Capítulo 2 - continuação

26/12/2020

Modelo: $Y_i \underset{iid}{\sim} \text{Normal}\left(\mu = \beta_0 + \beta_1 x_i \, ; \, \sigma^2\right)$

$$\hat{\beta}_1 = \sum_{i=1}^{n} c_i Y_i \longrightarrow \hat{\beta}_1 \sim \text{Normal}\left(\beta_1 \, ; \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x} \longrightarrow \hat{\beta}_0 \sim \text{Normal}\left(\beta_0 \, ; \, \sigma^2\left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]\right)$$

$$\hat{Y}_{(x_0)} = \hat{\mu}_{(x_0)} = \hat{\beta}_0 + \hat{\beta}_1 x_0$$

$$\hat{Y}_{(x_0)} \sim \text{Normal}\left(\beta_0 + \beta_1 x_0 \, ; \, \sigma^2\left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]\right)$$

Intervalo Preditivo:

$$Y_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0 + \varepsilon = \hat{\mu}_{(x_0)} + \varepsilon$$

$$Y_0 \sim \text{Normal}\left(\beta_0 + \beta_1 x_0 \, ; \, \sigma^2\left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]\right)$$

Análise de variância do modelo de regressão linear simples

$$\sum_i (y_i - \bar{y})^2 = \sum_i (\hat{y}_i - \bar{y})^2 + \sum_i (y_i - \hat{y}_i)^2$$

$$SQ_{\text{Total}} = SQ_{\text{Regressão}} + SQ_{\text{Resíduos}}$$

$$R^2 = \frac{SQ_{\text{Regressão}}}{SQ_{\text{Total}}} = \frac{SQ_{\text{Total}} - SQ_{\text{Resíduos}}}{SQ_{\text{Total}}}$$

$$= 1 - \frac{SQ_{\text{Resíduos}}}{SQ_{\text{Total}}}$$

Correlação linear

4/1/2021

2

$$\rho = \frac{E\{[X-E(X)][Y-E(Y)]\}}{\sqrt{Var(X) \cdot Var(Y)}} = \frac{E(XY) - E(X)E(Y)}{\sqrt{Var(X) \cdot Var(Y)}}$$

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}}$$

$$Z_x = \frac{x - \bar{x}}{S_x} \quad e \quad Z_y = \frac{y - \bar{y}}{S_y}$$

$$S_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

$$\hat{\beta}_1 = \hat{\rho} \cdot \frac{S_y}{S_x}$$

$$\boxed{\hat{\rho} = \sum \frac{Z_x \cdot Z_y}{n-1}} \; *$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = \hat{\mu}$$

Use of the coefficient of correlation assumes that there is a linear relation between the two variables - that is, that a given change in one variable always involves a certain constant change in the corresponding average value of the other.

* $\hat{\rho}$ gives the average deviation of either variable from its mean value corresponding to a given deviation of the other variable, provided that the standard deviation is the unit of measurement in both cases.

No caso da regressão linear simples:

$$R^2 = \hat{\rho}^2.$$

Sewall Wright
"Correlation and Causation"

Análise dos resíduos

O modelo:

$$Y_i \sim \text{Normal}(\mu = \beta_0 + \beta_1 x_i; \sigma^2)$$

ou

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \longleftrightarrow \varepsilon_i = Y_i - (\beta_0 + \beta_1 x_i)$$

erro

onde $\varepsilon_i \sim \text{Normal}(0, \sigma^2)$

$$\hat{\varepsilon}_i = Y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) = r_i \longrightarrow \text{é o resíduo!}$$

o resíduo (r_i) é o estimador do erro (ε_i).

ou seja, o resíduo também é uma variável aleatória: $E(r_i) = 0$ se o parâmetro β_0 estiver no modelo.

Gráfico dos resíduos: