

- São conhecidas uma ou mais séries de dados ou observações, *série_x*, tendo presente que:
 - Quando existe apenas uma variável independente, a forma dos intervalos das duas variáveis y e x podem ter qualquer forma.
 - Quando existe mais de uma variável independente, deve ser informado o intervalo abrangendo todas as variáveis independentes juntas.
 - Quando a informação da variável independente é omitida, a função assume que x é um *array* de números {1, 2, 3, ...} com o mesmo comprimento que a variável y .
- Definindo o argumento *constante* como:
 - VERDADEIRO (ou omitido) a função fornece todos os coeficientes b e m 's da curva exponencial de regressão.
 - FALSO a função fornece apenas os coeficientes m 's da reta de regressão

$$\hat{y} = m_1 \times m_2^{x_1} \times \dots \times m_n^{x_n}, \text{ isto é } b=1.$$

- Definindo o argumento *estatísticas* como:
 - FALSO a função fornece somente os coeficientes b e m 's.
 - VERDADEIRO (ou omitido) a função fornece os coeficientes b e m 's e os resultados seguintes: erros padrões dos coeficientes b e m 's; o coeficiente de determinação r^2 ; o erro padrão da estimativa S_e ; o valor da estatística F ; o número de graus de liberdade gl da regressão; a soma dos quadrados dos desvios explicados SSR ; a soma dos quadrados dos desvios não-explicados SST .

Para compreender o uso da função PROJ.LOG, serão realizadas aplicações gradativas para o caso de regressão linear simples $\hat{y} = b \times m^x$, acompanhando o Exemplo 16.1, reproduzido na planilha Função PROJ.LOG incluída na pasta CAP_16, conforme apresentado na Figura 16.6.

	A	B	C	D	E	F	G
1	Função PROJ.LOG						
2							
3						Constante	Estatísticas
4						VERDADEIRO	FALSO
5						Função PROJ.LIN	1,0298
6						$b =$	1,0298
7						$a =$	162,7431
8							
9						Constante	Estatísticas
10						VERDADEIRO	VERDADEIRO
11						Função PROJ.LIN	1,0298
12						1	2
13						1,0298	162,7431
14						0,0061	0,1771
15						0,7411	0,1955
16						22,9029	8
17						0,8754	0,3058
18							

Figura 16.6 • A função PROJ.LOG resolvendo o Exemplo 16.1

PROJ.LOG(*série_y*; *série_x*; VERDADEIRO; FALSO)

Informando o argumento *constante* como VERDADEIRO estamos definindo que a equação de regressão é $\hat{y} = b \times m^x$. Quando o argumento *estatísticas* é FALSO, a função PROJ.LOG fornece somente os coeficientes da reta de regressão. No intervalo de células F3:G7 da planilha Função PROJ.LOG tratamos deste caso. Na célula G5 digitamos a fórmula: =PROJ.LOG(B4:B13;C4:C13;F4;G4) obtendo como resultado o valor 9,7381.

Na realidade a função PROJ.LOG apresentou na célula G5 uma *matriz* com dois resultados, apresentando na célula apenas um dos dois resultados. Para ver a *matriz* com os resultados registrados na célula G5 procedemos da seguinte maneira:

- Selecionamos com o cursor a célula G5.
- Depois, pressionando primeiro a tecla de função F2 e depois a tecla de função F9 aparecerá: = {1,02984763144694;162,743118210196}.

O primeiro valor, já apresentado na célula G5, corresponde ao valor do coeficiente m e o segundo valor, separado por um ponto, símbolo (.), corresponde ao valor do coeficiente b . Para extrair cada um desses valores de forma separada devemos usar a função ÍNDICE¹⁹, como realizado nas células G6 e G7, com as fórmulas:

- Célula G6: =ÍNDICE(PROJ.LOG(B4:B13;C4:C13;F4;G4);1;1) obtendo o valor do coeficiente m igual a 1,02984763144694.

- Célula G7: =ÍNDICE(PROJ.LOG(B4:B13;C4:C13;F4;G4);1;2) obtendo o valor do coeficiente b igual a 162,743118210196.

Finalmente, a equação da reta de regressão é $\hat{y} = 162,743118 \times 1,029848^x$.

PROJ.LOG(*série_y*; *série_x*; VERDADEIRO; VERDADEIRO)

Quando o argumento *constante* é VERDADEIRO definimos a equação de regressão $\hat{y} = b \times m^x$. Se o argumento *estatísticas* for VERDADEIRO, a função PROJ.LOG fornecerá uma *matriz* com 10 resultados, os dois coeficientes da reta de regressão mais 8 resultados de interesse na regressão linear.

