O modelo estatístico de uma regressão linear múltipla

- Temos uma regressão linear múltipla quando admitimos que o valor da variável dependente (Y) é função linear de duas ou mais variáveis explanatórias ($X_1, X_2, \ldots$).

- O modelo estatístico de uma regressão linear múltipla com k variáveis explanatórias é:

$$Y_j = \alpha + \beta_1 X_{1j} + \beta_2 X_{2j} + \ldots + \beta_k X_{kj} + \mu_j, \quad j = 1, \ldots, n \text{ observações}$$

ou

$$Y_j = \alpha + \sum_{i=1}^{k} \beta_i X_{ij} + \mu_j, \quad i = 1, \ldots, k \text{ regressores}$$

- Utilizando notação matricial o modelo fica $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu}$

$$\underset{(n,1)}{\mathbf{y}} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} \qquad \underset{(n,k+1)}{\mathbf{X}} = \begin{pmatrix} 1 & X_{11} & X_{21} & \cdots & X_{k1} \\ 1 & X_{12} & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{kn} \end{pmatrix} \qquad \underset{(k+1,1)}{\boldsymbol{\beta}} = \begin{pmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} \qquad \underset{(n,1)}{\boldsymbol{\mu}} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}$$

- O modelo estimado é $\hat{\mathbf{y}} = \mathbf{X}\mathbf{b}$, em que $\underset{(n,1)}{\hat{\mathbf{y}}} = \begin{pmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{pmatrix} \qquad \underset{(k+1,1)}{\mathbf{b}} = \begin{pmatrix} a \\ b_1 \\ b_2 \\ \vdots \\ b_k \end{pmatrix}$

Pressupostos do modelo

H.1 – a variável dependente é (Y) função linear das variáveis explanatórias;

H.2 – as variáveis explanatórias não guardam correlação com o termo de erro do modelo: $\text{cov}(X_{ij}, \mu_j) = 0$;

H.3 – $E(\mu_j) = 0$, ou seja, $E(\boldsymbol{\mu}) = \mathbf{0}$, em que $\mathbf{0}$ representa um vetor de zeros;

H.4 – $V(\mu_j) = E(\mu_j^2) = \sigma^2 = \text{constante}$, ou seja, os erros são homocedásticos;

H.5 – $\text{cov}(\mu_j, \mu_h) = E\{[\mu_j - E(\mu_j)][\mu_h - E(\mu_h)]\} = E(\mu_j \mu_h) = 0$ para $j \neq h$, isto é, os erros são não-correlacionados entre si;

H.6 – os erros têm distribuição normal.

- Além disso, como $(k+1)$ é o número de parâmetros a serem estimados $(\alpha, \beta_1, ..., \beta_k)$, devemos ter $n > k+1$.
- Combinando H.4 e H.5 temos

$$E(\boldsymbol{\mu\mu}') = E\left[\underbrace{\begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}}_{(n,1)} \underbrace{\begin{pmatrix} \mu_1 & \mu_2 & \cdots & \mu_n \end{pmatrix}}_{(1,n)}\right] = E\left(\underbrace{\begin{matrix} \mu_1^2 & \mu_1\mu_2 & \cdots & \mu_1\mu_n \\ \mu_2\mu_1 & \mu_2^2 & \cdots & \mu_2\mu_n \\ \vdots & \vdots & \ddots & \vdots \\ \mu_n\mu_1 & \mu_n\mu_2 & \cdots & \mu_n^2 \end{matrix}}_{(n,n)}\right) = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{pmatrix} = \mathbf{I}\sigma^2$$

H.1, H.2 e H.3 são necessárias para demonstrar que os estimadores de mínimos quadrados são não tendenciosos. H.1 a H.5 permitem demonstrar que tais estimadores são estimadores lineares não tendenciosos de variância mínima. H.6 é necessária para realizar testes de hipóteses e para construir intervalos de confiança para os parâmetros.

Estimativa dos parâmetros segundo o método dos mínimos quadrados

- Seja $\mathbf{b}$ o vetor das estimativas dos parâmetros, $\mathbf{e}$ o vetor de desvios e $\hat{\mathbf{y}}$ o vetor de valores estimados de Y, isto é:

$$\underset{(k+1,1)}{\mathbf{b}} = \begin{pmatrix} a \\ b_1 \\ b_2 \\ \vdots \\ b_k \end{pmatrix} \qquad \underset{(n,1)}{\mathbf{e}} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix} \qquad \underset{(n,1)}{\hat{\mathbf{y}}} = \begin{pmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{pmatrix}$$

