Análise de Regressão Linear Múltipla I

Aula 04

Gujarati e Porter, 2011 – Capítulos 7 e 10 – tradução da 5ª ed.

Heij et al., 2004 – Capítulo 3

Wooldridge, 2011 – Capítulo 3 – tradução da 4ª ed.

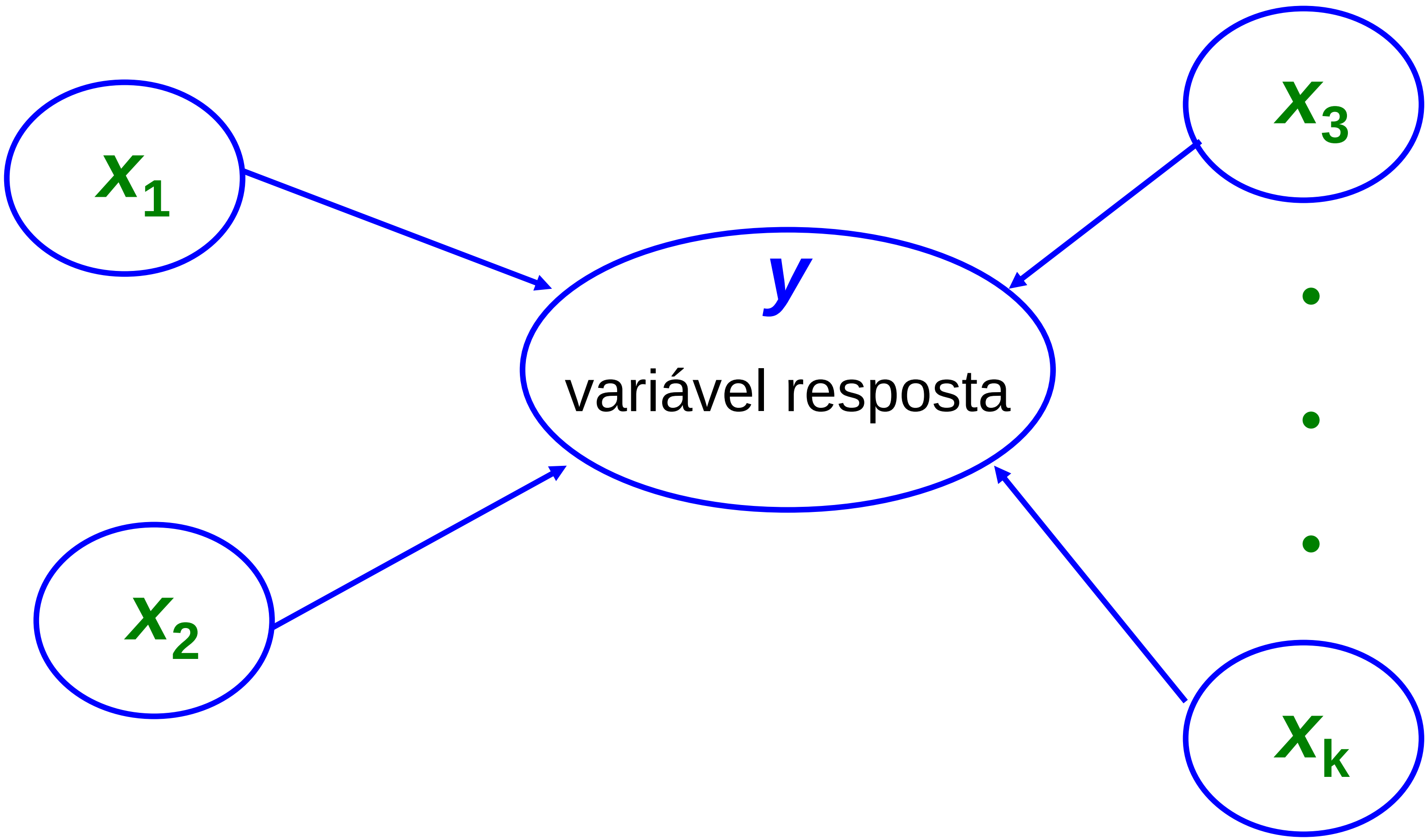
Introdução

Como pode ser visto anteriormente, o modelo de regressão linear simples, com uma variável explicativa (regressor), aplica-se a várias situações.

