/ /

# Estatística Inferencial Aula 21

## Regressão Linear Simples

Dado n pares de valores de duas variáveis $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ podemos estabelecer um modelo de regressão linear simples

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,$$

em que $y_i$ é o i-ésimo valor observado da variável resposta, $\beta_0$ é o parâmetro do modelo chamado intercepto, $\beta_1$ é o parâmetro chamado coeficiente angular, $x_i$ é o valor realizado da variável explicativa, $\varepsilon_i$ é o erro aleatório do modelo.

## Pressuposições do Modelo

$\varepsilon_1, \ldots, \varepsilon_n$ é uma amostra aleatória

$\varepsilon_i \sim N(0, \sigma^2)$, p/ $i = 1, \ldots, n$.

## Estimativa dos Parâmetros – Método dos Mínimos Quadrados

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \; ; \quad \hat{\beta}_1 = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}$$

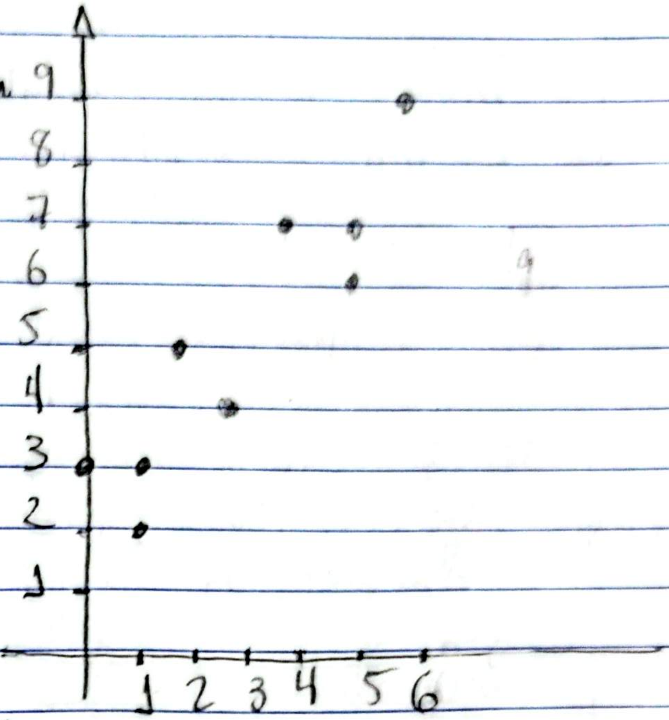
$$= \frac{s_{xy}}{s_x^2}$$

DATAPEL

/ /

Exemplo

| $y_i^2$ | $x_i$ | $y_i$ | $x_i^2$ | $x_i y_i$ |
|---|---|---|---|---|
| 9 | 0 | 3 | 0 | 0 |
| 4 | 1 | 2 | 1 | 2 |
| 9 | 1 | 3 | 1 | 3 |
| 25 | 2 | 5 | 4 | 10 |
| 16 | 3 | 4 | 9 | 12 |
| 16 | 3 | 4 | 9 | 12 |
| 49 | 4 | 7 | 16 | 28 |
| 36 | 5 | 6 | 25 | 30 |
| 49 | 5 | 7 | 25 | 35 |
| 81 | 6 | 9 | 36 | 54 |
| 294 | 30 | 50 | 126 | 186 |




$$\hat{B}_1 = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} = \frac{186 - \frac{30 \cdot 50}{10}}{126 - \frac{(30)^2}{10}}$$

$$\sum y_i^2 - \frac{(\sum y_i)^2}{n} = 294 - \frac{(50)^2}{10} = 44$$

$$= \frac{36}{36} = 1$$

$$\hat{B}_0 = \bar{y} - \hat{B}_1 \bar{x} = 5 - 1(3) = 2$$

$$\hat{y}_i = \hat{B}_0 + \hat{B}_1 x_i$$

$$R^2 = \frac{\left(\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}\right)^2}{\left(\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right)\left(\sum y_i^2 - \ldots\right)}$$

$$\hat{y}_i = 2 + x_i$$

$$= \frac{(36)^2}{(36)(44)} = \frac{36}{44} = \frac{9}{11} = ,81$$

$$\frac{\sum \hat{y}_i}{n} = \frac{1}{n}\sum(\bar{y} - \hat{B}_1 \bar{x} + \hat{B}_1 x_i)$$

