Regressão Linear Simples

A reta dos mínimos quadrados é dada por

$$\hat{y} = \hat{\alpha} + \hat{\beta}x$$

Com

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

$$\hat{\beta} = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}$$

A expressão acima pode ser reescrita como

$$\hat{\beta} = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}$$

Lembrando da expressão para o coeficiente de correlação linear de Pearson

$$r_{XY} = \frac{\text{Cov}(X, Y)}{DP(X)DP(Y)}$$

Portanto, temos a seguinte relação entre r_{XY} e $\hat{\beta}$

$$r_{XY} = \frac{DP(X)}{DP(Y)}\hat{\beta}$$

Regressão Linear Simples

Foi visto em aulas anteriores a decomposição das somas de quadrados

$$SQ_{tot} = SQ_{Reg} + SQ_{Res}$$

Também é possível obter uma relação entre a SQ_{Reg} com o $\hat{\beta}$.

$$SQ_{Reg} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

Mas

$$\hat{y}_i - \bar{y} = \hat{\alpha} + \hat{\beta}x_i - \bar{y} = \bar{y} - \hat{\beta}\bar{x} + \hat{\beta}x_i - \bar{y}$$

$$\hat{y}_i - \bar{y} = \hat{\beta}(x_i - \bar{x})$$

Resulta na seguinte relação

$$SQ_{Reg} = (\hat{\beta})^2 \sum_{i=1}^n (x_i - \bar{x})^2$$

Lembre também que

$$SQ_{Tot} = \sum_{i=1}^n (y_i - \bar{y})^2$$

$$SQ_{Res} = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Exemplo 1: Considere Y o tempo de reação de um indivíduo a um determinado estímulo e X a idade do indivíduo.

Para uma amostra de $n = 20$ indivíduos foi ajustada a seguinte reta de regressão (MQ)

$$\hat{y} = 80.5 + 0.9x$$

Sabendo que $Var(X) = 50$, $Var(Y) = 68.65$ e $\sum_{i=1}^n x_i^2 = 19000$.

Determine:

(a) O coeficiente de correlação linear e seu quadrado. Interprete.

Vamos usar a relação obtida anteriormente

$$r_{XY} = \frac{DP(X)}{DP(Y)} \hat{\beta} = \sqrt{\frac{50}{68.65}} (0.9) = 0.768$$

O quadrado do coeficiente de Pearson, também denotado por R^2 e denominado coeficiente de determinação é dado por

$$R^2 = 0.59$$

Interpretação:

- Existe uma correlação linear positiva (0.77) entre a idade e o tempo de reação. Portanto, quanto maior a idade maior o tempo para reação ao estímulo, em média.
- Aproximadamente 59 % da variação do tempo é explicada pela idade.

(b) Determine as somas de quadrados.

$$SQ_{Tot} = (20)Var(Y) = (20)(68.65) = 1373$$

$$SQ_{Reg} = (\hat{\beta})^2 \sum_{i=1}^n (x_i - \bar{x})^2 = (0.9)^2(20)(50) = 810$$

$$SQ_{Res} = 1373 - 810 = 563$$

(c) Construa a tabela de análise de variância e conclua a respeito da significância do uso da variável idade no modelo.