No intervalo de células F9:G17 da planilha Função PROJ.LOG tratamos deste caso. Na célula G11 digitamos a fórmula: =PROJ.LOG(B4:B13;C4:C13;F10;G10) obtendo como resultado o valor 1,029848. Na realidade, a função PROJ.LIN registrou na célula G11 uma *matriz* de 5 linhas e 2 colunas, apresentando, apenas, um dos dez resultados disponíveis na célula. Para ver os dez resultados da *matriz* registrada na célula G11 procedemos da seguinte maneira:

- Selecionamos com o cursor a célula G11.
- Depois, pressionando primeiro a tecla de função F2 e depois a tecla de função F9 aparecerá na célula e/ou na barra de fórmulas a seguinte expressão:

```
= {1,02984763144694;162,743118210196;
0,00614558038423411;0,177073784018667;
0,741124258554309;0,195502880178764;
22,9028569278991;8;
0,875378709738488;0,305771009265536}
```

¹⁹ Em inglês, a função de procura e referência ÍNDICE é INDEX. A sintaxe desta função é ÍNDICE(*matriz*; *linha*; *coluna*).

Nesta expressão existem 5 grupos de resultados separados pelo ponto e vírgula, símbolo (;). Cada um dos 5 grupos está composto por dois resultados separados pelo ponto, símbolo (.). O valor apresentado pela função PROJ.LIN na célula G11, é o primeiro resultado do primeiro grupo de resultados, que corresponde ao valor do coeficiente b , sendo que o segundo, separado pelo símbolo (.), corresponde ao valor do coeficiente a . Para extrair cada um desses valores de forma separada devemos usar a função ÍNDICE, formando a tabela limitada pelas células E12 e G17. A primeira coluna da tabela (com valores 1, 2, 3, 4, 5) identifica cada grupo de resultados, e a primeira linha da tabela (com valores 1 e 2) identifica o primeiro e o segundo resultado de cada grupo de resultados. Para extrair todos os resultados, registramos na célula F13 a fórmula:

=ÍNDICE(PROJ.LOG(\$B\$4:\$B\$13;\$C\$4:\$C\$13;\$F\$10;\$G\$10);\$E13;\$F\$12)

Depois copiamos esta fórmula até a célula G17, obtendo a tabela do intervalo F13:G17 da planilha **Função PROJ.LOG**, Figura 16.6. Cada uma das 10 células da tabela, separadas em 5 grupos (as 5 linhas da tabela) tem o seguinte significado:

Coefficiente $m=1,029848$	Coefficiente $b=162,7431$
Erro padrão do coeficiente m $S_m=0,006146$	Erro padrão do coeficiente b $S_b=0,177074$
Coefficiente de determinação $r^2=0,7385$	Erro padrão da estimativa $S_e=65,17343$
F observado =22,90286	Graus de liberdade da regressão 8
Soma dos quadrados dos desvios explicados $SSR=0,875379$	Soma dos quadrados dos desvios não-explicados $SSE=0,305771$

Análise da Condição *constante* = FALSO

Quando o argumento *constante* da função PROJ.LOG é FALSO, a curva exponencial de regressão usada, aplicando o método dos quadrados mínimos, é $\hat{y} = m_1^x \times m_2^x \times \dots \times m_n^x$, isto é, $b=1$. Mudando o valor dos argumentos *constante* nas células F4 e F10 da planilha **Função PROJ.LOG** o leitor pode ver os resultados para os dois valores desse argumento.

A partir da coluna J da planilha **Função PROJ.LOG** foram repetidos os mesmos cálculos informando as séries de informações como matrizes.

A seguir apresentaremos a função estatística CRESCIMENTO.

CRESCIMENTO(*série_y*; *série_x*; *x's*; *constante*)

A função estatística CRESCIMENTO²⁰ dá como resultado o valor projetado \hat{y} da curva exponencial de regressão, para um único ou um grupo de valores de x , denominado como $x's$, quando são conhecidas as duas séries de dados ou observa-

ções, *série_y* e *série_x*. Ao usar esta função deve-se tomar cuidado em fornecer os dados na ordem correta, o primeiro argumento *série_y* se refere à série de valores da variável dependente y , e o argumento *série_x* se refere à série de valores da variável independente x . Se o argumento *constante* for:

- VERDADEIRO (ou omitido) a função fornece um único ou um grupo de valores da curva exponencial de regressão $\hat{y} = b \times m^x$.
- FALSO, a função fornece apenas o coeficiente b da reta de regressão $\hat{y} = m^x$, isto é $b=1$.

