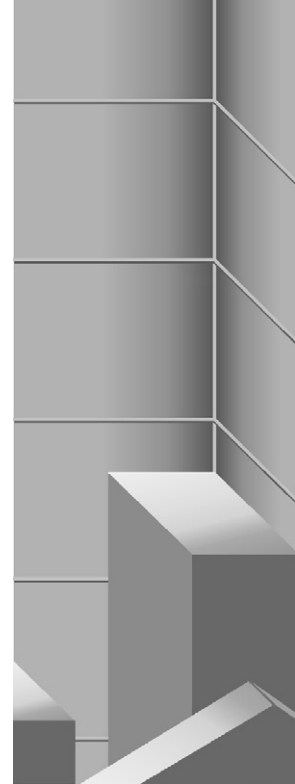


Capítulo 6

CORRELAÇÃO



Até este momento, foram analisados os dados de uma amostra ou variável pertencente a uma população. Outra análise importante é determinar como uma variável se relaciona com outras variáveis da mesma população. Neste capítulo, será mostrada uma forma de medir quanto e de que maneira se relacionam duas variáveis. Há muitos exemplos de relações entre amostras, por exemplo, geralmente os meios de comunicação divulgam informações de variáveis relacionadas obtidas de resultados de pesquisas como:

- Nas Instituições de Ensino Superior – IES há uma relação direta entre a qualidade do ensino e a taxa de inadimplência. A taxa de inadimplência das IES que obtiveram conceitos A e B no Provão é 12,1%, nas que obtiveram C é 16% e nas que obtiveram D e E a inadimplência é de 21,9%.¹
- O frio está para o setor farmacêutico como o Dia das Mães está para o comércio. As vendas de medicamentos não controlados, como analgésicos, antigripais e vitaminas, disparam.²
- O faturamento das empresas de energia nos Estados Unidos é diretamente influenciado pela temperatura, especialmente no inverno. Um inverno brando reduz a demanda de energia para calefação e pode diminuir drasticamente o lucro.³

A partir desses exemplos, você poderá encontrar outras relações como, por exemplo, reduzindo o custo, o preço do produto será reduzido e será possível aumentar a quantidade vendida, ou funcionário com maior escolaridade terá mais chance de crescer na empresa etc. Em qualquer caso, é importante lembrar que a informação recebida nem sempre é corretamente exposta, como no caso da relação direta entre qualidade do ensino das IES e a taxa de inadimplência que, na realidade, mantém uma relação inversa, pois a instituição com maior conceito corresponde à menor taxa de inadimplência.

Os gráficos de dispersão da Figura 6.1 mostram dois tipos de relação entre as variáveis X e Y. O gráfico de dispersão da esquerda mostra uma relação direta ou positiva, tendência destacada pela declividade positiva da elipse tracejada, enquanto o gráfico de dispersão da direita mostra uma relação inversa ou negativa, tendência também destacada pela declividade negativa da elipse tracejada.

¹ A Pressão da inadimplência, artigo de P. de Athayde publicado na revista *Carta Capital* de 15/10/2003.

² Frio chega e venda de remédios dispara, artigo de C. Silva publicado no jornal *O Estado de São Paulo* de 25 de maio de 2004.

³ Alugue o sol artigo de Cláudio Gradilone publicado na revista *Exame* em 6/2/2002. Divulga a proteção financeira com derivativos de clima ou *weather derivatives*.

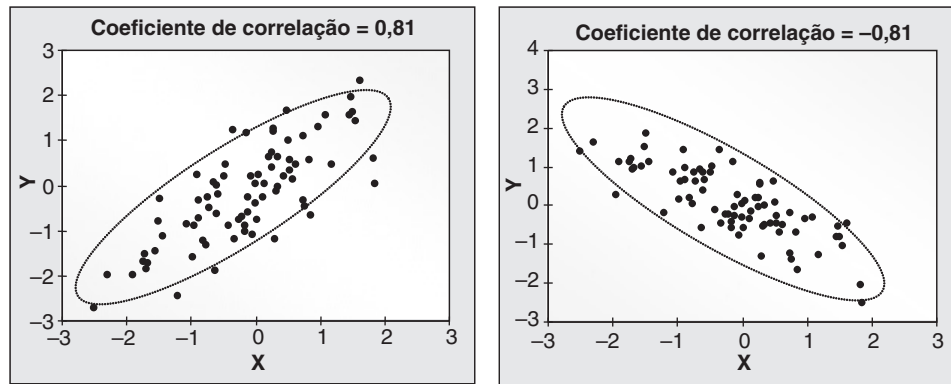
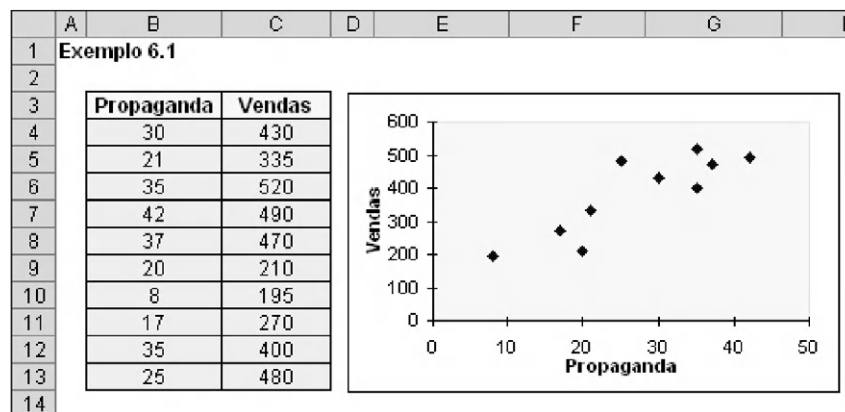


FIGURA 6.1 Dois tipos de relações entre duas variáveis.

EXEMPLO 6.1

O diretor de vendas da rede de varejo nacional está analisando a relação entre o investimento em propaganda e as vendas da empresa utilizando os dados registrados no intervalo B3:C13, incluindo os títulos, da planilha **Exemplo 6.1**, incluída na pasta **Capítulo 6**. Analise a relação entre essas duas amostras.

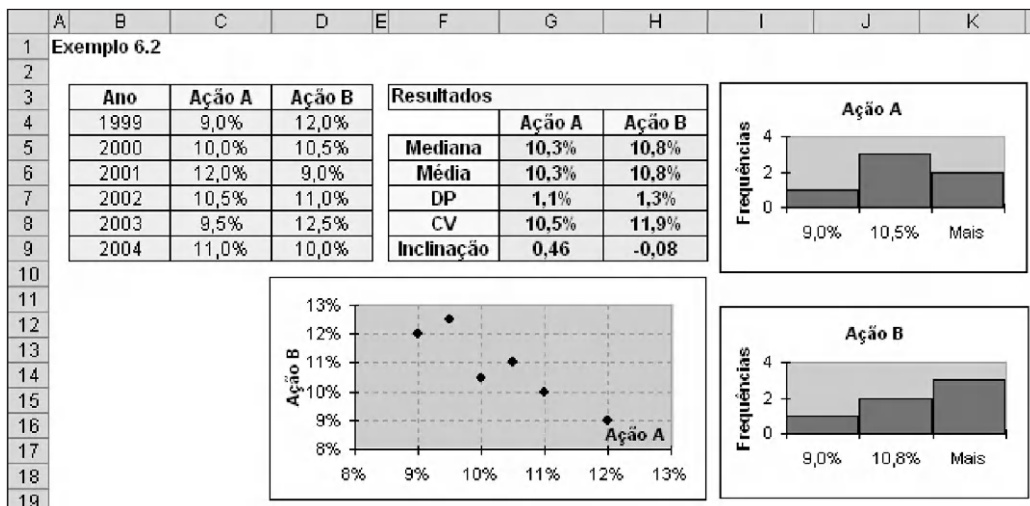
Solução. Com os dados da tabela, foi construído o gráfico de dispersão como, mostra a figura a seguir. Os registros dos dez pares de valores investimento e vendas mostram uma clara relação direta ou positiva, pois à medida que o investimento em propaganda aumenta, as vendas também aumentam, e vice-versa. Esse gráfico mostra que as duas variáveis estão correlacionadas de forma positiva.



EXEMPLO 6.2

Os retornos anuais durante os últimos seis anos da Ação A e da Ação B negociadas na Bolsa de Valores estão registrados na planilha **Exemplo 6.2**, incluída na pasta **Capítulo 6**. Realize uma análise estatística desses retornos e da relação entre eles.

Solução. No intervalo B3:D9 foram registrados os retornos das ações. No intervalo F4:H9 foram calculadas e registradas as medidas estatísticas mediana, média, desvio padrão, coeficiente de variação e coeficiente de inclinação, como mostra a próxima figura.