Entretanto, diversos problemas envolvem dois ou mais regressores influenciando o comportamento da variável resposta (dependente), *y*.

Chamamos Modelo de Regressão Linear Múltipla a qualquer modelo de regressão linear com duas ou mais variáveis explicativas.

Introdução



x_1, x_1, ..., x_k: variáveis explicativas (regressores)

Modelo de regressão linear múltipla

Vamos admitir que X_1, X_2, ..., X_k sejam as variáveis independentes e Y a variável dependente.

Dada uma amostra de *n* observações,

$$(x_{1i}, x_{2i}, ..., x_{ki}, y_i), \quad i = 1, 2, ..., n,$$

o modelo de regressão linear múltipla será dado por:

Modelo de regressão linear múltipla

$$E[y_i \mid x_{1i}, x_{2i}, \ldots, x_{ki}] = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki}\,,$$

i = 1, 2, ..., *n*

ou

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + \varepsilon_i\,,$$

i = 1, 2, ..., *n*.

em que *n* > (k+1).

Mínimos Quadrados Ordinários

Método dos Mínimos Quadrados

Para determinarmos os estimadores de mínimos quadrados de β_0, β_1, ..., β_k, devemos minimizar o erro quadrático total ($\Sigma\varepsilon_i^2$):

$$\boxed{\sum_{i=1}^{n}(\varepsilon_i)^2 = \sum_{i=1}^{n}\left(y_i - \beta_0 - \beta_1 x_{1i} - \cdots - \beta_k x_{ki}\right)^2}$$

Método dos Mínimos Quadrados

O mínimo da função

$$S(\beta_0,\ \beta_{1,}\ \ldots, \beta_k) = \sum_{i=1}^{n} (\varepsilon_i)^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{1i} - \cdots - \beta_k x_{ki})^2$$

é obtido derivando-a em relação a *β_0, β_1, ..., β_k*, e igualando o resultado a zero. Ou seja,

$$\frac{\partial}{\partial \beta_0} S(\beta_0,\ \beta_{1,}\ \ldots, \beta_k) = 0 \qquad \cdots \qquad \frac{\partial}{\partial \beta_k} S(\beta_0,\ \beta_{1,}\ \ldots, \beta_k) = 0$$

Equações Normais

$$\frac{\partial}{\partial \beta_0} S(\beta_0,\ \beta_{1,}\ \ldots, \beta_k) = -2\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \cdots - \hat{\beta}_k x_{ki}\right) = 0$$

$$\frac{\partial}{\partial \beta_1} S(\beta_0,\ \beta_{1,}\ \ldots, \beta_k) = -2\sum_{i=1}^{n}\left[\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \cdots - \hat{\beta}_k x_{ki}\right) x_{1i}\right] = 0$$

$\vdots$

$$\frac{\partial}{\partial \beta_k} S(\beta_0,\ \beta_{1,}\ \ldots, \beta_k) = -2\sum_{i=1}^{n}\left[\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \cdots - \hat{\beta}_k x_{ki}\right) x_{ki}\right] = 0$$

Abordagem Matricial

Devido à complexidade das fórmulas envolvidas, utilizaremos a abordagem matricial, que nos permitirá, entre outras coisas:

i. encontrar o vetor de estimadores;

ii. verificar as propriedades estatísticas de (i);

iii. obter a distribuição de probabilidades de (i);

qualquer que seja o número de regressores presentes no modelo.

