	42	37	36

Solução. Depois de ordenar os valores da amostra de forma crescente, foi associada a série de números 1, 2, ..., 26 aos valores como mostra a tabela seguinte.

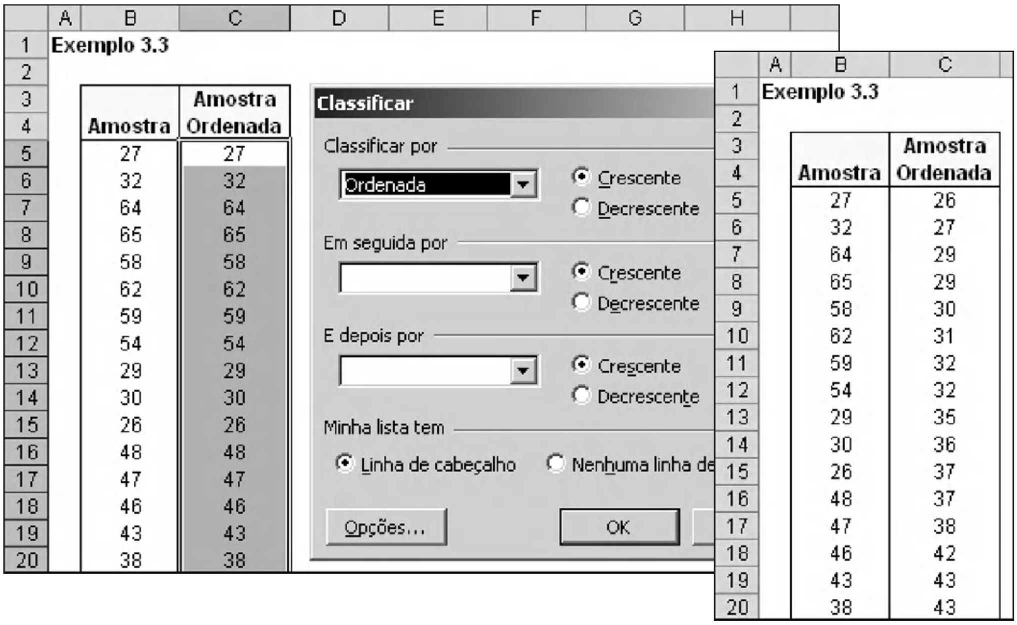
Amostra	26	27	29	29	30	31	32	32	35	36	37	37	38
Ordem	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*
Amostra	42	43	43	45	46	47	48	54	58	59	62	64	65
Ordem	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*	*25*	*26*

O procedimento de ordenamento em ordem crescente utilizado no Exemplo 3.2 foi o mesmo que o do Exemplo 3.1. No primeiro exemplo, o trabalho manual foi facilitado pelo pequeno tamanho da amostra. No último exemplo, o ordenamento manual é menos eficiente, pois é mais trabalhoso e está sujeito a erro de seleção dos valores da amostra. O comando de classificação do Excel ajudará a ordenar séries de valores em ordem crescente ou decrescente.

EXEMPLO 3.3

Ordene de forma crescente os dados do Exemplo 3.2 utilizando o Excel.

Solução. Primeiro, os dados da amostra do Exemplo 3.2 foram registrados na coluna B da planilha **Exemplo 3.3**, incluída na pasta **Capítulo 3**. A seguir, o intervalo B4:B30 foi copiado no intervalo C4:C30, adicionando o título **Amostra ordenada** como se pode ver na figura a seguir. O ordenamento dos valores da amostra pode ser realizado na própria coluna B; entretanto, a amostra foi copiada na coluna C para manter a amostra inicial e destacar o procedimento de ordenamento do Excel.



Para ordenar a amostra da coluna C procedemos assim:

- Selecione o intervalo C4:C30, incluindo o título **Ordenada** da célula C4.
- Depois de escolher **Classificar** no menu **Dados**, o Excel apresentará a caixa de diálogo que detecta outros valores ao redor do intervalo selecionado, mostrando duas opções **Expandir a seleção** e **Continuar com a seleção atual**; selecione esta última opção e depois pressione o botão **Classificar...**.
  - Em vez de utilizar o menu **Dados – Classificar**, é possível pressionar o ícone para classificar em ordem crescente, e o ícone para classificar em ordem decrescente.
- Em sequência, o Excel exibirá a caixa de diálogo **Classificar** com as seguintes escolhas: no grupo **Minha lista tem** a opção **Linha de cabeçalho**, na caixa **Classificar por** foi selecionado **Ordenada**, a opção **Crescente** e o intervalo C5:C30 estará selecionado, como mostra a figura à esquerda. Verifique que a célula C4 foi retirada da seleção do intervalo, pois informamos que o intervalo C4:C30 contém uma linha de cabeçalho. Essas escolhas estão de acordo com o intervalo da amostra informado.
  - Como teste, se no grupo **Minha lista tem** for selecionada a opção **Nenhuma linha de cabeçalho**, mantendo a opção **Crescente**, na caixa **Classificar por** aparecerá **Coluna C** e o intervalo C4:C30 estará selecionado. Nesse caso, a célula C4 foi incluída na seleção do intervalo, pois informamos que o intervalo C4:C30 não contém uma linha de cabeçalho.

Depois de pressionar o botão **OK**, os valores da amostra são ordenados de forma crescente no mesmo intervalo C5:C30 da planilha, como mostra a figura à direita. Para obter mais informações sobre o comando classificar, na ajuda do Excel, procure *Classificar uma lista*, onde encontrará suporte para realizar classificações em mais de uma coluna, classificando valores numéricos ou nomes e assemelhados na ordem crescente (A até Z ou 0 até 9) ou ordem decrescente (Z até A) ou (9 até 0).

Percentil

Os Exemplos 3.1 e 3.2 mostram o mesmo procedimento de ordenamento para duas listas de valores numéricos com quantidade de valores diferentes, sendo que há amostras com quantidades maiores de dados. É conveniente dispor de um procedimento que, mantendo o ordenamento crescente dos dados da amostra e a associação com os números naturais, tenha uma mesma medida e permita realizar comparações. A Figura 3.1 mostra uma relação entre a série de números naturais 0, 1, 2, ... n no eixo de abscissas com uma escala de 0% a 100% no eixo de ordenadas, sendo que 0% corresponde ao primeiro dado da amostra ordenada de forma crescente, e 100% ao último dado da amostra ordenada.

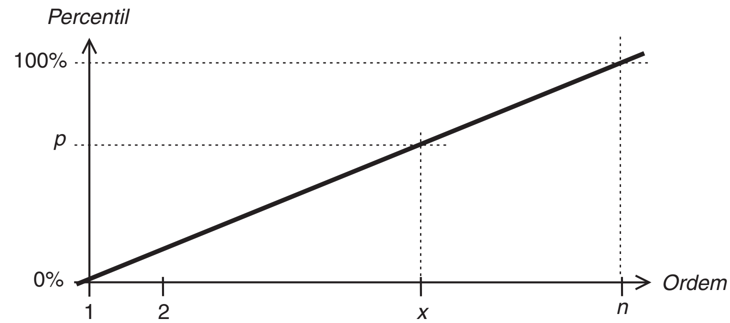


FIGURA 3.1 Ordenamento dos *n* valores de uma amostra.

Os valores da escala de ordenadas são denominados *percentil*, sendo que o menor valor do percentil é 0% e o maior valor 100%; dessa maneira, qualquer dado da amostra estará sempre entre o percentil 0% e 100%, como se pode ver na Figura 3.1, na qual o valor com *ordem x* corresponde ao percentil *p*. A relação entre as ordens dos *n* dados da amostra ou variável e todos os valores de percentil entre 0% a 100% é regida pela seguinte relação geométrica:

$$\frac{n-1}{100\%-0\%}=\frac{x-1}{p-0\%}$$

Nessa relação, *n* é a quantidade de dados da amostra, *x* é a ordem de um determinado dado da amostra ordenada de forma crescente, e *p* é o percentil correspondente em porcentagem. Dessa relação, obtemos as fórmulas de *p* e *x*.

- O percentil *p* em porcentagem do dado da amostra ou variável com ordem *x* é obtido com a fórmula: $p=\frac{x-1}{n-1}\times 100\%$. Qual é o significado do resultado *p*? O dado de ordem *x* é maior do que os primeiros *p* dados da amostra e, ao mesmo tempo, menor do que os restantes (1–*p*) dados da amostra.
- Da mesma maneira, conhecido o percentil *p* de um dado da amostra, sua ordem *x* é calculada com a fórmula: $x=(n-1)\times\frac{p}{100}+1$.

Resumindo, agora dispomos de uma relação entre uma escala de 0% a 100% (eixo de ordenadas) e a série de números naturais 0, 1, 2, ... *n* que representam uma série de dados quantitativos ou amostra ordenada de forma crescente (eixo de abscissas), sendo que 0% (percentil 0%) corresponde ao primeiro dado da amostra, e 100% (percentil 100%) corresponde ao último dado da amostra.

EXEMPLO 3.4

Calcule o percentil dos dados da amostra do Exemplo 3.1.

Solução. A partir da ordem de cada dado da amostra do Exemplo 3.1 foi calculado o percentil correspondente. Por exemplo, o dado 18 tem ordem *x*=2 e percentil *p*=10%, resultado obtido com a fórmula:

$$p=\frac{x-1}{n-1}\times 100\%$$

$$p=\frac{2-1}{11-1}\times 100\%=10\%$$

Repetindo esse procedimento de cálculo, foi construída a tabela a seguir:

Amostra	15	18	19	24	27	31	32	38	39	42	43
Percentil	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%

O percentil do dado 32 do Exemplo 3.4 é 60%. Qual o significado do resultado p=60%? O percentil 60% significa que o dado ordenado 32 é maior do que os primeiros 60% dos dados ordenados de forma crescente da amostra e, ao mesmo tempo, menor do que os demais 40% dos dados da amostra. Sem dúvida que a quantidade exata de dados da amostra do Exemplo 3.1 facilitou o cálculo do percentil de cada dado, pois é um múltiplo de 10%.

EXEMPLO 3.5

Determine a ordem do percentil 10%, 50% e 80% da amostra do Exemplo 3.1.

Solução. Para p=50%, obtemos a ordem x=6 como resultado da fórmula:

$$x = (n-1) \times \frac{p}{100} + 1$$

$$x = (11-1) \times \frac{50}{100} + 1 = 6$$

Portanto, consultando a tabela de dados ordenados do Exemplo 3.1, a posição 6 está ocupada pelo valor 31. Continuando com o exemplo:

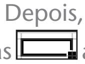
- Para p=10%, a ordem é x=2, que se refere ao valor 18.
- Para p=80%, a ordem é x=9, que se refere ao valor 39.

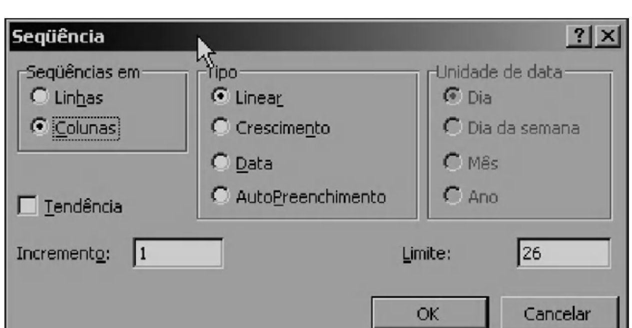
Tenha em mente que há diversas formas de relacionar um conjunto de dados ordenados de forma crescente com o respectivo percentil. A forma apresentada é a utilizada pelas funções estatísticas do Excel.

EXEMPLO 3.6

Determine a ordem dos dados da amostra do Exemplo 3.2, depois, para cada ordem, calcule o percentil correspondente e, por último e a partir desse resultado, obtenha a ordem utilizando o Excel e as fórmulas apresentadas.

Solução. Primeiro foi feita uma cópia da planilha **Exemplo 3.3** que recebeu o nome **Exemplo 3.6**. A seguir:

- Na coluna D, foi registrada a ordem de cada dado ordenado da coluna C, do número um até o 26. Esse preenchimento pode ser realizado de duas formas:
  - Registre os números 1 e 2, respectivamente, nas células D5 e D6. Depois, com o mouse, selecione as duas células e arraste a alça de preenchimento das células selecionadas  até a célula D30. Essa alternativa pode provocar mudanças das formatações de células que receberão a cópia dos valores.
  - A alternativa é a seguinte: registrar o número 1 na célula D5, no menu **Editar**, selecionar **Preencher** e, a seguir, **Sequência** que apresentará a caixa de diálogo **Sequência**, cuja figura é mostrada a seguir depois de preencher os dados necessários para registrar os números 1 a 26. Depois de pressionar **OK**, esse comando preenche os valores solicitados. Essa alternativa também pode provocar mudanças das formatações de células que receberão a cópia dos valores.



