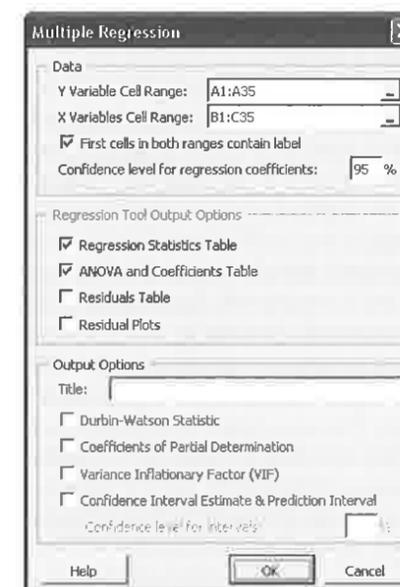


**GE14.1 DESENVOLVENDO um MODELO de REGRESSÃO MÚLTIPLA**

Interpretando os Coeficientes da Regressão

**PHStat2** Utilize o procedimento **Multiple Regression (Regressão Múltipla)** para realizar uma análise de regressão múltipla. Por exemplo, para realizar a análise para os dados sobre vendas de OmniPower da Tabela 14.1 ilustrada na Figura 14.1, abra a planilha **DADOS** da pasta de trabalho **OmniPower**. Selecione **PHStat** → **Regression** → **Multiple Regression (PHStat** → **Regressão** → **Regressão Múltipla**). Na caixa de diálogo do procedimento (ilustrada a seguir),



1. Insira **A1:A35** na caixa **Y Variable Cell Range (Intervalo de Células da Variável Y)**.
2. Insira **B1:C35** na caixa **X Variables Cell Range (Intervalo de Células das Variáveis X)**.
3. Marque a opção **First cells in both ranges contain label (Primeiras células em ambos os intervalos contêm rótulos)**.
4. Insira **95** na caixa para **Confidence level for regression coefficients (Nível de confiança para coeficientes da regressão)**.

5. Clique em **Regression Statistics Table (Tabela de Estatísticas da Regressão)** e em **ANOVA and Coefficients Table (ANOVA e Tabela de Coeficientes)**.
6. Insira um título na caixa correspondente a **Title** e clique em **OK**.

O procedimento cria uma planilha que contém uma cópia de seus dados, além da planilha com os resultados da regressão ilustrados na Figura 14.1. Para mais informações sobre essas planilhas, leia a seção do *Excel Avançado* a seguir.

**Excel Avançado** Utilize a função **PROJ.LIN** para ajudar a realizar uma regressão múltipla. Insira **PROJ.LIN (intervalo de células da variável Y, intervalo de células da variável X, Verdadeiro, Verdadeiro)** como parte de uma fórmula sequencial posicionada em um intervalo de células de 5 linhas por *k* colunas, em que *k* corresponde ao número de variáveis independentes, para calcular as estatísticas associadas à análise da regressão múltipla.

Utilize a planilha **CÁLCULO** da pasta de trabalho **Regressão Múltipla** ilustrada na Figura 14.1, como modelo para realizar regressões múltiplas. As colunas de A a I dessa planilha duplicam o modelo visual da planilha de regressão no suplemento Ferramentas de Análise. A planilha utiliza os dados da regressão na planilha **DADOSRM** para realizar a análise da regressão para os dados da Tabela 14.1 sobre vendas de OmniPower.

A Figura 14.1 não mostra a Área de Cálculos nas colunas K a N (veja a Figura GE14.1 a seguir). Essa área contém uma fórmula sequencial com a função **PROJ.LIN** no intervalo de células L2:N6 e cálculos para o teste *t* para a inclinação (veja a Seção 13.7). A fórmula sequencial calcula os coeficientes  $b_2$ ,  $b_1$  e  $b_0$  nas células L2, M2 e N2; o erro-padrão de  $b_2$ ,  $b_1$  e  $b_0$  nas células L3, M3 e N3;  $r^2$  e o erro-padrão da estimativa nas células L4 e M4; a estatística do teste *F* e os *gl* do erro nas células L5 e M5; e *SQReg* e *SQR* nas células L6 e M6. (O restante do intervalo de células, N4, N5 e N6, exibe a mensagem **#N/D**. Não se trata de um erro.)

Abra a planilha **CÁLCULO\_FÓRMULAS** para examinar todas as fórmulas na planilha, algumas das quais discutidas nas seções do *Excel Avançado* no Guia do Excel para o Capítulo 13. A fórmula da célula L9, no formato **=INVT(1 - nível de confiança, Graus de liberdade do erro)** calcula o valor crítico para o teste *t*.