Abordagem Matricial

Assim, a equação

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + \varepsilon_i \text{, } i = 1, 2, \ldots, n.$$

também pode ser escrita como

$$y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{21} + \ldots + \beta_k x_{k1} + \varepsilon_1$$

$$y_2 = \beta_0 + \beta_1 x_{12} + \beta_2 x_{22} + \ldots + \beta_k x_{k2} + \varepsilon_2$$

$$y_3 = \beta_0 + \beta_1 x_{13} + \beta_2 x_{23} + \ldots + \beta_k x_{k3} + \varepsilon_3$$

$$\vdots$$

$$y_n = \beta_0 + \beta_1 x_{1n} + \beta_2 x_{2n} + \ldots + \beta_k x_{kn} + \varepsilon_n$$

Abordagem Matricial

As igualdades anteriores podem ser alocadas facilmente em dois vetores colunas ($n\times 1$), descritos a seguir:

$$\underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}}_{(n\times 1)} = \underbrace{\begin{pmatrix} \beta_0 + \beta_1 x_{11} + \ldots + \beta_k x_{k1} + \varepsilon_1 \\ \beta_0 + \beta_1 x_{12} + \ldots + \beta_k x_{k2} + \varepsilon_2 \\ \vdots \\ \beta_0 + \beta_1 x_{1n} + \ldots + \beta_k x_{kn} + \varepsilon_n \end{pmatrix}}_{(n\times 1)}$$

Abordagem Matricial

Ainda,

$$\underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}}_{(n\times 1)} = \underbrace{\begin{pmatrix} \beta_0 + \beta_1 x_{11} + \ldots + \beta_k x_{k1} \\ \beta_0 + \beta_1 x_{12} + \ldots + \beta_k x_{k2} \\ \vdots \\ \beta_0 + \beta_1 x_{1n} + \ldots + \beta_k x_{kn} \end{pmatrix}}_{(n\times 1)} + \underbrace{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}}_{(n\times 1)}$$

Abordagem Matricial

Finalmente,

$$\underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}}_{(n\times 1)} = \underbrace{\begin{pmatrix} 1 & x_{11} & \cdots & x_{k1} \\ 1 & x_{12} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & \cdots & x_{kn} \end{pmatrix}}_{(n\times(k+1))} \cdot \underbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}}_{((k+1)\times 1)} + \underbrace{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}}_{(n\times 1)}$$

Abordagem Matricial

Vamos definir:

$$\underset{\sim}{\mathbf{y}} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

$$\underset{\sim}{\mathbf{X}} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{k1} \\ 1 & x_{12} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & \cdots & x_{kn} \end{pmatrix}$$

$$\underset{\sim}{\omega_{\mathbf{i}}} = \begin{pmatrix} 1 & x_{1i} & \cdots & x_{ki} \end{pmatrix}$$

$$\underset{\sim}{\boldsymbol{\beta}} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$$

$$\underset{\sim}{\boldsymbol{\varepsilon}} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Modelo de regressão linear múltipla

Assim, utilizando os resultados do *slide* anterior, podemos escrever o modelo de regressão linear múltipla como:

$$\underset{\sim}{\mathbf{y}} = \underset{\sim}{\mathbf{X}}\,\underset{\sim}{\boldsymbol{\beta}} + \underset{\sim}{\boldsymbol{\varepsilon}},$$

que é chamado Modelo Linear Geral.

Estimação

Método dos Mínimos Quadrados Ordinários

(MQO)

Estimação

Método dos Mínimos Quadrados Ordinários

Para determinarmos os estimadores de MQO de β_0, β_1, ..., β_k, devemos minimizar

$$S = \sum_{i=1}^{n} \left(\varepsilon_i\right)^2 = \varepsilon_1^2 + \ldots + \varepsilon_n^2 = \underset{\sim}{\boldsymbol{\varepsilon}}' \underset{\sim}{\boldsymbol{\varepsilon}}$$

ou, ainda,

$$S = \underset{\sim}{\boldsymbol{\varepsilon}}' \underset{\sim}{\boldsymbol{\varepsilon}} = \left(\underset{\sim}{\mathbf{y}} - \underset{\sim}{\mathbf{X}}\underset{\sim}{\boldsymbol{\beta}}\right)' \left(\underset{\sim}{\mathbf{y}} - \underset{\sim}{\mathbf{X}}\underset{\sim}{\boldsymbol{\beta}}\right)$$