- Continuando, na célula E5, foi registrada a fórmula =(D5-1)/($D$30 -1) que calcula o percentil do dado da amostra com ordem igual a um. Depois, essa fórmula foi copiada até a célula E30, completando o cálculo do percentil da ordem dos dados restantes da amostra.
  Na coluna F, foi calculada a ordem de cada percentil registrado na coluna E. Na célula F5, foi registrada a fórmula: =($D$30-1)*E5+1, que, depois, foi copiada até a célula F30.
  A próxima figura mostra a **Planilha 3.6** depois de completar o registro das fórmulas.

	A	B	C	D	E	F
1	Exemplo 3.6					
2						
3			Amostra			
4		Amostra	Ordenada	Ordem	*p*	*x*
5		27	26	1	0,00%	1
6		32	27	2	4,00%	2
7		64	29	3	8,00%	3
8		65	29	4	12,00%	4
9		58	30	5	16,00%	5
10		62	31	6	20,00%	6
11		59	32	7	24,00%	7

EXEMPLO 3.7

Continuando com os dados e resultados do Exemplo 3.2, quais os dados da amostra com percentil 50% e 77%?

Solução. Para o percentil p=50%, obtemos a ordem x=13,50, resultado obtido com a fórmula:

$$x = (26-1) \times \frac{50}{100} + 1 = 13{,}50$$

Na tabela do Exemplo 3.2 ou do Exemplo 3.6 ou na planilha Excel correspondente, observa-se que não há ordem 13,50. Entretanto, tendo presente que na definição de percentil foi estabelecida uma relação linear com a ordem, é possível realizar uma interpolação linear entre as ordens definidas. Dessa maneira, se para x=13 o dado da amostra é 38 e para x=14 é 42, a ordem x=13,50 corresponderá ao dado 40=38+(42–38)×0,50, valor que também não pertence à amostra. Com o mesmo procedimento, para o percentil p=77%, obtém-se a ordem x=20,25 e o dado correspondente 49,50=48+(54–48)×0,25, que também não pertence à amostra.

EXEMPLO 3.8

Os retornos acumulados nos últimos doze meses dos primeiros vinte fundos de investimento estão registrados em ordem crescente na segunda coluna da tabela da figura a seguir. Nessa tabela, foi adicionada uma coluna com a ordem dos retornos, de um a vinte. Calcule o percentil de cada retorno e, a partir dele, verifique a ordem desse retorno utilizando funções estatísticas do Excel.

Solução. As funções estatísticas ORDEM.PERCENTUAL e PERCENTIL do Excel retornam, respectivamente, o percentil e a ordem. Comecemos por conhecer as sintaxes dessas duas funções

- **ORDEM.PORCENTUAL(*matriz; valor; núm_ decimais*)**
  A função estatística ORDEM.PORCENTUAL[1] retorna o percentil do argumento *valor*, considerando a *matriz* ordenada de forma crescente. Se a *matriz* tiver valores repetidos, a função informará o percentil do primeiro valor que encontrar. O argumento *núm_decimais* define o número de casas decimais do resultado; se omitido, o resultado terá três casas decimais. Perceba que não será necessário ordenar previamente os dados da amostra, pois a função ORDEM.PORCENTUAL ordena os dados da amostra de forma crescente antes de calcular.

1 Em inglês, a função ORDEM.PORCENTUAL é *PERCENTRANK*.

A fórmula =ORDEM.PORCENTUAL($C$4:$C$23;C4;6) foi registrada na célula D4 e depois copiada até a célula D23. Agora, no intervalo D4:D23 está registrado o percentil de cada retorno do intervalo C4:C23. Os cifrões no intervalo da matriz foram adicionados para poder copiar essa fórmula até o último dado da amostra, e o número seis de casas decimais foi para comparar esses resultados. A função ORDEM.PORCENTUAL também pode ser registrada como matriz em uma coluna da planilha:

- Selecione o intervalo G4:G23.
- Digite a fórmula =ORDEM.PORCENTUAL(C4:C23;C4:C23;6) sem pressionar a tecla Enter.
- Para inserir essa função como matriz, pressione simultaneamente as três teclas **Ctrl + Shift + Enter**; mantendo pressionada a tecla **Ctrl**, pressione e mantenha pressionada a tecla **Shift** e, por último, pressione a tecla **Enter**. Depois de pressionar as três teclas simultaneamente, obtemos os mesmos resultados do intervalo D4:D23 no qual as fórmulas receberam as chaves { }. As fórmulas matriciais não utilizam cifrões e ocupam menos memória da unidade de processamento comparada a com o registro individual de fórmulas.

	A	B	C	D	E
1	**Retorno anual de fundos de investimento**				
2					
3		**Ordem**	**Retorno**	**Percentil**	**Retorno**
4		1	16,3%	0,000%	16,30%
5		2	20,4%	5,263%	20,40%
6		3	21,0%	10,526%	21,00%
7		4	23,6%	15,789%	23,60%
8		5	24,7%	21,053%	24,70%
9		6	24,8%	26,316%	24,80%
10		7	26,2%	31,579%	26,20%
11		8	26,6%	36,842%	26,60%
12		9	27,0%	42,105%	27,00%
13		10	27,8%	47,368%	27,80%
14		11	28,6%	52,632%	28,60%
15		12	30,2%	57,895%	30,20%
16		13	30,3%	63,158%	30,30%
17		14	30,7%	68,421%	30,70%
18		15	32,0%	73,684%	32,00%
19		16	32,5%	78,947%	32,50%
20		17	32,6%	84,211%	32,60%
21		18	34,3%	89,474%	34,30%
22		19	35,5%	94,737%	35,50%
23		20	36,7%	100,000%	36,70%
24					

- **PERCENTIL(*matriz; k*)**
  A função estatística PERCENTIL[2] retorna o valor que divide a *matriz* em duas partes, uma menor do que o argumento *k* e a outra maior do que *k*. O argumento *k* é um valor entre 0 e 1, correspondendo respectivamente a 0% e 100% da quantidade de dados da *matriz*. Observe que não será necessário ordenar previamente os dados da amostra, pois a função PERCENTIL ordenará os dados da amostra de forma crescente antes de calcular. Nem sempre o resultado da função percentil é um valor da amostra. Por exemplo, o valor correspondente ao percentil 75% da amostra do Exemplo 3.1 é 38,50, resultado obtido por interpolação linear a partir da relação linear entre a ordem e o percentil de cada valor da amostra, como vimos no Exemplo 3.7.

A fórmula =PERCENTIL($C$4:$C$23;D4) foi registrada na célula E4 e depois copiada até a célula E23. Agora, no intervalo E4:E23 está registrado o retorno do percentil registrado no intervalo D4:D23. Os cifrões no intervalo da matriz foram adicionados para poder copiar essa fórmula até o último dado da amostra. A função PERCENTIL pode ser também registrada como matriz em uma coluna da planilha:

- Selecione o intervalo H4:H23.
- Digite a fórmula =PERCENTIL(C4:C23;D4:D23) sem pressionar a tecla Enter.

2 Em inglês, a função PERCENTIL é *PERCENTILE*.

- Para inserir essa função como matriz, pressione simultaneamente as três teclas **Ctrl + Shift + Enter.** Depois de pressionar as três teclas simultaneamente, obtemos os mesmos resultados do intervalo E4:E23, no qual as fórmulas receberam as chaves { }.

Outras funções estatísticas relacionadas com esse tema podem ser encontradas no Apêndice 1 deste capítulo.

Se o administrador de um fundo equivalente não listado na tabela afirma que o retorno acumulado nos últimos doze meses de seu fundo foi 30,2%, então seu percentil é p=57,9% e, consequentemente, o retorno do seu fundo é maior do que 57,9% dos primeiros fundos da tabela e menor do que os 42,1% dos demais fundos. Observe que um fundo com retorno de 32,52% tem percentil 80%; dessa maneira, o retorno desse fundo é maior do que 80% dos fundos da amostra e menor do que os restantes 20% dos fundos com seus retornos ordenados de forma crescente. Note que, enquanto o percentil 80% é uma medida relativa, pois somente avalia o desempenho do fundo em relação aos outros fundos, o retorno do fundo de 32,52% é uma medida absoluta. O ordenamento com percentil não representa uma escala intervalar constante, pois trata apenas com posições de valores ordenados.

Quartil

Na relação entre a escala de 0% a 100% e a série de números naturais 0, 1, 2, ... n que representam uma série de dados de uma amostra ordenada de forma crescente, o primeiro dado da amostra é o percentil 0%, e o último dado da amostra é o percentil 100%. Também há outras formas de definir referências fixas, por exemplo, cada 10% ou *decil*, ou cada 12,5% ou *octil*, ou cada 25% ou *quartil* que será apresentado a seguir. Dividindo os valores ordenados da variável em quatro quartos iguais, obtém-se um quartil para cada quarto definido desta forma:

- O primeiro quartil Q_1 é o percentil 25%. O valor da amostra do primeiro quartil Q_1 é maior do que 25% dos valores menores e menor do que 75% dos demais valores maiores.
- O segundo quartil[3] Q_2 é o percentil 50%. O valor da amostra do segundo quartil Q_2 é maior do que 50% dos valores menores e menor do que 50% dos demais valores maiores. O segundo quartil é também a *mediana* que divide a área da distribuição de frequências em duas partes iguais a 50%.
- O terceiro quartil Q_3 é o percentil 75%. O valor da amostra do terceiro quartil Q_3 é maior do que 75% dos valores menores e menor do que 25% dos demais valores maiores.

Da fórmula do percentil, obtêm-se as fórmulas dos três quartis utilizadas pelo Excel, como mostrado a seguir.

- Conhecido o percentil p de um dado da amostra ordenada, sua ordem x é calculada com a fórmula $x = (n-1) \times \frac{p}{100} + 1$. No primeiro quartil, p=25% ou 1/4, a fórmula passa a ser $x = (n-1) \times \frac{1}{4} + 1 = \frac{n+3}{4}$
- A fórmula da ordem no segundo quartil p=50% é $x = \frac{n+1}{2}$.
- A fórmula da ordem no terceiro quartil p=75% é $x = \frac{3 \times n+1}{4}$.

Se o resultado de x não for um número inteiro, o valor do dado da amostra ou variável será obtido com interpolação linear como já apresentado.

3 A *mediana* divide a área da distribuição de frequências em duas partes iguais a 50%.

EXEMPLO 3.9

Calcule o primeiro, segundo e terceiro quartis dos retornos do Exemplo 3.8.

Solução. A função estatística QUARTIL do Excel retorna o valor do quartil informado. Comecemos por conhecer a sintaxe dessa função.

- **QUARTIL(*matriz; quarto*)**
  A função estatística QUARTIL[4] retorna o dado da *matriz* ordenada correspondente ao argumento *quarto* identificado da seguinte maneira:
  - Se *quarto*=0, a função retornará o primeiro ou menor valor da *matriz*.
  - Se *quarto*=1, 2 ou 3, a função retornará o valor da matriz correspondente e respectivamente, ao primeiro, segundo ou terceiro quartis.
  - Se *quarto*=4, a função retornará o último ou maior valor da *matriz*.

Enquanto a função QUARTIL fornece resultados de posições definidas na amostra ordenada, a função PERCENTIL dá os resultados para qualquer posição de 0 a 1, ou 0% a 100%. No entanto, nem sempre o retorno da função QUARTIL é um dado da amostra.

A próxima figura mostra o cálculo de todos os resultados da função QUARTIL utilizando os retornos dos fundos de investimento da planilha **Exemplo 3.8** a partir da linha 26.

	A	B	C	D	E	F
26	Exemplo 3.9					
27						
28		Percentil	Quartil	Retorno		
29		0%		16,300%	=QUARTIL(C4:C23;0)	
30		25%	Primeiro	24,775%	=QUARTIL(C4:C23;1)	
31		50%	Segundo	28,200%	=QUARTIL(C4:C23;2)	
32		75%	Terceiro	32,125%	=QUARTIL(C4:C23;3)	
33		100%		36,700%	=QUARTIL(C4:C23;4)	
34						

Analisemos os cinco resultados da função estatística QUARTIL, lembrando que nem sempre o retorno é um dado da amostra.