Para realizar regressão múltipla para outros dados, cole os dados da regressão na planilha **DADOSRM**. Cole os valores da variável *Y* na coluna A. (Isso é o oposto da operação de colagem para a regressão linear simples, discutida na Seção GE13.2.) Cole os valores da variável *X* nas colunas consecutivas, iniciando com

	K	L	M
1			
2	Interseção $b_2$ , $b_1$ e $b_0$	=PROJ.LIN(DADOSRM!A2:A35, DADOSRM!B2:C35, VERDADEIRO, VERDADEIRO)	=PROJ.LIN(DADOSRM!A2:A35, DADOSRM!B2:C35, VERDADEIRO, VERDADEIRO)
3	Erro-padrão de $b_1$ e $b_0$	=PROJ.LIN(DADOSRM!A2:A35, DADOSRM!B2:C35, VERDADEIRO, VERDADEIRO)	=PROJ.LIN(DADOSRM!A2:A35, DADOSRM!B2:C35, VERDADEIRO, VERDADEIRO)
4	R-quadrado, Erro-padrão	=PROJ.LIN(DADOSRM!A2:A35, DADOSRM!B2:C35, VERDADEIRO, VERDADEIRO)	=PROJ.LIN(DADOSRM!A2:A35, DADOSRM!B2:C35, VERDADEIRO, VERDADEIRO)
5	F, gl dos Resíduos	=PROJ.LIN(DADOSRM!A2:A35, DADOSRM!B2:C35, VERDADEIRO, VERDADEIRO)	=PROJ.LIN(DADOSRM!A2:A35, DADOSRM!B2:C35, VERDADEIRO, VERDADEIRO)
6	SQ Regressão, SQ Resíduos	=PROJ.LIN(DADOSRM!A2:A35, DADOSRM!B2:C35, VERDADEIRO, VERDADEIRO)	=PROJ.LIN(DADOSRM!A2:A35, DADOSRM!B2:C35, VERDADEIRO, VERDADEIRO)
7			
8	Nível de confiança	0,95	
9	Valor Crítico de <i>t</i>	=INVT(1-B, B13)	
10	$b_0$ de metade da amplitude	=L9*C17	
11	$b_1$ de metade da amplitude	=L9*C18	
12	$b_2$ de metade da amplitude	=L9*C19	

**FIGURA GE14.1** Área de cálculos da planilha da regressão múltipla (colunas N e O não ilustradas)

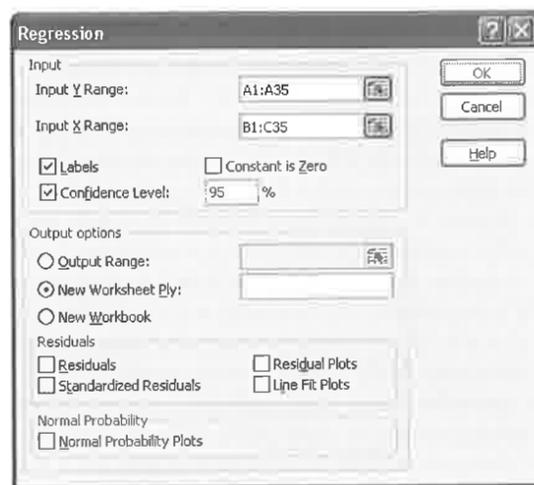
a coluna B. Abra a planilha CÁLCULO. Inicialmente, insira o nível de confiança na célula L8. Depois, edite a sequência correta de 5 linhas por  $k$  colunas, que tem início com a célula L2. Edite os intervalos de células na fórmula sequencial e, depois disso, mantendo pressionadas as teclas **Control** e **Shift** (ou a tecla **Apple** em um Mac) pressione a tecla **Enter**.

Para fazer a associação com a planilha que é criada pelo procedimento Regressão no suplemento Ferramentas de Análise, a planilha CÁLCULO (desnecessariamente) duplica os valores para os intervalos de confiança nas colunas H e I. Esses valores duplicados não estão ilustrados nas figuras apresentadas no capítulo.

