

# Lista 4 - Estatística

Gustavo Lopes Rodrigues

2021

## Fórmulas

$$IC(1-\alpha)\% = \hat{y} \pm \tau_{\frac{\alpha}{2}} ; \times n-2 \times S_e \times \sqrt{1 + \frac{1}{n} + \frac{n(X_0 - \bar{X})^2}{S_{xx}}}$$

$$IC(1-\alpha)\% = \hat{y} \pm \tau_{\frac{\alpha}{2}} ; \times n-2 \times S_e \times \sqrt{\frac{1}{n} + \frac{n(X_0 - \bar{X})^2}{S_{xx}}}$$

$$\text{Corr}(x,y) = \frac{S_{xy}}{\sqrt{S_{xx} \times S_{yy}}}$$

$$S_{xx} = n \times \sum x^2 - (\sum x)^2$$

$$S_{yy} = n \times \sum y^2 - (\sum y)^2$$

$$S_{xy} = n \times \sum xy - (\sum x)(\sum y)$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$$

$$S_e = \sqrt{\frac{\sum y^2 - \hat{\beta}_0 \sum y - \hat{\beta}_1 \sum xy}{n-2}}$$

Questão 1)

$$S_x = 30$$

$$S_y = 11240$$

$$S_x^2 = 128$$

$$S_y^2 = 20353600$$

$$S_{xy} = 50480$$

$$\text{a) } S_{xx} = n \times \sum x^2 - (\sum x)^2$$

$$S_{xx} = 8 \times 128 - (30)^2$$

$$S_{xx} = 1024 - 900$$

$$S_{xx} = 124$$

$$S_{yy} = n \times \sum y^2 - (\sum y)^2$$

$$S_{yy} = 8 \times 20353600 - (11240)^2$$

$$S_{yy} = 162828800 - 126337600$$

$$S_{yy} = 36491200$$

$$S_{xy} = n \times \sum xy - (\sum x)(\sum y)$$

$$S_{xy} = 8 \times 50480 - 337200$$

$$S_{xy} = 403840 - 337200$$

$$S_{xy} = 66640$$

$$\text{Corr}(x,y) = \frac{S_{xy}}{\sqrt{S_{xx} \times S_{yy}}}$$

$$\text{Corr}(x,y) = \frac{66640}{\sqrt{124 \times 36491200}}$$

$$\text{Corr}(x,y) = \frac{66640}{67267}$$

$$\text{Corr}(x,y) = 0.9907$$

**Interpretação do resultado:** Há uma forte correlação positiva.

$$\text{b) } \hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$$

$$\hat{\beta}_1 = \frac{66640}{124} = 537.42$$

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

$$\hat{\beta}_0 = 1405 - (537.42) \times 3.75$$

$$\hat{\beta}_0 = -610.325$$

**Interpretação do resultado:**

$\hat{\beta}_0$  : Não há interpretação prática.

$\hat{\beta}_1$  : Para cada ano, existe um acréscimo na média em 537.42 palavras no vocabulário de cada criança.

c)  $R^2 = (0.9907)^2 = 0.9815$  ou 98.15%

**Interpretação do resultado:**

A variabilidade no número de palavras no vocabulário de uma criança de 98.15%, pode ser explicado pela idade. Os restantes 1.85% não possuem explicação clara, logo pode ser resultado de um possível erro, ou outra variável que não foi incluída no estudo da questão.

d) 
$$S_e = \sqrt{\frac{\sum y^2 - \hat{\beta}_0 \sum y - \hat{\beta}_1 \sum xy}{n-2}}$$

$$S_e = \sqrt{\frac{20353600 - (610.325 \times 11240) - (537.42 \times 50480)}{6}}$$

$$S_e = \sqrt{\frac{84691.4}{6}}$$

$$S_e = 118.81$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \times n$$

$$\hat{y} = -610.325 + (537.42) \times 7$$

$$\hat{y} = 3151.615$$

$$IC(95\%) = \hat{y} \pm \tau_{\frac{\alpha}{2}} ; \times n-2 \times S_e \times \sqrt{1 + \frac{1}{n} + \frac{n(X_0 - \bar{X})^2}{S_{xx}}}$$

$$IC(95\%) = 3151.615 \pm 2.4469 \times (118.81 \times \sqrt{1 + \frac{1}{8} + \frac{8(7-3.75)^2}{124}})$$