Estimação

Método dos Mínimos Quadrados Ordinários

Abrindo a expressão anterior, vem que

$$S = \left(\underset{\sim}{\mathbf{y}} - \underset{\sim}{\mathbf{X}}\underset{\sim}{\boldsymbol{\beta}}\right)'\left(\underset{\sim}{\mathbf{y}} - \underset{\sim}{\mathbf{X}}\underset{\sim}{\boldsymbol{\beta}}\right) = \left(\underset{\sim}{\mathbf{y}}' - \underset{\sim}{\boldsymbol{\beta}}'\underset{\sim}{\mathbf{X}}'\right)\left(\underset{\sim}{\mathbf{y}} - \underset{\sim}{\mathbf{X}}\underset{\sim}{\boldsymbol{\beta}}\right) =$$

$$= \underset{\sim}{\mathbf{y}}'\underset{\sim}{\mathbf{y}} - \underset{\sim}{\mathbf{y}}'\underset{\sim}{\mathbf{X}}\underset{\sim}{\boldsymbol{\beta}} - \underset{\sim}{\boldsymbol{\beta}}'\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{y}} + \underset{\sim}{\boldsymbol{\beta}}'\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}}\underset{\sim}{\boldsymbol{\beta}}$$

Estimação

Métodos dos Mínimos Quadrados Ordinários

Como

$$\underset{\sim}{y}'\underset{\sim}{X}\underset{\sim}{\beta} \qquad \text{e} \qquad \underset{\sim}{\beta}'\underset{\sim}{X}'\underset{\sim}{y}$$

são escalares e

$$\underset{\sim}{y}'\underset{\sim}{X}\underset{\sim}{\beta} = \left(\underset{\sim}{\beta}'\underset{\sim}{X}'\underset{\sim}{y}\right)'$$

então

$$\underset{\sim}{y}'\underset{\sim}{X}\underset{\sim}{\beta} = \underset{\sim}{\beta}'\underset{\sim}{X}'\underset{\sim}{y}$$

Estimação

Método dos Mínimos Quadrados Ordinários

Assim

$$S = \underset{\sim}{\mathbf{y}}'\underset{\sim}{\mathbf{y}} - 2\underset{\sim}{\mathbf{y}}'\underset{\sim}{\mathbf{X}}\underset{\sim}{\boldsymbol{\beta}} + \underset{\sim}{\boldsymbol{\beta}}'\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}}\underset{\sim}{\boldsymbol{\beta}}$$

Logo, nosso interesse, agora, é encontrar o resultado para

$$\frac{\partial S}{\partial \underset{\sim}{\boldsymbol{\beta}}}$$

Derivadas

de

Formas Lineares

e

Formas Quadráticas

Derivadas de Formas Lineares e Quadráticas

Definição.

Considere o vetor coluna $\underset{\sim}{\mathbf{x}} = \left(x_1, x_2, \ldots, x_n\right)'$ *e* $f\left(\underset{\sim}{\mathbf{x}}\right)$, *uma função real de* $x_1, x_2, \ldots, x_n$. *Assim, a derivada parcial de* $f\left(\underset{\sim}{\mathbf{x}}\right)$ *com relação a* $\underset{\sim}{\mathbf{x}}$ *é dada por:*

$$\frac{\partial f\left(\underset{\sim}{\mathbf{x}}\right)}{\partial \underset{\sim}{\mathbf{x}}} = \begin{pmatrix} \dfrac{\partial f\left(\underset{\sim}{\mathbf{x}}\right)}{\partial x_1} \\ \dfrac{\partial f\left(\underset{\sim}{\mathbf{x}}\right)}{\partial x_2} \\ \vdots \\ \dfrac{\partial f\left(\underset{\sim}{\mathbf{x}}\right)}{\partial x_n} \end{pmatrix}$$

Derivadas de Formas Lineares e Quadráticas

Teorema 1. (forma linear)