- Temos $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\mathbf{b}$

- A soma dos quadrados dos desvios é dada por

$$Z = \sum_{j=1}^{n} e_j^2$$ (medida agregada dos desvios para as n observações da amostra)

$$Z = \mathbf{e}'\mathbf{e}$$ ou seja, $$\underbrace{\begin{pmatrix} e_1 & e_2 & \cdots & e_n \end{pmatrix}}_{(1,n)} \underbrace{\begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}}_{(n,1)} = \sum_{j=1}^{n} e_j^2$$

Substituindo $\mathbf{e} = \mathbf{y} - \mathbf{Xb}$

$$Z = \underbrace{(\mathbf{y} - \mathbf{Xb})'}_{\mathbf{e}'} \underbrace{(\mathbf{y} - \mathbf{Xb})}_{\mathbf{e}}$$

Das propriedades da transposta: $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$

$$Z = (\mathbf{y}' - \mathbf{b}'\mathbf{X}')(\mathbf{y} - \mathbf{Xb})$$

$$Z = \underset{(1,1)}{\mathbf{y}'\mathbf{y}} - \underset{(1,1)}{\mathbf{y}'\mathbf{X}\mathbf{b}} - \underset{(1,1)}{\mathbf{b}'\mathbf{X}'\mathbf{y}} + \underset{(1,1)}{\mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}}$$

As matrizes $\mathbf{y}'\mathbf{X}\mathbf{b}$ e $\mathbf{b}'\mathbf{X}'\mathbf{y}$ são iguais, pois uma é a transposta da outra e cada uma tem apenas um elemento. Então:

$$Z = \underset{(1,1)}{\mathbf{y}'\mathbf{y}} - 2\underset{(1,1)}{\mathbf{b}'\mathbf{X}'\mathbf{y}} + \underset{(1,1)}{\mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}}$$

O método dos mínimos quadrados consiste em adotar como estimativas dos parâmetros os valores que minimizam a soma dos quadrados dos desvios.

A função Z apresenta ponto de mínimo para os valores de $\mathbf{b}$ que tornem sua diferencial identicamente nula, isto é:

$$dZ = -2(d\mathbf{b}')\mathbf{X}'\mathbf{y} + (d\mathbf{b}')\mathbf{X}'\mathbf{X}\mathbf{b} + \mathbf{b}'\mathbf{X}'\mathbf{X}(d\mathbf{b}) = 0$$

Como $(d\mathbf{b}')\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{b}'\mathbf{X}'\mathbf{X}(d\mathbf{b})$, por serem matrizes com apenas um elemento e uma ser a transposta da outra, segue-se que

$$-2(d\mathbf{b}')\mathbf{X}'\mathbf{y} + 2(d\mathbf{b}')\mathbf{X}'\mathbf{X}\mathbf{b} = 0$$

$$(d\mathbf{b}')(\mathbf{X}'\mathbf{X}\mathbf{b} - \mathbf{X}'\mathbf{y}) = 0$$

Portanto, a diferencial de Z será identicamente nula para

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y}$$

Desde que $\mathbf{X}'\mathbf{X}$ seja não singular (ou seja, tenha determinante não nulo), existe a matriz inversa $(\mathbf{X}'\mathbf{X})^{-1}$.

Pré-multiplicando os dois membros de $\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y}$ por $(\mathbf{X}'\mathbf{X})^{-1}$, obtemos

$$\underbrace{\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{X}}_{\mathbf{I}}\mathbf{b} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$$

$$\boxed{\mathbf{b} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}}$$

em que

$$\underset{(k+1,k+1)}{\mathbf{X}'\mathbf{X}} = \begin{pmatrix} n & \sum_{j=1}^{n} X_{1j} & \sum_{j=1}^{n} X_{2j} & \cdots & \sum_{j=1}^{n} X_{kj} \\ \sum_{j=1}^{n} X_{1j} & \sum_{j=1}^{n} X_{1j}^{2} & \sum_{j=1}^{n} X_{1j}X_{2j} & \cdots & \sum_{j=1}^{n} X_{1j}X_{kj} \\ \sum_{j=1}^{n} X_{2j} & \sum_{j=1}^{n} X_{2j}X_{1j} & \sum_{j=1}^{n} X_{2j}^{2} & \cdots & \sum_{j=1}^{n} X_{2j}X_{kj} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^{n} X_{kj} & \sum_{j=1}^{n} X_{kj}X_{1j} & \sum_{j=1}^{n} X_{kj}X_{2j} & \cdots & \sum_{j=1}^{n} X_{k}^{2} \end{pmatrix} \text{ e } \underset{(k+1,1)}{\mathbf{X}'\mathbf{y}} = \begin{pmatrix} \sum_{j=1}^{n} Y_{j} \\ \sum_{j=1}^{n} X_{1j}Y_{j} \\ \sum_{j=1}^{n} X_{2j}Y_{j} \\ \vdots \\ \sum_{j=1}^{n} X_{kj}Y_{j} \end{pmatrix}$$

Exemplo

São dados os valores de X_1, X_2 e Y da tabela a seguir:

X_1	X_2	Y
0	0	-1
0	2	3
0	4	5
0	6	5
2	0	4
2	2	10
2	4	12
2	6	10

Admite-se que as variáveis estão relacionadas de acordo com o modelo $Y_j = \alpha + \beta_1 X_{1j} + \beta_2 X_{2j} + \mu_j$, em que os μ_j são variáveis aleatórias independentes, homocedásticas, com média zero e distribuição normal.