As medidas estatísticas dos retornos das duas ações são parecidas, exceto o coeficiente de inclinação, que indica formas diferentes das distribuições dos retornos como fica confirmado pelos histogramas construídos com a ferramenta de análise *Histograma*. Os histogramas dos retornos mostram que essas duas amostras têm particularidades que as medidas estatísticas não conseguem capturar. Para ver a diferença entre os retornos das duas ações, na mesma planilha foi construído o gráfico de dispersão que mostra a relação negativa entre os retornos das duas ações. Analisemos o gráfico de dispersão:

- Partindo do ano 1999, retornos 9% e 12%, respectivamente Ação A e Ação B, no ano 2000, o retorno da Ação A aumentou para 10%, enquanto o retorno da Ação B diminuiu para 10,50%. No ano 2001, os retornos mantiveram a mesma tendência do ano 2000.
- Nos anos 2002 e 2003, os retornos inverteram a tendência anterior. Enquanto o retorno da Ação A diminuiu, o retorno da Ação B aumentou.
- No ano 2004, os retornos das ações A e B inverteram novamente a tendência dos dois anos anteriores.

As medidas estatísticas dos retornos anuais das ações A e B do Exemplo 6.2 são parecidas, porém o gráfico de dispersão mostra que os retornos têm tendências opostas. A covariância e o coeficiente de correlação medem a tendência e a força da relação linear entre as duas variáveis ou amostras.

Covariância

O coeficiente de correlação pode ser calculado diretamente com a fórmula do coeficiente de Pearson; entretanto, preferimos iniciar este assunto definindo primeiro a covariância de duas variáveis, apresentação parecida à realizada com o desvio padrão, definindo primeiro a variância.

A covariância σ_{XY} das variáveis $X = X_1, X_2, \dots, X_N$ e $Y = Y_1, Y_2, \dots, Y_N$, consideradas como população é:⁴

$$\sigma_{XY} = \frac{1}{N} \sum_{i=1}^N (X_i - \mu_X) \times (Y_i - \mu_Y)$$

⁴ Quando necessário, as variáveis são separadas com vírgula, $\sigma_{X,Y}$ e $S_{X,Y}$.

A covariância S_{XY} das variáveis $X = X_1, X_2, \dots, X_n$ e $Y = Y_1, Y_2, \dots, Y_n$, consideradas como amostra é:

$$S_{XY} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}) \times (Y_i - \bar{Y})$$

EXEMPLO 6.3

Calcule a covariância das variáveis do Exemplo 6.1.

Solução. Os dados desse exemplo foram registrados na planilha **Exemplo 6.3**, incluída na pasta **Capítulo 6**, como mostra a figura seguinte incluindo os resultados. O primeiro resultado necessário e o valor das médias das duas variáveis, calculadas e registradas nas células D16 e D17. Depois:

- Na célula D4, foi registrada a fórmula =B4-\$D\$16 que calcula o desvio do primeiro dado da variável *Propaganda*. Essa fórmula foi copiada até a célula D13. O mesmo procedimento foi utilizado para calcular e registrar os desvios da variável *Vendas*.
- Na célula F4 foi registrada a fórmula =D4*E4 que retorna o produto dos desvios do primeiro dado. Essa fórmula foi copiada até a célula D13.

	A	B	C	D	E	F
1	Exemplo 6.3					
2						
3		Propaganda	Vendas	X-Méd.X	Y-Méd.Y	(X-Méd.X)*(Y-Méd.Y)
4		30	430	3,00	50,00	150
5		21	335	(6,00)	(45,00)	270
6		35	520	8,00	140,00	1120
7		42	490	15,00	110,00	1650
8		37	470	10,00	90,00	900
9		20	210	(7,00)	(170,00)	1190
10		8	195	(19,00)	(185,00)	3515
11		17	270	(10,00)	(110,00)	1100
12		35	400	8,00	20,00	160
13		25	480	(2,00)	100,00	-200
14						
15		Resultados				
16		Média Propaganda	27,00			
17		Média Vendas	380,00			
18		Soma produtos dos desvios	9.855,00			
19		Covariância população	985,50	=D18/CONT.NÚM(B4:B13)		
20		Covariância amostra	1.095,00	=D18/(CONT.NÚM(B4:B13)-1)		
21						
22		Covariância população	985,50	=COVAR(B4:B13;C4:C13)		
23						

No intervalo de resultados:

- Na célula D18 foi registrada a fórmula =SOMA(F4:F13) que calcula a soma dos produtos dos desvios.
- Com a fórmula =D18/CONT.NÚM(B4:B13), registrada na célula D19, é calculada a covariância da população $\sigma_{XY}=985,50$. E na célula D20 foi registrada a fórmula =D18/(CONT.NÚM(B4:B13)-1) que calcula a covariância da amostra $S_{XY}=1.095,00$.

O resultado da covariância da população também pode ser obtido com a função COVAR do Excel, registrando a fórmula =COVAR(B4:B13;C4:C13) na célula D22. A sintaxe da função COVAR é a seguinte:

• COVAR(matriz1; matriz2)

A função estatística COVAR⁵ retorna a covariância da população dos valores registrados nos argumentos *matriz1* e *matriz2*. Esses argumentos podem ser registrados como intervalos de uma planilha, como já mos-

⁵ Em inglês, COVAR é COVAR.

trado, tomando o cuidado de verificar que as duas variáveis tenham a mesma quantidade de dados. Também é possível registrar os argumentos como *matriz* na própria fórmula da função, evitando registrar os valores da amostra num intervalo de células da planilha como foi feito na célula D29, registrando a fórmula =COVAR({30;21;35;42;37;20;8;17;35;25};{430;335;520;490;470;210;195;270;400;480})

Características da covariância

A covariância mede a tendência e a força da relação linear entre duas variáveis. Das expressões da covariância para população e para amostra temos seguintes características:

- As duas amostras ou variáveis devem ter o mesmo número de dados.
- Os pares de dados ocorrem ao mesmo tempo, são pares casados. Embora possa parecer redundante, tenha presente que não se pode mudar a ordem de uma única variável; a mudança de ordem deverá ser realizada nas duas amostras sem descartar os pares de dados.
- A covariância é a média dos produtos dos desvios das duas amostras ou variáveis, obtida como resultado da divisão:
 - No caso de população, da soma dos produtos dos desvios pela quantidade de dados das variáveis.
 - No caso de amostra, da soma dos produtos dos desvios pela quantidade de dados das variáveis menos um.⁶
- Os numeradores das expressões da covariância para população e para amostra são iguais, o resultado da soma dos produtos dos desvios.

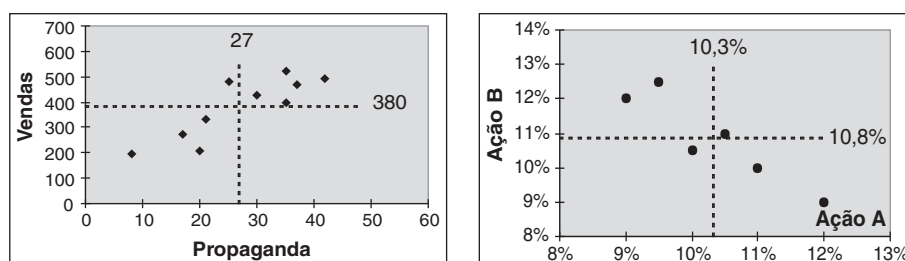


FIGURA 6.2 Análise dos gráficos de dispersão dos Exemplos 6.1 e 6.2.

- A covariância pode assumir qualquer valor do conjunto dos números reais, pois pode ser nula, negativa ou positiva. Baseada na definição dos produtos dos desvios,⁷ uma explicação intuitiva é que a covariância é a medida do afastamento simultâneo das respectivas médias. Se ambas as variáveis aleatórias tendem a estar simultaneamente acima, ou abaixo, de suas respectivas médias, então a covariância tenderá a ser positiva e, nos outros casos, poderá ser negativa, como mostram os gráficos de dispersão da Figura 6.2.
- O gráfico da esquerda mostra que a maioria dos pares de valores do Exemplo 6.1 tem os dois valores acima de sua média correspondente, provocando a covariância positiva, resultado que pode ser confirmado no Exemplo 6.3.
- O gráfico da direita mostra que a maioria dos pares de valores do Exemplo 6.2 tem um valor acima da média e o outro abaixo da média correspondente, provocando a covariância negativa, como poderá ser confirmado no Exemplo 6.4.
- Da mesma forma que a variância, a covariância é afetada pelos valores extremos da variável, ela não é uma medida resistente.

⁶ Equivalente ao caso da variância da amostra S^2 , Capítulo 4.

⁷ Copeland T. – *Opções Reais*, Editora Campus 2001.

- A unidade de medida é o resultado do produto das unidades dos valores das variáveis; no caso do Exemplo 6.1, a unidade é o binômio *vendas e investimento* e, no Exemplo 6.2, o binômio % e %, ambas sem nenhum significado prático.