Se $\underset{\sim}{\mathbf{a}} = \left(a_1, a_2, \ldots, a_n\right)'$ *um vetor coluna de constantes e se* $f\left(\underset{\sim}{\mathbf{x}}\right) = \underset{\sim}{\mathbf{a}}'\underset{\sim}{\mathbf{x}}, = \underset{\sim}{\mathbf{x}}'\underset{\sim}{\mathbf{a}}$

então $\dfrac{\partial f\left(\underset{\sim}{\mathbf{x}}\right)}{\partial \underset{\sim}{\mathbf{x}}} = \underset{\sim}{\mathbf{a}}\,.$

Teorema 2. (forma quadrática)

Se $f\left(\underset{\sim}{\mathbf{x}}\right) = \underset{\sim}{\mathbf{x}}'\underset{\sim}{\mathbf{A}}\,\underset{\sim}{\mathbf{x}}$ *(forma quadrática, e* **A** *matriz simétrica de constantes),*

então $\dfrac{\partial f\left(\underset{\sim}{\mathbf{x}}\right)}{\partial \underset{\sim}{\mathbf{x}}} = \mathbf{2}\,\underset{\sim}{\mathbf{A}}\,\underset{\sim}{\mathbf{x}}$, *que é um vetor coluna de n elementos.*

Voltando à Estimação (MQO)

Lembrando que objetivamos minimizar

$$S\left(\underset{\sim}{\boldsymbol{\beta}}\right) = \underset{\sim}{\mathbf{y}}'\underset{\sim}{\mathbf{y}} - 2\underset{\sim}{\mathbf{y}}'\underset{\sim}{\mathbf{X}}\underset{\sim}{\boldsymbol{\beta}} + \underset{\sim}{\boldsymbol{\beta}}'\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}}\underset{\sim}{\boldsymbol{\beta}}$$

e, utilizando os resultados vistos anteriormente, temos que

$$\frac{\partial S\left(\underset{\sim}{\boldsymbol{\beta}}\right)}{\partial \underset{\sim}{\boldsymbol{\beta}}} = -2\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{y}} + 2\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}}\underset{\sim}{\boldsymbol{\beta}}$$

Voltando à Estimação (MQO)

E, igualando o resultado anterior a zero, vem que

$$-2\,\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{y}} + 2\,\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}}\,\underset{\sim}{\hat{\boldsymbol{\beta}}} = \underset{\sim}{\mathbf{0}} \Leftrightarrow \underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}}\,\underset{\sim}{\hat{\boldsymbol{\beta}}} = \underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{y}}$$

que é o sistema de equações normais na forma matricial.

Para encontrarmos o resultado de interesse, precisaremos supor que a matriz X'X admite inversa (ou seja, precisaremos supor que X'X é não-singular). Para tanto, assumiremos que os regressores não apresentam relação linear perfeita.

Estimação (MQO)

Assim, assumindo que X'X é não-singular, a solução do sistema de equações normais é dada por

$$\hat{\underset{\sim}{\beta}} = \left(\underset{\sim}{X}'\underset{\sim}{X}\right)^{-1}\underset{\sim}{X}'\underset{\sim}{y}$$

que é o vetor de estimadores de mínimos quadrados do vetor de parâmetros de interesse.

Regressão Múltipla

$$E\left(Y \mid X_1 = x_1, \ldots, X_k = x_k\right) = \beta_0 + \beta_1 x_1 + \ldots \beta_k x_k$$

Modelo Estimado

$$\underbrace{\hat{y}}_{\text{abuso de notação}} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots \hat{\beta}_k x_k$$

Interpretação do Intercepto

Valor médio estimado para a variável resposta, condicionado a $x_1 = x_2 = \ldots = x_k = 0$.

Muitas vezes pode não ter significado!!!

Interpretação dos demais parâmetros

Considerando

$$\Delta\hat{y} = \hat{\beta}_1 \Delta x_1 + \hat{\beta}_2 \Delta x_2 + \ldots + \hat{\beta}_k \Delta x_k$$

se Δx_2 = ... = Δx_k = 0 (ou seja, as outras variáveis são mantidas constantes), então o efeito parcial de x_1 no valor médio estimado para a variável resposta é dado por

$$\frac{\Delta\hat{y}}{\Delta x_1} = \hat{\beta}_1$$

Aplicação

O departamento de RH da empresa TEMCO objetiva estudar o comportamento dos salários dos funcionários dos mais diversos setores da empresa.