- Os resultados da função QUARTIL para o argumento *quarto* igual a zero ou quatro coincide, respectivamente, com o primeiro (menor) ou último (maior) dado da amostra ordenada.
- O retorno do primeiro quartil é 24,775%, valor que não consta na série de retornos. Nesse caso, o valor do quartil foi obtido com a interpolação linear 0,24775=0,2470+(0,2480–0,2470)×(0,25–0,21053)/(0,26316–0,21053).
- Os retornos do segundo e do terceiro quartil foram obtidos da mesma forma que o do segundo quartil.

Ferramenta de análise *Ordem e Percentil*

A partir de uma amostra quantitativa discreta registrada em uma planilha Excel, uma série de valores registrados em uma ou mais colunas contíguas, a ferramenta de análise *Ordem e percentil* retornará, a partir do endereço informado pelo usuário, uma tabela com a posição ordinal e percentual de cada dado da amostra, permitindo analisar a posição relativa dos valores em um conjunto de dados.

Para utilizar a ferramenta de análise *Histograma Ordem e Percentil*,[5] a amostra que será analisada deve estar registrada em uma planilha como a **Ferramenta Ordem e Percentil** incluída na pasta **Capítulo 3**, sendo que:

- No intervalo B3:B29 foram registrados os valores numéricos da amostra do Exemplo 3.2, incluindo o nome *Amostra* na célula B3. Os valores da amostra podem ser registrados em uma linha, uma coluna ou combinando linhas e colunas, contanto que sejam contíguos e possíveis de identificá-los com um único intervalo.

4 Em inglês, a função QUARTIL é *QUARTILE*.

5 Em inglês, a ferramenta ORDEM E PERCENTIL é *RANK AND PERCENTILE*.

- Selecione o intervalo B3:B29.
- 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 do livro.
- Escolhendo a ferramenta **Ordem e percentil**, depois de pressionar o botão OK, você receberá a caixa de diálogo **Ordem e percentil** mostrada na Figura 3.2, depois de selecionadas algumas opções.
  - Pressionando o botão **Ajuda** dessa caixa de diálogo, o Excel apresentará a página *Sobre a caixa de diálogo Ordem e percentil* pertencente à *Ajuda do Excel*.

FIGURA 3.2 Caixa de diálogo da ferramenta *Ordem e percentil*.

As informações que devem ser registradas no quadro **Entrada** da caixa de diálogo da ferramenta *Ordem e percentil* são:

- **Intervalo de entrada**. Informe o intervalo de células da planilha no qual os dados estão registrados; nesse caso, o intervalo B3:B29 que inclui a célula onde foi registrado o título *Amostra*, ou rótulo no Excel.
- **Agrupado por**. Selecionamos **Colunas**, pois a amostra foi registrada em uma coluna. Em geral, o Excel selecionará automaticamente depois de ter informado o 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 3.3 Ferramenta Ordem e Percentil resolvendo o Exemplo 3.6.

	A	B	C	D	E	F	G
1	Ferramenta de Análise ORDEM E PERCENTIL						
2							
3		Amostra		*Ponto*	*Amostra*	*Ordem*	*Porcentagem*
4		27		4	65	1	100,00%
5		32		3	64	2	96,00%
6		64		6	62	3	92,00%
7		65		7	59	4	88,00%
8		58		5	58	5	84,00%
9		62		8	54	6	80,00%
10		59		12	48	7	76,00%
11		54		13	47	8	72,00%

No quadro **Opções de saída**, deve ser obrigatoriamente informado um endereço a partir do qual a ferramenta de análise registrará os resultados. Há três alternativas excludentes de 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, que é o endereço 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. Podem ser encontradas mais informações no Capítulo 1 ou na *Ajuda* do Excel.

- **Nova planilha**. Os resultados serão apresentados a partir da célula A1 de uma nova planilha da mesma pasta.
- **Nova pasta de trabalho**. Os resultados serão apresentados em uma nova pasta e a partir da célula A1 da planilha **Plan1**.

Depois de pressionar o botão **OK**, a ferramenta *Ordem e percentil* apresentará os resultados solicitados nas seleções realizadas, como mostra a Figura 3.3. A partir da célula D3 da planilha, a ferramenta registra a tabela de resultados cuja análise é realizada a seguir.

- Na coluna E (*Amostra*) da tabela, a ferramenta registrou os dados da *amostra* ordenados de forma decrescente.
- Na coluna D (*Ponto*), foi registrada a posição de cada dado da coluna E registrado na coluna B. Por exemplo, o valor 62 registrado na célula E6 tem a posição 6 (célula D6) na amostra da coluna B, ou o valor 62 é o sexto dado da amostra da coluna B, célula B9.
- Na coluna F (*Ordem*), foi registrada a ordem de cada dado da amostra registrada na coluna E da tabela. Se na amostra há valores repetidos, a classificação manterá *ordem* do primeiro valor não repetido. A *ordem* é calculada com a função estatística ORDEM, apresentada no Apêndice 1 do Capítulo 3.
- Na coluna G (*Porcentagem*), foi registrado o *percentil* de cada dado da amostra ordenada de forma decrescente. Esses valores foram calculados com a função estatística ORDEM.PORCENTUAL já apresentada.

Medidas de tendência central

No Capítulo 2, mostramos como apresentar dados numéricos de forma agrupada utilizando tabelas de frequências e histogramas. A parte inicial deste Capítulo 3 mostrou como trabalhar com as posições relativas dos dados ordenados de uma amostra utilizando percentil e quartil. Os exemplos desenvolvidos no Capítulo 2 mostram que os dados tendem a se agrupar ao redor de um ponto central, mostrando a oportunidade de definir novas medidas que podem representar toda a amostra ou variável. A *mediana* é uma das medidas de tendência central cuja definição coincide com o percentil 50%, ou o segundo quartil, de uma série de dados ordenados de forma crescente. As outras medidas de tendência central são a *moda* e a *média aritmética* ou simplesmente *média*.

Mediana

A *mediana Md* é uma medida de tendência central cuja definição coincide com o percentil 50%, ou o segundo quartil, de uma série de dados ordenados de forma crescente. A mediana *Md* é um valor localizado na posição central, tal que 50% dos valores são menores do que *Md*, e os demais 50% são maiores.

Depois de ordenar os n valores da variável de forma crescente, a Md é determinada de acordo com o tipo do número n:

Se n for um número ímpar, a Md será o valor da variável situado na posição $(n+1)/2$.

Se n for um número par, a Md será igual ao resultado da divisão por dois da soma dos valores das posições $(n/2)$ e $(n/2)+1$. Nesse caso, a Md poderá não ser um valor da variável.

Note que a quantidade de dados da amostra acima de *Md* é igual à quantidade de dados da amostra abaixo dele, seja *n* par ou ímpar. De outra maneira, a mediana *Md* divide a área da distribuição de frequências em duas partes iguais a 50%.

EXEMPLO 3.10

Calcule a *mediana* da amostra do Exemplo 3.1.

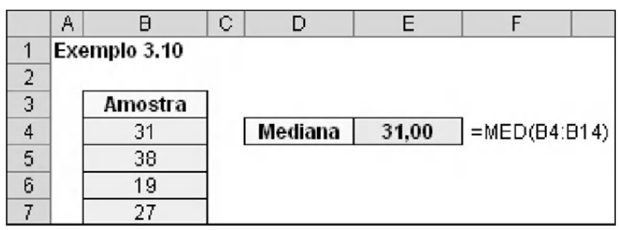
Solução. Para facilitar o trabalho, os dados da amostra são repetidos a seguir.

31	38	19	27	24	42	32	18	43	15	39

A tabela a seguir mostra os 11 valores da amostra ordenados de forma crescente, identificando o valor da mediana dentro de um círculo.

15	18	19	24	27	31	32	38	39	42	43

Como a quantidade de dados da amostra n=11 é um número ímpar, o valor da mediana é *Md*=31, que corresponde ao dado da posição 6=(11+1)/2. O mesmo resultado foi obtido com a função MED do Excel, como mostra a figura a seguir, referente à planilha **Cálculo da Mediana** da pasta **Capítulo 3**.



	A	B	C	D	E	F
1	Exemplo 3.10					
2						
3		Amostra				
4		31		Mediana	31,00	=MED(B4:B14)
5		38				
6		19				
7		27				

A mediana foi obtida com a fórmula =MED(B4:B14) registrada na célula E4.

- **MED(*núm1; núm2; ... ; núm30*)**
  A função estatística MED(*núm1; núm2; ... ; núm30*) retorna a mediana 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:B14 foi registrada no primeiro argumento *núm1*. Mais informações sobre essa função e outras formas de utilizá-la estão disponíveis no Apêndice 1 deste capítulo.

EXEMPLO 3.11

Calcule a *mediana* da amostra do Exemplo 3.2.

Solução. Para facilitar o trabalho, os dados da amostra são repetidos a seguir.

27	32	64	65	58	62	59	54	29	30	26	48	47
46	43	38	29	32	35	37	31	43	45	42	37	36

A tabela a seguir mostra os 26 valores da amostra ordenados de forma crescente, identificando os valores que fazem parte do cálculo da mediana dentro de um círculo.

26	27	29	29	30	31	32	32	35	36	37	37	38
42	43	43	45	46	47	48	54	58	59	62	64	65

Como a quantidade de dados n=26 é um número par, o valor da mediana será igual ao resultado da divisão por dois da soma dos valores das posições $(n/2)=13$ e $(n/2)+1=14$. O valor da mediana é Md=40, resultado obtido de (38+42)/2. O mesmo resultado foi obtido com a função MED do Excel, como mostra a figura a seguir referente à planilha **Cálculo da Mediana** da pasta **Capítulo 3**.

	A	B	C	D	E	F	G
17	**Exemplo 3.11**						
18							
19		**Amostra**					
20		27		**Mediana**	**40,00**	=MED(B20:B45)	
21		32					
22		64					
23		65					

Analisando os resultados dos exemplos anteriores, podemos chegar a algumas conclusões interessantes:

- Na amostra do Exemplo 3.10, acima da Md=31, há cinco dados da amostra e, abaixo dela, também há cinco dados, e a mediana é um valor da amostra.
- Da mesma forma, na amostra do Exemplo 3.11, acima da Md=40, há 13 dados da amostra e, abaixo dela, também há 13 dados; entretanto, a Md não é um valor da amostra.
- A mediana divide a distribuição de frequências em duas áreas iguais, ou duas áreas com a mesma quantidade de valores ordenados da amostra ou variável ou, de outra maneira, a mediana Md divide a área da distribuição de frequências em duas partes iguais a 50%.
- Se o maior valor da amostra for duplicado, o valor Md não será alterado, pois está relacionado apenas com a ordem da série de valores. A mediana é uma medida, resistente, ela é menos sensível à presença de valores suspeitos, dados bastante diferentes da maioria dos dados coletados na mesma amostra. A eliminação de dados suspeitos não deverá afetar a mediana, o que não ocorrerá com a média que será afetada.

Moda

A tabela de frequências absolutas do Exemplo 2.1 do Capítulo 2 mostra que o número de operações diárias fechadas pelo Operador B com maior frequência da série de dados dessa amostra é 14 operações. Essa é a medida de tendência central denominada *moda Mo*, nesse exemplo Mo=14.

> *Moda* é o valor da amostra ou variável que mais se repete; ou valor com mais frequência.

EXEMPLO 3.12

Calcule a moda Mo da amostra do número de operações fechadas diariamente pelo Operador B do Exemplo 2.1, cujos dados repetimos.

14	12	13	11	12	13	16	14	14	15	17	14	11
13	14	15	13	12	14	13	14	13	15	16	12	12

Solução. A tabela de frequências absolutas do Exemplo 2.1 mostra que *Mo*=14, o número de operações diárias fechadas pelo Operador B com maior frequência. O mesmo resultado foi obtido com a função MODO do Excel, como mostra a figura a seguir referente à planilha **Cálculo da Moda** da pasta **Capítulo 3**.