Regras operacionais da covariância

As propriedades⁸ mais importantes da covariância são:

- Outra forma de calcular a covariância é com a seguinte fórmula para a população que tem a vantagem de não ter de calcular os desvios. No Apêndice 1 você encontra a demonstração desta fórmula:

$$\sigma_{XY} = \frac{1}{N} \sum_{i=1}^N X_i Y_i - \mu_X \mu_Y$$

- O Exemplo 6.3 mostra como calcular a covariância da população e a covariância da amostra, procedimentos que diferem apenas no valor do divisor da soma dos produtos dos desvios. De forma equivalente ao realizado com as variâncias da população e da amostra, a partir das expressões das covariâncias, pode-se estabelecer a seguinte igualdade:

$$\sum_{i=1}^N (X_i - \mu_X) \times (Y_i - \mu_Y) = \sigma_{XY} \times N = S_{XY} \times (n - 1)$$

Portanto, a expressão de equivalência entre as duas covariâncias é:

$$S_{XY} = \frac{N}{n - 1} \sigma_{XY}$$

A covariância da amostra do Exemplo 6.3 pode ser obtida a partir do valor da covariância da população com a fórmula:

$$S_{XY} = \frac{10}{10 - 1} (985,5) = 1.095,00$$

- A covariância de uma variável, e ela mesma, é a própria variância da variável, seja no caso de população ou amostra. Como $Y = X$, então:

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

- A permutação das variáveis não altera o resultado da covariância, se os pares de valores não forem alterados: $\sigma_{XY} = \sigma_{YX}$.
- Há outras propriedades operacionais muito práticas. Por exemplo, representando a covariância como $Cov(X, Y)$ e sendo a , b e c constantes, sempre se verifica:
 - $Cov(X, a) = 0$
 - $Cov(X, -Y) = -Cov(X, Y)$
 - $Cov(aX, Y) = a Cov(X, Y)$

⁸ Estas propriedades aplicadas com a covariância considerada como população também se aplicam com a covariância considerada como amostra.

- [illegible]

Para mostrar a diferença de procedimento entre resultados de população e amostra dos dados:

- No intervalo G5:H5 foram calculados e registrados os desvios padrão, dos retornos considerando-os como população. Depois, no intervalo G6:H6 os desvios padrão considerando os retornos como amostras.
- No intervalo G7:G8 foi calculada a covariância, primeiro, como população utilizando a função COVAR do Excel, e depois como amostra, a partir do primeiro resultado.
- No intervalo G9:G10 foi calculado o coeficiente de correlação, primeiro, utilizando os resultados parciais de população, e depois como amostra, utilizando também os resultados parciais, porém como amostra. Com esse cálculo, verifica-se o mesmo valor de coeficiente de correlação $r=-0,9203$.
- Utilizando os dados como a população, temos $r_{AB}=-0,9203$ da seguinte forma:

$$r_{AB} = \frac{-0,00010694}{0,009860 \times 0,011785} = -0,9203$$

- Utilizando os dados como a amostra, também se obtém $r_{AB}=-0,9203$ da seguinte forma:

$$r_{AB} = \frac{-0,00012833}{0,010801 \times 0,012910} = -0,9203.$$

O coeficiente de correlação pode ser obtido com a função CORREL do Excel registrando na célula G12 a fórmula =CORREL(C4:C9;D4:D9). A sintaxe da função CORREL é a seguinte:

• **CORREL(matriz1; matriz2)**

A função estatística CORREL⁹ retorna o coeficiente de correlação dos valores registrados nos argumentos *matriz1* e *matriz2*. Esses argumentos podem ser intervalos de uma planilha, como mostrado anteriormente, tomando o cuidado de verificar que as duas variáveis tenham a mesma quantidade de dados. Também se podem registrar os argumentos como *matriz* na própria fórmula da função, evitando registrar os valores da amostra em um intervalo de células da planilha, como feito na célula G14, registrando a fórmula =CORREL({0,09;0,1;0,12;0,105;0,095;0,11}; {0,12;0,105;0,09;0,11;0,125;0,1})

O coeficiente de correlação pode também ser obtido com a função PEARSON do Excel como foi obtido com a fórmula = PEARSON(C4:C9;D4:D9) registrada na célula G16. A sintaxe da função PEARSON é a seguinte:

• **PEARSON(matriz1; matriz2)**

A função estatística PEARSON¹⁰ retorna o coeficiente de correlação dos valores registrados nos argumentos *matriz1* e *matriz2*. Esses argumentos podem ser intervalos de uma planilha, como mostrado anteriormente, tomando o cuidado de verificar que as duas variáveis tenham a mesma quantidade de dados. Também se podem registrar os argumentos como *matriz* na própria fórmula da função, evitando registrar os valores da amostra em um intervalo de células da planilha, como feito na célula G18 registrando a fórmula =CORREL({0,09;0,1;0,12;0,105;0,095;0,11}; {0,12;0,105;0,09;0,11;0,125;0,1})

O resultado da função PEARSON é o mesmo que o da função CORREL, porém utilizando os valores das variáveis como mostra a fórmula seguinte.

$$r = \frac{n \times \sum_{i=1}^n X_i \times Y_i - \sum_{i=1}^n X_i \times \sum_{i=1}^n Y_i}{\sqrt{n \times \sum_{i=1}^n X_i^2 - \left(\sum_{i=1}^n X_i\right)^2} \times \sqrt{n \times \sum_{i=1}^n Y_i^2 - \left(\sum_{i=1}^n Y_i\right)^2}}$$

⁹ Em inglês, CORREL é CORREL.

¹⁰ Em inglês, PEARSON é PEARSON.

Características do coeficiente de correlação

Vejamos algumas características importantes do coeficiente de correlação:

- A fórmula do coeficiente de correlação pode ser apresentada sem incluir a covariância, como mostram as fórmulas a seguir, que dão o mesmo resultado do coeficiente de correlação, embora utilizem medidas estatísticas diferentes:

- População: $r_{XY} = \frac{1}{N} \sum_{i=1}^n \left(\frac{X_i - \mu_X}{\sigma_X} \right) \left(\frac{Y_i - \mu_Y}{\sigma_Y} \right)$

- Amostra: $r_{XY} = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{S_X} \right) \left(\frac{Y_i - \bar{Y}}{S_Y} \right)$

Essas duas fórmulas se assemelham à fórmula de Pearson.

- Se a variável Y é a mesma variável X, então o coeficiente de correlação é igual a um, como mostramos a seguir.

$$r_{XX} = \frac{\sigma_{XX}}{\sigma_X \times \sigma_X} = \frac{\sigma_X^2}{\sigma_X^2} = 1$$

- A permutação das variáveis não altera o resultado do coeficiente de correlação, se os mesmos pares de valores forem mantidos $r_{XY} = r_{YX}$.
- Da mesma forma que a covariância, o coeficiente de correlação é afetado pelos valores extremos da variável, ele não é uma medida resistente.
- Se as variáveis X e Y forem estatisticamente independentes, então o coeficiente de correlação dessas variáveis será igual a zero. Entretanto, se o resultado do coeficiente de correlação das variáveis X e Y for igual a zero, não se poderá afirmar que as duas variáveis sejam estatisticamente independentes. Para confirmar essa independência, deve-se verificar se todos os pares de valores das variáveis X e Y cumprem a condição: $P(XeY) = P(X) \times P(Y)$.

Análise dos valores do coeficiente de correlação

Na planilha **Análise**, incluída na pasta **Capítulo 6**, é analisada a tendência e a força da relação linear entre duas variáveis ou amostras X e Y medida pelo coeficiente de correlação.

Variáveis perfeitamente correlacionadas de forma positiva

A Figura 6.3 mostra o comportamento de duas amostras X e Y perfeitamente correlacionadas em sentido positivo. O coeficiente de correlação dessas amostras calculado na célula F3 é igual a $r=+1$.

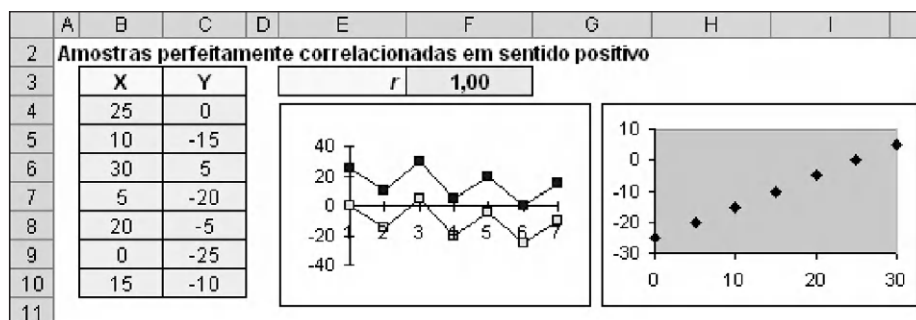


FIGURA 6.3 Amostras perfeitamente correlacionadas em sentido positivo, $r=+1$.

