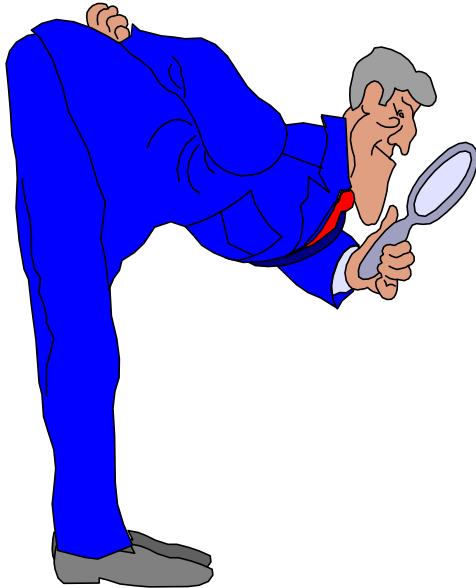


# UTILIZAÇÃO DE RECURSOS ESTATÍSTICOS AVANÇADOS DO EXCEL – PROJ.LIN



↳ *Fazendo regressão linear e logarítmica*

↳ *Relacionando variáveis e criando uma equação para explicá-las*

↳ *Como checar se as variáveis estão relacionadas*

↳ *Quando utilizar o PROJ.LIN*

## **Autores: 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

PÁG

◆ APRESENTAÇÃO	03
◆ REGRESSÃO LINEAR	04
◆ PROJ.LIN	05

## APRESENTAÇÃO

*"Muitas vezes sabemos que determinadas variáveis de um problema estão relacionadas entre si e utilizamos este conhecimento para fazer simulações. Por exemplo, podemos relacionar a renda com as vendas de determinado produto e, sabendo que haverá determinada variação na renda, estimar uma variação no volume de vendas deste produto.*

*O objetivo deste **Up-To-Date**® é apresentar recursos estatísticos do Excel que auxiliam na determinação de uma relação entre variáveis permitindo a criação de uma equação que explique o comportamento destas variáveis. Poderemos, por exemplo, criar uma equação que explique, com uma certa margem de erro, a relação entre a renda e as vendas de um produto evitando “chutes” acerca de qual será a variação do volume de vendas dada uma mudança na renda.”*

## REGRESSÃO LINEAR

Neste **Up-To-Date®** trataremos da regressão linear.

Você pode acompanhar a execução deste exercício através da planilha enviada junto com este **Up-To-Date®**.

A regressão linear irá apresentar uma equação de uma reta.

O objetivo é criar uma equação que vai explicar a movimentação da(s) variável(is) envolvida(s).

Chegaremos a uma equação que diga, por exemplo, que as vendas seguem a seguinte relação com a renda:  **$0,30 \times \text{crescimento da renda} + 1,5$** .

Você já pode notar que ao criarmos uma equação que explique como uma variável se comporta eliminamos a "técnica do chute".

## PROJ.LIN

A fórmula PROJ.LIN é a que faz a regressão linear.

Para utilizá-la, vamos usar o seguinte exemplo:

Comissão (%)	Gastos em propaganda (R\$)	Vendas totais (R\$)
5,00%	R\$ 1.000	R\$ 10.020
6,00%	R\$ 1.200	R\$ 12.500
7,00%	R\$ 1.400	R\$ 13.500
5,50%	R\$ 1.250	R\$ 11.000
4,75%	R\$ 1.150	R\$ 10.500
4,00%	R\$ 1.300	R\$ 10.500
5,50%	R\$ 1.350	R\$ 11.750
4,50%	R\$ 1.000	R\$ 10.000
4,40%	R\$ 900	R\$ 8.500
5,00%	R\$ 1.100	R\$ 11.050
5,25%	R\$ 1.150	R\$ 10.150
6,75%	R\$ 1.300	R\$ 13.250
6,00%	R\$ 1.275	R\$ 10.200
4,10%	R\$ 950	R\$ 9.700
6%	R\$ 1.300	R\$ 13.000

Nesta tabela, colocamos uma série histórica dos gastos de uma empresa com propaganda, comissão dos vendedores e as vendas totais.

Creemos haver uma relação entre as 2 variáveis (comissão e gastos com propaganda) e as vendas totais da empresa.

Vamos utilizar o PROJ.LIN para:

- a) Verificar se realmente existe relação entre as variáveis e as vendas e
- b) Determinar a equação que explique esta relação.