Determine as estimativas dos parâmetros da regressão linear múltipla de Y em relação a X_1 e X_2.

$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \mu$ ou $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu}$, em que

$$\underset{(8,1)}{\mathbf{y}} = \begin{pmatrix} -1 \\ 3 \\ 5 \\ 5 \\ 4 \\ 10 \\ 12 \\ 10 \end{pmatrix} \qquad \underset{(8,3)}{\mathbf{X}} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 2 \\ 1 & 0 & 4 \\ 1 & 0 & 6 \\ 1 & 2 & 0 \\ 1 & 2 & 2 \\ 1 & 2 & 4 \\ 1 & 2 & 6 \end{pmatrix} \qquad \underset{(3,1)}{\boldsymbol{\beta}} = \begin{pmatrix} \alpha \\ \beta_1 \\ \beta_2 \end{pmatrix} \qquad \underset{(8,1)}{\boldsymbol{\mu}} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \\ \mu_5 \\ \mu_6 \\ \mu_7 \\ \mu_8 \end{pmatrix}$$

Estimativa dos parâmetros: $\mathbf{b} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$

$$\mathbf{X}'\mathbf{X} = \underbrace{\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 2 & 2 & 2 & 2 \\ 0 & 2 & 4 & 6 & 0 & 2 & 4 & 6 \end{pmatrix}}_{(3,8)} \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 2 \\ 1 & 0 & 4 \\ 1 & 0 & 6 \\ 1 & 2 & 0 \\ 1 & 2 & 2 \\ 1 & 2 & 4 \\ 1 & 2 & 6 \end{pmatrix}}_{(8,3)} =$$

$$\mathbf{X}'\mathbf{X} = \underbrace{\begin{pmatrix} 8 & 8 & 24 \\ 8 & 16 & 24 \\ 24 & 24 & 112 \end{pmatrix}}_{(3,3)} = \underbrace{\begin{pmatrix} n & \sum X_1 & \sum X_2 \\ \sum X_1 & \sum X_1^2 & \sum X_1 X_2 \\ \sum X_2 & \sum X_2 X_1 & \sum X_2^2 \end{pmatrix}}_{(3,3)}$$

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} 0{,}475 & -0{,}125 & -0{,}075 \\ -0{,}125 & 0{,}125 & 0 \\ -0{,}075 & 0 & 0{,}025 \end{pmatrix}$$

$$\mathbf{X}'\mathbf{y} = \underbrace{\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 2 & 2 & 2 & 2 \\ 0 & 2 & 4 & 6 & 0 & 2 & 4 & 6 \end{pmatrix}}_{(3,8)} \underbrace{\begin{pmatrix} -1 \\ 3 \\ 5 \\ 5 \\ 4 \\ 10 \\ 12 \\ 10 \end{pmatrix}}_{(8,1)} = \begin{pmatrix} 48 \\ 72 \\ 184 \end{pmatrix} = \begin{pmatrix} \sum Y \\ \sum X_1 Y \\ \sum X_2 Y \end{pmatrix}$$

$$\mathbf{b} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} = \underbrace{\begin{pmatrix} 0,475 & -0,125 & -0,075 \\ -0,125 & 0,125 & 0 \\ -0,075 & 0 & 0,025 \end{pmatrix}}_{(3,3)} \underbrace{\begin{pmatrix} 48 \\ 72 \\ 184 \end{pmatrix}}_{(3,1)} = \begin{pmatrix} 0 \\ 3 \\ 1 \end{pmatrix} \qquad \therefore \hat{Y} = 3X_1 + X_2$$

Dicas para inverter matriz:

- inversa de matriz 2x2;

- inversa de matriz diagonal;

- calculadoras on-line: http://www.mathportal.org/calculators/matrices-calculators/matrix-calculator.php

Análise de variância da regressão linear múltipla

Causas de variação	Graus de liberdade	Soma de quadrados
Regressão	k	$\mathbf{b}'\mathbf{X}'\mathbf{y} - \dfrac{\left(\sum_{j=1}^{n} Y_j\right)^2}{n}$
Resíduo	$n-k-1$	$\mathbf{y}'\mathbf{y} - \mathbf{b}'\mathbf{X}'\mathbf{y}$
Total	$n-1$	$\mathbf{y}'\mathbf{y} - \dfrac{\left(\sum_{j=1}^{n} Y_j\right)^2}{n}$

em que

k = número de regressores

n = tamanho da amostra

Exemplo

Faça a análise de variância da regressão com 5% de significância.

