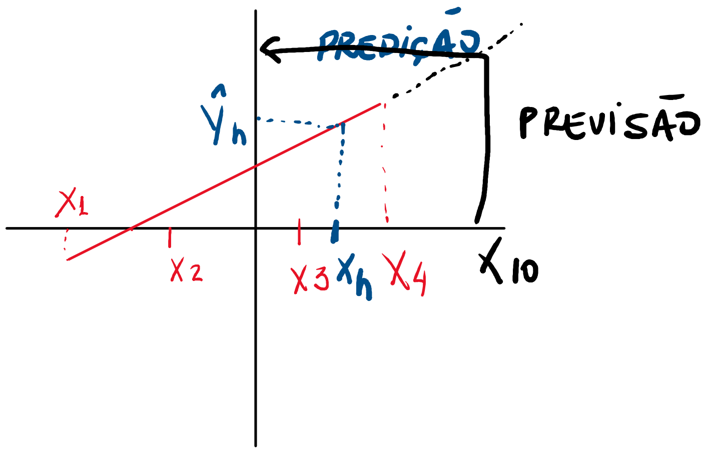
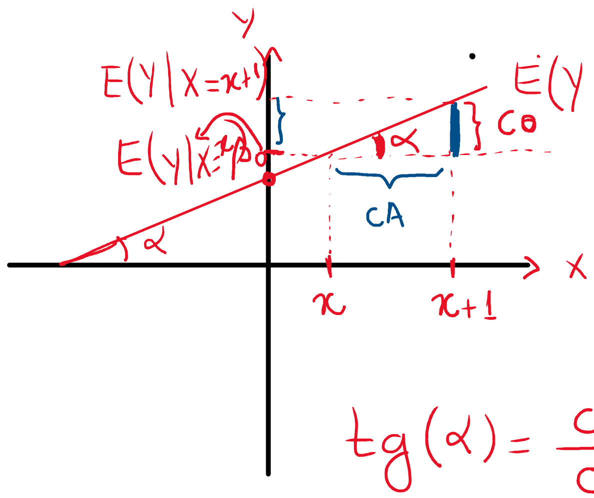
Regressão

segunda-feira, 21 de agosto de 2023 14:05





$$E(Y|X) = \beta_0 + \beta_1 X$$

$\beta_0 \rightarrow$ intercepto-y

$$E(Y|X=0) = \beta_0 + \beta_1(0) = \beta_0$$

$$tg(\alpha) = \frac{CO}{CA} = \frac{E(Y|X=x+1) - E(Y|X=x)}{x+1-\cancel{x}}$$

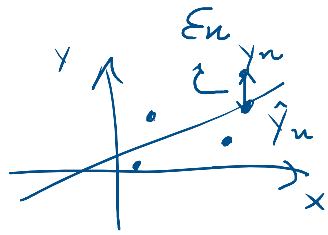
$$= \beta_0 + \beta_1(x+1) - (\beta_0 + \beta_1 x)$$

$$= \cancel{\beta_0} + \cancel{\beta_1 x} + \beta_1 - \cancel{\beta_0} - \cancel{\beta_1 x} = \beta_1$$

n pares $(x_1, y_1) \ldots (X_n, Y_n)$ e assumindo relação linear entre X e Y $\rightarrow$ RLS

$$\underline{\underline{Y_i}} = \underbrace{\beta_0 + \beta_1 X_i}_{E(Y_i)} + \varepsilon_i, \quad i = 1, \ldots, n$$

$X_i \uparrow$ EF. FIXO



1) Relação X e Y é linear;
2) $X \rightarrow$ Fixo;
3) A média dos erros é nula $\rightarrow E(\varepsilon_i) = 0$
4) ... variância do ε_i é sempre σ^2;

3) A média dos erros é nula ⇒

4) Para um dado X, a variância do ε_i é sempre σ^2;