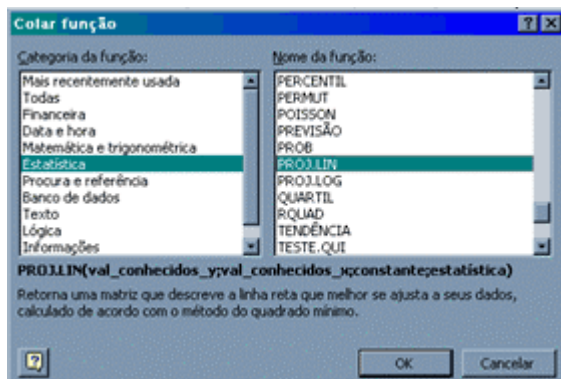
Primeiramente, vamos determinar a relação entre as variáveis.

Na pasta de trabalho que você recebeu, vá até a planilha Exemplo Proj.Lin.

Vá até a célula A21.

Agora, vamos inserir a fórmula PROJ.LIN.

Na célula A21, clique em Inserir, Função. Vamos selecionar do lado esquerdo (Categoria), Estatística, uma vez que a função é estatística. Do lado direito, selecione PROJ.LIN.



Ir aparecer a seguinte tela:

No campo **Val\_conhecidos\_y**, vamos marcar toda a rea determinada pelas vendas (que vai de C4 at C18).

A frmula de regresso linear   $y = mx + b$ , isto : vendas = varivel x comisso + varivel x gasto em propaganda + constante b (falaremos dela daqui a pouco).

Ento os valores de y correspondem aos valores das vendas.

Em **Val\_conhecidos\_x**, temos as variveis que iro determinar o resultado de y (as vendas). Eles so 2 em nosso exemplo: comisso e gastos em propaganda. Vamos marcar ento a rea que vai de A4 at B18.

Em **constante**, colocamos 1 ou 2. Colocando 1, a planilha ir calcular o chamado coeficiente **b**, que  o valor que ir auxiliar no ajuste da reta de regresso linear. Por isso, recomendamos que ele sempre seja calculado, permitindo que se encontre uma reta mais precisa. Colocando 2, o **b** no ser calculado.

Em **estatística**, também colocamos 1 ou 2. Se colocarmos 1, a planilha calcula alguns indicadores estatísticos. Colocando 2, estes indicadores não serão calculados.

Como precisamos destes valores para saber se nossa relação de fato existe, então vamos calcular estes indicadores colocando 1 no campo especificado.

Depois disso, teremos a seguinte tela:

PROJ.LIN

Val\_conhecidos\_y C4:C18 = {10020;12500;13500}

Val\_conhecidos\_x A4:B18 = {0,05.1000;0,06.1200}

Constante 1 = VERDADEIRO

Estatística 1 = VERDADEIRO

Retorna uma matriz que descreve a linha reta que melhor se ajusta a seus dados, calculado de acordo com o método do quadrado mínimo.

**Estatística** é um valor lógico: retorna estatística de regressão adicional = VERDADEIRO; retorna coeficientes-m e a constante 'b' = FALSO ou não especificado.

Resultado da fórmula = 3,237249797

OK Cancelar

Agora, clicamos em OK. Vai aparecer na célula A21 o valor de 3,23. A fórmula de PROJ.LIN tem um característica, que é o fato de ser matricial. Portanto, é preciso “abrir” esta matriz:

Ainda na célula A21, vamos marcar a área entre as células A21 e C25.

Depois de marcá-la, clique em F2.



Agora, clique nesta sequência, sem soltar os botões:

CTRL

SHIFT

ENTER

Você terá a seguinte tabela:

	A	B	C
21	3,237249797	92999,60447	2317,885746
22	1,837561491	30267,03421	1507,920022
23	0,779618179	740,5883437	#N/D
24	21,22547607	12	#N/D
25	23283120,19	6581653,139	#N/D

O formato da tabela será:

5 linhas x (número de variáveis +1). Isto é, se tivermos 5 variáveis, nossa matriz será de 5 linhas por 6 colunas. Se tivermos 1 variável, nossa matriz terá 5 linhas por 2 colunas, e assim por diante.

Antes de montarmos nossa equação, vamos checar se as variáveis estão relacionadas.

A 1ª informação que precisamos olhar é a que está na 3ª linha da 1ª coluna (sempre): é o chamado R2, ou coeficiente de correlação (no nosso exemplo o valor está na célula A23). Ele nos diz se há relação entre as variáveis e o que queremos calcular. Quanto mais próximo de 1, melhor.

