

Regressão Linear Múltipla

Rodrigo R. Pescim

Universidade Estadual de Londrina

30 de junho de 2020

Regressão Linear Múltipla

- Tem-se uma regressão linear múltipla quando se admite que a variável resposta (Y) é função de duas ou mais variáveis explicativas (regressoras).
- Assim, o objetivo da análise de regressão múltipla é estabelecer uma equação que possa ser utilizada para prever os valores de (Y_i) considerando as diversas variáveis explicativas (X_i).
- O modelo estatístico de um regressão linear múltipla com k variáveis regressoras (X_1, X_2, \dots, X_k) é dado por

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \epsilon_i$$

em que, $\epsilon_i \sim N(0, \sigma^2)$

- Em notação matricial, o modelo de regressão linear múltipla pode ser representado por

$$Y = X\beta + \epsilon$$

em que

Y é o vetor de dimensões $n \times 1$ da variável resposta Y ;

X é a matriz $n \times p$, conhecida como matriz de delineamento, com $p = k + 1$ (número de parâmetros);

β é o vetor de dimensão $p \times 1$ de parâmetros desconhecidos;

ϵ é o vetor de dimensão $n \times 1$ de erros aleatórios, isto é,
 $\epsilon \sim N(0, I\sigma^2)$

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{bmatrix} ; X = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1k} \\ 1 & X_{21} & X_{22} & \dots & X_{2k} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & X_{n1} & X_{n2} & \dots & X_{nk} \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \dots \\ \beta_k \end{bmatrix} ; \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \dots \\ \epsilon_n \end{bmatrix}$$

$$\begin{aligned}
 Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{bmatrix} &= \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1k} \\ 1 & X_{21} & X_{22} & \dots & X_{2k} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & X_{n1} & X_{n2} & \dots & X_{nk} \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \dots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \dots \\ \epsilon_n \end{bmatrix} \\
 &= \begin{bmatrix} \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + \dots + \beta_k X_{1k} + \epsilon_1 \\ \beta_0 + \beta_1 X_{21} + \beta_2 X_{22} + \dots + \beta_k X_{2k} + \epsilon_2 \\ \dots \\ \beta_0 + \beta_1 X_{n1} + \beta_2 X_{n2} + \dots + \beta_k X_{nk} + \epsilon_n \end{bmatrix}
 \end{aligned}$$

Estimação dos Parâmetros - Método dos Mínimos Quadrados

- O número de parâmetros a serem estimados é $p = k + 1$
- Observa-se também que $n > p$
- Utilizando o Método dos Mínimos Quadrados tem-se que

$$\epsilon = Y - X\beta$$

- De forma geral, esse método propõe minimizar o tamanho do comprimento do vetor ϵ
- Portanto,

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

Análise de Variância e Teste F

Tabela: Quadro da análise de variância para regressão linear múltipla

C.V.	S.Q.	gl	Q.M.	F_{calc}	F_{tab}
Regressão	$SQReg$	$k = p - 1$	$\frac{SQReg}{p-1}$	$\frac{QMReg}{QMRes}$	$F_{\alpha;p-1,n-p}$
Res	$SQRes$	$n - p$	$\frac{SQRes}{n-p}$		
Total	$SQTotal$	$n - 1$			

A variação total presente nos valores de Y pode ser dividida em dois componentes: a variação explicada (que mede a quantidade de **variação total** que é explicada pela superfície de regressão) e, **a variação não explicada** (resíduo). A hipótese a seguir determina se uma quantidade significativa de variação é explicada pelo modelo:

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_k$$

$$H_1 : \beta_i \neq \beta_j$$

Exemplo 1

Os dados a seguir foram coletados a partir de um estudo em uma planta de fabricação de semicondutores. Nessa planta, o semicondutor final é um arame colado a uma estrutura. As variáveis reportadas são a resistência à tração (uma medida da quantidade de força requerida para romper a cola), o comprimento do arame e a altura da matriz. O objetivo do estudo é encontrar um modelo que explique a resistência à tração por meio das variáveis comprimento do arame e altura da matriz.

Exemplo 1

Dados sobre a Resistência de Tração da Cola no Arame

Observação	Resistência à Tração	Comprimento do Arame	Altura do Molde
1	9,95	2	50
2	24,45	8	110
3	31,75	11	120
4	35,00	10	550
5	25,02	8	295
6	16,86	4	200
7	14,38	2	375
8	9,60	2	52
9	24,35	9	100
10	27,50	8	300
11	17,08	4	412
12	37,00	11	400
13	41,95	12	500
14	11,66	2	360
15	21,65	4	205
16	17,89	4	400
17	69,00	20	600
18	10,30	1	585
19	34,93	10	540
20	46,59	15	250
21	44,88	15	290
22	54,12	16	510
23	56,63	17	590
24	22,13	6	100
25	21,15	5	400