A função CRESCIMENTO é equivalente à função TENDÊNCIA apresentada. A partir da coluna L1 da planilha **Função PROJ.LOG** incluída na pasta CAP_16, o leitor tem uma aplicação desta função.

FERRAMENTA DE ANÁLISE REGRESSÃO

A ferramenta *Regressão* realiza a *análise da regressão linear múltipla*, incluindo o caso de *regressão linear simples*. Para compreender seu uso, vamos resolver o Exemplo 16.1 com essa ferramenta, conforme registrado na planilha **Ferramenta Regressão** a partir da célula E4. Começamos por abrir o menu **Ferramentas** na versão 7 do Excel, ou **Utilitários** na versão 5 em português, onde escolhemos **Analisar Dados**. Recebemos a caixa de diálogo com todas as ferramentas de análise disponíveis; escolhemos *Regressão*, recebendo em seqüência a caixa de diálogo desta ferramenta, Figura 16.7. As informações necessárias são as seguintes:

- **Intervalo Y de entrada:** Informamos o intervalo da série de observações dependentes, incluindo os títulos.
- **Intervalo X de entrada:** Informamos o intervalo da série de observações independentes, incluindo os títulos. O número máximo de observações independentes é 16.
- **Constante é zero:** Seleccionamos esta opção quando queremos que a linha de regressão passe pela origem.
- **Rótulos:** Seleccionamos este item pois os intervalos das séries incluem os nomes das séries.
- **Nível de confiança:** Informamos o valor do nível de significância 0,05.
- **Em Opções de saída** devemos informar o endereço da célula inicial a partir da qual a ferramenta de análise apresentará os resultados. Para isso temos três formas de informar esse endereço:
 - Para receber os resultados na mesma planilha onde foram informados os dados da série escolhemos **Intervalo de saída**. A seguir registramos o endereço E4, digitando ou *clikando* com o botão esquerdo do *mouse* na célula E4.
 - Para receber os resultados em outra planilha da mesma pasta escolhemos **Nova planilha**, registrando a seguir o endereço completo da célula a partir da qual a ferramenta de análise apresentará os resultados.

²⁰ Em inglês, a função CRESCIMENTO é GROWTH.

- Para receber os resultados em outra planilha de outra pasta, escolhemos **Nova pasta de trabalho**, registrando a seguir o endereço completo da célula a partir da qual a ferramenta de análise apresentará os resultados.

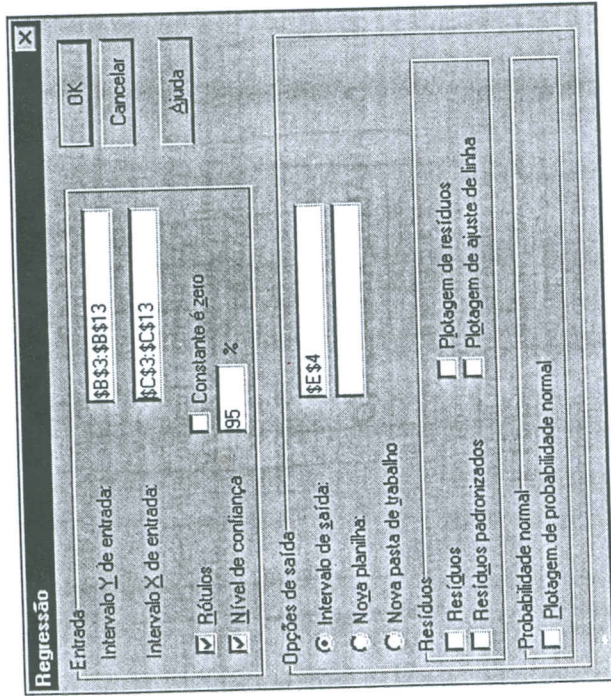


Figura 16.7 • Caixa de diálogo da ferramenta Regressão

As restantes seleções disponíveis na caixa de diálogo da ferramenta *Regressão* serão apresentadas mais adiante. Completadas as informações na caixa de diálogo da Figura 16.7, pressionamos o botão **OK** e recebemos a partir da célula E4 a tabela *Regressão* com os resultados divididos em três partes, com os significados apresentados a seguir explicados, na sua maioria, pelos desenvolvimentos teóricos desenvolvidos neste capítulo, Figura 16.8.

