

Ciência dos Dados, projeto 3

a)

Entrega 2 22/11/16

d) $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ / $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$

$\epsilon_i = y_i - \hat{y}_i \rightarrow$ Soma dos erros Quadrados = SEQ

$$SEQ = \sum_{i=1}^m \epsilon_i^2 = \sum_{i=1}^m (y_i - \hat{y}_i)^2 = \sum_{i=1}^m (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

$\frac{dSEQ}{d\hat{\beta}_0} = 0$ (A) $\frac{dSEQ}{d\hat{\beta}_1} = 0$ (B)

(A) $\frac{dSEQ}{d\hat{\beta}_0} = -2 \cdot \sum_{i=1}^m (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)$

$$\frac{dSEQ}{d\hat{\beta}_0} = -2 \left\{ \sum_{i=1}^m y_i - m \hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^m x_i \right\} \div m$$

$$= -2m \left\{ \frac{\sum y_i}{m} - \hat{\beta}_0 - \hat{\beta}_1 \cdot \frac{\sum x_i}{m} \right\} = -2m \left\{ \bar{y} - \hat{\beta}_0 - \hat{\beta}_1 \bar{x} \right\}$$

$$= 0 \rightarrow \hat{\beta}_0 = -\hat{\beta}_1 \bar{x} + \bar{y} \quad (1)$$

(B) $\frac{dSEQ}{d\hat{\beta}_1} = -2 \sum_{i=1}^m x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$

$$= -2 \left\{ \sum x_i y_i - \hat{\beta}_0 \sum x_i - \hat{\beta}_1 \sum x_i^2 \right\} = 0 \quad (2)$$

Substituindo 1 em 2:

$$\sum_{i=1}^m x_i y_i + (\hat{\beta}_1 \bar{x} - \bar{y}) \cdot \sum_{i=1}^m x_i - \hat{\beta}_1 \cdot \sum_{i=1}^m x_i^2 = 0$$

$$\hat{\beta}_1 = \frac{\bar{y} \sum x_i - \sum x_i y_i}{\sum x_i - \sum x_i^2} = \frac{\sum x_i y_i - m \bar{y} \bar{x}}{\sum x_i^2 - m \bar{x}^2} = \frac{\text{cov}(x, y)}{\text{var}(x)}$$

b) Os erros são aleatórios e independentes entre si, e assim seguem uma distribuição normal, com variância σ^2 . Seu valor ideal (esperado) é de ser zero. Essas propriedades podem ser observadas a partir de uma visualização gráfica de ambas as variáveis.

c) $\begin{cases} H_0: \beta_1 = 0 \\ H_1: \beta_1 \neq 0 \end{cases}$

Caso H_0 não seja rejeitado, isso significará que não há relação entre as variáveis, do mesmo modo que a sua rejeição significa uma relação entre as variáveis.

d) Sim, é possível realizar uma regressão múltipla, a partir de determinadas condições:

- 1) O erro tem média zero e variância σ^2 , desconhecida;
- 2) Os erros são não correlacionados;
- 3) Os erros têm distribuição normal;
- 4) As variáveis regressoras x_1, x_2, \dots, x_p assumem valores fixos.

Além disso, o teste de hipóteses permanecerá o mesmo, com a aceitação ou rejeição de H_0 .