A tabela dos dados das variáveis X e Y mostra que qualquer par de valores dessas variáveis é obtido do par anterior adicionando ou subtraindo o mesmo valor. Por exemplo, o segundo par de valores (10,-15) é o resultado de subtrair 15 de cada valor do primeiro par (25,0). O terceiro par de valores é o resultado de adicionar 20 ao segundo par, e assim sucessivamente, até completar todos os pares de valores. Da análise dos dois gráficos da Figura 6.3:

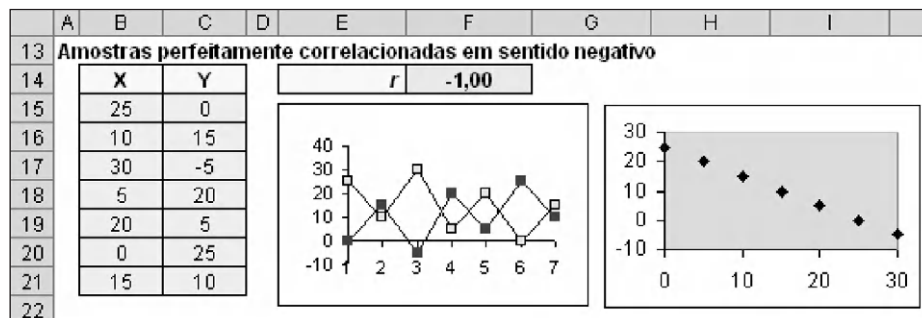
- O gráfico sequencial dos pares de valores, lado esquerdo da Figura 6.3, mostra que os valores das variáveis X e Y formam duas linhas paralelas, com acréscimos e decréscimos iguais e na mesma direção. As duas séries estão perfeitamente correlacionadas de forma positiva.
- O gráfico de dispersão, lado direito da Figura 6.3, mostra que os incrementos dos valores das duas variáveis X e Y são iguais e na mesma direção, sejam acréscimos ou decréscimos. Todos os pares de valores são pontos de uma reta com declividade 45° e, portanto, $r=+1$.

Se os incrementos entre pares são proporcionais e na mesma direção, sejam acréscimos ou decréscimos, os pontos formados pelos pares de valores fazem parte de uma reta com declividade positiva e, portanto, $r=+1$.

Variáveis perfeitamente correlacionadas de forma negativa

A Figura 6.4 mostra o comportamento de duas amostras X e Y perfeitamente correlacionadas em sentido negativo. O coeficiente de correlação dessas amostras calculado na célula F14 é igual a $r=-1$.

FIGURA 6.4 Amostras perfeitamente correlacionadas em sentido negativo, $r=-1$.



Neste caso, também, a tabela dos dados das variáveis X e Y mostra que qualquer par de valores das variáveis X e Y é obtido do par anterior. Por exemplo, o segundo par de valores (10,15) é o resultado de subtrair 15 do valor de X do primeiro par (25,0) e adicionar o valor 15 no primeiro valor de Y . Da mesma maneira, o terceiro par de valores (30,-5) é o resultado de adicionar 20 ao valor de X do segundo par e subtrair 20 do segundo valor de Y , e assim sucessivamente, até completar todos os pares de valores. Da análise dos dois gráficos da Figura 6.4:

- O gráfico sequencial dos pares, lado esquerdo da Figura 6.4, mostra que os valores das variáveis X e Y formam duas linhas opostas, os acréscimos e decréscimos são iguais, porém em direções opostas. As duas séries estão perfeitamente correlacionadas de forma negativa.
- O gráfico de dispersão, lado direito da Figura 6.4, mostra que os incrementos dos valores das duas variáveis X e Y são iguais e em direções opostas. Todos os pares de valores são pontos de uma reta com declividade 135° e, portanto, $r=-1$.

Se os incrementos são proporcionais e em direções opostas, os pontos formados pelos pares de valores fazem parte de uma reta com declividade negativa e, portanto, $r=-1$.

Variáveis não correlacionadas

A Figura 6.5 mostra o comportamento de duas amostras X e Y não correlacionadas. O coeficiente de correlação dessas amostras calculado na célula F25 é igual a $r=0$. Os pares de valores do gráfico de dispersão não apresentam nenhuma tendência.

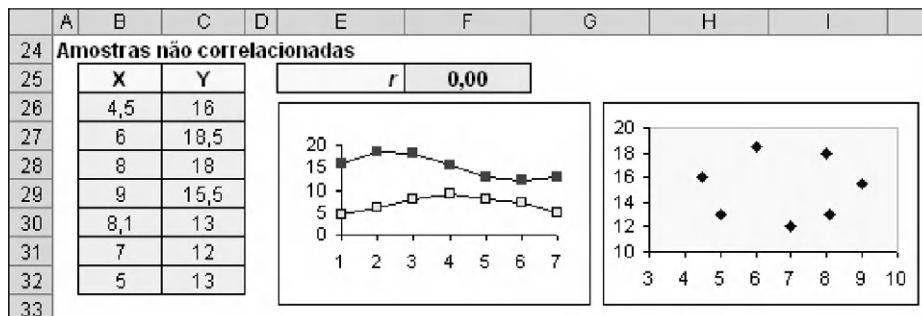


FIGURA 6.5
Amostras não correlacionadas, $r=0$.

Simulador coeficiente de correlação

As três análises apresentadas mostram que o coeficiente de correlação de duas variáveis X e Y com n pares de valores será um valor dentro do intervalo $-1 \leq r_{XY} \leq +1$. Dentro desse intervalo $(-1, +1)$, há um número muito grande de valores do coeficiente de correlação possíveis, que são consequência do afastamento simultâneo dos pares de valores de suas respectivas médias.

A planilha **Simulador de Correlação**, incluída na pasta **Capítulo 6**, ajudará a compreender a formação do coeficiente de correlação, como mostra a Figura 6.6 com o gráfico de dispersão de duas amostras aleatórias X e Y com 100 pares de dados. As características do simulador são:

- A análise pode ser realizada com amostras aleatórias contendo 50, 100 ou 150 pares de dados, ou pontos do gráfico de dispersão, opções que ajudam a compreender a formação do coeficiente de correlação. Cada vez que for escolhida uma *Quantidade de pares de valores*, selecionando o botão de opção correspondente, será ativada a macro que gera as séries aleatórias bivariadas normais.¹¹
- O acionamento da *Barra de rolagem*, localizada na parte inferior do gráfico, gera séries com novos coeficientes de correlação cujo valor é registrado na parte superior do gráfico. Por exemplo, o coeficiente de correlação das séries de dados da Figura 6.6 é $r=0,85$.
- O gráfico inclui a *reta Média X* e a *reta Média Y* que facilitam a visualização da formação do sinal do coeficiente de correlação, como apresentado anteriormente.
- O *modelo* pode gerar séries de valores para o valor de coeficiente de correlação informado na célula I5. Para isso, informe o valor do coeficiente em I5 e depois pressione o botão *r*. Verifique que a barra de rolagem se posicionou no valor registrado em I5.

A Figura 6.7 mostra seis gráficos com coeficientes de correlação diferentes e dentro do intervalo $(-1,1)$.

- No gráfico esquerdo da primeira linha, as amostras têm uma correlação positiva perfeita, $r=1$. Os pontos estão em uma mesma reta com declividade positiva. No gráfico da Figura 6.6, os pontos mostram uma correlação positiva, e as amostras têm correlação positiva, $r=0,85$. A maioria dos pontos está próxima de uma reta com declividade positiva.
- No gráfico direito da primeira linha, as amostras têm uma fraca correlação positiva, $r=0,32$. Os pares de valores formam uma nuvem com ligeira tendência de declividade positiva.

¹¹ *Discrete-Event System Simulation*, Banks J. et al – Prentice Hall, 2ª ed., 1996.

- No gráfico esquerdo da segunda linha, não há relação entre as variáveis, $r=-0,01$. Os pontos formam uma nuvem sem nenhuma tendência.
- No gráfico direito da segunda linha, as amostras têm uma fraca correlação negativa, $r=-0,32$, apresentando uma nuvem de pontos com ligeira tendência de declividade negativa.
- Na terceira e última linha, o gráfico da esquerda mostra que as amostras têm uma boa correlação negativa, $r=-0,85$. Coeficientes de correlação inferiores a esse valor e se aproximando de -1 mostrariam uma forte correlação negativa. A maioria dos pontos está próxima de uma reta com declividade negativa.
- No gráfico direito da terceira linha, as amostras têm uma correlação negativa perfeita, $r=-1$. Os pontos estão em uma mesma reta com declividade negativa.

