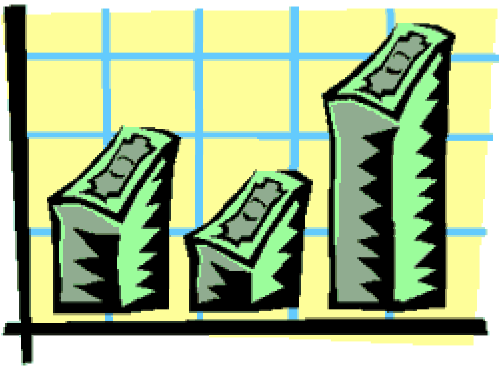
# REGRESSÃO LINEAR SIMPLES – PARTE II



- Erro Padrão de Estimativa.
- Correlação.



## Francisco Cavalcante(f_c_a@uol.com.br)

- Administrador de Empresas graduado pela EAESP/FGV.
- É Sócio-Diretor da Cavalcante & Associados, empresa especializada na elaboração de sistemas financeiros nas áreas de projeções financeiras, preços, fluxo de caixa e avaliação de projetos. A Cavalcante & Associados também elabora projetos de capitalização de empresas, assessora na obtenção de recursos estáveis e compra e venda de participações acionárias.
- O consultor Francisco Cavalcante já desenvolveu mais de 100 projetos de consultoria, principalmente nas áreas de planejamento financeiro, formação do preço de venda, avaliação de empresas e consultoria financeira em geral.

## Fábio Vianna(f_c_a@uol.com.br)

- Administrador de Empresas graduado pela EAESP/FGV.
- Há três anos é consultor da Cavalcante & Associados, especializado na elaboração de sistemas de projeções financeiras sempre com o apoio do microcomputador
- Também é responsável pelo planejamento/coordenação de cursos e seminários, tendo sido responsável pelo planejamento de mais de 150 cursos/seminários realizados nacionalmente tanto pela Cavalcante & Associados como por empresas parceiras.

# ÍNDICE

# APRESENTAÇÃO

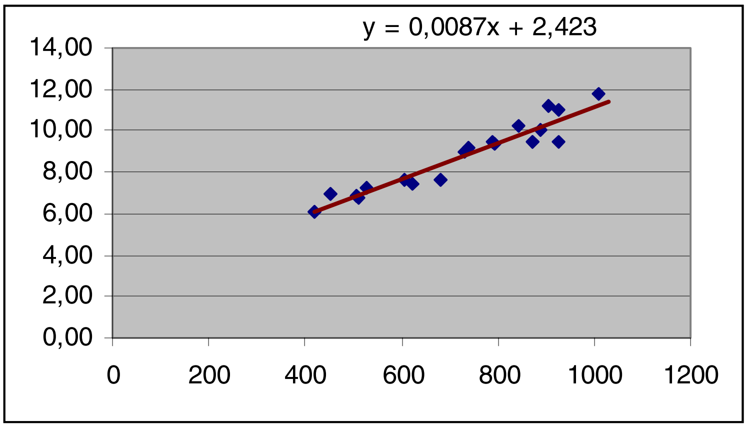
*Damos continuidade neste **Up-To-Date®** ao iniciado na semana passada. Nele, vamos abordar alguns indicadores estatísticos relevantes para a elaboração de uma regressão linear. Para facilitar o entendimento, vamos manter o mesmo exemplo apresentado no **Up-To-Date® 151**.*

# ERRO PADRÃO DE ESTIMATIVA

Quando calculamos uma regressão linear, é preciso que saibamos que é uma **previsão.** Isto quer dizer que nem todos os pontos do gráfico estarão sobre a reta calculada.

Então, é necessário sabermos calcular o quanto os valores calculados estão variando em torno da linha de regressão calculada.

Veja o gráfico abaixo:




Os pontos são os valores de vendas realizadas e a reta foi calculada com base nestas observações.

