O modelo de Regressão Linear Múltipla

19/01/2021

Capítulo 3

Representação Matricial

$$Y_{n\times 1} = X_{n\times p} \cdot \beta_{p\times 1} + \epsilon_{n\times 1}$$

$$\epsilon \sim Normal\,(0_{n\times 1}\,;\,\sigma^2 I_{n\times n})$$

$$Y \sim Normal\,(X\beta\,;\,\sigma^2 I)$$

$$E(\epsilon) = 0_{n\times 1} \qquad Cov(\epsilon) = \sigma I_{n\times n}$$

os erros são independentes : $cov(\epsilon_i, \epsilon_j) = 0 \;\forall\; i \neq j$

$var(\epsilon_i) = \sigma^2$

Detalhe:

$$\rho_{xy} = \frac{Cov(X,Y)}{\sqrt{Var(X) \cdot Var(Y)}}$$

$$cov(X,Y) = E\{[X - E(X)][Y - E(Y)]\}$$

$$Var(Y) = cov(Y,Y)$$

$$= E[Y - E(Y)]^2$$

Por curiosidade, suponha $n = 2$:

$$\epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix} \qquad Cov(\epsilon) = \begin{bmatrix} Var(\epsilon_1) & cov(\epsilon_1, \epsilon_2) \\ cov(\epsilon_2, \epsilon_1) & Var(\epsilon_2) \end{bmatrix}$$

Representação Matricial da Soma dos Quadrados dos erros

$$\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k = X\beta$$

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

$$\sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)]^2 = \sum_{i=1}^{n} \epsilon_i^2$$

$$= \epsilon^T \epsilon = (Y - X\beta)^T (Y - X\beta)$$

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

Cuja solução é :

$$\frac{\partial (Y - X\beta)^T (Y - X\beta)}{\partial \beta} =$$

$$= -2X^T (Y - X\hat{\beta})$$

$$= -2X^T Y + 2X^T X \hat{\beta} = 0$$

$$X^T X \hat{\beta} = X^T Y$$

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

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

Errado, mas interessante:

$$\frac{\partial (Y - X\beta)^2}{\partial \beta} = 2(Y - X\beta) \cdot X = 0$$

como estou derivando em relação ao vetor $\beta_{p\times 1}$, o resultado final deveria ter a mesma dimensão.

$$2\underbrace{X^T}_{p\times n} \underbrace{(Y - X\beta)}_{n\times 1} = 0$$