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

$$var(y) = E\{y - E(y)\}^2$$

$$Var(\varepsilon_i) = E\{\varepsilon_i - \underbrace{E(\varepsilon_i)}\}^2 = E\{\varepsilon_i\}^2 = E\{\varepsilon_i^2\} = \sigma^2$$

$$Var(y_i) = E\{\underbrace{y_i - E(y_i)}_{\varepsilon_i}\}^2 = E\{\varepsilon_i\}^2 = \sigma^2$$

y_i ou ε_i são HOMOCEDÁSTICOS

5) ERROS são independentes $\Rightarrow Cov(\varepsilon_i, \varepsilon_j) = 0$

$$cov(XY) = 0 \quad \Rightarrow \quad cov(XY) = E\{(X - E(X))(Y - E(Y))\}$$

$$Cov(\varepsilon_i, \varepsilon_j) = E\{[\varepsilon_i - E(\varepsilon_i)][\varepsilon_j - E(\varepsilon_j)]\}$$

$$= E\{\varepsilon_i\varepsilon_j - \varepsilon_i E(\varepsilon_j) - E(\varepsilon_i)\varepsilon_j + E(\varepsilon_i)E(\varepsilon_j)\}$$

$$= E(\varepsilon_i\varepsilon_j) - E(\varepsilon_i)E(\varepsilon_j) - E(\varepsilon_i)E(\varepsilon_j) + E(\varepsilon_i)E(\varepsilon_j)$$

$$= E(\varepsilon_i\varepsilon_j) - E(\underset{0}{\cancel{\varepsilon_i}})E(\underset{0}{\cancel{\varepsilon_j}}) = 0 \quad \Rightarrow \boxed{E(\varepsilon_i\varepsilon_j) = 0}$$

6) ERROS Tem Dist. Normal

$$\varepsilon_i \sim N(\underset{=}{0}, \sigma^2) \quad 0 \to E(\varepsilon_i)$$

$$\varepsilon_i \sim Poi(\lambda)$$

$$\varepsilon_i \sim N(\mu, \sigma^2)$$

$$y_i \sim N(\underbrace{\beta_0 + \beta_1 X_i}_{E(y|x)}; \sigma^2)$$

MMQ ⇒ minimizar $\underset{\sim}{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^T$

$$Z = \|\varepsilon\|^2 = \sum^n \varepsilon_i^2 \qquad \to y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

$$Z = \|\underset{\sim}{\varepsilon}\|^2 = \sum_{i=1}^{n} \varepsilon_i^2 \qquad \rightarrow \quad y_i = \underbrace{\beta_0 + \beta_1 X_i}_{E(y_i)} + \varepsilon_i$$

$$Z = \sum_{i=1}^{n} \left[y_i - E(y_i) \right]^2$$

$$Z = \sum_{i=1}^{n} \left[y_i - \beta_0 - \beta_1 X_i \right]^2 = \left[y_1 - \beta_0 - \beta_1 X_1 \right]^2 + \cdots + \left[y_n - \beta_0 - \beta_1 X_n \right]^2$$

$$\frac{\partial Z}{\partial \beta_0} = 2\left[y_1 - \beta_0 - \beta_1 X_1 \right](-1) + \cdots + 2\left[y_n - \beta_0 - \beta_1 X_n \right](-1)$$

$$= -2 \sum_{i=1}^{n} \left[y_i - \beta_0 - \beta_1 X_i \right]$$

$$\frac{\partial Z}{\partial \beta_1} = -2 \sum_{i=1}^{n} \left[y_i - \beta_0 - \beta_1 X_i \right](X_i)$$

SEN

$$\begin{cases} \dfrac{\partial Z}{\partial \beta_0} = 0 \\ \dfrac{\partial Z}{\partial \beta_1} = 0 \end{cases} \Rightarrow \begin{cases} \cancel{-2} \sum_{i=1}^{n} \left[y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i \right] = 0 \\ \cancel{-2} \sum_{i=1}^{n} \left[y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i \right](X_i) = 0 \end{cases} \Rightarrow$$

$$\begin{cases} \sum_{i=1}^{n} y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^{n} X_i = 0 & \text{I} \\ \sum_{i=1}^{n} X_i y_i - \hat{\beta}_0 \sum_{i=1}^{n} X_i - \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 = 0 & \text{II} \end{cases}$$

Isolar $\hat{\beta}_0$ na I

$$\Sigma y_i - n\hat{\beta}_0 - \hat{\beta}_1 \Sigma X_i = 0$$

$$-n\hat{\beta}_0 = \hat{\beta}_1 \Sigma X_i - \Sigma y_i$$

subst. na II $\leftarrow \hat{\beta}_0 = -\hat{\beta}_1 \underbrace{\frac{\Sigma X_i}{n}}_{\bar{X}} + \underbrace{\frac{\Sigma y_i}{n}}_{\bar{y}}$

