



## INTRODUÇÃO: ANÁLISE DE REGRESSÃO

Cavalcante Consultores > Biblioteca de Excel > Introdução: análise de regressão

### Introdução

O Excel conta com duas funções relacionadas a análise de regressão, que devolvem estatísticas diversas relacionadas ao ajuste de determinada curva a pontos dados. Estas informações podem ser usadas na criação e análise de um modelo de regressão adequado aos dados disponíveis nas suas planilhas. As funções são as seguintes:

- PROJ.LIN: regressão linear;
- PROJ.LOG: ajuste de curva exponencial.

Estas funções trabalham com pontos de coordenadas ( $x_1, x_2, x_3 \dots y$ ), resolvendo diferentes sistemas lineares pelo método dos mínimos quadrados. No caso da função PROJ.LIN, por exemplo, resolve-se o sistema linear  $Y = MX + b$ .

Já na função PROJ.LOG, o sistema desejado é  $y = b \cdot (m_1^{x_1}) \cdot (m_2^{x_2}) \cdot (m_3^{x_3}) \dots$ , mas como o método utilizado para a regressão é baseado na resolução de sistemas lineares, resolve-se a equação  $\ln(y) = x_1 \ln(m_1) + x_2 \ln(m_2) + x_3 \ln(m_3) \dots + \ln(b)$ .

Cada uma das funções acima devolve uma matriz de  $n$  colunas, onde  $n$  é o número de dimensões dos vetores utilizados, e duas ou cinco linhas - o número de linhas depende do usuário desejar obter apenas os valores de  $M$  e  $b$ , ou todas as informações da regressão.

Por devolver uma matriz, você deverá elaborar estas fórmulas em modo matricial. Leia o exemplo de aplicação de cada uma das funções para aprender a fazê-lo.

### Valores devolvidos pelas funções

Observe a tabela de valores devolvidos pelas funções PROJ.LIN e PROJ.LOG:

mn	mn-1	...	m1	b
Erro mn	Erro mn-1	...	Erro m1	Erro b
r2	Erro y	#N/D	#N/D	#N/D
Valor F	GL	#N/D	#N/D	#N/D
SQ regressão	SQ resíduos	#N/D	#N/D	#N/D

A primeira linha contém os coeficientes de  $M$  nas equações descritas anteriormente, ou seja, os valores que acompanham cada variável na curva de regressão. O último coeficiente é o valor independente  $b$ .

A segunda linha contém o erro padrão para cada variável. Por exemplo, se a variável  $b$  foi estimada como 5,00 com erro 0,25,  $b$  está no intervalo  $[5,00 - 0,25; 5,00 + 0,25]$ , ou seja,  $[4,75; 5,25]$ .

O valor em (3; 1) é o coeficiente de determinação, representado por  $r^2$ . Ele representa a qualidade da associação entre duas variáveis: trata-se de um número no intervalo  $[0; 1]$ , que indica associação fraca (quanto mais próximo de zero) ou forte (quanto mais próximo de um). Note que isto não indica uma relação de causalidade entre as variáveis.

O valor em (3; 2) é o erro padrão para a variável  $Y$ .

Os valores em (4; 1) e (4; 2) são, respectivamente, o *valor F observado* e o número de *graus de liberdade* da estimativa. A distribuição F é comumente utilizada em testes de hipóteses contra a hipótese de associação puramente aleatória dos dados em mãos: quando utilizadas com uma tabela para teste F, estas informações permitem a avaliação do modelo obtido de forma mais confiável que a pura análise de  $r^2$  e dos erros-padrão.

Os valores em (5; 1) e (5; 2) são, respectivamente, a *soma de quadrados da regressão* e a *soma de quadrados dos resíduos*. Estes valores são utilizados no cálculo dos parâmetros anteriores, e o processo de regressão está diretamente relacionado à minimização da soma dos quadrados dos resíduos.

### **Mais informações**

Procure um livro sobre cálculo numérico para informações sobre o método dos mínimos quadrados, e um sobre estatística para mais informações sobre a distribuição F. Para a análise estatística do modelo, você precisará saber manipular uma tabela de valores F críticos.

Cavalcante Consultores - soluções de treinamento e consultoria  
Av. Paulista, 2006, cj. 703/705, CEP 01310-926 - São Paulo, SP  
Telefone: (11) 3524 9222