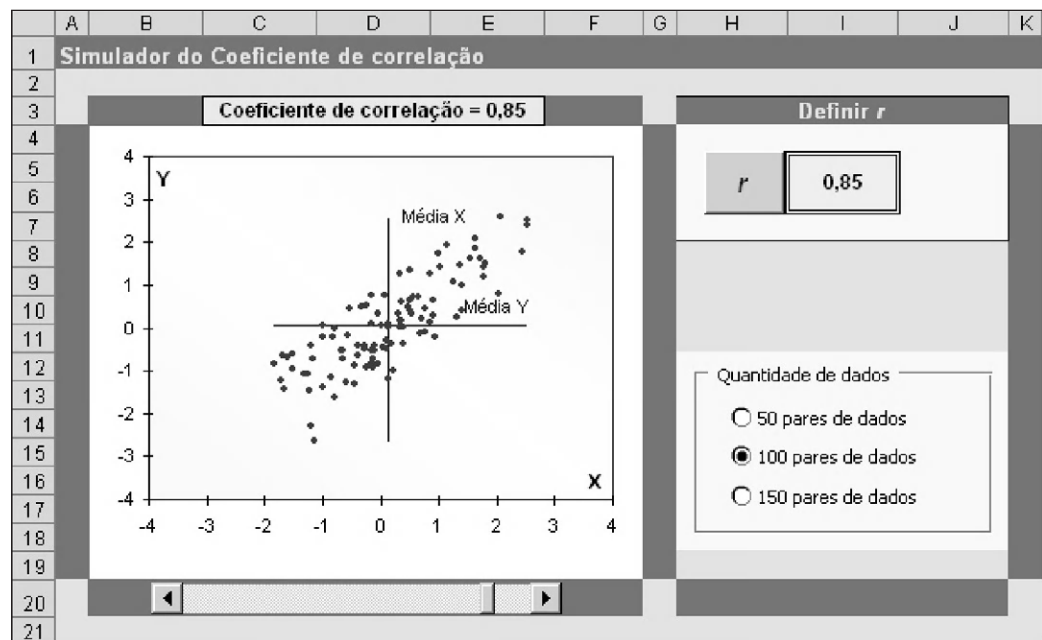


FIGURA 6.6
Simulador de
Correlação.

Alguns cuidados com os resultados

É importante ter em mente duas características do coeficiente de correlação:

- Mantendo os mesmos pares de valores, a permutação das variáveis não altera o resultado do coeficiente de correlação $r_{XY} = r_{YX}$.
- O valor r_{XY} é uma medida da tendência e da força da relação linear entre as variáveis X e Y .

Quando uma cozinheira varia a dosagem de fermento na produção de pão e consegue um melhor sabor, pode-se concluir que a melhoria do sabor foi causado pela nova dosagem de fermento. Nesse caso, há uma relação causa-efeito. Contudo, em geral, uma forte correlação não é sinônimo de uma relação causa-efeito entre as amostras ou variáveis. Há situações em que um coeficiente de correlação próximo de um ou de menos um não significa que a maioria dos pares de valores esteja contida em uma reta. Como será mostrado, o simples conhecimento do coeficiente de correlação não é suficiente devido a anomalias na dispersão dos dados, sendo recomendado construir o gráfico de dispersão das amostras para melhor compreender o resultado.

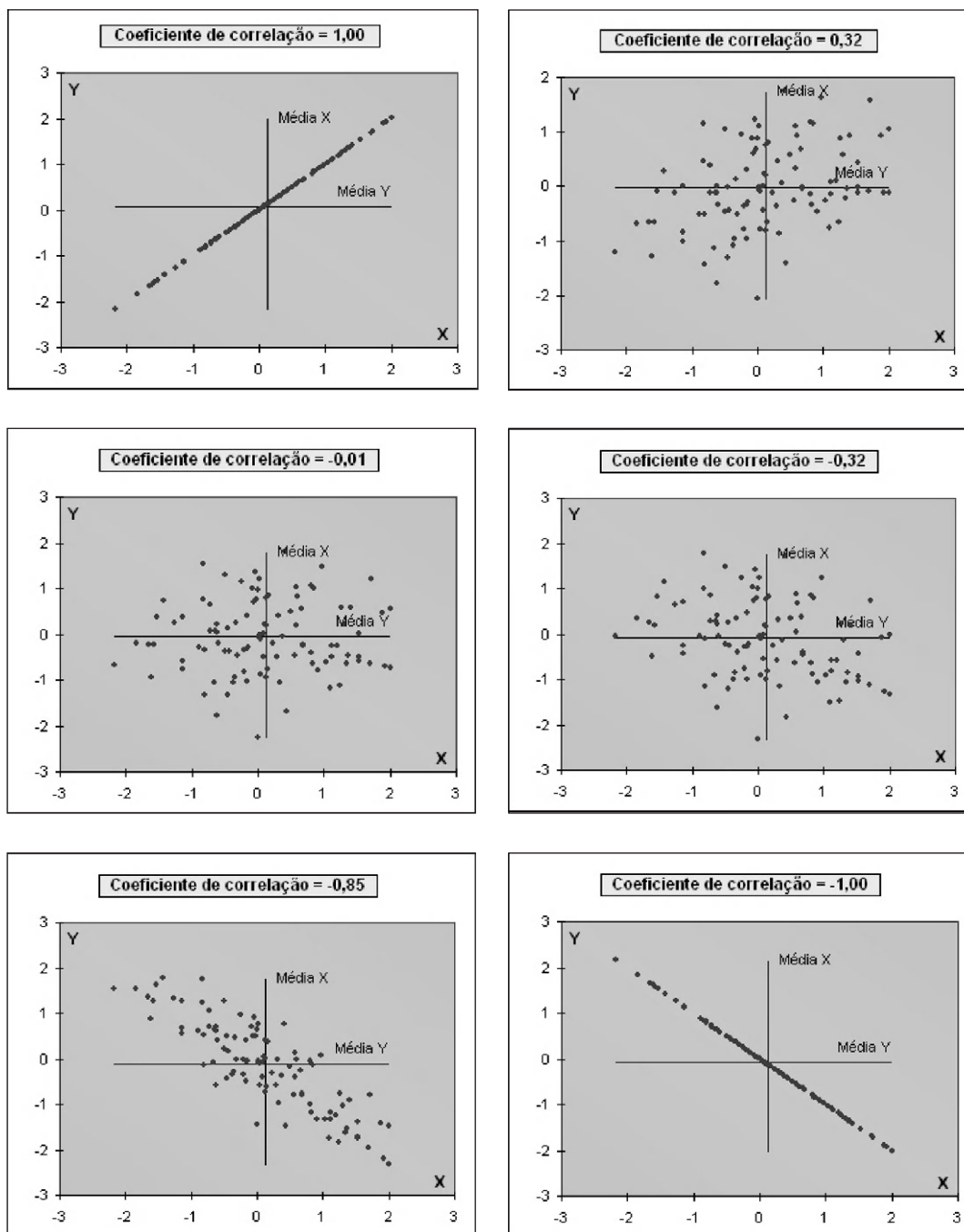


FIGURA 6.7 Simulação de valores do Coeficiente de Correlação de duas amostras.

Correlação e causalidade

As características descritas anteriormente mostram que o coeficiente de correlação não mede a relação causa-efeito entre as variáveis, apesar de essa relação poder estar presente. Por exemplo, uma correlação fortemente positiva entre as variáveis X e Y não autoriza afirmar que variações da variável X provocam variações na variável Y , ou vice-versa. O coeficiente de correlação sozinho não identifica a relação causa-efeito entre as duas variáveis; entretanto, na regressão linear, a relação causa efeito é definida no início da análise. Em alguns casos, a relação causa-efeito pode ser provocada por um ou mais fatores ocultos, uma variável não considerada na análise. Por exemplo, suponha que o número de vendas diá-

rias de um jornal e a produção diária de ovos tenham uma forte correlação positiva. Não se pode afirmar que o aumento da produção de ovos seja a causa do aumento do número de jornais vendidos, nem que o aumento do número de jornais vendidos resulte no aumento da produção de ovos! Para compreender a forte e positiva correlação, devem-se procurar fatores ocultos, por exemplo, o aumento de riqueza da população, que resulta em aumento de demanda dos dois produtos ao mesmo tempo, jornais e ovos.

Anomalias do coeficiente de correlação

Se o coeficiente de correlação for igual a mais um, os pares de valores das variáveis fazem parte de uma reta com declividade positiva. À medida que os pares de valores se afastam dessa reta, o coeficiente de correlação diminuirá de mais um em direção a menos um, passando pelo valor zero, simulação que pode ser facilmente realizada com o *modelo Simulador de Correlação* apresentado na seção anterior. Quanto a essa imagem de linearidade, você deve tomar alguns cuidados, pois há casos em que um coeficiente de correlação próximo de um ou de menos um não significa que a maioria dos pares de valores esteja contida em uma reta. A seguir, comentamos alguns casos registrados com mais detalhes na planilha *Anomalias*, incluída na pasta *Capítulo 6*.