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

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

$$\hookrightarrow \sum X_i Y_i - \left[\frac{\sum Y_i}{n} - \hat{\beta}_1 \frac{\sum X_i}{n}\right]\sum X_i - \hat{\beta}_1 \sum X_i^2 = 0$$

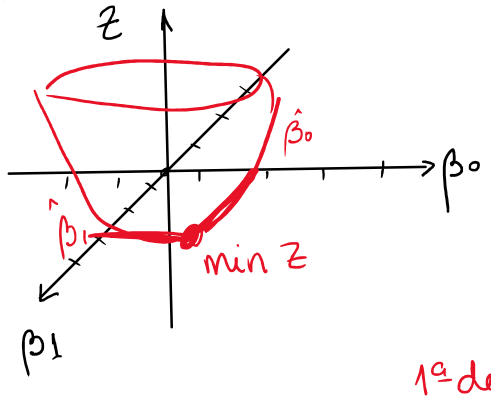
$$\sum X_i Y_i - \frac{(\sum Y_i)(\sum X_i)}{n} + \hat{\beta}_1 \frac{(\sum X_i)^2}{n} - \hat{\beta}_1 \sum X_i^2 = 0$$

$$-\hat{\beta}_1\left[\sum X_i^2 - \frac{(\sum X_i)^2}{n}\right] = -\sum X_i Y_i + \frac{(\sum Y_i)(\sum X_i)}{n}$$

$$-\hat{\beta}_1 = \frac{-\sum X_i Y_i + \frac{(\sum Y_i)(\sum X_i)}{n}}{\sum X_i^2 - \frac{(\sum X_i)^2}{n}}$$

1ª opção

$$\hat{\beta}_1 = \frac{\sum X_i Y_i - \frac{(\sum Y_i)(\sum X_i)}{n}}{\sum X_i^2 - \frac{(\sum X_i)^2}{n}} \quad // \quad \Rightarrow \quad \hat{\beta}_1 = \frac{SPXY}{SQX}$$



matriz Hessiana

$\det|H| > 0$

$$H = \begin{bmatrix} \frac{\partial^2 Z}{\partial \beta_0^2} & \frac{\partial^2 Z}{\partial \beta_0 \partial \beta_1} \\ \frac{\partial^2 Z}{\partial \beta_1 \partial \beta_0} & \frac{\partial^2 Z}{\partial \beta_1^2} \end{bmatrix}$$

(=)

1ª deriva β_0
2ª deriva β_0

1ª deriva β_0
2ª deriva β_1

1ª deriv β_1
2ª deriv β_0

1ª deriva β_1
2ª deriva β_1

e $\frac{\partial^2 Z}{\partial \beta_0^2} > 0$

Dado que $\frac{\partial Z}{\partial \beta_0} = -2\sum[Y_i - \beta_0 - \beta_1 X_i]$

$$= -2\sum Y_i + 2n\beta_0 + 2\beta_1 \sum X_i$$

$$\frac{\partial^2 Z}{\partial \beta_0^2} = \frac{\partial^2 Z}{\partial \beta_0 \partial \beta_0} = 2n$$

$$\frac{\partial^2 z}{\partial\beta_0\,\partial\beta_1} = 2\Sigma X_i$$

Dado que $\dfrac{\partial z}{\partial\beta_1} = -2\Sigma[Y_i - \beta_0 - \beta_1 X_i](X_i)$

$$= -2\Sigma X_i Y_i + 2\underline{\underline{\beta_0}}\Sigma X_i + 2\underline{\underline{\beta_1}}\Sigma X_i^2$$

$$\frac{\partial z^2}{\partial\beta_1\,\partial\beta_0} = 2\Sigma X_i$$

$$\frac{\partial z^2}{\partial\beta_1^2} = \frac{\partial z^2}{\partial\beta_1\,\partial\beta_1} = 2\Sigma X_i^2$$

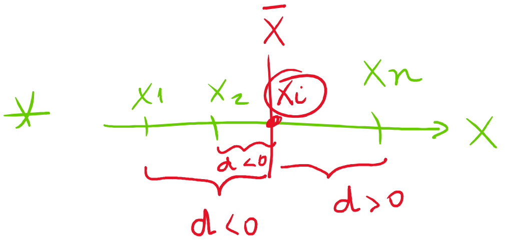
diag sec.