Para tanto, o gerente de RH, baseando-se numa amostra aleatória de 46 empregados, coletou informações sobre as seguintes variáveis:

Aplicação

id – número cadastral do funcionário;

salario – anual, em dólares;

anosemp – tempo (em anos) na empresa;

expprev – experiência anterior (em anos);

educ – anos de estudo após o segundo grau;

sexo – (feminino = 0, masculino = 1);

dept – departamento no qual atua (Compras = 1, Engenharia = 2, Propaganda = 3, Vendas = 4);

super – número de empregados sob responsabilidade do empregado.

Aplicação

EViews - [Group: UNTITLED Workfile: TEMCO::Temco\]

File Edit Object View Proc Quick Options Window Help

View Proc Object Print Name Freeze Default Sort Transpose Edit+/- Smpl+/- Title Sample

obs	ID	SALARIO	ANOSEMP	EXPPREV	EDUC	SEXO	DEPT	SUPER
1	972.0000	47536.00	15.00000	5.000000	6.000000	0.000000	3.000000	4.000000
2	539.0000	23654.00	0.000000	0.000000	0.000000	1.000000	3.000000	2.000000
3	649.0000	37548.00	19.00000	9.000000	4.000000	0.000000	3.000000	6.000000
4	824.0000	36578.00	4.000000	4.000000	8.000000	0.000000	3.000000	8.000000
5	649.0000	54679.00	20.00000	3.000000	6.000000	1.000000	3.000000	4.000000
6	624.0000	53234.00	25.00000	0.000000	6.000000	0.000000	3.000000	3.000000
7	891.0000	31425.00	7.000000	6.000000	5.000000	1.000000	3.000000	6.000000
8	974.0000	39743.00	9.000000	6.000000	5.000000	1.000000	2.000000	1.000000
9	648.0000	26452.00	1.000000	3.000000	2.000000	1.000000	2.000000	0.000000
10	321.0000	34632.00	5.000000	4.000000	4.000000	0.000000	2.000000	0.000000
11	264.0000	35631.00	6.000000	4.000000	4.000000	0.000000	2.000000	2.000000
12	291.0000	46211.00	14.00000	5.000000	6.000000	1.000000	2.000000	5.000000
13	267.0000	34231.00	6.000000	2.000000	6.000000	0.000000	2.000000	3.000000
14	548.0000	26548.00	5.000000	1.000000	0.000000	0.000000	2.000000	2.000000
15	555.0000	36512.00	6.000000	6.000000	4.000000	1.000000	2.000000	2.000000
16	366.0000	34869.00	7.000000	5.000000	4.000000	1.000000	2.000000	1.000000
17	246.0000	41255.00	9.000000	4.000000	6.000000	0.000000	2.000000	4.000000
18	215.0000	39331.00	9.000000	3.000000	6.000000	1.000000	2.000000	1.000000
19	814.0000	35487.00	8.000000	2.000000	2.000000	1.000000	2.000000	2.000000
20	212.0000	36487.00	6.000000	5.000000	2.000000	0.000000	2.000000	3.000000
21	526.0000	68425.00	25.00000	2.000000	12.00000	0.000000	2.000000	1.000000
22	778.0000	69246.00	22.00000	3.000000	10.00000	0.000000	2.000000	45.00000

Quadro 1 - Parte de uma planilha que contem informações sobre os empregados da empresa TEMCO.

Aplicação

Como parte do estudo, a gerente de RH propôs a estimação dos parâmetros do seguinte modelo de regressão múltipla:

$$salario = \beta_0 + \beta_1 educ + \beta_2 anosemp + \varepsilon$$

a) Em termos do problema, β_0 apresenta algum significado prático?

b) Qual o sinal esperado para β_1? E para β_2?

c) Encontre as estimativas dos parâmetros, via mínimos quadrados ordinários, escreva a equação estimada e interprete os resultados obtidos, em termos do problema de interesse.