Causas de Variação	Graus de liberdade (gl)	Soma de Quadrados (SQ)	Quadrados Médios (QM)
Regressão	$k=2$	$\mathbf{b}'\mathbf{X}'\mathbf{y}-\frac{\left(\sum Y\right)^2}{n}$	$\dfrac{\mathbf{b}'\mathbf{X}'\mathbf{y}-\frac{\left(\sum Y\right)^2}{n}}{2}$
Resíduo	$n-k-1=5$	$\mathbf{y}'\mathbf{y}-\mathbf{b}'\mathbf{X}'\mathbf{y}$	$\dfrac{\mathbf{y}'\mathbf{y}-\mathbf{b}'\mathbf{X}'\mathbf{y}}{5}$
Total	$n-1=7$	$\mathbf{y}'\mathbf{y}-\frac{\left(\sum Y\right)^2}{n}$	

$$\mathbf{b}'\mathbf{X}'\mathbf{y} = \begin{pmatrix} 0 & 3 & 1 \end{pmatrix} \begin{pmatrix} 48 \\ 72 \\ 184 \end{pmatrix} = 400$$

$$\mathbf{y}'\mathbf{y} = \begin{pmatrix} -1 & 3 & 5 & 5 & 4 & 10 & 12 & 10 \end{pmatrix} \begin{pmatrix} -1 \\ 3 \\ 5 \\ 5 \\ 4 \\ 10 \\ 12 \\ 10 \end{pmatrix} = 420$$

$$\sum Y = 48$$

Causas de Variação	Graus de liberdade (gl)	Soma de Quadrados (SQ)	Quadrados Médios (QM)
Regressão	2	$400-\left(\dfrac{48^2}{8}\right)=112$	56
Resíduo	5	20	4
Total	7	$420-\left(\dfrac{48^2}{8}\right)=132$	

Coeficiente de determinação do modelo

$$R^2=\frac{SQ\ \text{Regressão}}{SQ\ \text{Total}}=\frac{112}{132}=84{,}85\%$$

"84,85% das variações observadas em Y são explicadas pela regressão"

Teste *F*

$$\begin{cases} H_0 : \beta_1 = \beta_2 = 0 \\ H_A : \beta_1 \text{ e/ou } \beta_2 \neq 0 \end{cases} \quad \text{ao} \quad ns = 5\%$$

$$F_{calc} = \frac{\text{QM Regressão}}{\text{QM Resíduos}} = \frac{56}{4} = 14 \qquad\qquad F_0 = ?$$

	1	2	3	4	5	6	7	8	9
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49

Como $F_{calc} \geq F_0$, rejeita-se H_0.

Aplicação

O tamanho das escolas tem efeito sobre o desempenho dos estudantes?

Estudantes de escolas menores saem-se melhor do que aqueles de escolas maiores?

Arquivo ‘meap93.gdt’ traz uma amostra de 408 escolas de ensino médio em Michigan em 1993.

math10 = % de aprovados no teste de matemática

enroll = nº de estudantes matriculados (~ tamanho da escola)

Problema: como **isolar** o efeito do tamanho da escola (~ número de alunos) sobre o desempenho dos estudantes?

É preciso **controlar** todos os outros fatores que afetam o desempenho dos estudantes.

salary = salário anual médio dos professores (~ qualidade do professor)

staff = nº de funcionários por mil estudantes (~ atenção recebida pelos estudantes)

...

Modelo: MQO, usando as observações 1-408
Variável dependente: math10

	Coeficiente	*Erro Padrão*	*razão-t*	*p-valor*	
const	0,425914	6,05623	0,0703	0,94397	
enroll	-0,000253017	0,000215152	-1,1760	0,24029	
staff	0,0529116	0,0396184	1,3355	0,18245	
salary	0,000598902	0,000119156	5,0262	<0,00001	***

Média var. dependente	24,10686	D.P. var. dependente	10,49361
Soma resíd. quadrados	41961,94	E.P. da regressão	10,19148
R-quadrado	0,063709	R-quadrado ajustado	0,056756
F(3, 404)	9,163212	P-valor(F)	7,05e-06
Log da verossimilhança	-1524,110	Critério de Akaike	3056,220
Critério de Schwarz	3072,265	Critério Hannan-Quinn	3062,569

Por que obtivemos um R^2 de apenas 6,37%?