	A	B	C	D	E	F
1	Exemplo 3.12					
2						
3		Amostra				
4		14	14	14	13	
5		12	14	15	15	
6		13	15	13	16	
7		11	17	12	12	
8		12	14	14	12	
9		13	11	13		
10		16	13	14		
11						
12		Moda	14,00	=MODO(B4:B10;C4:C10;D4:D10;E4:E8)		
13						

O valor da moda foi obtido com a fórmula registrada na célula C12 =MODO(B4:B10;C4:C10;D4:D10;E4:E8).

- **MODO(*núm1; núm2; ... ; núm30*)**
  A função estatística MODO(*núm1; núm2; ... ; núm30*) retorna a moda 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 foi registrada nos quatro primeiros argumentos *núm1*, *núm2*, *núm3* e *núm4*. Mais informações sobre essa função e outras formas de utilizá-la estão disponíveis no Apêndice 1 deste capítulo.

EXEMPLO 3.13

Determine a moda *Mo* da amostra do Exemplo 3.2.

Solução. Para facilitar a determinação da moda, os dados ordenados de forma crescente da amostra são repetidos e identificados a seguir.

26	27	29	29	30	31	32	32	35	36	37	37	38
42	43	43	45	46	47	48	54	58	59	62	64	65

Na amostra da tabela apresentada detectamos quatro modas, com dois dados cada uma com áreas pintadas, *Mo*=29, 32, 37 e 43. O resultado obtido com a função MODO do Excel na planilha **Cálculo da Moda** da pasta **Capítulo 3** é 32.

As amostras ou variáveis com valores quantitativos contínuos costumam não apresentar moda; por exemplo, a série das 50 maiores empresas privadas por venda mostrada no Capítulo 1 não tem moda. A amostra ou variável com uma única *moda* é denominada *unimodal*, com duas modas é *bimodal* etc. A *moda* também é uma medida resistente, pois está relacionada apenas com a frequência de um ou mais dados da amostra. Por exemplo, a mudança de um dado da amostra poderá não afetar a moda *Mo*.

Média

A medida de posição mais utilizada é a *média aritmética* ou simplesmente *média* de uma amostra ou variável.

Média $\overline{X}$ é o resultado da divisão da soma dos valores das observações ou dados $X_1, X_2, \cdots, X_i, \cdots, X_n$ da amostra X pela quantidade de dados n:

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

As características importantes da *média* são:

- A unidade de medida da média é a mesma que a dos valores da amostra.
- O resultado da multiplicação da média $\overline{X}$ pela quantidade n de valores da amostra X é igual à soma dos n valores da amostra.

No Apêndice 3 você encontra informações e como utilizar o símbolo somatória $\sum$.

EXEMPLO 3.14

Calcule a média da amostra do Exemplo 3.1.

Solução. A média da amostra é igual a $\overline{X}$, resultado obtido com a fórmula e também resolvido na planilha **Cálculo da Média** da pasta **Capítulo 3**.

	A	B	C	D	E	F	G	H
1	Exemplo 3.14							
2								
3		Amostra						
4		31						
5		38						
6		19						
7		27						
8		24						
9		42						
10		32						
11		18						
12		43						
13		15		Média	29,82	=SOMA(B4:B14)/CONT.NÚM(B4:B14)		
14		39		Média	29,82	=MÉDIA(B4:B14)		
15								

$$\overline{X} = \frac{\sum_{i=1}^{11} X_i}{11} = \frac{31 + 38 + \cdots + 39}{11} = 29{,}82$$

O cálculo da média da amostra é realizado de três formas diferentes.

- De forma manual, utilizando a fórmula que define a média da amostra.
- Com funções do Excel equivalentes à fórmula que define a média da amostra utilizando a fórmula =SOMA(B4:B14)/CONT.NÚM(B4:B14) registrada na célula E13.
  - **SOMA(*núm1; núm2; ... ; núm30*)**
    A função matemática SOMA(*núm1; núm2; ... ; núm30*) retorna a soma dos valores numéricos *núm1; núm2; ... ; núm30*. Cada um desses *núm* pode ser um intervalo de células de uma planilha contendo valores numéricos ou assemelhados. Mais informações sobre essa função e outras formas de utilizá-la estão disponíveis no Apêndice 1 deste capítulo.
- Com a função estatística MÉDIA do Excel utilizando a fórmula =MÉDIA(B4:B14) registrada na célula E14.
  - **MÉDIA(*núm1; núm2; ... ; núm30*)**
    A função estatística MÉDIA(*núm1; núm2; ... ; núm30*) retorna a média aritmética 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:B14 foi registrada no primeiro argumento *núm1*. Se o nome da função MÉDIA for inserido com letras minúsculas ou maiúscu-

las ou sem o acento ortográfico, o Excel aceitará e registrará a função com letras maiúsculas e com o acento ortográfico. Mais informações sobre essa função e outras formas de utilizá-la estão disponíveis no Apêndice 1 deste capítulo.

EXEMPLO 3.15

Calcule a média da amostra de operações diárias fechadas pelo Operador B e explicar seu significado, Exemplo 2.1.

Solução. Aplicando a definição de média da população temos o resultado obtido com a seguinte fórmula.

$$\overline{X} = \frac{1}{26}\sum_{i=1}^{26} X_i = \frac{1}{26}\times(14+12+\cdots+12) = \frac{352}{26}$$

$$\overline{X} = 13{,}54$$

Qual o significado da média igual a 13,54?

- A média tem a mesma unidade de medida que os valores da amostra.
- A média 13,54 é a quantidade equivalente de operações fechadas diariamente pelo operador B, pois o resultado da multiplicação da média pelo número 26 é igual a 352, a soma dos 26 valores da variável.

Propriedades da média

A média é a medida de posição mais utilizada porque tem propriedades importantes, como as que serão apresentadas. Para mostrar essas propriedades, necessitamos utilizar algumas expressões matemáticas. Suponha uma amostra ou variável X com n dados ou observações, não necessariamente ordenados, e identificados pela sequência de valores $X_1, X_2, \cdots, X_i, \cdots, X_n$, onde X_1 é o primeiro dado, X_2 é o segundo dado, X_i é um dado qualquer da amostra, e assim sucessivamente até o último dado X_n. Denomina-se *desvio de um dado* X_i de uma amostra o resultado da diferença entre X_i e a média $\overline{X}$ da amostra X. Em termos matemáticos $= X_i - \overline{X}$.

Primeira propriedade

A soma dos desvios de uma amostra ou variável é sempre igual a zero.

$$\sum_{i=1}^{n}(X_i - \overline{X}) = 0$$

Essa propriedade é útil para verificar ou confirmar o resultado do cálculo da média de uma amostra ou variável, como também no desenvolvimento de provas matemáticas que apresentam a soma de desvios com relação à média. A primeira propriedade da média também pode ser utilizada para determinar a média de uma amostra, como mostra o Exemplo 3.16.

EXEMPLO 3.16

Determine o valor da média da amostra do Exemplo 3.1 aplicando a primeira propriedade da média e utilizando o Excel.

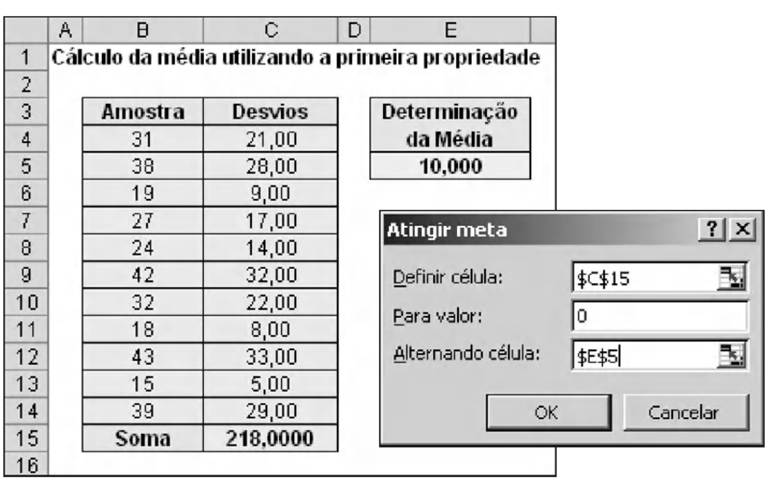
Solução. No intervalo B3:B14 da planilha **Média com Atingir Meta**, incluída na pasta **Capítulo 3**, foi registrada a amostra do Exemplo 3.1. Depois foram adicionados os registros mostrados na próxima figura.

	A	B	C	D	E
1	Cálculo da média utilizando a primeira propriedade				
2					
3		Amostra	Desvios		Determinação
4		31	21,00		da Média
5		38	28,00		10,000
6		19	9,00		
7		27	17,00		
8		24	14,00		
9		42	32,00		
10		32	22,00		
11		18	8,00		
12		43	33,00		
13		15	5,00		
14		39	29,00		
15		Soma	218,0000		
16					

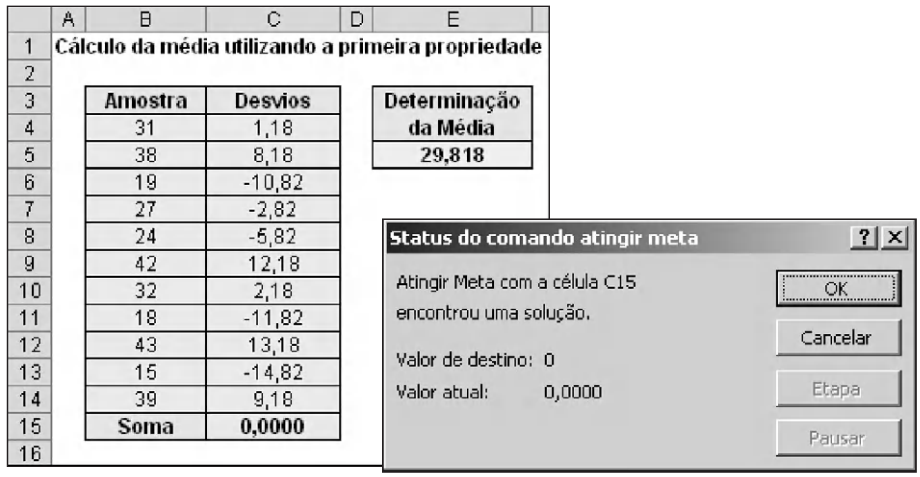
- Na célula E5 será determinado o valor da média da amostra.
- Na célula C4 foi registrada a fórmula =B4-$E$5 que calcula o desvio do dado da amostra registrado na célula B4 com relação à média registrada na célula E5. Depois essa fórmula foi copiada até a célula C14.
- Na célula C15 foi registrada a fórmula =SOMA(C4:C14) que retorna a soma de todos os desvios.

Pela primeira propriedade da média, verificamos que o valor 10 registrado na célula E5 não é o valor da média da amostra, pois a soma dos desvios é diferente de zero. Da forma como foi preparada a planilha, poderemos encontrar o valor da média de forma manual, registrando diferentes valores na célula E5 até conseguir zerar o valor da célula E5, procedimento trabalhoso e cansativo. Essa resposta pode ser encontrada rapidamente utilizando o comando *Atingir Meta* da seguinte forma:

- Posicione o cursor do Excel na célula C15.
- No menu **Ferramentas** do Excel, selecione **Atingir meta**. Será exibida a caixa de diálogo **Atingir meta**.
- Nessa caixa de diálogo, informe os dados, como mostra a figura a seguir.
  - **Definir célula**. Nessa caixa é registrado o endereço da célula que contém a fórmula cujo resultado será definido na caixa seguinte. Posicionando o cursor do Excel na célula C15, nessa caixa aparecerá esse endereço. A célula C15 deve obrigatoriamente conter uma fórmula.
  - **Para valor**. Nessa caixa, registramos o resultado desejado na célula C15 endereço definido em *Definir célula*, nesse caso o valor 0. Para acessar a caixa **Para valor**, basta pressionar a tecla **Tab** ou clicar na caixa.
  - **Alternando célula**. Nessa caixa é registrado o endereço da célula que deverá ser alterada para que a célula C15 atinja o valor desejado 0, ou o endereço da célula que contém o valor que se deseja ajustar. Esse dado pode ser registrado, depois de posicionar o cursor nesta caixa, clicando na própria célula E5, ou digitando o endereço da célula E5 na própria caixa.
- Depois de completar as informações, clique em **OK**, e o comando Atingir Meta inicia o processo de busca da



solução desejada. Concluído o processo de busca, o Excel apresentará a caixa de diálogo **Status do comando atingir meta**, informando que foi encontrada uma solução, o *Valor de destino* 0 registrado na caixa **Para valor** e o *Valor atual* encontrado na célula C15.