Coeficiente de correlação próximo de +1

A Figura 6.8 mostra alguns casos comentados a seguir.

- Os pares de valores estão contidos numa curva crescente, por exemplo, como a função matemática $Y = 0,10 \times X^2$ mostrada no gráfico esquerdo da primeira linha da Figura 6.8. Nesse caso, o coeficiente de correlação das variáveis X e Y no intervalo $1 \leq X \leq 10$ é igual a 0,97.

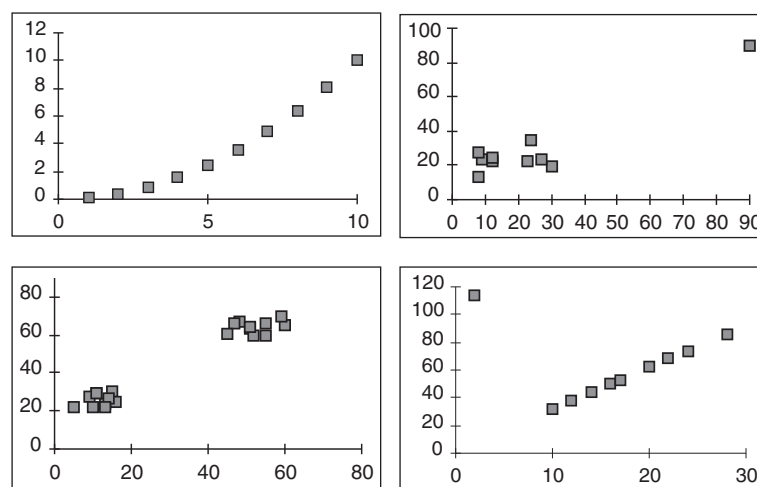


FIGURA 6.8 Anomalias no resultado do coeficiente de correlação.

- O coeficiente de correlação dos pontos do gráfico da direita da primeira linha da Figura 6.8 é 0,92. Um dos pares de valores é um dado suspeito, (90, 90), pois ele está bastante afastado dos demais pares que estão distribuídos, formando uma nuvem de pontos com coeficiente de correlação pequeno, próximo de zero. Essa forte correlação positiva é provocada pelo dado suspeito que gera uma forte tendência positiva.
- Se o primeiro par de valores (90, 90) registrado no intervalo B18:C18 da planilha *Anomalias* for substituído pelo novo par (9, 32) mais próximo da nuvem de pontos, o coeficiente de correlação diminuirá para próximo de zero.

- No gráfico esquerdo da segunda linha da Figura 6.8, os dados estão distribuídos em dois grupos com $r=0,98$. Em geral, amostras de populações diferentes podem provocar essa situação, os dois grupos geram uma tendência de declividade positiva nesse caso.

Coeficiente de correlação próximo de zero

O coeficiente de correlação das variáveis do gráfico direito da segunda linha é igual a zero com um dado suspeito, pois apenas um par está completamente afastado dos demais pares que estão contidos numa reta com declividade positiva. Removendo o dado suspeito, o coeficiente de correlação será igual a +1.

Analise as consequências dessas anomalias:¹²

- Embora seja recomendado excluir os dados suspeitos, esses dados não devem ser simplesmente desprezados. Deve-se dar a devida atenção à causa de tais anomalias, pois esses dados suspeitos podem ser úteis para descobrir a causa dessa ocorrência.
- A escala dos eixos dos gráficos deve ser escolhida adequadamente. Se a escala for mal escolhida, isso poderá prejudicar a interpretação dos resultados. Para evitar que a dispersão apresente tendência diferente, levando a conclusões incorretas, sugere-se que o limite inferior e superior da escala do eixo sejam próximos dos correspondentes valores mínimo e máximo dos dados.

Como conclusão, o simples conhecimento do coeficiente de correlação não é suficiente se não for construído o gráfico de dispersão e, em alguns casos, obtidas outras medidas estatísticas.

Tabelas de covariâncias e de coeficientes de correlação

A covariância e o coeficiente de correlação sempre se referem a duas variáveis ou amostras. Quando há mais de duas variáveis¹³, é possível aplicar os conceitos estatísticos considerando as variáveis duas a duas. Nesse caso, as covariâncias e os coeficientes de correlação são registrados em uma tabela ou matriz de tamanho definido pelo número de variáveis.¹⁴ Para três variáveis A , B e C , as possíveis covariâncias das três variáveis tomadas duas a duas estão registradas na tabela da Figura 6.9.

	A	B	C
A	$\sigma_{A,A}$	$\sigma_{A,B}$	$\sigma_{A,C}$
B	$\sigma_{B,A}$	$\sigma_{B,B}$	$\sigma_{B,C}$
C	$\sigma_{C,A}$	$\sigma_{C,B}$	$\sigma_{C,C}$

FIGURA 6.9 Tabela de covariâncias de três variáveis.

A tabela da Figura 6.9 pode ser simplificada, considerando que:

- A covariância $\sigma_{A,A}$ é a própria variância de A . Portanto, as covariâncias da diagonal principal da tabela são as variâncias das variáveis.
- A permutação das variáveis não altera o resultado da covariância, pois $\sigma_{A,B} = \sigma_{B,A}$. Como as covariâncias da tabela são simétricas com relação à diagonal principal, a tabela de covariâncias pode ser representada com a diagonal principal e apenas uma das duas metades, por exemplo, a parte inferior como mostra a Figura 6.10.

¹² Métodos Estatísticos para Melhora da Qualidade de Kume H. – Editora Gente, 1993.

¹³ Como é o caso da combinação linear de variáveis aleatórias, tema do Capítulo 9.

¹⁴ Para n variáveis, o número de covariâncias ou coeficientes de correlação diferentes é igual a $C(n,2) = \frac{n!}{2!(n-2)!}$

	A	B	C
A	σ_A		
B	$\sigma_{B,A}$	σ_B	
C	$\sigma_{C,A}$	$\sigma_{C,B}$	σ_C

FIGURA 6.10 Tabela de covariâncias, simplificada.

A tabela dos coeficientes de correlação da Figura 6.11 é obtida da tabela de covariâncias da Figura 6.10, substituindo $\sigma_{B,A}$ pelo seu equivalente $r_{B,A}$ e, da mesma forma, as outras duas covariâncias. As variâncias registradas na diagonal da tabela devem ser substituídas pelo valor um, pois para a variável A,

por exemplo, verifica-se que $r_{A,A} = \frac{\sigma_{A,A}}{\sigma_A \times \sigma_A} = \frac{\sigma_A^2}{\sigma_A^2} = 1$.

	A	B	C
A	1		
B	$r_{B,A}$	1	
C	$r_{C,A}$	$r_{C,B}$	1

FIGURA 6.11 Tabela dos coeficientes de correlação.

Ferramenta de análise *Covariância*

O Excel dispõe da ferramenta de análise *Covariância* para construir tabelas de covariâncias, como mostrada na planilha *Ferramenta Covariância*, incluída na pasta **Capítulo 6**, com os dados do Exemplo 6.1 registrados no intervalo B3:C13 incluindo os nomes das variáveis. O procedimento da ferramenta de análise *Covariância* é:

- No menu **Ferramentas**, escolha **Análise de Dados** e na caixa de diálogo **Análise de dados** escolha *Covariância* na lista de **Ferramentas de análise**. Depois pressione o botão **OK**.
- Para calcular a covariância das duas amostras, preencha a caixa de diálogo **Covariância** como mostra a Figura 6.12.
 - Pressionando o botão **Ajuda** dessa caixa de diálogo, o Excel apresentará a página *Sobre a caixa de diálogo Covariância* pertencente à *Ajuda do Excel*.

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

- **Intervalo de entrada.** Informe o intervalo de células da planilha onde os dados estão registrados, nesse caso o intervalo B3:C13, que inclui as células nas quais foram registrados os títulos *Propaganda* e *Vendas*.
- **Agrupado por.** Selecione **Colunas**, pois as amostras foram registradas em 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 nas primeiras células das séries foram registrados os títulos *Propaganda* e *Vendas*.

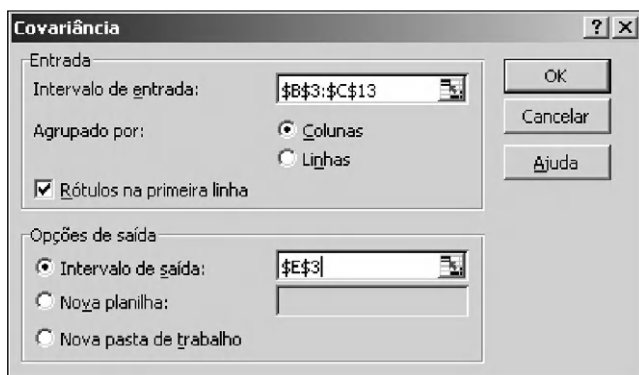


FIGURA 6.12 Caixa de diálogo da ferramenta Covariância.