Em nosso exemplo, ele é de 0,779 (célula A23), o que podemos considerar como bom.

Agora vamos olhar outras informações para checar se nosso exemplo está bem “amarrado”.

Isto será feito analisando-se outras informações estatísticas.

A próxima é o chamado **F**, que irá dizer se o valor de **R2** ocorreu por acaso. Pode acontecer de termos um **R2** alto, mas as variáveis não estarem relacionadas, ou por estarmos usando uma amostra pequena, ou por casualidade.

Na tabela estatística gerada pela função **Proj.Lin**, o **F** está localizado sempre na 4ª linha da 1ª coluna. No nosso caso o valor é de 21,22 (célula A24 no nosso caso).

Para sabermos se ele é bom ou ruim, precisamos de uma tabela estatística chamada **de F - crítico**, encontrada em qualquer livro de estatística. Na pasta de trabalho que você recebeu há uma cópia desta tabela.

Para encontrarmos o valor, precisamos conhecer a interseção das linhas e colunas da tabela.

Na coluna, procuraremos o valor 2, que é o número de variáveis que estamos usando.

Na linha, procuramos o valor 12, que é calculado pela fórmula:  $n - (k+1)$ , onde  $n$  é o número de pontos (em nosso caso, 15) e  $k$  é o número de variáveis (em nosso caso, 2), então temos  $15 - (2+1) = 12$ .

Então, na tabela, vamos procurar o número que está na linha 12, coluna 2, que é de 3,88. O valor calculado foi de 21,22, que é maior que 3,88. Isto significa que a relação entre as variáveis não ocorreu por acaso.

Outro indicador que vamos calcular é o que determina a relevância de cada variável no resultado final. Isto nos permite identificar qual a variável que mais “pesou” na determinação do resultado final.

Para isto, vamos procurar as informações contidas sempre nas linhas 1 e 2. Vamos calcular para a variável 1 a razão entre 92,999 (célula B21) e 30,367 (célula B22), que será igual a 3,07, e para a variável 2 a razão entre 3,237 (célula A21) e 1,837 (célula A22) que será igual a 1,761.

Note que na tabela, as informações relacionadas à variável 2 estão na 1ª coluna, e as da variável 1 na 2ª coluna. Isto porque a tabela é sempre composta da seguinte forma:

Coluna 1 = variável n (última)

Coluna 2 = variável n-1 (penúltima)

Coluna 3 = variável n-2 (antepenúltima), etc.

Para sabermos a relevância de cada variável, precisamos analisar outra tabela, a tabela chamada de **teste-t**.

Na tabela de teste-t na 1ª linha, procuramos o nível de confiança que queremos (em nosso caso, será de 5%, significa a probabilidade de concluirmos de forma errada quanto à relevância de cada variável, neste caso, de 5%).

Na 1ª coluna, procuramos os chamados graus de liberdade, que estão indicados na tabela calculada na 4ª linha da 2ª coluna, e indica 12 (célula B24). Na tabela de teste-t, temos o valor de 1,78.(localizado na interseção entre a linha com o valor 12 e a coluna 5%).

Para a 1ª variável, temos o valor de 3,07, que é superior ao valor de 1,78 localizado na tabela **teste-t**. Já na 2ª variável, temos o valor de 1,76, que é inferior ao valor localizado na tabela **teste-t** e portanto, significa ser pouco útil na determinação do resultado final.

Depois de analisarmos estas informações, vamos montar nossa reta de regressão.

Nossa fórmula será a seguinte:

$$\text{Vendas} = \text{comissão} \times \text{variável 1} + \text{gastos em propaganda} \times \text{variável 2} + b$$

$$\text{Vendas} = \text{comissão} \times 92999,60 + \text{gastos com propagada} \times 3,237 + 2317,88$$

Na tabela criada, teremos sempre na 1ª linha os valores das variáveis, sendo que **sempre** na última coluna teremos o valor de b.

Para estimarmos as vendas dada uma comissão de 6,25% e gasto com propaganda de R\$1.100, teríamos:

$$\text{Vendas} = 0,0625 \times 92999,60 + 1100 \times 3,237 + 2317,88 = \text{R}\$11.691.$$

Se considerarmos os mesmos valores apresentados na tabela, teremos algumas diferenças entre o valor calculado e o apresentado. Isto acontece porque não há uma correlação perfeita entre as variáveis. Lembre-se que estamos colocando sobre uma nuvem de pontos uma reta, e que ela será o mais precisa possível, porém não passará por todos os pontos.