**Ferramentas de Análise** Utilize o procedimento **Regressão** para realizar uma análise da regressão múltipla. Por exemplo, para realizar a análise da Figura 14.1 (na Seção 14.1) para os dados sobre vendas de OmniPower na Tabela 14.1, abra a **planilha DADOS** da pasta de trabalho OmniPower e:

1. Selecione **Dados** → **Análise de Dados** (Excel 2007) ou **Ferramentas** → **Análise de Dados** (Excel 2003).
2. Na caixa de diálogo Análise de Dados, selecione **Regressão** da lista de **Ferramentas de Análise** e, depois, clique em **OK**.

Na caixa de diálogo Regressão (veja a tela do procedimento a seguir):



	A	B	C	D	E	F
1	Regressão Múltipla					
2						
3	<b>Estadística de Regressão</b>					
4	R Múltiplo	=RAIZ(C12/C14)				
5	R-Quadrado	=L4				
6	R-Quadrado Ajustado	=1 - (B14/B13) * (C13/C14)				
7	Erro-padrão	=M4				
8	Observações	=CONT.NÚM(DADOSR1A:A)				
9						
10	<b>ANOVA</b>					
11		gl	SQ	MQ	F	F de significação
12	Regressão	=B14 - B13	=L6	=C12/B12	=L5	=DISTF(E12, B12, B13)
13	Resíduo	=M5	=M6	=C13/B13		
14	Total	=B8 - 1	=C12 + C13			

FIGURA GE14.2

Fórmulas para o intervalo de células de A1:F14 da planilha de resultados da regressão múltipla

3. Insira **A1:A35** na caixa para **Intervalo Y de Entrada** e insira **B1:C35** como **Intervalo X de Entrada**.
4. Marque as opções **Rótulos**, marque a opção **Nível de Confiança** e insira **95** como valor para a caixa respectiva.
5. Clique em **Nova planilha**.
6. Clique em **OK**.

Preveno a Variável Dependente Y

**PHStat2** Utilize as instruções do *PHStat2* na Seção GE14.1 “Interpretando os Coeficientes da Regressão”, mas substitua a etapa 6 pela nova etapa 6 e a etapa 7 apresentadas a seguir:

6. Marque a opção **Confidence Interval Estimate & Prediction Interval (Estimativa do Intervalo de Confiança & Intervalo de Previsão)** e digite **95** como o percentual para a caixa **Confidence level for intervals (Nível de confiança para os intervalos)**.
7. Insira um título em **Title** e clique em **OK**.

A planilha adicional criada é explicada na seção do *Excel Avançado* apresentada a seguir.

**Excel Avançado** Utilize as funções **INVT**, **MATRIZ.MULT** e **MATRIZ.INVERSO** para ajudar a calcular a estimativa para o intervalo de confiança e um intervalo de previsão. Digite **INVT(1-nível de confiança, Graus de liberdade do erro)** para calcular a estatística  $t$ .

Examine a planilha **EICeIP** da pasta de trabalho **Regressão Múltipla** ilustrada na Figura 14.2, como um modelo para calcular estimativas de intervalos de confiança e intervalos de previsão. Em uma fórmula sequencial, **MATRIZ.MULT** calcula o produto da matriz  $X'X$  pelo fato de tratar a planilha **RMSequencia** como uma matriz. Em uma outra fórmula sequencial, **MATRIZ.INVERSO** calcula o inverso da matriz  $X'X$  criada por **MATRIZ.MULT**. Outros usos de **MATRIZ.MULT** calculam  $X'G$  multiplicado pelo inverso de  $X'X$  e  $[X'G$  multiplicado pelo inverso de  $X'X]$  multiplicado por  $XG$ .

A planilha contém os dados e fórmulas para os dados sobre vendas de OmniPower da Tabela 14.1. A modificação dessa planilha (e da planilha **RMSequencia**) para outros problemas requer um conhecimento que está além do escopo deste livro.

### GE14.2 $r^2$ , $r^2$ AJUSTADO e o TESTE F GERAL

Os três tópicos discutidos na Seção GE14.2 são calculados na planilha de regressão múltipla criada utilizando-se as instruções da Seção GE14.1 “Interpretando os Coeficientes da Regressão”.

A Figura GE14.2 mostra as fórmulas utilizadas para calcular o coeficiente de determinação múltipla,  $r^2$ , o  $r^2$  ajustado e o teste  $F$  geral na pasta de trabalho **CÁLCULO** da pasta de trabalho **Regressão Múltipla**, criada quando se utilizam as instruções para o *PHStat2* ou para o *Excel Avançado* da Seção GE14.1. As fórmulas nas células B5, B7, B13, C12, C13, D12 e E12 copiam valores calculados pela fórmula sequencial no intervalo de células L2:N6 que utilizam a função **PROJ.LIN**. (Veja as instruções do *Excel Avançado* da Seção GE14.1 para mais informações.) A fórmula na célula F12, no formato **=DISTF(estatística do teste F, 1, graus de liberdade do erro)** calcula o valor- $p$  para o teste  $F$  para a inclinação, discutido na Seção 13.7.