Note que nem todos os pontos do gráfico estão "colados" na reta. Isto comprova o que dissemos acima, ou seja, que estamos fazendo uma previsão.

Para calcularmos, usamos a função EPADYX, que calcula o erro padrão da estimativa.

Sua fórmula é: EPADYX(valores_conhecidos_y,valores_conhecidos_x), onde valores_conhecidos_y corresponde à variável dependente, isto é, vendas e valores_conhecidos_x corresponde à variável independente, isto é, clientes.

Supondo-se a tabela a seguir:

| | **A** | **B** | **C** |
|---|---|---|---|
| **1** | **Semana** | **Clientes** | **Vendas** |
| **2** | | **X** | **Y** |
| **3** | | | |
| **4** | 1 | 907 | 11,20 |
| **5** | 2 | 926 | 11,05 |
| **6** | 3 | 506 | 6,84 |
| **7** | 4 | 741 | 9,21 |
| **8** | 5 | 789 | 9,42 |
| **9** | 6 | 889 | 10,08 |
| **10** | 7 | 874 | 9,45 |
| **11** | 8 | 510 | 6,73 |
| **12** | 9 | 529 | 7,24 |
| **13** | 10 | 420 | 6,12 |
| **14** | 11 | 679 | 7,63 |
| **15** | 12 | 872 | 9,43 |
| **16** | 13 | 924 | 9,46 |
| **17** | 14 | 607 | 7,64 |
| **18** | 15 | 452 | 6,92 |
| **19** | 16 | 729 | 8,95 |
| **20** | 17 | 794 | 9,33 |
| **21** | 18 | 844 | 10,23 |
| **22** | 19 | 1010 | 11,77 |
| **23** | 20 | 621 | 7,41 |

Nossa fórmula seria: =EPADYX(C4:C23;B4:B23)

O que retornaria 0,501.

O valor de erro padrão está relacionado à medida da variável dependente y. Isto quer dizer que o erro padrão é de aproximadamente 0,501 unidades de vendas, isto é $0,501. Em outras palavras, a variação média ao redor da reta é de 0,501.

# CORRELAÇÃO

É preciso calcular também o quanto uma variável independente prevê bem a dependente, isto é, a intensidade com que a variável dependente determina a independende. Isto é feito pelo chamado coeficiente de correlação.

Para calcular o coeficiente de correlação, pé preciso antes calcular a chamada soma total dos quadrados, que nada mais é do que a soma da variação dos valores Y em relação à sua média aritmética. Esta soma ainda é dividida entre variações explicadas e variações não explicadas.

Então temos:

Soma dos quadrados = Soma dos quadrados explicados (ou devidos á regressão) + soma dos quadrados inexplicados (ou dos resíduos)

Ou então; STQ = SQReg + SQR

Para calcularmos STQ precisamos somar a diferença entre os valores de y observados e sua média, como a tabela a seguir:

Exemplo: no 1º ítem temos $(11{,}20 - 8{,}81)^2 = 5{,}73$ e daí por diante.

A soma destes quadrados dá 51,36 que é o STQ.

| Semana | Clientes | Vendas | |
|---|---|---|---|
| | **X** | **Y** | **(Y - Média Y)^2** |
| 1 | 907 | 11,20 | 5,73 |
| 2 | 926 | 11,05 | 5,04 |
| 3 | 506 | 6,84 | 3,86 |
| 4 | 741 | 9,21 | 0,16 |
| 5 | 789 | 9,42 | 0,38 |
| 6 | 889 | 10,08 | 1,62 |
| 7 | 874 | 9,45 | 0,42 |
| 8 | 510 | 6,73 | 4,31 |
| 9 | 529 | 7,24 | 2,45 |
| 10 | 420 | 6,12 | 7,21 |
| 11 | 679 | 7,63 | 1,38 |
| 12 | 872 | 9,43 | 0,39 |
| 13 | 924 | 9,46 | 0,43 |
| 14 | 607 | 7,64 | 1,36 |
| 15 | 452 | 6,92 | 3,56 |
| 16 | 729 | 8,95 | 0,02 |
| 17 | 794 | 9,33 | 0,28 |
| 18 | 844 | 10,23 | 2,03 |
| 19 | 1010 | 11,77 | 8,79 |
| 20 | 621 | 7,41 | 1,95 |
| **SOMA** | **14.623,00** | **176,11** | **51,36** |
| | **MÉDIA:** | **8,81** | |