$$IC(95\%) = 3151.615 \pm 390.73$$

$$IC(95\%) = [2760.885 ; 3542.345]$$

e) 
$$IC(95\%) = \hat{y} \pm \tau_{\frac{\alpha}{2}} ; \times n-2 \times S_e \times \sqrt{\frac{1}{n} + \frac{n(X_0 - \bar{X})^2}{S_{xx}}}$$

$$IC(95\%) = 3151.615 \pm 2.4469 \times (118.81 \times \sqrt{\frac{1}{8} + \frac{8(7-3.75)^2}{124}})$$

$$IC(95\%) = 3151.615 \pm 261.07$$

$$IC(95\%) = [2890.545 ; 3412.685]$$

Questão 2)

$$S_x = 60$$

$$S_y = 891$$

$$S_x^2 = 346$$

$$S_y^2 = 65451$$

$$S_{xy} = 4620$$

$$\text{a) } S_{xx} = n \times \sum x^2 - (\sum x)^2$$

$$S_{xx} = 13 \times 346 - (60)^2$$

$$S_{xx} = 4498 - 3600$$

$$S_{xx} = 898$$

$$S_{yy} = n \times \sum y^2 - (\sum y)^2$$

$$S_{yy} = 13 \times 65451 - (891)^2$$

$$S_{yy} = 850863 - 793881$$

$$S_{yy} = 56982$$

$$S_{xy} = n \times \sum xy - (\sum x)(\sum y)$$

$$S_{xy} = 13 \times 4620 - 53460$$

$$S_{xy} = 60060 - 53460$$

$$S_{xy} = 6600$$

$$\text{Corr}(x,y) = \frac{S_{xy}}{\sqrt{S_{xx} \times S_{yy}}}$$

$$\text{Corr}(x,y) = \frac{6600}{\sqrt{898 \times 56982}}$$

$$\text{Corr}(x,y) = \frac{6600}{7153.30}$$

$$\text{Corr}(x,y) = 0.9226$$

**Interpretação do resultado:** Há uma forte correlação positiva.

$$\text{b) } \hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$$

$$\hat{\beta}_1 = \frac{6600}{898} = 7.35$$

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

$$\hat{\beta}_0 = 68.54 - (7.35) \times 4.62$$

$$\hat{\beta}_0 = 34.58$$

**Interpretação do resultado:**

$\hat{\beta}_0$  : Para os alunos que não possui horas de estudo é esperado uma nota média de 34.58 pontos.

$\hat{\beta}_1$  : Para cada hora de estudo adicional, terá um aumento na média de 7.35 pontos na prova.

c)  $R^2 = (0.9226)^2 = 0.8512$  ou 85.12%

**Interpretação do resultado:**

A variabilidade da pontuação do teste de 85.12%, pode ser explicado pelas horas de estudo. Os restantes 14.88% não possuem explicação clara, logo pode ser resultado de um possível erro, ou outra variável que não foi incluído no estudo da questão.

d) 
$$S_e = \sqrt{\frac{\sum y^2 - \hat{\beta}_0 \sum y - \hat{\beta}_1 \sum xy}{n-2}}$$

$$S_e = \sqrt{\frac{65451 - (34.58 \times 891) - (7.35 \times 4620)}{11}}$$

$$S_e = \sqrt{\frac{683.22}{11}}$$

$$S_e = 7.88$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \times n$$

$$\hat{y} = 34.58 + (7.35) \times 3$$

$$\hat{y} = 56.63$$

$$IC(99\%) = \hat{y} \pm \tau_{\frac{\alpha}{2}} ; \times n-2 \times S_e \times \sqrt{1 + \frac{1}{n} + \frac{n(X_0 - \bar{X})^2}{S_{xx}}}$$

$$IC(99\%) = 56.63 \pm 3.1058 \times (7.88 \times \sqrt{1 + \frac{1}{13} + \frac{13(3-4.62)^2}{898}})$$

$$IC(99\%) = 56.63 \pm 25.84$$

$$IC(99\%) = [30.79 ; 82.47]$$

e) 
$$IC(99\%) = \hat{y} \pm \tau_{\frac{\alpha}{2}} ; \times n-2 \times S_e \times \sqrt{\frac{1}{n} + \frac{n(X_0 - \bar{X})^2}{S_{xx}}}$$