ESTATÍSTICA DE REGRESSÃO

A partir da célula E6, a ferramenta *Regressão* apresenta o primeiro grupo de resultados, que a seguir comentamos.

- **R múltiplo**: é o *coeficiente de correlação r* igual a 0,859366. É o valor positivo da raiz quadrada do coeficiente de determinação, combinado com o sinal do coeficiente *b*.
- **R-Quadrado**: este resultado é o *coeficiente de determinação r²* da regressão linear, que pode ser calculado com a função estatística **RQUAD**, cujo valor é 0,73851.

- **R-quadrado ajustado**: este valor é o *coeficiente de determinação ajustado \bar{r}^2* , medida usada em regressão linear múltipla. Partindo da regressão linear simples, com uma única variável independente (ou *explicatória*), sabemos que o *coeficiente de determinação* dá a porcentagem de explicação dessa regressão. Ao adicionar uma nova (ou mais) variável independente, se demonstra que o valor de r^2 não pode diminuir, aumentando em alguns casos. O *coeficiente de determinação ajustado \bar{r}^2* tenta compensar o aumento natural de explicação r^2 ao aumentar o número de variáveis independentes, sendo calculado:

$$\bar{r}^2 = r^2 - \frac{k}{n - k - 1} \times (1 - r^2)$$

onde, *n* é o número de observações, e *k* é o número de variáveis independentes. Substituindo os dados do nosso exemplo teremos:

$$\bar{r}^2 = 0,73851 - \frac{1}{10 - 1 - 1} \times (1 - 0,73851) = 0,705824$$

Pode-se ver que na medida que *n* aumenta, \bar{r}^2 se aproxima de r^2 .

D	E	F	G	H	I	J	K
1	Ferramenta de Análise						
2	RESUMO DOS RESULTADOS						
3	Estatística de regressão						
4	R múltiplo	0,85936613					
5	R-Quadrado	0,73851014					
6	R-quadrado ajustado	0,7058239					
7	Erro padrão	65,1734299					
8	Observações	10					
9	ANOVA						
10		gl	SQ	MQ	F	F de significação	
11	Regressão	1	95969,3923	95969,3923	22,59392	0,001439122	
12	Resíduo	8	33980,6077	4247,57596			
13	Total	9	129950				
14	Coeficientes Erro padrão Stat valor-P 95% inferiores 95% superiores						
15	Interseção	117,070158	59,0236508	1,98323656	0,0826337	-19,05300997	253,1933262
16	x	9,73814229	2,04870922	4,75330624	0,0014391	5,013607314	14,46247727

Figura 16.8 • Resultados do Exemplo 16.1 com a ferramenta de análise Regressão

- **Erro padrão**: é o *erro padrão da estimativa S_e* , definido como:

$$S_e = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - 2}} = \sqrt{\frac{SSE}{n - 2}}$$

O erro padrão da estimativa pode ser calculado com a função estatística EPADYX, obtendo o resultado é 65,17343. Atendidas as premissas da regressão linear, se espera que aproximadamente 95% das observações y se encontrem dentro do intervalo $\pm 2S_e$ de seus respectivos valores projetados \hat{y} da reta de regressão.

- **Observações:** é o número de observações das séries que têm o mesmo tamanho.

ANOVA

A partir da célula E13 da planilha **Ferramenta Regressão**, está registrada a tabela ANOVA já estudada, acrescida do valor F de significação, célula J15, que é o p -value do F observado ou a probabilidade $P(F \geq F \text{ observado})$, cujo valor é igual a 0,00143912 no Exemplo 16.1. Sabemos que esse valor da distribuição F , pode ser obtido com a função FDIST, digitando em qualquer célula de uma planilha a fórmula =DISTF(22,59392;1;8) que dá o resultado 0,00143912.

Portanto, a partir da definição do nível de significância a decisão de aceitar ou não a regressão pode ser realizada comparando o valor F de significação=0,00143912 com o nível de significância adotado=0,05. Como o valor do F de significação é menor que o valor do nível de significância escolhido, devemos aceitar a regressão.

INTERSEÇÃO e X

A partir da célula E19, a ferramenta **Regressão** apresenta o último grupo de resultados, que a seguir comentamos. Para compreender os resultados seguintes, a partir da célula E23 são apresentados os mesmos resultados usando somente fórmulas, baseadas nos conceitos desenvolvidos no livro.