### GE14.3 ANÁLISE de RESÍDUOS para o MODELO de REGRESSÃO MÚLTIPLA

**PHStat2** Utilize as instruções do *PHStat2* da Seção GE14.1 “Interpretando os Coeficientes da Regressão”. Modifique a etapa 5 marcando as opções **Residuals Table (Tabela de Resíduos)** e **Residuals Plot (Gráfico de Resíduos)**, além de marcar **Regression Statistics Table (Tabela de Estatísticas da Regressão)** e **ANOVA and Coefficients Table (ANOVA e Tabela de Coeficientes)**.

**Excel Avançado** Crie uma planilha que calcule os resíduos e, depois disso, crie um gráfico de dispersão da variável  $X$  com relação aos resíduos (inseridos no gráfico como a variável  $Y$ ).

Utilize a **planilha RESÍDUOS** da pasta de trabalho **Regressão Múltipla** ilustrada na Figura GE14.3 como modelo para criar uma planilha de resíduos. As fórmulas nessa planilha calculam os resíduos da análise da regressão para os dados sobre vendas de OmniPower na Tabela 14.1 utilizando os dados da regressão na **planilha DADOSRM** na mesma pasta de trabalho.

Na coluna D, a planilha calcula os valores de  $Y$  previstos (com a legenda **Barras Previstas** na Figura GE14.3) ao multiplicar os valores de  $X_1$  pelo coeficiente  $b_1$  e os valores de  $X_2$  pelo coeficiente  $b_2$  e somando esses produtos ao coeficiente  $b_0$ . Na coluna F, a planilha calcula os resíduos subtraindo dos valores de  $Y$  os

	A	B	C	D	E	F
1	Observação	Preço	Promoção	Barras Previstas	Barras	Resíduos
2	1	59	200	=CÁLCULOIB18 * B2 + CÁLCULOIB19 * C2 + CÁLCULOIB17	4141	=E2-D2
3	2	59	200	=CÁLCULOIB18 * B3 + CÁLCULOIB19 * C3 + CÁLCULOIB17	3842	=E3-D3
4	3	59	200	=CÁLCULOIB18 * B4 + CÁLCULOIB19 * C4 + CÁLCULOIB17	3056	=E4-D4
33	32	99	400	=CÁLCULOIB18 * B33 + CÁLCULOIB19 * C33 + CÁLCULOIB17	1602	=E33-D33
34	33	99	600	=CÁLCULOIB18 * B34 + CÁLCULOIB19 * C34 + CÁLCULOIB17	3354	=E34-D34
35	34	99	600	=CÁLCULOIB18 * B35 + CÁLCULOIB19 * C35 + CÁLCULOIB17	2927	=E35-D35

FIGURA GE14.3

Fórmulas para a planilha de resíduos (ilustrando apenas as linhas de 1 a 4 e de 33 a 35)

valores de  $Y$  previstos. Para outros problemas, modifique essa planilha do seguinte modo:

1. Se o número de variáveis independentes for maior do que 2, selecione a coluna D, clique à direita e clique em **Inserir** a partir do menu de atalhos. Repita essa etapa tantas vezes quanto necessário para criar as colunas adicionais para alocar todas as variáveis  $X$ .
2. Cole os dados das variáveis  $X$  em colunas, iniciando pela coluna B.
3. Cole os valores de  $Y$  na coluna E (ou a antepenúltima coluna caso existam mais de duas variáveis  $X$ ).
4. Para tamanhos de amostra menores do que 34, exclua as linhas adicionais; Para tamanhos de amostra maiores do que 34, copie as fórmulas para  $Y$  previsto e para resíduos (colunas D e F na Figura GE14.3) para baixo até a linha que contém o último par de valores de  $X$  e  $Y$ . Também, acrescente os novos números de observações na coluna A.