$$H = \begin{bmatrix} 2n & 2\Sigma X_i \\ 2\Sigma X_i & 2\Sigma X_i^2 \end{bmatrix}$$

diag principal

$$\det|H| = 2n.2\Sigma X_i^2 - (2\Sigma X_i)^2$$
$$= 4n\Sigma X_i^2 - 4(\Sigma X_i)^2$$
$$= 4\Big[\underbrace{n\Sigma X_i^2 - (\Sigma X_i)^2}\Big]$$

$SQX = \underbrace{\Sigma(X_i - \bar{X})^2}_{\geq 0}$



$$SQX = \Sigma(X_i - \bar{X})^2$$
$$= \Sigma(X_i^2 - 2X_i\bar{X} + \bar{X}^2)$$
$$= \Sigma X_i^2 - 2\bar{X}\Sigma X_i + n\bar{X}^2$$
$$= \Sigma X_i^2 - 2\frac{\Sigma X_i}{n}\Sigma X_i + \cancel{n}\frac{(\Sigma X_i)^2}{n^{\cancel{2}}}$$
$$= \Sigma X_i^2 - \frac{2(\Sigma X_i)^2}{n} + \frac{(\Sigma X_i)^2}{n}$$
$$= \left[\Sigma X_i^2 - \frac{(\Sigma X_i)^2}{n}\right] \cdot n$$
$$= n\Sigma X_i^2 - (\Sigma X_i)^2$$

OPÇÕES $\hat{\beta}_1$

1) $$\hat{\beta}_1 = \frac{\Sigma X_i Y_i - \frac{(\Sigma X_i)(\Sigma Y_i)}{n}}{\Sigma X_i^2 - (\Sigma X_i)^2}$$

OBS

$$\Sigma X_i^2 = X_1^2 + X_2^2 + \dots + X_n^2$$

1) $p_1 = \dfrac{n}{\sum X_i^2 - \dfrac{(\sum X_i)^2}{n}}$

$\sum X_i^2 = X_1^2 + X_2^2 + \cdots + X_n^2$

$(\sum X_i)^2 = (X_1 + X_2 + \cdots + X_n)^2$

2) $$\hat{\beta}_1 = \frac{n \sum X_i Y_i - (\sum X_i)(\sum Y_i)}{n \sum X_i^2 - (\sum X_i)^2}$$

3) $\hat{\beta}_1 \Rightarrow$ variável X centrada na sua média

$\frac{a}{b} \cdot c = \frac{ac}{b} = a \cdot \frac{c}{b}$

$x_i = X_i - \bar{X}$

$$\hat{\beta}_1 = \frac{\sum X_i Y_i - \frac{(\sum X_i)(\sum Y_i)}{n}}{\sum X_i^2 - \frac{(\sum X_i)^2}{n}} = \frac{\sum X_i Y_i - \bar{X} \sum Y_i}{\sum (X_i - \bar{X})^2} = \frac{\sum Y_i \overbrace{(X_i - \bar{X})}^{x_i}}{\sum \underbrace{(X_i - \bar{X})}_{x_i}^2}$$

$$\hat{\beta}_1 = \frac{\sum Y_i x_i}{\sum x_i^2}$$

4) $\hat{\beta}_1 = \dfrac{\sum y_i X_i}{\sum x_i^2}$

5) $\hat{\beta}_1 = \dfrac{\sum x_i y_i}{\sum x_i^2}$

Propriedades do SEN

1) O ponto $(\bar{X}, \bar{Y})$ é um ponto da reta estimada

se $X_i = \bar{X}$

$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$

$\hat{Y}_i = \bar{Y} - \hat{\beta}_1 \cancel{\bar{X}} + \hat{\beta}_1 \cancel{\bar{X}}$

$\hat{Y}_i = \bar{Y}$

2) Usando o fato de que $\sum [Y_i - \underbrace{\hat{\beta}_0 - \hat{\beta}_1 X_i}] = 0$

$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$

temos que: $\sum [Y_i - \hat{Y}_i] = 0$

temos que: $\sum [y_i - \hat{y}_i] = 0$ $\quad y_i = \beta_0 + \beta_1 x_i$

y_i	$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$
10	10,3
⋮	
11	9,8
$\sum y_i$	$\sum \hat{y}_i$