- **Coefficientes:** Na coluna F , a partir da célula F20 são registrados todos os valores dos coeficientes da reta de regressão linear. O valor da célula F20 corresponde ao valor do coeficiente a , e o resultado da célula F21 corresponde ao valor do coeficiente b . No caso de regressão linear múltipla, os outros coeficientes serão apresentados na coluna F , células F22 em diante.
- **Erro padrão:** são os erros padrões dos coeficientes. Com os dados do exemplo que estamos desenvolvendo, os erros padrões são:
 - Do coeficiente a é obtido da expressão $S_a = S_e \sqrt{\frac{1}{n} + \frac{\bar{X}^2}{(n-1) \times S_x^2}} = 59,0298$.
 - Do coeficiente b é obtido da expressão $S_b = \frac{S_e}{\sqrt{(n-1) \times \text{Var}(x)}} = 2,04871$.
- **Stat t** (deve-se entender como Estatística t ou t observado): são obtidos das seguintes relações:

- Do coeficiente a : $t_a = \frac{a}{S_a} = 1,983237$.
- Do coeficiente b : $t_b = \frac{b}{S_b} = 4,753306$.

- **Valor -P** (deve-se entender como p -value): sabemos que o p -value é igual ao valor de probabilidade $P(t \geq t \text{ observado})$. Usando a função estatística DISTT do Excel obtemos os seguintes resultados:
 - Para o coeficiente a =DISTT(1,98323656;9;2)=0,0786487. Este valor é diferente do obtido pela ferramenta **Regressão** igual a 0,0826337, que corresponde ao caso da distribuição com 8 graus de liberdade.
 - Para o coeficiente b =DISTT(4,75330624;9;2)=0,0010394. Este valor é diferente do obtido pela ferramenta **Regressão** igual a 0,0014391, que corresponde ao caso da distribuição com 8 graus de liberdade.
- **95% inferiores:** é o valor mínimo de cada coeficiente de regressão. Para o intervalo de confiança escolhido igual a 95%, os valores mínimos dos coeficientes de regressão são:
 - Coeficiente a : $a_{\min} = a - t \times S_a = 117,070158 - 0,07864787 \times 590298508 = 0$.
 - Coeficiente b : $b_{\min} = b - t \times S_b = 9,73814229 - 4,75330624 \times 2,04870922 = 0$.
- **95% superiores:** é o valor máximo de cada coeficiente de regressão. Para o intervalo de confiança escolhido igual a 95%, os valores máximos dos coeficientes de regressão são:
 - Coeficiente a : $a_{\max} = a + t \times S_a = 234,1403162$.
 - Coeficiente b : $b_{\max} = b + t \times S_b = 19,47628458$.

COMPLETANDO OS RESULTADOS

Vamos completar os resultados que a ferramenta de análise **Regressão** tem disponíveis, escolhendo as restantes alternativas da caixa de diálogo dessa ferramenta. As informações necessárias são as seguintes, Figura 16.9:

- **Resíduos:** para cada valor de x , é a diferença entre o valor observado e o valor projetado pela reta de regressão.
- **Resíduos padronizados:** para cada valor de x , é o resultado de dividir o valor do resíduo pelo erro da estimativa da regressão.
- **Plotagem de resíduos:** é o gráfico dos resíduos para cada valor de x .
- **Plotagem de ajuste de linha:** é o gráfico de dispersão xy contendo as observações e a reta de regressão, equivalente ao gráfico construído no Exemplo 16.2.
- **Plotagem de probabilidade normal:** é o gráfico de cada observação y em função de seu correspondente percentil. O percentil de cada observação da série y ordenada de forma crescente é obtido com a expressão: $p = \frac{x-1}{x-1} \times 95\% + 5\%$.

Nessa relação, n é o número total de observações da série, x é a ordem de uma determinada observação, e p é o percentil em percentos dessa observação numa escala de 5 a 95%. Esta expressão é equivalente à expressão apresentada no Capítulo 4 quando estudamos medidas de ordenamento.