No quadro **Opções de saída**, deve ser obrigatoriamente informado um endereço a partir do qual a ferramenta *Covariância* 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 E3. Depois de clicar com o botão esquerdo do mouse dentro da caixa correspondente, o endereço pode ser registrado digitando E3, ou *clcando* com o botão esquerdo do *mouse* na célula E3. Nesse caso, será registrado o endereço com os dois cifrões, \$E\$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 6**, a ferramenta inserirá a planilha **Plan1**.
 - Há a alternativa de informar o nome da planilha na caixa desta 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 numa nova pasta e a partir da célula A1 da planilha **Plan1**.

	A	B	C	D	E	F	G
1	Ferramenta de análise Covariância						
2							
3		Propaganda	Vendas				
4		30	430				
5		21	335				
6		35	520				
7		42	490				
8		37	470				
9		20	210				
10		8	195				
11		17	270				
12		35	400				
13		25	480				
14							

		<i>Propaganda</i>	<i>Vendas</i>
Propaganda		101,2	
Vendas		985,5	12995

FIGURA 6.13 Resolução do Exemplo 6.1 com a ferramenta Covariância.

Depois de realizar as escolhas e pressionar o botão **OK**, a ferramenta registra a *tabela de covariâncias* a partir da célula E3, Figura 6.13. Verifique que a covariância como as variâncias obtidas com a ferramenta de análise *Covariância* referem-se à população.

Ferramenta de análise *Correlação*

A ferramenta de análise *Correlação* tem o mesmo formato e procedimento operacional da ferramenta *Covariância*. Dessa maneira, serão apresentadas apenas as diferenças importantes. Para utilizar a ferramenta de análise *Correlação*, foi preparada a planilha **Ferramenta Correlação**, incluída na pasta **Capítulo 6**, com os dados do Exemplo 6.1 registrados no intervalo B3:C13, incluindo os nomes das variáveis. O procedimento da ferramenta de análise *Correlação* é:

- No menu **Ferramentas**, escolha **Análise de Dados** e, na caixa de diálogo **Análise de dados**, escolha *Correlação* na lista de **Ferramentas de análise**. Depois pressione o botão **OK**.
- Para calcular a correlação das duas amostras, preencha a caixa de diálogo *Correlação*, como mostra a Figura 6.14.
- Pressionando o botão **Ajuda** dessa caixa de diálogo, o Excel apresentará a página *Sobre a caixa de diálogo Correlação* pertencente à *Ajuda do Excel*.

Depois de realizar as escolhas e pressionar o botão **OK**, a ferramenta registra a *tabela de correlações* a partir da célula E3, Figura 6.15.

FIGURA 6.14 Caixa de diálogo da ferramenta *Correlação*.

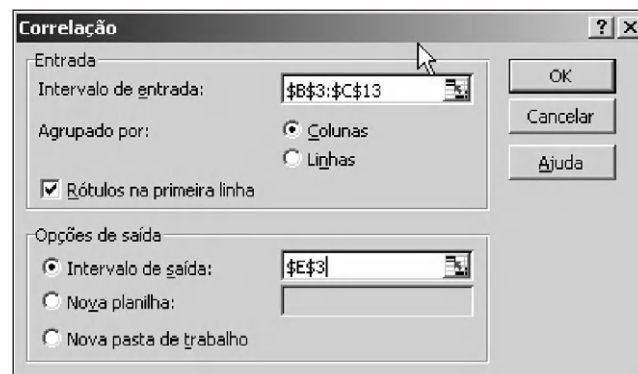


FIGURA 6.15 Resolução do Exemplo 6.1 com a ferramenta *Correlação*.

	A	B	C	D	E	F	G
1	Ferramenta de análise Correlação						
2							
3		Propaganda	Vendas			Propaganda	Vendas
4		30	430			Propaganda	1
5		21	335			Vendas	0,85936613
6		35	520				
7		42	490				
8		37	470				
9		20	210				
10		8	195				
11		17	270				
12		35	400				
13		25	480				
14							

EXEMPLO 6.5

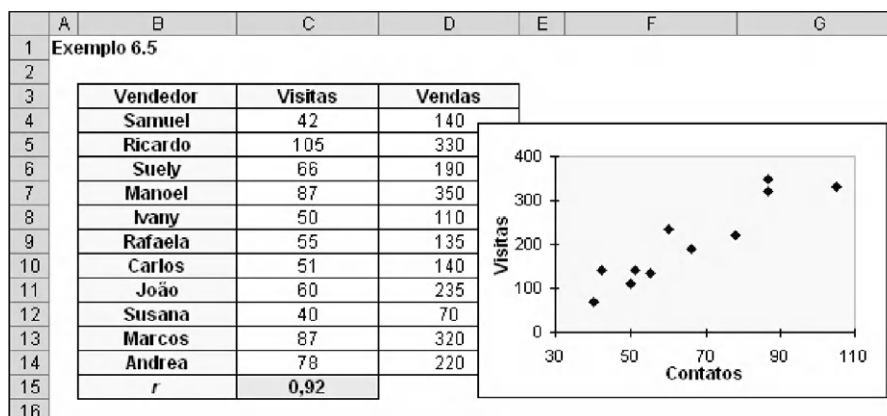
A venda dos produtos tem forte relação com as visitas realizadas pelos vendedores aos seus clientes, é o que afirma o gerente de vendas. A empresa tem onze vendedores e, como regra, eles visitam seus clientes uma vez por mês. Para tentar confirmar sua afirmativa, o gerente preparou a tabela com as visitas realizadas e as vendas de cada vendedor durante o mês passado. O objetivo é verificar se os dados confirmam a crença do gerente.

Solução. Na planilha **Exemplo 6.5**, incluída na pasta **Capítulo 6**, foram registradas as séries de dados, construído o gráfico de dispersão e calculado o coeficiente de correlação das duas amostras, como mostra a figura seguinte.

- O gráfico de dispersão mostra que a maioria dos pares de valores das amostras *Visitas* e *Vendas* se aproxima de uma reta com declividade positiva, confirmando a existência de uma relação forte entre as visitas dos

vendedores e as vendas dos produtos oferecidos. Podemos aceitar que mais visitas geram mais vendas, mas sem poder definir o número de visitas.

- O coeficiente de correlação 0,92 mostra uma forte correlação positiva entre as duas variáveis e parece que confirma a crença do gerente de vendas.



EXEMPLO 6.6

Construa a tabela dos coeficientes de correlação da Tabela de Índices de Preços¹⁵ registrada na planilha **Índices de preço**, incluída na pasta **Capítulo 6**. Embora esses índices tenham metodologias e períodos de coletas de preços diferentes e as séries sejam pequenas, apenas dez dados cada uma, é uma oportunidade interessante para aplicar a ferramenta de análise *Correlação* e analisar as relações entre as séries de índices.

Solução. Com a ferramenta de análise *Correlação*, foi construída a tabela dos coeficientes de correlação a partir da célula B16 da planilha. Analisemos os resultados de maior destaque:

- A menor correlação $r=0,26$ ocorre entre os índices IPCA-E e ICV, que mostra uma fraca correlação entre esses dois índices de preço.
- O índice IPCA-E mantém uma baixa correlação com os outros índices.
- As duas maiores correlações ocorrem com os índices IGPM e IGP-DI e IPA-M e IGP-DI.