	A	B	C	D	E
1	Cálculo da média utilizando a primeira propriedade				
2					
3		Amostra	Desvios		Determinação
4		31	1,18		da Média
5		38	8,18		29,818
6		19	-10,82		
7		27	-2,82		
8		24	-5,82		
9		42	12,18		
10		32	2,18		
11		18	-11,82		
12		43	13,18		
13		15	-14,82		
14		39	9,18		
15		Soma	0,0000		
16					

Segunda propriedade

A soma dos quadrados dos desvios com relação à própria média de uma variável ou amostra é sempre um valor mínimo.

$$\sum_{i=1}^{n}(X_i - \overline{X})^2 \Rightarrow \textit{mínimo}$$

No Capítulo 4 será mostrado como medir a variabilidade dos dados de uma amostra utilizando os desvios dos dados com relação à média, onde a soma dos quadrados dos desvios é utilizada na definição de variância.

Visualização das propriedades

No caminho ficou a pergunta: qual é o significado de *mínimo*? A resposta está, inicialmente, na própria declaração da propriedade. Que a soma dos quadrados dos desvios com relação à média da própria variável ou amostra seja um valor mínimo significa que se os desvios fossem calculados com relação a qualquer outro valor diferente da média da amostra, a nova soma dos quadrados dos desvios seria maior do que a primeira. Demonstra-se que somente a própria média da amostra ou variável satisfaz à condição de mínimo, como se pode ver no Apêndice 3 deste capítulo. Também há a possibilidade de compreender essa propriedade de forma visual com a planilha **Visualização Propriedades** incluída na pasta **Capítulo 3**, como mostra a Figura 3.4, utilizando a amostra do Exemplo 3.1.

- No intervalo B5:B15 foi registrada a amostra do Exemplo 3.1.
- Na célula D26 foi calculada e registrada a verdadeira média da amostra utilizando a função estatística MÉDIA.
- No intervalo D21:D24, foi incluído o controle giratório, recurso disponível no Excel para aumentar ou diminuir o valor da célula D17, as possíveis médias da amostra. Para aumentar o valor do parâmetro da célula D17, clique na seta para cima do controle e, para diminuir, clique na seta para baixo.
- No intervalo C5:C15, foram calculados os desvios dos valores da amostra com relação ao valor registrado na célula D17. A soma dos desvios foi registrada na célula D18.

- No intervalo D5:D15 foram calculados os quadrados dos desvios cuja soma foi registrada na célula D19.

Na planilha, foram construídos dois gráficos que permitem visualizar o que ocorre quando informamos valores arbitrários da média da amostra. No primeiro gráfico, *Ajuste manual do valor da Média*, clicando na seta para cima ou na seta para baixo do controle giratório, a reta se desloca no sentido vertical do gráfico. Esse ajuste pode ser visualizado de duas formas:

- O primeiro procedimento é comparar os comprimentos das retas tracejados verticais entre os pontos acima e os pontos abaixo da reta horizontal, que representa a possível média da amostra. Considerando positivos os comprimentos dos valores situados acima da reta horizontal, e negativos os valores abaixo da mesma reta, a soma desses comprimentos tem de ser igual a zero, de acordo com a primeira propriedade.
- O segundo procedimento é acompanhar a variação do valor da soma dos quadrados dos desvios registrada na célula D19.

O segundo gráfico, *Soma quadrado dos desvios Vs. Médias*, mostra a parábola dos valores da soma dos quadrados dos desvios para diversos valores arbitrários da média. O valor de média registrado na célula

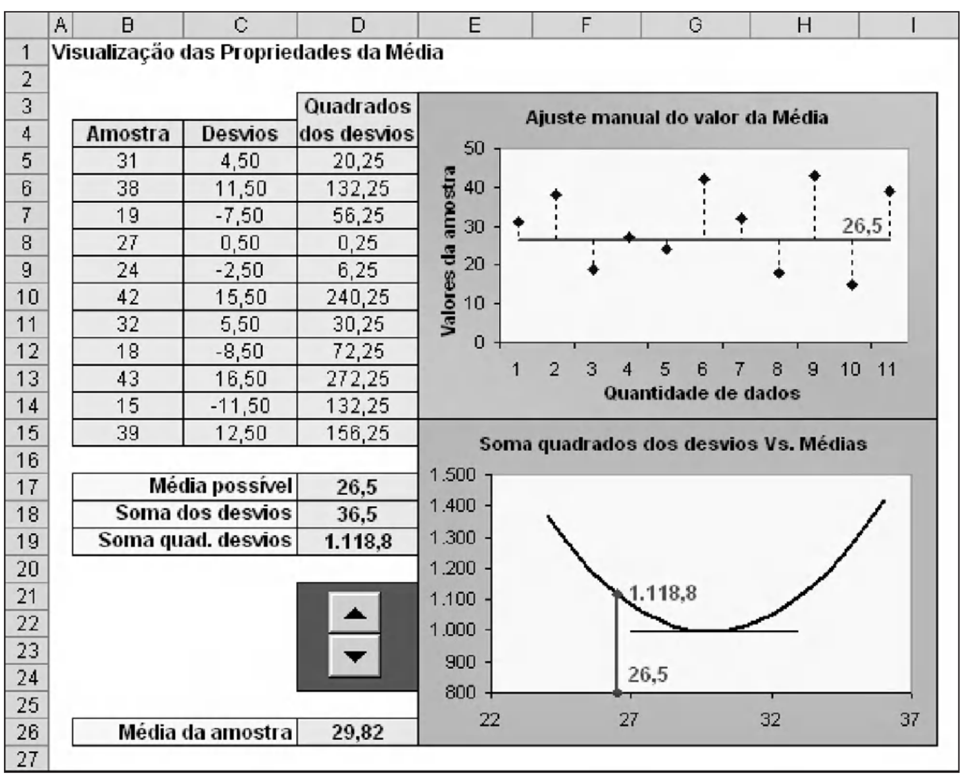


FIGURA 3.4 Visualização das propriedades da média.

D17 é destacado nessa parábola, facilitando a compreensão do procedimento de procura do mínimo. Resumindo, ao clicar na seta para cima ou na seta para baixo do controle giratório, um novo valor arbitrário de média é registrado, a reta do primeiro gráfico se desloca na vertical, o ponto que representa o novo valor arbitrário de média se desloca na parábola e os valores dos desvios mudam, intervalo D18:D19. Uma reta horizontal de espessura fina localizada na parte inferior da parábola é a tangente à curva no ponto de mínimo.

Análise do resultado da média

Analisando o procedimento de cálculo da média, pode-se concluir que:

- Todos os valores da variável são incluídos no cálculo da média.
- A média é um valor único.
- A média está posicionada de forma equilibrada entre os valores ordenados da amostra. De outra maneira, os valores da amostra se distribuem ao redor da média. Os gráficos da planilha **Visualização Propriedades** ajudam a compreender o que descrevemos.
- A média não é uma medida resistente, como a mediana ou a moda, pois ela é sensível à presença de dados suspeitos ou extremos; dados com valores bastante diferentes da maioria dos dados coletados na mesma amostra. Nesse caso, a média será uma medida distorcida da tendência dos valores da amostra, como mostra o Exemplo 3.17. Ao mesmo tempo, a eliminação de dados suspeitos deverá também afetar a média.
- Nas amostras ou variáveis com histograma simétrico, os valores da mediana, a moda e a média, coincidem, seus valores são iguais. Sugerimos que você tenha em mente essa representação ao analisar a formação da média e as variações ou dispersões dos valores da variável ao redor da média, tema que será apresentado no Capítulo 4.

Você deve ter percebido que alguns termos foram utilizados como sinônimos, ou quase. Por exemplo, dados e observações, amostra e variável etc. Poucas vezes nos referimos à amostra e à população como sinônimos, embora o procedimento de cálculo e o resultado da média, e apenas ela, sejam os mesmos. Entretanto, no caso de população e amostra deve-se manter essa separação para identificar a origem das variáveis, pois:

- *Parâmetros* são as medidas numéricas de uma população, identificados com letras gregas, μ para a média e σ para o desvio padrão (tema do próximo capítulo).
- *Estatísticas* são as medidas numéricas de uma amostra, identificadas com letras do nosso alfabeto, $\overline{X}$ para a média e S para o desvio padrão.

Média da população X é o resultado da divisão da soma dos valores $X_1, X_2, \cdots, X_i, \cdots, X_n$ da variável X pela quantidade de valores N:

$$\mu_X = \frac{\sum_{i=1}^{N} X_i}{N} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

EXEMPLO 3.17

A tabela a seguir registra uma amostra ordenada de 28 retornos de diversos investimentos no mesmo período. Analise a média dessa amostra e detecte dados suspeitos.

–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 3.17** incluída na pasta **Capítulo 3**, foram calculadas a mediana e a média dos retornos de diversos investimentos no mesmo período, respectivamente, $Md = 28{,}17\%$ e $\overline{X} = 32{,}31\%$. Analisando a série de retornos desses diversos investimentos ordenados verificamos que:

- A série de retornos tem valores extremos, por exemplo, o primeiro retorno igual a –2,1% e o último igual a +114,7%. Recalculando a média sem considerar os dois valores extremos, temos $\overline{X}$=30,27%, retorno mais próximo da mediana.
- Recalculando a média sem considerar o penúltimo valor da série 76,2%, temos $\overline{X}$=28,44%, próximo da mediana.

Análise das medidas de tendência central

Embora média, mediana e moda sejam medidas importantes de tendência central por serem fáceis de serem obtidas e úteis para obter informações sobre uma amostra, elas devem ser utilizadas de acordo com a análise desejada. Analisemos, primeiro, as principais vantagens e desvantagens dessas medidas.

MODA

Vantagens	Desvantagens
Fácil de calcular.	Pode estar afastada do centro dos dados.
Não é afetada pelos dados extremos da amostra.	Difícil de incluir em funções matemáticas.
Pode ser aplicada em qualquer escala: nominal, ordinal, intervalar e proporcional.	Não utiliza todos os dados da amostra.
	A amostra pode ter mais de uma moda.
	Algumas amostras podem não ter moda.

MEDIANA

Vantagens	Desvantagens
Fácil de calcular.	Difícil de incluir em funções matemáticas.
Não é afetada pelos dados extremos da amostra.	Não utiliza todos os dados da amostra.
É um valor único.	
Pode ser aplicada nas escalas: ordinal, intervalar e proporcional.	

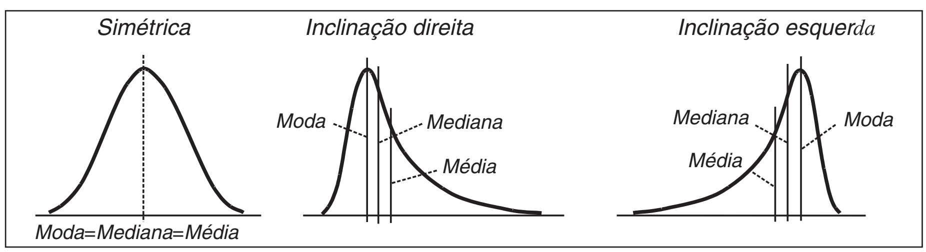
MÉDIA

Vantagens	Desvantagens
Fácil de compreender e aplicar.	É afetada pelos dados extremos da amostra.
Utiliza todos os dados da amostra.	É necessário conhecer todos os dados da amostra.
É um valor único.	
Fácil de incluir em funções matemáticas.	
Pode ser aplicada nas escalas: intervalar e proporcional.	