Aplicação

Interpretação dos parâmetros do modelo proposto, em termos do problema:

β_0 – salário médio dos funcionários da empresa TEMCO, que acabaram de entrar na empresa (ou que ainda não completaram um ano) e que não apresentam nenhum ano de escolaridade após o segundo grau;

β_1 – efeito no salário médio dos funcionários da empresa TEMCO, dada a variação de um ano no tempo de escolaridade após o segundo grau, mantendo constante a variável *anosemp*; e

β_2 – efeito no salário médio dos funcionários da empresa TEMCO, dada a variação de um ano no tempo de empresa, mantendo constante a variável *educ*.

Aplicação

Dependent Variable: SALARIO
Method: Least Squares
Date: 08/26/12 Time: 15:45
Sample: 1 46
Included observations: 46
SALARIO=C(1)+C(2)*EDUC+C(3)*ANOSEMP

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	23177.47	1769.732	13.09660	0.0000
C(2)	1916.489	379.2670	5.053139	0.0000
C(3)	672.3250	141.6725	4.745629	0.0000

R-squared	0.739927	Mean dependent var	39827.39
Adjusted R-squared	0.727830	S.D. dependent var	10999.24
S.E. of regression	5738.291	Akaike info criterion	20.21070
Sum squared resid	1.42E+09	Schwarz criterion	20.32996
Log likelihood	-461.8462	Hannan-Quinn criter.	20.25538
F-statistic	61.16907	Durbin-Watson stat	1.229794
Prob(F-statistic)	0.000000		

Aplicação

Modelo estimado

$$\hat{salário} = 23177{,}47 + 1916{,}49\, educ + 672{,}32\, anosemp$$

Pergunta: qual o salário médio estimado para pessoas com 3 anos de escolaridade após o 2º grau e com 5 anos na empresa?

$$\hat{salario} = 23.177{,}47 + 1.916{,}49 * 3 + 672{,}33 * 5$$

$$\hat{salario} = 32288{,}54$$

Exercício (para entrega na próxima aula)

Considere o seguinte modelo de regressão linear:

$$y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon_i$$

Ainda, sabendo que o vetor de estimadores de MQO para o vetor de parâmetros do modelo de interesse é dado por

$$\hat{\underset{\sim}{\boldsymbol{\beta}}} = \left(\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}}\right)^{-1}\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{y}}$$

Encontre as expressões analíticas para o estimador do intercepto e do coeficiente angular.

Leitura Complementar

Posto de uma Matriz

Posto de uma Matriz: Seja **A** uma matriz *n x k*. O posto de **A**, denotado por r(**A**), é a ordem da “maior” submatriz quadrada não singular de **A**.

Propriedades:

(i) O determinante de uma matriz quadrada é diferente de zero se, e somente se, a matriz tiver posto completo;

(ii) Caso uma matriz quadrada apresente determinante igual a zero, então tal matriz será dita singular;

(iii) r(**A**) = r(**AA’**) = r(**A’A**).

Posto de uma Matriz

Conseqüências:

(i) Se r(A) = p, então A contém pelo menos um menor p x p não nulo e nenhum menor não nulo de dimensão maior que p;

(ii) Se $n > k$, r(A) é o número de colunas linearmente independentes de A, portanto r(A) $\leq k$:

- Se r(A) = k, A é de posto completo;

(iii) Se $n < k$, r(A) é o número de linhas linearmente independentes de A, portanto r(A) $\leq n$:

- Se r(A) = n, A é de posto completo.

Observação

Tecnicamente, foi feita a suposição de que não existe colinearidade perfeita entre os regressores, que formam as colunas da matriz X. Ou seja, nenhum dos regressores pode ser expresso como uma combinação linear exata dos demais regressores do modelo. Assim, o posto de X é igual a k (como k é o número de colunas da matriz X, estamos supondo que a matriz X é de posto completo, ou seja, as colunas de X são linearmente independentes).