	A	B	C	D	E	F	G	H	I	J
1	Índices de preços (%)									
2										
3										
4		Mês	IBGE INPC	Fipe IPC	DIEESE ICV	IBGE IPCA	FGV IGP-DI	FGV IGPM	FGV IPA-M	IBGE IPCA-E
5		jun/03	-0,06	-0,16	-0,26	-0,15	-0,70	-1,00	-1,82	0,22
6		jul/03	0,04	-0,08	0,35	0,20	-0,20	-0,42	-0,75	-0,18
7		ago/03	0,18	0,63	-0,15	0,34	0,62	0,38	0,20	0,27
8		set/03	0,82	0,84	1,26	0,78	1,05	1,18	1,54	0,57
9		out/03	0,39	0,63	0,47	0,29	0,44	0,38	0,36	0,66
10		nov/03	0,37	0,27	0,26	0,34	0,48	0,49	0,57	0,17
11		dez/03	0,54	0,42	0,32	0,52	0,60	0,61	0,64	0,46
12		jan/04	0,83	0,65	1,46	0,76	0,80	0,88	0,98	0,68
13		fev/04	0,39	0,19	-0,18	0,61	1,08	0,69	0,79	0,90
14		mar/04	0,57	0,12	0,47	0,47	0,93	1,13	1,33	0,40
15										
16			INPC	IPC	ICV	IPCA	IGP-DI	IGPM	IPA-M	IPCA-E
17		INPC	1							
18		IPC	0,70	1						
19		ICV	0,82	0,60	1					
20		IPCA	0,91	0,68	0,69	1				
21		IGP-DI	0,80	0,65	0,42	0,90	1			
22		IGPM	0,89	0,67	0,57	0,90	0,97	1		
23		IPA-M	0,90	0,66	0,59	0,91	0,96	1,00	1	
24		IPCA-E	0,62	0,52	0,26	0,62	0,68	0,59	0,59	1
25										

15 Tabela publicada no jornal *O Estado de São Paulo*, 02 de junho de 2004.

Problemas

Problema 1

O dono da oficina mecânica regulou seu carro e, em um dia sem muito movimento na estrada, realizou as medições de consumo de combustível registradas na tabela seguinte para seis velocidades diferentes. Construa o gráfico de dispersão e analise o comportamento das duas variáveis, *Velocidade* e *Consumo*.

Velocidade	Consumo – Km/l
70	10,2
80	9,7
90	9,1
100	8,3
110	7,8
120	7

Problema 2

Calcule a covariância da população e da amostra da relação *Velocidade* e *Consumo* do Problema 1.

R: $Cov(Vel, Con) = -18,75$ e $Cov(Vel, Con) = -22,5$

Problema 3

Calcule o coeficiente de correlação da relação *Velocidade* e *Consumo* do Problema 1 utilizando a fórmula para os dois casos, população e amostra.

R: $r = -0,9975$

Problema 4

O seguinte trecho foi extraído do jornal:¹⁶ “Para cada 1% de aumento no PIB o impacto no nível de emprego é de 0,4%. ... Entretanto, a criação de vagas formais ainda não é suficiente para reduzir significativamente o desemprego no País ... Para isso o PIB precisaria crescer em média 5% ao ano”. Responda às seguintes perguntas:

- Qual é o tipo de relação entre PIB e Emprego?
- Qual deve ser o impacto no emprego de um crescimento de PIB de 5% ao ano?

R: a) Correlação positiva. b) 2%

Problema 5

Supondo que durante cinco anos se mantenha a relação do Problema 4, para cada 1% de aumento no PIB, o impacto no nível de emprego é de 0,4%.

- Construa a tabela com o crescimento do PIB, começando por 100, e o crescimento do Emprego, começando por 60, mantendo ambas as variáveis com crescimento anual constante.
- Construa o gráfico de dispersão correspondente.
- Calcule o valor do coeficiente de correlação.

R: c) $r = 0,999991$ arredondando $r = 1$

Problema 6

Analisando o resultado do Problema 5, você concorda com as seguintes afirmações?

- As variáveis PIB e Emprego têm uma perfeita correlação positiva.
- Os pares de dados são pontos de uma linha reta com declividade positiva.

¹⁶ Criação de emprego acompanha alta do PIB, artigo de Cleide Silva publicado no jornal O Estado de São Paulo de 2 de junho de 2004.

Problema 7

A tabela seguinte registra os retornos das ações tipo ON e PN de um grupo de empresa. Com esses dados:

- Construa o gráfico de dispersão correspondente.
- Calcule a covariância da população e o coeficiente de correlação.
- Analise os resultados e verifique se há algum dado suspeito.

ON%	PN%	ON%	PN%
37,5	20,9	212,5	367,1
-45,0	5,4	46,3	6,9
0,0	49,4	11,1	45,4
31,5	31,1	43,0	27,8
-1,0	30,0	67,0	43,1
20,1	28,0	9,4	13,4

R: $Cov(ON, PN) = 5.083,84$ e $r = 0,8942$

Problema 8

Repita o Problema 7, porém sem considerar os retornos da empresa com ON% de 212,5 e PN% de 367,1.

Problema 9

O professor passou para os alunos uma folha com duas séries de dados para calcular o coeficiente de correlação e explicar o tipo de relação entre as duas séries. Seu colega rapidamente calculou o valor do coeficiente de correlação igual a zero e afirmou que as duas séries não apresentam nenhuma relação. Concorde com a afirmação de seu colega de que as duas séries não apresentam nenhuma relação? Por quê?

Problema 10

Na tabela seguinte, foram registrados sete pares de valores correspondentes aos resultados de um teste de aptidão. Com esses dados e sem construir o gráfico de dispersão nem calcular o coeficiente de correlação:

- Você conseguiria definir a relação e a tendência dessas duas séries?
- Você conseguiria determinar o valor desse coeficiente?
- Se for possível, qual o valor do coeficiente de correlação?

A	0	6	4	12	8	2	10
B	35	26	29	17	23	32	20

Problema 11

Continuando com o Problema 10.

- Construa o gráfico de dispersão correspondente.
- Calcule o coeficiente de correlação.

Problema 12

A diferença de idades dos irmãos Ana e João é de 5 anos. Considerando a série de dados dos anos de Ana, começando por 10 e terminando com 15, e a série de dados de João, começando com 6 e terminando com 11. Sem construir o gráfico de dispersão nem calcular o coeficiente de correlação:

- Você conseguiria definir a relação e a tendência dessas duas séries?
- Você conseguiria determinar o valor desse coeficiente?
- Se for possível, qual o valor do coeficiente de correlação?

Problema 13

Analise a relação entre as amostras X e Y registradas na tabela seguinte, sugerindo começar pela construção do gráfico de dispersão.

X	10	15	18	12	9
Y	21	15	12	18	20

Problema 14

Os *prêmios e preços de exercícios* de cinco séries de opções de compra com mesmo vencimento estão registrados na tabela seguinte. Construa o gráfico de dispersão, calcule o coeficiente de correlação e analise os resultados.

Prêmios	Preços de exercício
\$257,52	\$2.100
\$99,25	\$2.200
\$38,17	\$2.300
\$14,65	\$2.400
\$5,61	\$2.500

R: $r=-0,8933$

Problema 15

Na planilha **Problemas**, incluída na pasta **Capítulo 6**, está registrada a tabela com a relação dos dez maiores e melhores grupos de supermercados no ano 1991, porém sem indicar os nomes das empresas. Construa e analise a tabela dos coeficientes de correlação.

Problema 16

Na planilha **Problemas**, incluída na pasta **Capítulo 6**, está registrada a tabela com a relação dos dez maiores e melhores grupos de supermercados no ano 1998, porém sem indicar os nomes das empresas. Construa e analise a tabela dos coeficientes de correlação.

Problema 17

Tomando como base os resultados e as análises das empresas do Problema 15, analise a evolução dessas empresas entre os anos 1991 e 1998, comparando os resultados de 1998 com os de 1991.

Apêndice 1

Outra forma de calcular a covariância

Partindo da fórmula da covariância que repetimos em seguida:

$$\sigma_{XY} = \frac{1}{N} \sum_{i=1}^N (X_i - \mu_X) \times (Y_i - \mu_Y)$$

Desenvolvendo o produto indicado temos:

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N (X_i - \mu_X) \times (Y_i - \mu_Y) &= \frac{1}{N} \sum_{i=1}^N (X_i Y_i - X_i \mu_Y - \mu_X Y_i + \mu_X \mu_Y) \\ \frac{1}{N} \sum_{i=1}^N (X_i \mu_X) \times (Y_i - \mu_Y) &= \frac{1}{N} \left(\sum_{i=1}^N X_i Y_i - \sum_{i=1}^N X_i \mu_Y - \sum_{i=1}^N \mu_X Y_i + \sum_{i=1}^N \mu_X \mu_Y \right) \end{aligned}$$

Simplificando as parcelas do segundo membro temos:

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N (X_i - \mu_X) \times (Y_i - \mu_Y) &= \frac{1}{N} \left(\sum_{i=1}^N X_i Y_i - \mu_Y \sum_{i=1}^N X_i - \mu_X \sum_{i=1}^N Y_i + N \mu_X \mu_Y \right) \\ \frac{1}{N} \sum_{i=1}^N (X_i - \mu_X) \times (Y_i - \mu_Y) &= \frac{1}{N} \sum_{i=1}^N X_i Y_i - \mu_Y \mu_X - \mu_X \mu_Y + \mu_X \mu_Y \end{aligned}$$

Depois de simplificar as três últimas parcelas do segundo membro temos:

$$\sigma_{XY} = \frac{1}{N} \sum_{i=1}^N (X_i - \mu_X) \times (Y_i - \mu_Y) = \frac{1}{N} \sum_{i=1}^N X_i Y_i - \mu_Y \mu_X$$

Essa última expressão será utilizada no Apêndice 1 do Capítulo 9.