Y : variável aleatória de interesse (variável resposta)

X : variável regressora ou independente

Atenção: X não é variável aleatória

Y é descrita como uma soma de uma parcela determinística e outra aleatória.

A parte determinística é uma reta em função de X.

informação sobre Y que só depende de conhecermos X.

A parte aleatória, chamada de erro, representa os inúmeros fatores que podem influenciar Y.

O erro provoca variações ou flutuações ou mesmo distorções na parte determinística.

→ erros positivos ou negativos podem ocorrer e que $\mathbb{E}(\varepsilon) = 0$

→ $Var(\varepsilon)$ não depende de X.

Modelo de Regressão Linear Simples (MRLS)

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

β_0, β_1 e x : constantes

$$\mathbb{E}(\varepsilon) = 0$$

$$Var(\varepsilon) = \sigma^2$$

MRLS (Normal)

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

β_0, β_1 e x : constantes

$$\varepsilon \sim N(0; \sigma^2).$$

Resultado 1: para um valor pré-fixado, x, de X

$$Y \sim N(\beta_0 + \beta_1 x ; \sigma^2)$$

ou seja: $\mathbb{E}(Y | x) = \beta_0 + \beta_1 \cdot x$

$$Var(Y | x) = \sigma^2$$

Exercício 1: verifique o resultado 1.

Estimadores de mínimos quadrados

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

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

Sabemos que $Y|x \sim N(\beta_0 + \beta_1 x ; \sigma^2)$

Então $Y_i \sim N(\beta_0 + \beta_1 x_i ; \sigma^2)$

independentes para $i = 1, 2, \ldots, n$.

E os estimadores $\hat{\beta}_0$ e $\hat{\beta}_1$

são combinações lineares de $Y_1, Y_2 \ldots Y_n$, logo tb tem distribuição normal.

Exercício 2: verifique que

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

b) $\mathbb{E}(\hat{\beta}_0) = \beta_0$

ou seja, que $\hat{\beta}_0$ e $\hat{\beta}_1$ são estimadores não viciados para β_0 e β_1.

Exercício 3: Verifique que

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

$$Var(\hat{\beta}_0) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})} \right]$$

Exercício 4: Verifique que

$Cov(Y_i, Y_j) = 0$ para $i \neq j$

Exercício 5: Verifique que:

a) $Cov(\bar{Y}, \hat{\beta}_1) = 0$ (use o exercício 4)

b) $$Cov(\hat{\beta}_0, \hat{\beta}_1) = \frac{-\bar{x}\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2},$$

o que significa que o sinal de $Cov(\hat{\beta}_0, \hat{\beta}_1)$ depende do valor de $\bar{x}$.

A partir dos resultados dos exercícios 2 e 3, temos que as distribuições de $\hat{\beta}_0$ e $\hat{\beta}_1$ são

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

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

$Y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ é escrito como

$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ o que pode confundir bastante.

Observar que $\mathbb{E}(Y|x_i) = y_i$

E falta apresentar um estimador para $Var(\varepsilon) = \sigma^2$.

Seu estimador de mínimos quadrados é

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

, um estimado não viciado.

Sabe-se ainda

$$\frac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{(n-2)}$$

(qui-quadrado c/ n-2 graus de liberdade)