$\sum y_i - \sum \hat{y}_i = 0 \iff \sum y_i = \sum \hat{y}_i$

3) Usando o fato de que $\sum [y_i - \underbrace{\hat{\beta}_0 - \hat{\beta}_1 x_i}_{\hat{y}_i}](x_i) = 0$

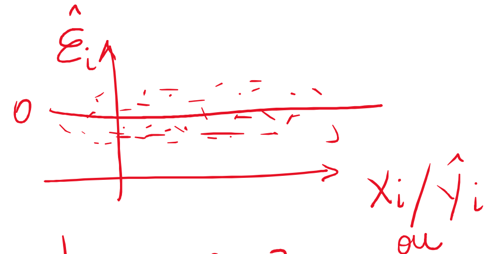
temos que:

$$\sum [y_i - \hat{y}_i](x_i) = 0$$

$$\sum y_i x_i - \sum \hat{y}_i x_i = 0 \Rightarrow \sum y_i x_i = \sum \hat{y}_i x_i$$

$$\sum [\hat{\varepsilon}_i](x_i) = 0$$

Análise de resíduo



4) Usando os resultados 2 e 3

$$\sum \hat{y}_i \hat{\varepsilon}_i = 0$$

$$\sum \hat{y}_i \hat{\varepsilon}_i = \sum (\hat{\beta}_0 + \hat{\beta}_1 x_i) \hat{\varepsilon}_i$$

$$= \hat{\beta}_0 \sum \hat{\varepsilon}_i + \hat{\beta}_1 \sum x_i \hat{\varepsilon}_i$$

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

$$= \hat{\beta}_0 \underbrace{[\sum y_i - \sum \hat{y}_i]}_{0} + \hat{\beta}_1 \underbrace{\sum x_i \hat{\varepsilon}_i}_{0}$$

$$= 0$$

5) Os estimadores $\hat{\beta}_0$ e $\hat{\beta}_1$ são combinações lineares das obs $y_i \to \hat{\beta}_0 = \sum y_i c_i = c_1 y_1 + c_2 y_2 + \dots + c_n y_n$

lineares das obs y_i $\to \hat{\beta}_0 = \sum y_i c_i = c_1 y_1 + c_2 y_2 + \dots + c_n y_n$

$\to \hat{\beta}_1 = \sum y_i d_i$

$$\hat{\beta}_1 = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} = \frac{\sum y_i x_i}{\sum x_i^2} \to c_i$$

$\sum x_i^2 = (x_1^2 + x_2 + \dots + x_n^2) = n^{\circ}$

$$= \frac{y_1 x_1 + y_2 x_2 + \dots + y_n x_n}{cte = \sum x_i^2}$$

$$\frac{a+b}{c} = \frac{a}{c} + \frac{b}{c}$$

$$= y_1 \frac{x_1}{K} + y_2 \frac{x_2}{K} + \dots + y_n \frac{x_n}{K} = y_1 c_1 + y_2 c_2 + \dots + y_n c_n$$

$$= \boxed{\sum c_i y_i = \hat{\beta}_1}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \frac{\sum y_i}{n} - \hat{\beta}_1 \bar{x}$$

$\hat{\beta}_1 \to \sum c_i y_i$

$$= \frac{\sum y_i}{n} - \sum c_i y_i \bar{x} = \sum y_i \underbrace{\left(\frac{1}{n} - c_i \bar{x}\right)}_{d_i}$$

$$\hat{\beta}_0 = \sum y_i d_i$$

$\sum c_i = 0$ $\qquad$ $\sum d_i = 1$

$\sum c_i x_i = 1$ $\qquad$ $\sum d_i x_i = 0$

$$\sum c_i = \sum \left(\frac{x_i}{\sum x_i^2}\right) = \frac{1}{\sum x_i^2} \sum x_i = \frac{1}{\sum x_i^2} \sum (x_i - \bar{x}) =$$

$$= \frac{1}{\sum x_i^2} (\sum x_i - n\bar{x}) = \frac{1}{\sum x_i^2} \left(\sum x_i - \cancel{n} (\sum x_i / \cancel{n})\right)$$

$$= \frac{1}{\sum x_i^2} (\sum x_i - \sum x_i) = 0$$

$$= \frac{1}{\sum x_i^2} \left(\sum X_i - \sum X_i \right) = \underline{0} //$$