Para calcularmos o valor de SQR, isto é, as variações inexplicadas, precisamos somar o quadrado da diferença entre os valores de y calculados com base na fórmula de regressão e os valores de y observados. Veja a tabela:

Para o 1º valor o cálculo é de: $(10,34 - 11,20)^2 = 0,74$

| Semana | Clientes | Vendas | | (Y calculado - Y observado)^2 |
|---|---|---|---|---|
| | X | Y | | |
| | | OBSERVADO | CALCULADO | |
| 1 | 907 | 11,20 | 10,34 | 0,74 |
| 2 | 926 | 11,05 | 10,51 | 0,30 |
| 3 | 506 | 6,84 | 6,84 | 0,00 |
| 4 | 741 | 9,21 | 8,89 | 0,10 |
| 5 | 789 | 9,42 | 9,31 | 0,01 |
| 6 | 889 | 10,08 | 10,18 | 0,01 |
| 7 | 874 | 9,45 | 10,05 | 0,36 |
| 8 | 510 | 6,73 | 6,88 | 0,02 |
| 9 | 529 | 7,24 | 7,04 | 0,04 |
| 10 | 420 | 6,12 | 6,09 | 0,00 |
| 11 | 679 | 7,63 | 8,35 | 0,52 |
| 12 | 872 | 9,43 | 10,04 | 0,37 |
| 13 | 924 | 9,46 | 10,49 | 1,06 |
| 14 | 607 | 7,64 | 7,72 | 0,01 |
| 15 | 452 | 6,92 | 6,37 | 0,30 |
| 16 | 729 | 8,95 | 8,79 | 0,03 |
| 17 | 794 | 9,33 | 9,35 | 0,00 |
| 18 | 844 | 10,23 | 9,79 | 0,19 |
| 19 | 1010 | 11,77 | 11,24 | 0,28 |
| 20 | 621 | 7,41 | 7,84 | 0,19 |
| **SOMA** | **14.623,00** | **176,11** | **176,11** | **4,53** |

Então o valor de SQR é de 4,53.

Por diferença temos o valor de SQReg = 51,36 – 4,53 = 46,83.

Então temos:

| | | |
|---|---|---|
| STQ = | SQReg + | SQR |
| 51,36 = | 46,83 + | 4,53 |

Para calcularmos o coeficiente de determinação, fazemos SQReg / STQ = 46,83 / 51,66 = 0,911

Isto quer dizer que 91% da variação de vendas da empresa pode ser explicada pela variação de clientes, e aproximadamente 9% da variação de vendas não é explicada por clientes e sim por fatores outros.

Outro indicador é o coeficiente de correlação, que dirá o quanto as variáveis dependentes e independentes estão associadas.

Isto quer dizer que, se o coeficiente de corrrelação for positivo, que as variáveis dependentes e independentes estão positivamente correlacionadas; se a dependente cresce, a independente cresce. Se o coeficiente de correlação for negativo as variáveis estão negativamente correlacionadas; se uma cresce a outra decresce. Se for zero, não há correlação entre elas.

Para calcular, basta extrair a raiz quadrada do coeficiente de determinação, o que em nosso exemplo dá 0,9549. O sinal é o mesmo do coeficiente b1 (calculado no último Up-To-Date).

Outra forma de calcular é através da fórmula CORREL do Excel;

CORREL(valores_y;valores_x).

Usando a mesma tabela da página 6, teríamos =CORREL(C4:C23;B4:B23).