$$= \bar{y} - \hat{B}_1 \bar{x} + \hat{B}_1 \bar{x}$$

$$= \bar{y}$$

DATAPEL

$\hat{y}_i = 2 + x_i$ / /

| $x_i$ | $y_i$ | $y_i - \bar{y}$ | $(y_i-\bar{y})^2$ | $\hat{y}_i$ | $(\hat{y}_i-\bar{y})$ | $(\hat{y}_i-\bar{y})$ |
|---|---|---|---|---|---|---|
| 0 | 3 | $3-5=-2$ | 4 | 2 | -3 | 9 |
| 1 | 2 | $2-5=-3$ | 9 | 3 | -2 | 4 |
| 1 | 3 | $3-5=-2$ | 4 | 3 | -2 | 4 |
| 2 | 5 | $5-5=0$ | 0 | 4 | -1 | 1 |
| 3 | 4 | $4-5=-1$ | 1 | 5 | 0 | 0 |
| 3 | 4 | $4-5=-1$ | 1 | 5 | 0 | 0 |
| 4 | 7 | $7-5=2$ | 4 | 6 | 1 | 1 |
| 5 | 6 | $6-5=1$ | 1 | 7 | 2 | 4 |
| 5 | 7 | $7-5=2$ | 4 | 7 | 2 | 4 |
| 6 | 9 | $9-5=4$ | 16 | 8 | 3 | 9 |
| | 50 | | | 50 | | |

$$\sum_{i=1}^{n} (y_i-\bar{y})^2 = 44 \qquad \sum_{i=1}^{n} (\hat{y}_i-\bar{y})^2 = 36$$

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i-\bar{y})^2}{\sum_{i=1}^{n} (y_i-\bar{y})^2} = \frac{36}{44} = \frac{9}{11} = 81{,}81\%$$

# Regressão linear Simples

$$r = \frac{S_{xy}}{\sqrt{S_x^2}\sqrt{S_y^2}}$$, sendo

$$S_{xy} = \frac{1}{n-1}\left[\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\right] = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}\right]$$

$$S_x^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\right] = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right]$$

$$S_y = \frac{1}{n-1}\left[\sum_{i=1}^{n}(y_i-\bar{y})^2\right] = \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}\right]$$

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

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

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

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

$$= \frac{\sum_{i=1}^{n}\left[\bar{y}-\hat{\beta}_1\bar{x} - (\bar{y} + \hat{\beta}_1 x_i)\right]^2}{\sum(y_i-\bar{y})^2}$$

$$= \frac{\sum_{i=1}^{n}\left(\hat{\beta}_0 + \hat{\beta}_1 x_i - \bar{y}\right)^2}{\sum(y_i-\bar{y})^2} = \frac{\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$$

Exercícios

1- O coeficiente de determinação da regressão $R^2$ é igual a 0,36. Sabendo que desvio padrão de X é igual a 1,50, desvio padrão de Y é igual a 2,00, $\bar{X} = 10$ e $\bar{Y} = 20$, determine a equação de regressão de Y em relação a X.

2- Para duas variáveis negativamente correlacionadas, foram obtidos: $\bar{X} = 0$, $\bar{Y} = 12$, desvio padrão de X é igual a 8, desvio padrão Y é igual a 10 e $r^2 = 0,64$.

Determine a equação de regressão de Y em relação a X.

1):

$$r^2 = \frac{\left(\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}\right)^2}{\sum x_i^2 - \frac{(\sum x_i)^2}{n} \sum y_i^2 - \frac{(\sum y_i)^2}{n}} \qquad 12 + 0,8x$$

$$S_x = \sqrt{\frac{1}{n-1}\left(\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right)} \qquad S_y = \sqrt{\frac{1}{n-1}\left(\sum y_i^2 - \frac{(\sum y_i)^2}{n}\right)}$$

$$\hat{\beta}_1 = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$S_x = 1,50 \Rightarrow S_x^2 = 2,25$ $\qquad r^2 = \dfrac{\left[\frac{1}{n-1} \sum x_i y_i - \frac{\sum x_i \sum y_i}{n}\right]^2}{S_x^2 S_y^2} = \dfrac{S_{xy}^2}{S_x^2 S_y^2}$

$S_y = 2,00 \Rightarrow S_y^2 = 4,00$

$S_{xy}^2 = r^2 S_x^2 S_y^2 \Rightarrow S_{xy}^2 = 0,36 \cdot 4 \cdot 2,25 \Rightarrow S_{xy}^2 = 3,24 \Rightarrow S_{xy} = 1,8$

$\hat{\beta}_1 = \frac{Sxy}{Sx^2} \Rightarrow \hat{\beta}_1 = \frac{1,8}{2,25} = 0,80$

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \Rightarrow \hat{\beta}_0 = 20 - 0,80 \cdot 10$
$= 20 - 8$

$\hat{y}_i = 12,0 + 0,80 x_i$

2-

$r^2 = 0,64$ $\quad \bar{x} = 0$ $\quad r^2 = \frac{Sxy^2}{Sx^2 Sy^2}$

$Sx = 8$ $\quad \bar{y} = 12$

$Sy = 10$

$Sxy^2 = r^2 Sx^2 Sy^2$
$= 0,64 \cdot 64 \cdot 100$

$Sxy^2 = 64^2 \Rightarrow Sxy = \pm 64$

No entanto, como ele disse que é negativamente correlacionado, $Sxy = -64$

$\hat{\beta}_1 = \frac{Sxy}{Sx^2} = \frac{-64}{64} = -1$

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \Rightarrow 12$

$\hat{y}_i = 12 - x_i$