$$IC(99\%) = 56.63 \pm 3.1058 \times (7.88 \times \sqrt{\frac{1}{13} + \frac{13(3-4.62)^2}{898}})$$

$$IC(99\%) = 56.63 \pm 8.29$$

$$IC(99\%) = [48.34 ; 64.92]$$

Questão 3)

$$S_x = 54$$

$$S_y = 908$$

$$S_x^2 = 332$$

$$S_y^2 = 70836$$

$$S_{xy} = 3724$$

$$\text{a) } S_{xx} = n \times \sum x^2 - (\sum x)^2$$

$$S_{xx} = 12 \times 332 - (54)^2$$

$$S_{xx} = 3984 - 2916$$

$$S_{xx} = 1068$$

$$S_{yy} = n \times \sum y^2 - (\sum y)^2$$

$$S_{yy} = 12 \times 70836 - (908)^2$$

$$S_{yy} = 850032 - 824464$$

$$S_{yy} = 25568$$

$$S_{xy} = n \times \sum xy - (\sum x)(\sum y)$$

$$S_{xy} = 12 \times 3724 - 49032$$

$$S_{xy} = 44688 - 49032$$

$$S_{xy} = -4344$$

$$\text{Corr}(x,y) = \frac{S_{xy}}{\sqrt{S_{xx} \times S_{yy}}}$$

$$\text{Corr}(x,y) = \frac{-4344}{\sqrt{1068 \times 25568}}$$

$$\text{Corr}(x,y) = \frac{-4344}{5225.57}$$

$$\text{Corr}(x,y) = -0.8313$$

**Interpretação do resultado:** Há uma forte correlação negativa.

$$\text{b) } \hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$$

$$\hat{\beta}_1 = \frac{-4344}{1068} = -4.07$$

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

$$\hat{\beta}_0 = 75.66 - (-4.07) \times 4.5$$

$$\hat{\beta}_0 = 93.97$$

**Interpretação do resultado:**

$\hat{\beta}_0$  : Para os alunos que não gastam tempo nenhum na frente da televisão é esperado uma nota média de 93.97 pontos na prova de segunda.

$\hat{\beta}_1$  : Para cada hora adicional na frente da televisão, é esperado um decréscimo na média em 4.07 pontos na nota.

c)  $R^2 = (0.8313)^2 = 0.691$  ou 69.10%

**Interpretação do resultado:**

A variabilidade da pontuação do teste de 69.10%, pode ser explicado pelas horas na frente da televisão. Os restantes 30.90% não possuem explicação clara, logo pode ser resultado de um possível erro, ou outra variável que não foi incluída no estudo da questão.

$$\begin{aligned} \text{d) } S_e &= \sqrt{\frac{\sum y^2 - \hat{\beta}_0 \sum y - \hat{\beta}_1 \sum xy}{n-2}} \\ S_e &= \sqrt{\frac{70836 - (93.97 \times 908) - (4.07 \times 3724)}{10}} \\ S_e &= \sqrt{\frac{667.92}{10}} \\ S_e &= 8.17 \end{aligned}$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \times n$$

$$\hat{y} = 93.97 + (-4.07) \times 9$$

$$\hat{y} = 57.34$$

$$\text{IC}(95\%) = \hat{y} \pm \tau_{\frac{\alpha}{2}} ; \times n-2 \times S_e \times \sqrt{1 + \frac{1}{n} + \frac{n(X_0 - \bar{X})^2}{S_{xx}}}$$

$$\text{IC}(95\%) = 57.34 \pm 2.2281 \times (8.17 \times \sqrt{1 + \frac{1}{12} + \frac{12(9-4.5)^2}{1068}})$$

$$\text{IC}(95)\% = 57.34 \pm 20.84$$

$$\text{IC}(95)\% = [36.5 ; 78.18]$$

e)  $\text{IC}(95\%) = \hat{y} \pm \tau_{\frac{\alpha}{2}} ; \times n-2 \times S_e \times \sqrt{\frac{1}{n} + \frac{n(X_0 - \bar{X})^2}{S_{xx}}}$

$$\text{IC}(95\%) = 57.34 \pm 2.2281 \times (8.17 \times \sqrt{\frac{1}{12} + \frac{12(9-4.5)^2}{1068}})$$

$$\text{IC}(95)\% = 57.34 \pm 10.15$$

$$\text{IC}(95)\% = [47.19 ; 67.49]$$