Agora comparemos os valores dessas medidas em três formas diferentes do contorno de uma distribuição de frequências de uma amostra ou variável. A distribuição da esquerda da Figura 3.5 mostra uma distribuição de frequências simétrica ao redor da média. Na distribuição simétrica de frequências, os valores de média, mediana e moda coincidem. As outras duas distribuições da Figura 3.5 não são simétricas, e as medidas de tendência central têm posições relativas diferentes entre si, antecipando a forma da distribuição de frequências da amostra ou variável:

- Na figura do meio a distribuição tem inclinação para a direita, simplesmente *inclinação direita* ou *positiva*. A moda está na posição do pico da distribuição, e a mediana, que divide a distribuição em duas áreas iguais, situa-se à direita da moda, pois a distribuição tem inclinação para a cauda direita. Como a média é uma medida afetada pelos dados extremos da amostra, ela estará situada à direita da mediana. Utilizando os valores das medidas, verifica-se a seguinte relação *Média* > *Mediana* > *Moda*. Como nem sempre uma amostra ou variável terá moda, a análise da forma de distribuição poderá ser realizada com as outras duas medidas, *Média* > *Mediana*. Ou seja, se a média é maior do que a mediana, a distribuição deve ter inclinação para a direita.
- De forma equivalente, na distribuição da direita da Figura 3.5, a distribuição tem inclinação para a esquerda, simplesmente *inclinação esquerda* ou *negativa*. A moda está na posição do pico da distribuição, e a mediana, que divide a distribuição em duas áreas iguais, está situada à esquerda da moda, pois a distribuição tem inclinação para a cauda esquerda. Como a média é uma medida afetada pelos dados extremos da amostra, ela estará situada à esquerda da mediana. Utilizando os valores das medidas, verifica-se a seguinte relação *Média* < *Mediana* < *Moda*. Como nem sempre uma amostra ou variável terá moda, a análise da forma de distribuição poderá ser realizada com as outras duas medidas, *Média* < *Mediana*. Ou seja, se a média é menor do que a mediana, a distribuição deve ter inclinação esquerda.

FIGURA 3.5
Distribuições de frequências, simétrica e inclinada.



Qual das três medidas de tendência central utilizar? A escolha da medida depende da aplicação.

- Quando procuramos conhecer valores totais, será utilizada a média. Por exemplo, em controle de qualidade, a média é utilizada para determinar se o processo opera ao redor de um valor esperado ou alvo. Dá-se preferência à média pelas suas propriedades matemáticas.
- Se a amostra apresentar valores extremos, uma distribuição com acentuada inclinação, a mediana será mais adequada, pois não é afetada pelos dados extremos, como a média. Se quisermos conhecer o valor típico dos salários de uma determinada categoria de trabalhadores, será utilizada a mediana. Por exemplo, se os salários pesquisados da categoria são $500, $1.800, $2.000, $2.200 e $2.500, a mediana é $2.000 e a média $1.800. Portanto, o valor da média tende na direção dos valores extremos e a mediana não é afetada por esses valores extremos.
- A moda é um valor típico de uma amostra ou variável. Por exemplo, na distribuição do consumo de um mesmo produto com diferentes apresentações, a moda mostra a apresentação mais consumida, como é o caso do número de calçados, o tamanho de calças etc.

Média ponderada

O cálculo da média de uma amostra é realizado com todos os dados da amostra. Todos os dados recebem a mesma importância ou o mesmo peso; eles têm uma distribuição uniforme e discreta. Contudo, os valores repetidos poderiam ser agrupados como mostra o cálculo da média do Exemplo 3.15 que repetimos.

$$\overline{X} = \frac{1}{26} \times (2\times 11 + 5\times 12 + 6\times 13 + 7\times 14 + 3\times 15 + 2\times 16 + 1\times 17)$$

Realizando a operação indicada nessa expressão:

$$\overline{X} = \frac{2}{26}\times 11 + \frac{5}{26}\times 12 + \frac{6}{26}\times 13 + \frac{7}{26}\times 14 + \frac{3}{26}\times 15 + \frac{2}{26}\times 16 + \frac{1}{26}\times 17$$

$$\overline{X} = 0{,}0769\times 11 + 0{,}1923\times 12 + 0{,}2308\times 13 + 0{,}2692\times 14 + \cdots + 0{,}0385\times 17$$

$$\overline{X} = 13{,}54$$

O agrupamento dos dados repetidos formam a *média ponderada*, que é a distribuição de frequências relativas de X, veja Exemplo 2.4 do Capítulo 2.

A *média ponderada* $\overline{X}$ da amostra ou variável X é obtida com:

$$\overline{X} = \frac{\sum_{i=1}^{n} w_i \times X_i}{\sum_{i=1}^{n} w_i}$$

Nessa expressão, X_i é o dado repetido e w_i seu peso ou frequência.

Algumas conclusões importantes:

- O cálculo da média ponderada é um caso particular do cálculo da média aritmética.
- Os pesos formam a distribuição de frequências relativas da variável.
- No cálculo da média aritmética, a quantidade de dados da variável é conhecida; entretanto, no caso da média ponderada, a quantidade de valores da variável não é explícita.
- Uma vantagem do procedimento da média ponderada é poder definir os pesos de cada dado numa previsão, lembrando que a soma dos pesos deve ser sempre igual a um ou 100%.

EXEMPLO 3.18

O capital da empresa foi captado de três fontes, ações, financiamentos de longo prazo e debêntures, cada um com seu próprio custo definido por uma taxa anual de juros. O objetivo é calcular o custo médio ponderado do capital captado pela empresa, considerando as informações na tabela a seguir:

Capital da empresa	Participação	Taxa de juros
Acionistas	\$1.000.000	12%
Financiamentos	\$600.000	8%
Debêntures	\$400.000	14%

Solução. O capital da empresa é \$2.000.000, obtido como resultado da soma dos três capitais. O custo médio anual *CM* do capital da empresa é 11,20%.

$$CM = \frac{\sum_{i=1}^{3} w_i \times X_i}{\sum_{i=1}^{3} w_i}$$

$$CM = \frac{1.000.000 \times 12\% + 600.000 \times 8\% + 400.000 \times 14\%}{1.000.000 + 600.000 + 400.000} = 11{,}20\%$$

Na planilha **Exemplo 3.18**, incluída na pasta **Capítulo 3,** são apresentadas outras formas de cálculo utilizando diversos recursos do Excel.

	A	B	C	D	E
1	Exemplo 3.18				
2					
3		Pesos	Taxas		
4		50%	12%		
5		30%	8%		
6		20%	14%		
7					
8		*CM*	11,20%		
9					
10		Utilizando a função SOMARPRODUTO			
11		*CM*	11,20%		
12					
13		Utilizando a função SOMA registrada como matriz			
14		*CM*	11,20%		
15					
16		Utilizando o produto de matrizes			
17		Taxas / Pesos	50%	30%	20%
18		12%			
19		8%			
20		14%		*CM*	11,20%
21					

O resultado do custo médio de capital *CM* foi obtido da seguinte forma:

- Na célula C8, foi registrada a fórmula =B4*C4+B5*C5+B6*C6
- Na célula C11, foi registrada =SOMARPRODUTO(B4:B6;C4:C6)[6].
- Na célula C14, foi registrada a fórmula =SOMA(B4:B6*C4:C6), inserida como matriz.
- A fórmula =MATRIZ.MULT(C17:E17;B18:B20)[7] foi registrada na célula E20.

Problemas

Problema 1

Determine a quantidade de valores e os valores mínimo e máximo da amostra:

5	7	3	4	2	8	9	12

R: n=8, *Mínimo*=2 e *Máximo*=12

6 Em inglês, a função SOMARPRODUTO é *SUMPRODUCT*.
7 Em inglês, a função MATRIZ.MULT é *MMULT*.

Problema 2
Continuando com o Problema 1, determine a ordem e o percentil do valor 7.
R: Ordem=5 e *Percentil*=57,1%

Problema 3
Continuando com o Problema 1, qual o valor da amostra com percentil 85,7%?
R: Valor=9

Problema 4
Continuando com o Exemplo 3.2, determine o percentil das observações cujas ordens são x=1, 4, 10 e 22.
R: p=0%, 12% e 84%.

Problema 5
Continuando com o Problema 4, qual o valor da amostra com p=32%?
R: x=9.

Problema 6
Repita os Problemas 1, 2 e 3 considerando a amostra a seguir: você escolhe o valor do segundo.

15	16	12	18	22	21	17	16
12	16	18	21	19	18	16	

Problema 7
Continuando com o Problema 6, quais os valores do primeiro quartil, do segundo quartil e do terceiro quartil?
R: Q_1=16; Q_2=17 e Q_3=18,50

Problema 8
Calcule os *quartis* da amostra registrada na próxima tabela.

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

R: Q_1=12 Q_2=15 e Q_3=18

Problema 9
Continuando com o Problema 8, qual o percentil do valor 15?
R: p(15)=50%

Problema 10
Continuando com o Problema 8, qual o percentil dos valores 10 e 21?
R: p(10)=0% e p(21)=83%

Problema 11
Continuando com o Problema 8, qual o valor com percentil 35% e 63%?
R: X(p=35%)=14 e X(p=63%)=15

Problema 12
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. Os *quartis* da amostra.
b. Quais os percentis dos valores 8 e 11?
c. Quais os valores com percentis 40% e 75%?

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: a) Q_1=7,75 Q_2=9 e Q_3=9,75 *b)* $p(8)$=26% e $p(11)$=82%
c) (p=40%)=8 e $X(p$=75%)=10

Problema 13

A tabela a seguir registra os retornos das aplicações mais tradicionais do mercado financeiro. Calcule a *ordem* e o *percentil* de cada retorno.

Aplicação	Retorno mensal %
Ouro	–1,74%
Inflação	0,10%
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%

Problema 14

Continuando com o Problema 13. No mesmo mês, o retorno do produto financeiro *FourA* foi 1,85% ao mês. Qual o percentil do retorno 1,85%? Explique o significado desse percentil.

R: O produto *FourA* tem percentil p=83,3%. O retorno desse produto é maior do que os 83,3% primeiros retornos da tabela, e menor do que os 16,7% restantes.

Problema 15

Continuando com o Problema 13. Para que o gerente de produtos do *Banco* possa afirmar que o retorno de fundo *TREAL* é maior do que os 75% primeiros produtos da tabela, qual deve ser o retorno desse produto?

R: 1,54% ao mês

Problema 16

A tabela seguinte registra o salário bruto mensal dos operadores de oito empresas do mesmo ramo. Qual o percentil e o significado do salário $1.050?

$1.250	$980	$1.050	$1.165	$1.175	$1.220	$1.100	$1.050

R: p=14,0%

Problema 17

Continuando com o Problema 16. Quando Carlos reivindicou aumento de salário o chefe afirmou que nada podia fazer, pois seu salário está entre o segundo e o terceiro quartis de sua categoria. Qual deve ser o salário de Carlos?

R: O salário de Carlos está no intervalo de $1.132,50 ($Q_2$) até $1.186,30 ($Q_3$).

Problema 18

Calcular a média da variável do Exemplo 3.2 considerada como população.

R: μ=42,11

Problema 19

Calcule a média, a moda e a mediana da amostra registrada na tabela seguinte.

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

R: $\overline{X}$ =15,62 *Mo*=15 e *Md*=15

Problema 20

Calcule a média, a moda e a mediana das notas finais da *Turma C* da disciplina Estatística registradas na tabela a seguir.

89,5	74,7	99,4	84,9	96,5	82,1	77,7	92,7	59,1	74,7	91,0	100,0	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: $\overline{X}$=78,1 *Mo*=74,7 e *Md*=83,7

Problema 21

Calcule a média, a moda e a mediana da série de dados do Problema 13.

R: $\overline{X}$=1,05% *Mo*=Não tem e *Md*=1,16%

Problema 22

Calcule a média, a moda e a mediana dos dados da relação das 50 maiores empresas listadas na pasta **Capítulo 1**.

R: $\overline{X}$ =$2.550,5 *Mo*=Não tem e *Md*=$2.119,7

Problema 23

A tabela a seguir registra o lucro bruto em $milhares no primeiro trimestre do ano dos vinte maiores hotéis. Calcular a média, a moda e a mediana do lucro.

619,7	475,5	356,5	338,5	336	310,5	258	223	209,7	198,4
190,5	189,3	176,9	162,4	155,5	155,5	149	143	141,9	136,6

R: $\overline{X}$=$246,3 *Mo*=$155,5 e *Md*=$194,5

Problema 24

Continuando com o Problema 23, calcule os três *quartis*.

R: *Q1*=$155,5 *Q2*=$194,5 *Q3*=$316,9

Problema 25

Com os resultados do Problema 23, explique a forma da distribuição do lucro bruto dos vinte maiores hotéis.

R: Como os resultados do Problema 23 verificam a condição $\mu > Md$, a distribuição do lucro tem inclinação positiva.

Problema 26

O hotel TRI não participa do grupo de hotéis do Problema 23. Se no mesmo período o lucro bruto foi igual a $190, determine o percentil do lucro dessa empresa e explique o significado desse valor.

