Capítulo 2 – Regressão Linear Simples
(resumo)

antes : $Y_i \sim \text{Normal}(\mu ; \sigma^2)$ $\quad \hat{\mu} = \arg\min_{\mu} \sum_{i=1}^{n} (y_i - \mu)^2$

agora : $Y_i \sim \text{Normal}(\mu = \beta_0 + \beta_1 x_i ,\ \sigma^2)$ [o modelo gerador dos dados]

Forma alternativa:

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

$$\varepsilon \sim N(0, \sigma^2)$$

Propriedades:

$$E(Y) = \beta_0 + \beta_1 x$$

$$\text{Var}(Y) = \sigma^2$$

$$\hat{\beta}_0, \hat{\beta}_1 = \arg\min_{\beta_0, \beta_1} \sum_{i=1}^{n} \left[y_i - (\beta_0 + \beta_1 x_i) \right]^2$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} y_i (x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Obs:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

ou

$$n\bar{x} = \sum x_i$$

Um ponto interessante:

seja $z_i = x_i - \bar{x}$

$$\hat{\beta}_1 = \frac{\langle z, y \rangle}{\langle z, z \rangle} \rightarrow \text{produto interno entre dois vetores}$$

onde $\langle z, y \rangle = \sum_i x_i y_i$

Detalhe:

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

$$\hat{\beta}_1 = \sum_{i=1}^{n} Y_i \frac{(x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

$$\hat{\beta}_1 = \sum_{i=1}^{n} Y_i \cdot c_i$$

Propriedades :

$$E(\hat{\beta}_1) = \beta_1$$

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

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

$$FR(\hat{\beta}_1) : \beta_1 \pm Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}}$$

$$IC(\beta_1) : \hat{\beta}_1 \pm Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}}$$

se σ for desconhecido :

$$IC(\beta_1) : \hat{\beta}_1 + t_{\alpha/2, n-2} \cdot \frac{S}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}}$$

$$S^2 = \frac{\sum_{i=1}^{n} \hat{\varepsilon}_i^2}{n-2} = \frac{\sum r_i^2}{n-2}$$

$$r_i = \hat{\varepsilon}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$$

Correlação Linear e Regressão Linear Simples

$$Z_x = \frac{x-\bar{x}}{S_x} \qquad Z_y = \frac{y-\bar{y}}{S_y} \qquad S_x = \sqrt{\frac{\sum (x_i-\bar{x})^2}{n-1}}$$

$$\hat{\rho} = \frac{\sum Z_x \cdot Z_y}{n-1} \qquad y = \beta_0 + \beta_1 x$$

$$\hat{\beta}_1 = \hat{\rho} \cdot \frac{S_y}{S_x} \quad \text{e} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \cdot \bar{x}$$

Obs: $\hat{\rho}$ é invariante a transformações lineares em x e y: $\hat{\rho} = r = \dfrac{\sum_i (x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_i (x_i-\bar{x})^2 \cdot \sum_i (y_i-\bar{y})^2}}$

Forma geral: $y = \hat{\beta}_0 + \hat{\beta}_1 x$ onde $\hat{\beta}_1 = \hat{\rho} \cdot \dfrac{\sigma_y}{\sigma_x}$

$$\boxed{\hat{\beta}_0 = \hat{\mu}_y - \hat{\beta}_1 \cdot \hat{\mu}_x}$$

Teste de Hipóteses sobre o parâmetro β_1:

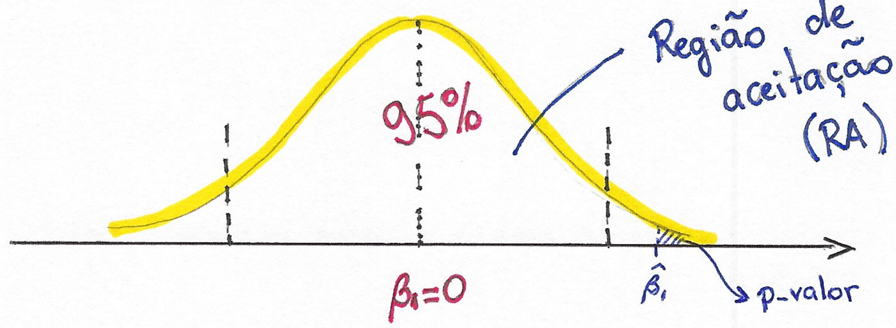
$Y = \beta_0 + \beta_1 x + \varepsilon \qquad \varepsilon \sim N(0, \sigma^2)$ (o modelo)

$\hat{\beta}_1 \sim \text{Normal}(\beta_1 ; \sigma^2_{\beta_1})$ onde $Var(\hat{\beta}_1) = \sigma^2_{\beta_1}$

$H_0 : \beta_1 = 0 \longrightarrow$ Sob H_0 (se H_0 for verdadeira)

$H_1 : \beta_1 \neq 0$

$$\boxed{Var(\hat{\beta}_1) = \sigma^2_{\beta_1} = \frac{\sigma^2}{\sum_i (x_i-\bar{x})^2}}$$



RA: $\beta_1^{H_0} \pm Z_{\alpha/2} \cdot \sigma_{\beta_1} \longrightarrow \pm Z_{\alpha/2} \cdot \sigma_{\beta_1}$

Se σ_{β_1} é desconhecido:

Intervalo de confiança p/ β_1:

$IC(\beta_1)$: $\hat{\beta}_1 \pm Z_{\alpha/2} \cdot \sigma_{\beta_1} \longrightarrow \hat{\beta}_1 \pm t_{\alpha/2, n-2} \cdot \dfrac{\sqrt{S^2}}{\sqrt{\sum_i (x_i-\bar{x})^2}}$