Para criar gráficos de resíduos, utilize os valores para copiar e colar especiais (veja a Seção C.12 no Apêndice C) para colar valores de dados em uma nova planilha na ordem apropriada antes de aplicar as instruções para gráficos de dispersão na Seção GE2.7

**Ferramentas de Análise** Utilize as instruções para *Ferramentas de Análise* da Seção GE14.1. Modifique a etapa 5 marcando as opções **Resíduos** e **Plotar Resíduos** antes de clicar em **Nova Planilha** e, depois disso, em **OK**. (Observe que a opção **Plotar Resíduos** cria gráficos de resíduos somente para cada uma das variáveis independentes.

### GE14.4 INFERÊNCIAS RELATIVAS aos COEFICIENTES de REGRESSÃO da POPULAÇÃO

As planilhas com os resultados da regressão criadas utilizando-se as instruções da Seção GE14.1 incluem as informações necessárias para realizar as inferências discutidas na Seção 14.4.

### GE14.5 TESTANDO PARTES do MODELO de REGRESSÃO MÚLTIPLA

Coeficientes de Determinação Parcial

**PHStat2** Utilize as instruções do *PHStat2* na Seção GE14.1 “Interpretando os Coeficientes da Regressão”, mas substitua a etapa 6 marcando a opção **Coefficients of Partial Determination (Coeficientes de Determinação Parcial)** antes de clicar em **OK**.

**Excel Avançado** Você calcula os coeficientes de determinação parcial utilizando um processo de duas etapas. Você utiliza inicialmente as instruções do *Excel Avançado* na Seção GE14.1 para criar todas as planilhas possíveis de resultados da regressão em uma cópia da **pasta de trabalho Regressão Múltipla**. Por exemplo, se você tem duas variáveis independentes, você realiza três análises de regressão:  $Y$  com  $X_1$  e  $X_2$ ;  $Y$  com  $X_1$  e  $Y$  com  $X_2$  para criar três planilhas de resultados da regressão. Depois disso, abra a **planilha CDP** para o número de variáveis independentes (as **planilhas CPD\_2**, **CPD\_3** e **CPD\_4** estão incluídas) e siga as instruções em itálico para os valores de copiar e colar especiais a partir de planilhas de resultados da regressão. A **planilha CPD\_2** contém os dados para calcular os coeficientes de determinação parcial para o modelo de regressão utilizado como um exemplo na Seção 14.5.

## GE14.6 UTILIZANDO VARIÁVEIS BINÁRIAS (DUMMY) e TERMOS de INTERAÇÃO em MODELOS de REGRESSÃO

### Criando Variáveis Binárias (Dummy)

Para criar variáveis binárias (*dummy*), acrescente novas colunas à planilha que contém os dados de sua regressão.

Para o caso de uma variável categórica que contenha dois níveis, abra a planilha com os dados da regressão. Copie e cole em uma nova planilha os valores da variável categórica de dois níveis. Selecione todos os valores recentemente colados e, em seguida, pressione **Ctrl + L**, o atalho de teclado para **Localizar e Substituir**. Na caixa de diálogo Localizar e Substituir:

1. Insira o primeiro valor categórico na caixa ao lado de **Localizar**.
2. Insira a variável binária (1 ou 0) que representará esse valor categórico na caixa ao lado de **Substituir por**.
3. Clique em **Substituir Tudo**. Caso apareça uma mensagem para confirmar a substituição, clique em **OK** para continuar.

Repita a operação localizar e substituir para os outros valores categóricos e insira o outro valor binário (*dummy*) na caixa ao lado de **Substituir por**. Depois, clique em **Fechar**.

Variáveis categóricas que têm mais de dois níveis requerem o uso de fórmula em colunas múltiplas. Por exemplo, para criar as variáveis binárias para o Exemplo 14.3, na Seção 14.6, são necessárias duas colunas. Suponha que a variável categórica de três níveis mencionada no exemplo esteja na Coluna D da planilha aberta. Uma primeira coluna nova que contenha fórmulas

no formato **=SE(célula da coluna D = primeiro nível, 1,0)** e uma segunda coluna nova que contenha fórmulas no formato **=SE(célula da coluna D = segundo nível, 1,0)** criariam apropriadamente as duas variáveis binárias (*dummy*) que o exemplo requer.

### Criando Interações

Para criar uma interação, acrescente uma coluna de fórmulas que multipliquem uma variável independente pela outra. Por exemplo, se a primeira variável independente tiver aparecido na coluna B e a segunda variável independente tiver aparecido na coluna C, insira a fórmula **=B2 \* C2** na célula da linha 2 de uma coluna em branco e, depois, copie a fórmula para baixo ao longo de todas as linhas de dados para criar a interação.