R: O lucro da empresa TRI tem percentil *p*=45%; portanto, o lucro da empresa é maior do que as 45% primeiras empresas listadas em ordem crescente de lucro, e menor do que as 55% demais empresas listadas.

Problema 27

Continuando com o Problema 23. Para que seja possível afirmar que o lucro bruto de um hotel foi maior do que o lucro das 60% primeiras empresas listadas, qual deverá ser o lucro desse hotel?

R: Lucro=$215,1 milhares

Problema 28

Mensalmente a empresa fabrica 40 lotes de 100.000 parafusos cada um. Ao escolher uma amostra aleatória de oito lotes, o controle de qualidade verificou o seguinte número de parafusos com defeito em cada lote:

Amostra	1	2	3	4	5	6	7	8
# Defeitos	300	550	480	980	1.050	350	450	870

Estime o número de parafusos com defeito em um mês de trabalho.

R: A média de defeitos é 628,75 parafusos por lote, isto é, 0,62875% de cada lote de 100.000 parafusos. Como durante um mês de produção serão produzidos 4.000.000 de parafusos, a projeção mensal do número de parafusos com defeitos será igual a 25.150 por mês.

Problema 29

A revista de negócios de maior circulação informou que os salários anuais de seus leitores têm média de $2.200.000 e mediana $800.000.

a. Desenhe a distribuição de frequências dos salários anuais dos leitores.
b. Explique a forma dessa curva.

Problema 30

Na empresa de contabilidade trabalham sete funcionários e o dono da empresa. No ano passado, o rendimento anual dos dois contadores *seniores* foi de $60.000 cada um e dos cinco contadores *juniores* foi de $25.000 cada um. Se o rendimento anual do dono da empresa de contabilidade foi $255.000:

a. Calcule a média, a moda e a mediana dos rendimentos anuais.
b. Desenhe a curva da distribuição das frequências dos rendimentos anuais e explique sua forma.

R: Média=R$62.500 e Mediana=R$25.000

Problema 31

As duas tabelas seguintes registram a remuneração total dos executivos das empresas brasileiras incluindo o salário fixo, a remuneração variável e os seguintes benefícios quantificados: assistência médica, assistência odontológica, automóvel, previdência privada e alimentação.[8]

	Empresas com faturamento mensal acima de $100 milhões					
	Presidente	Dir. financeiro	Dir. comercial	Dir. industrial	Dir. de RH	Dir. marketing
Primeiro quartil	$30.911	$18.973	$14.750	$15.084	$13.944	$12.703
Mediana	$37.328	$20.521	$17.974	$19.991	$15.235	$18.026
Terceiro quartil	$40.538	$21.663	$20.116	$20.638	$19.118	$18.582

	Empresas com faturamento mensal entre $25 e $100 milhões					
	Presidente	Dir. financeiro	Dir. comercial	Dir. industrial	Dir. de RH	Dir. marketing
Primeiro quartil	$25.998	$13.305	$12.746	$13.523	- -	$11.250
Mediana	$29.654	$15.225	$14.762	$13.940	- -	$12.765
Terceiro quartil	$31.282	$18.026	$15.801	$15.902	- -	$16.579

Analise os resultados registrados acima e responda às seguintes perguntas:

a. Que percentagem dos entrevistados de cada categoria pesquisada se encontram entre o primeiro e o terceiro quartis?
b. Por que o intervalo entre a mediana e o primeiro quartil de remuneração da categoria Presidente é diferente do intervalo entre o terceiro quartil e a mediana? Explique essa diferença.
c. Repita a comparação anterior com as outras categorias.
d. Apresente os resultados das empresas com faturamento mensal acima de $100 milhões em um gráfico e analise sua forma.

Problema 32

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 a seguir 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

Pede-se realizar uma análise dos dados, considerando que a experiência no gerenciamento desse tipo de negócio mostra que o ponto de equilíbrio de uma loja é de 250 refeições por dia.

8 Revista *Exame* de 15/12/1999 de uma pesquisa da consultoria Arthur Andersen.

Apêndice 1

Funções de procura e ordenamento do Excel

O cálculo das medidas de ordenamento utilizando o Excel pode ser realizado utilizando expressões matemáticas e procedimentos combinados com os recursos da planilha, as funções estatísticas e a ferramenta de análise *Ordem e Percentil* do Excel. Na planilha **Funções de Ordenamento**, incluída na pasta **Capítulo 3**, está registrada a utilização de cada função utilizando a amostra do Exemplo 3.1, como se pode ver na Figura 3.6. As sintaxes dessas funções estatísticas são apresentadas a seguir.

CORRESP(*valor; matriz; tipo*)

A função de procura e referência CORRESP[9] retorna a posição relativa do argumento *valor* especificado no argumento *matriz* de valores em uma ordem específica. A procura é realizada conforme o argumento *tipo*:

- Se *tipo*=1, então a função CORRESP selecionará o menor valor da *matriz* que for maior ou igual ao *valor* em uma *matriz* previamente ordenada de forma decrescente.
- Se *tipo*=0, então a função CORRESP selecionará o primeiro valor da *matriz* que for exatamente igual ao *valor* especificado, sem necessidade de a *matriz* estar ordenada.
- Se *tipo*=1, então a função CORRESP selecionará o maior valor da *matriz* que for menor ou igual ao *valor* especificado, em uma *matriz* previamente ordenada de forma crescente.

A função CORRESP é parecida com as funções PROCV e PROCH com a diferença de retornar a posição de um valor em um intervalo em vez do valor propriamente dito. O argumento *matriz* pode ser informado como um intervalo de células no qual foi registrada previamente a amostra, por exemplo, o intervalo B4:B14 da Figura 3.6; ou pode ser informado declarando todos os valores da amostra {31;38;19;27;24;42;32;18;43;15;39}.

ORDEM(*valor; amostra; ordem*)

A função estatística ORDEM[10] retorna a posição do argumento *valor* da *amostra*, considerando a *ordem* informada:

- Se *ordem* for igual a 0 ou omitida, os valores da amostra serão classificados em ordem decrescente.
- Se *ordem* for diferente de 0, igual a 1, os valores da amostra serão classificados em ordem crescente.

Se o argumento *amostra* tiver valores repetidos, a função ORDEM informará a posição do primeiro valor que encontrar na sua procura, considerando o ordenamento escolhido.

9 Em inglês, a função CORRESP é *MATCH*.
10 Em inglês, a função ORDEM é *RANK*.

ORDEM.PORCENTUAL(*matriz; valor; núm_ decimais*)

A função estatística ORDEM.PORCENTUAL[11] retorna o *percentil* do argumento *valor*, considerando a *matriz* ordenada de forma crescente. Se a matriz tiver valores repetidos, a função informará o percentil do primeiro valor que encontrar. O argumento *núm_decimais* define o número de casas decimais do resultado; se for omitido, o resultado terá três casas decimais. O argumento matriz pode ser informado em qualquer ordem, pois a função ORDEM.PORCENTUAL ordena os valores da amostra de forma crescente antes de calcular. O argumento *matriz* pode ser informado como um intervalo de células onde previamente foi registrada a amostra, por exemplo, o intervalo B4:B14 da Figura 3.6; ou pode ser informado declarando todos os valores da amostra {31;38;19;27;24;42;32;18;43;15;39}.

PERCENTIL(*matriz; k*)

A função estatística PERCENTIL[12] retorna o valor que divide a *matriz* em duas partes, uma menor do que o argumento *k* e a outra maior do que *k*. O argumento *k* é um valor entre 0 e 1,0% e 100%, ou o valor do percentil em que a matriz ordenada será dividida. A função PERCENTIL ordena os valores da *matriz* de forma crescente antes de calcular. Nem sempre o resultado da função percentil é um valor da amostra. O argumento *matriz* pode ser informado como um intervalo de células no qual previamente foi registrada a amostra, por exemplo, o intervalo B4:B14 da Figura 3.6; ou pode ser informado como {31;38;19;27;24;42;32;18;43;15;39}, declarando todos os valores da amostra.

QUARTIL(*matriz; quarto*)

A função estatística QUARTIL[13] retorna o valor da *matriz* correspondente ao argumento *quarto* identificado da seguinte maneira:

- Se *quarto*=0, a função retornará o primeiro ou menor valor da matriz.
- Se *quarto*=1, 2 ou 3, a função retornará o valor da matriz correspondente e, respectivamente, ao primeiro, segundo ou terceiro *quartil*.
- Se *quarto*=4, a função retornará o último ou maior valor da matriz.

A função QUARTIL ordena os valores da *matriz* de forma crescente antes de calcular. Enquanto a função QUARTIL fornece resultados de posições definidas na amostra ordenada, a função PERCENTIL retorna os resultados para qualquer posição de 0 a 1, ou 0% a 100%. No entanto, nem sempre o retorno da função QUARTIL é um dado da amostra. O argumento *matriz* pode ser informado como um intervalo de células no qual previamente foi registrada a amostra, por exemplo, o intervalo B4:B14 da Figura 3.6; ou pode ser informado declarando todos os valores da amostra {31;38;19;27;24;42;32;18;43;15;39}.

MENOR(*matriz; k-ésimo*)

A função estatística MENOR[14] retorna o *k-ésimo* menor valor da *matriz* ordenada de forma crescente. Para uma mesma *matriz*, o resultado dessa função dependerá do valor do argumento *k-ésimo*:

- Se *k-ésimo*=1, então o menor valor será o primeiro valor da *matriz* ordenada de forma crescente.
- Se *k-ésimo*=2, então o menor valor será o segundo valor da *matriz* ordenada de forma crescente e assim sucessivamente, até o último valor da *matriz*.

11 Em inglês, a função ORDEM.PORCENTUAL é *PERCENTRANK*.
12 Em inglês, a função PERCENTIL é *PERCENTILE*.
13 Em inglês, a função QUARTIL é *QUARTILE*.
14 Em inglês, a função MENOR é *SMALL*.

	A	B	C	D	E
1	**Funções de ordenamento**				
2					
3		**Amostra**		**CORRESP(*valor; matriz; tipo*)**	
4		31		***valor***	**18**
5		38		***tipo***	**0**
6		19		**CORRESP**	**8**
7		27		**Como matriz**	
8		24		**CORRESP**	**8**
9		42			
10		32			
11		18		**ORDEM(*valor; amostra; ordem*)**	
12		43		***valor***	**38**
13		15		***tipo***	**1**
14		39		**ORDEM**	**8**
15				**Como matriz**	
16				**ORDEM**	**Não se aplica**
17					
18					
19				**ORDEM.PORCENTUAL(*matriz; valor; núm_ decimais*)**	
20				***valor***	**38**
21				***núm_ decimais***	**1**
22				**ORDEM.PORCENTUAL**	**70,00%**
23				**Como matriz**	
24				**ORDEM.PORCENTUAL**	**70,00%**
25					

FIGURA 3.6 Como utilizar as funções de ordenamento.

Na função MENOR, não é necessário informar a série ordenada de forma crescente. O argumento *matriz* pode ser informado como um intervalo de células no qual previamente foi registrada a amostra, por exemplo, o intervalo B4:B14 da Figura 3.6; ou pode ser informado declarando todos os valores da amostra {31;38;19;27;24;42;32;18;43;15;39}.

MAIOR(*matriz; k-ésimo*)

A função estatística MAIOR[15] dá o *k-ésimo* maior valor da *matriz* ordenada de forma crescente. Para uma mesma *matriz*, o resultado dessa função dependerá do valor do argumento *k-ésimo*:

- Se *k-ésimo*=1, então o maior valor da *matriz* será o último valor da *matriz* ordenada de forma crescente.
- Se *k-ésimo*=2, então o maior valor da *matriz* será o penúltimo valor da *matriz* e assim sucessivamente, até o primeiro valor da *matriz*.

O argumento *matriz* pode ser informado como um intervalo de células no qual previamente foi registrada a amostra, por exemplo, o intervalo B4:B14 da Figura 3.6; ou pode ser informado declarando todos os valores da amostra {31;38;19;27;24;42;32;18;43;15;39}. Na função MAIOR, não é necessário informar a série ordenada de forma crescente.

15 Em inglês, a função MAIOR é *LARGE*.

Apêndice 2

O símbolo somatório

Suponha uma amostra ou variável X com n dados ou observações identificados pela sequência de valores $X_1, X_2, \cdots, X_i, \cdots, X_n$, onde X_1 é o primeiro dado, X_2 é o segundo dado, X_i é um dado qualquer da amostra, e assim sucessivamente, até o último dado X_n.