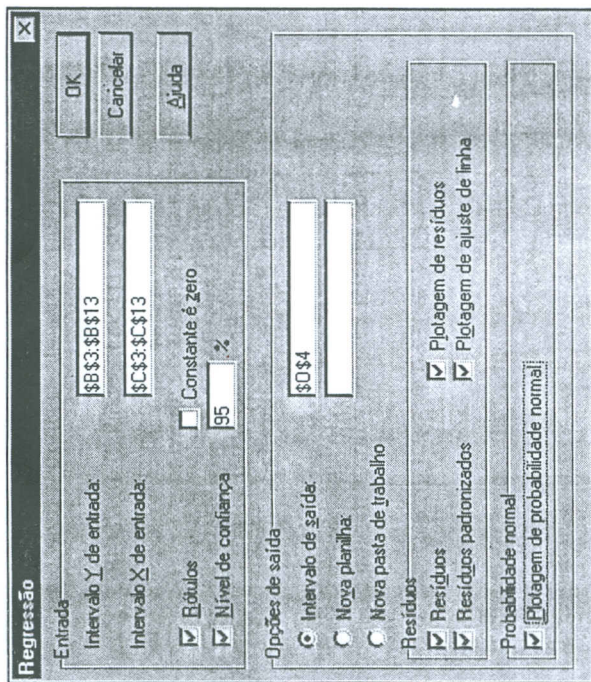


Figura 16.9 • Caixa de diálogo da ferramenta Regressão, escolha completa

A partir da célula O4 a ferramenta Regressão apresentou todos os resultados, e a Figura 16.10 apresenta somente os resultados de resíduos e probabilidades. Na Figura 16.11 são apresentadas dois dos três gráficos construídos pela ferramenta regressão.

	N	O	P	Q	R	S	T	U
24	RESULTADOS DE RESÍDUOS							
25	Observação	Previsto(a) y	Resíduos	Resíduos padrão	Percentil	y		
26	1	409,2144269	20,78557312	0,318927102	5	195		
27	2	321,5711462	13,42885375	0,206047983	15	210		
28	3	457,9051383	62,09486166	0,952763446	25	270		
29	4	526,0721344	-36,0721344	-0,553479147	35	335		
30	5	477,3814229	-7,38142292	-0,113258163	45	400		
31	6	311,8330004	-101,8330004	-1,562492631	55	430		
32	7	194,9752964	0,024703557	0,000379043	65	470		
33	8	282,6185771	-12,6185771	-0,19361536	75	480		
34	9	457,9051383	-57,9051383	-0,888477689	85	490		
35	10	360,5237154	119,4762846	1,833205415	95	520		
36	RESULTADOS DE PROBABILIDADE							
37								
38								

Figura 16.10 • Resíduos do Exemplo 16.1 com a ferramenta de análise Regressão

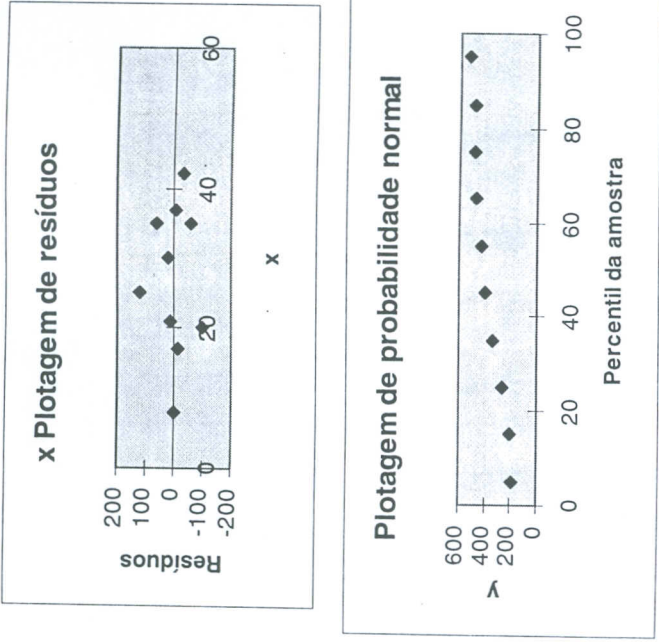


Figura 16.11 • Gráficos da ferramenta de análise Regressão

Para compreender os resultados da ferramenta, a partir da célula O40 são apresentados os mesmos resultados usando somente fórmulas, baseadas nos conceitos desenvolvidos no livro.

PROBLEMAS

Problema 1

Conhecidos os valores das rentabilidades das ações tipo ON e PN de um grupo de empresas registradas na tabela seguinte, pede-se obter a reta de regressão linear das rentabilidades das ações PN em função das ações ON.

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

R: $PN = 4,7014 + 1,4155 \times ON$

Problema 2

Continuando com o Problema 1. Considerando um nível de significância igual a 0,05, pede-se verificar se a regressão deve ser aceita aplicando a distribuição F. R: Aceitar regressão linear. Porém, analise o diagrama de dispersão das rentabilidades das ações tipo PN em função das ações ON.