A soma desses valores representada com $X_1 + X_2 + \cdots + X_i + \cdots + X_n$ se pode expressar simbolicamente com $\sum_{i=1}^{n} X_i$, pois $\sum_{i=1}^{n} X_i = X_1 + X_2 + \cdots + X_i + \cdots + X_n$. A letra grega Σ, sigma maiúscula, indica que devem ser somadas expressões da forma X_i começando com i=1 até i=n.

Outro exemplo: a expressão simbólica da soma $R = 2^1 + 2^2 + 2^3 + 2^4$ é $R = \sum_{i=1}^{4} X^i = 2^1 + 2^2 + 2^3 + 2^4$.

Vejamos algumas propriedades de interesse, tendo presente que as propriedades se aplicam sempre nos dois sentidos da igualdade.

- O resultado de somar n vezes a constante c é o resultado do produto de n vezes a constante c. Com o símbolo somatório $\sum_{i=1}^{n} c = n \times c$.
- Se cada valor da sequência $X_1, X_2, \cdots, X_i, \cdots, X_n$ for multiplicado pela constante c, o resultado dessa soma será $\sum_{i=1}^{n} cX_i = c \times \sum_{i=1}^{n} X_i$.
- A soma algébrica das sequências $X_1, X_2, \cdots, X_i, \cdots, X_n$ e $Y_1, Y_2, \cdots, Y_i, \cdots, Y_n$ é $\sum_{i=1}^{n} (X_i \pm Y_i) = \sum_{i=1}^{n} X_i \pm \sum_{i=1}^{n} Y_i$. Há casos em que as propriedades anteriores do somatório são combinadas $\sum_{i=1}^{n} (cX_i + Y^{2i}) = \sum_{i=1}^{n} cX_i + \sum_{i=1}^{n} Y^{2i} = c\sum_{i=1}^{n} X_i + \sum_{i=1}^{n} Y^{2i}$.
- Somatórios múltiplos. A seguinte expressão é formada por três somatórios.

$$\sum_{i=1}^{3}\sum_{j=1}^{3} X_{i,j} = \sum_{i=1}^{3} X_{i,1} + \sum_{i=1}^{3} X_{i,2} + \sum_{i=1}^{3} X_{i,3}$$

Essa expressão desenvolvida é:

$$\sum_{i=1}^{3}\sum_{j=1}^{3} X_{i,j} = X_{1,1} + X_{1,2} + X_{1,3} + X_{2,1} + X_{2,2} + X_{2,3} + X_{3,1} + X_{3,2} + X_{3,3}$$

Essas expressões representam a soma dos dados da seguinte tabela, onde i representa a linha e j a coluna.

i/j	Coluna 1	Coluna 2	Coluna 3
Linha 1	$X_{1,1}$	$X_{1,2}$	$X_{1,3}$
Linha 2	$X_{2,1}$	$X_{2,2}$	$X_{2,3}$
Linha 3	$X_{3,1}$	$X_{3,2}$	$X_{3,3}$

Apêndice 3

Prova do mínimo da soma dos quadrados dos desvios

Denominando o desvio como D e z a qualquer número possível de ser a média da amostra X, a soma dos quadrados dos desvios será medida com a expressão $D = \sum_{i=1}^{n} (X_i - z)^2$. Para calcular o mínimo dessa função, primeiro deve-se calcular a primeira derivada da função D.

$$D = \sum_{i=1}^{n} (X_i^2 - 2zX_i + z^2)$$

$$\frac{dD}{dz} = \frac{d}{dz}\sum_{i=1}^{n} X_i^2 - \frac{d}{dz}\sum_{i=1}^{n} 2zX_i + \frac{d}{dz}\sum_{i=1}^{n} z^2$$

Depois, a primeira derivada deve ser igualada a zero.

$$\frac{dD}{dz} = 0 - 2\sum_{i=1}^{n} X_i + 2\sum_{i=1}^{n} z = 0$$

Na última expressão simplificada $-\sum_{i=1}^{n} X_i + nz = 0$, reconhecemos que a segunda parcela é a soma dos dados da amostra. O valor de z é o próprio valor da amostra de X já definido como $z = \frac{\sum_{i=1}^{n} X_i}{n}$. O valor encontrado é realmente um mínimo, pois sua segunda derivada é positiva, como mostrado a seguir:

$$\frac{d^2 D}{dz^2} = \frac{d^2}{dz^2}\left(-2\sum_{i=1}^{n} X_i + 2nz\right)$$

$$\frac{d^2 D}{dz^2} = 2n > 0$$

Apêndice 4

Funções de tendência central do Excel

O cálculo das medidas de tendência central utilizando o Excel pode ser realizado utilizando expressões matemáticas e procedimentos combinados com os recursos da planilha e funções estatísticas. Na planilha **Funções de Tendência Central**, incluída na pasta **Capítulo 3**, está registrada a utilização de cada função utilizando a amostra do Exemplo 3.15, como se pode ver na Figura 3.7. Uma característica comum das funções a seguir, exceto a função MÉDIA.INTERNA, são os 30 argumentos (*núm1; núm2; ... ; núm30*) utilizados para registrar os valores de intervalos. Na apresentação da primeira função SOMA, 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.

SOMA(*núm1; núm2; ... ; núm30*)

A função matemática SOMA[16] retorna a *soma* dos valores numéricos *núm1; núm2; ... ; núm30*. Cada um desses *núm* pode ser um intervalo de células de uma planilha contendo valores numéricos ou assemelhados.[17] Por exemplo, a função SOMA aplicada aos valores da amostra do Exemplo 3.15 dá como resultado 352. Para obter esse resultado, a função SOMA pode ser utilizada das seguintes maneiras, Figura 3.7:

- 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 F6 foi registrada a fórmula =SOMA(B4:C17), Figura 3.7.
  - 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, a fórmula =SOMA(B4:C8;B9:B17;C9:C15) registrada na célula F7 tem três intervalos nos três primeiros argumentos da função SOMA *núm1; núm2; núm3*.
- 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 G6, os valores foram registrados em uma única matriz:
    =SOMA({14;12;13;11;12;13;16;14;14;15;17;14;11;13;14;15;
    13;12;14;13;14;13;15;16;12;12})
  - Na célula G7, os valores foram registrados em quatro matrizes:
    =SOMA({14;12;13;11};{12;13;16;14;14;15;17;14;11;13};
    {14;15;13;12;14;13;14;13;15};{16;12;12})
    correspondentes aos quatro primeiros argumentos da função SOMA *núm1; núm2; núm3; núm4*.

16 Em inglês, a função SOMA é *SUM*.
17 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.

MÉDIA(*núm1; núm2; ... ; núm30*)

A função estatística MÉDIA[18] retorna a *média aritmética* dos valores numéricos *núm1; núm2; ... ; núm30*. Cada um desses *núm* pode ser um intervalo de células de uma planilha contendo valores numéricos ou assemelhados. Um detalhe importante: se o nome da função MÉDIA for inserido com letras minúsculas ou maiúsculas sem o acento ortográfico, o Excel aceitará e registrará a função com letras maiúsculas e com o acento ortográfico. A função MÉDIA pode ser registrada de diversas formas equivalentes às descritas na função SOMA menciona anteriormente, Figura 3.7.

MÉDIAA(*núm1; núm2; ... ; núm30)*

A função estatística MEDIAA[19] é equivalente à função anterior MÉDIA. 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. Deixamos para o leitor pesquisar na Ajuda do Excel.

	A	B	C	D	E	F	G
1	**Funções de tendência central**						
2							
3		**Amostra**				**Dados informados como**	
4		14	14		**Função Matemática**	**Intervalo**	**Matriz**
5		12	15				
6		13	13		**SOMA**	**352,00**	**352,00**
7		11	12			**352,00**	**352,00**
8		12	14				
9		13	13		**Funções Estatísticas**		
10		16	14				
11		14	13		**MÉDIA**	**13,54**	**13,54**
12		14	15			**13,54**	**13,54**
13		15	16				
14		17	12		**MED**	**13,50**	**13,50**
15		14	12			**13,50**	**13,50**
16		11					
17		13			**MODO**	**14,00**	**14,00**
18						**14,00**	**14,00**
19							

FIGURA 3.7 Como utilizar as funções de tendência central.

MED(*núm1; núm2; ... ; núm30*)

A função estatística MED[20] retorna a *mediana* dos valores numéricos *núm1; núm2; ... ; núm30*. Cada um dos *núm* pode ser um intervalo de células de uma planilha contendo valores numéricos ou assemelhados. A função MED pode ser registrada de diversas formas equivalentes às descritas na função SOMA anteriormente, Figura 3.7.

MODO(*núm1; núm2; ... ; núm30*)

A função estatística MODO[21] retorna o *modo* dos valores numéricos *núm1; núm2; ... ; núm30*. Cada um desses *núm* pode ser um intervalo de células de uma planilha que contém valores numéricos ou *asseme-*

18 Em inglês, a função MÉDIA é *AVERAGE*.
19 Em inglês, a função MEDIAA é *AVERAGEA*.
20 Em inglês, a função MED é *MEDIAN*.
21 Em inglês, a função MODO é *MODE*.

lhados. Quando a série tem mais de uma moda, a função reconhece apenas uma delas. A função MOD pode ser registrada de diversas formas equivalentes às descritas na função SOMA anteriormente, Figura 3.7.

MÉDIA.GEOMÉTRICA(*núm1; núm2; ... ; núm30*)

A função estatística MÉDIA.GEOMÉTRICA[22] retorna a *média geométrica* dos valores da amostra. Cada um dos *núm* pode ser um intervalo de células de uma planilha que contém valores numéricos ou *assemelhados*. A média geométrica *Mg* é definida como $Mg = (X_1 \times X_2 \times \cdots \times X_n)^{1/n}$ com os valores X_i maiores do que zero. Comparando com a *média aritmética*:

- A média geométrica é menos afetada por valores extremos.
- A média geométrica é uma medida mais central quando os valores da variável apresentam uma taxa constante de crescimento.
- Para um mesmo grupo de valores, a média geométrica é sempre menor do que a média aritmética.

A função MÉDIA.GEOMÉTRICA pode ser registrada de diversas formas equivalentes às descritas na função SOMA anteriormente, Figura 3.7. Uma aplicação frequente da *média geométrica* é o cálculo da taxa equivalente de juros de uma operação financeira formada por *n* operações com taxas de juros diferentes, como mostrado no Capítulo 16, utilizando a fórmula:

$$Mg = \left((1+i_1)\times(1+i_2)\times\cdots\times(1+i_n)^{1/n}\right)$$
$$i = Mg - 1$$

MÉDIA.HARMÔNICA(*núm1; núm2; ... ; núm30*)

A função estatística MÉDIA.HARMÔNICA[23] retorna a *média harmônica* dos valores da amostra. Cada um dos *núm* pode ser um intervalo de células de uma planilha que contém valores numéricos ou *assemelhados*. A *média harmônica* é uma medida útil quando os valores se referem a mudanças de uma magnitude, e seu valor é sempre menor do que o da média geométrica do mesmo conjunto de valores.

- A *média harmônica* é a inversa da média aritmética das inversas dos valores da amostra: $Mh = \dfrac{1}{\dfrac{1}{n}\times\displaystyle\sum_{i=1}^{n}\dfrac{1}{X_i}}$.
- De outra maneira, a inversa da *média harmônica Mh* é a média da inversa dos valores da amostra: $\dfrac{1}{Mh} = \dfrac{1}{n}\times\displaystyle\sum_{i=1}^{n}\dfrac{1}{X_i}$.

A função MÉDIA.HARMÔNICA pode ser registrada de diversas formas equivalentes às descritas na função SOMA anteriormente, Figura 3.7.

MÉDIA.INTERNA(*matriz*; *porcentagem*)

A função estatística MÉDIA.INTERNA[24] retorna a *média aritmética* da *matriz* de valores, tendo previamente excluído, de ambos extremos da *matriz*, uma *porcentagem* de valores informada como valor unitário. É uma média reduzida útil para remover dados extremos, *suspeitos*, de uma amostra.

22 Em inglês, MÉDIA.GEOMÉTRICA é *GEOMEAN*.
23 Em inglês, MÉDIA.HARMÔNICA é *HARMEAN*.
24 Em inglês, MÉDIA.INTERNA é *TRIMMEAN*.