Exemplo (modelo RLS)

Para explorar a relação entre a massa muscular e a idade (no género feminino) um nutricionista selecionou aleatoriamente 16 mulheres com idades compreendidas entre os 40 e os 79 anos. Os resultados observados encontram-se na tabela seguinte (x representa a idade e y é um índice de massa muscular):

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
x_i	71	64	43	67	56	73	68	56	76	65	45	58	45	53	49	78
y_i	82	91	100	68	87	73	78	80	65	84	116	76	97	100	105	77

$$\sum_{i=1}^{16} x_i = 967, \quad \sum_{i=1}^{16} y_i = 1379, \quad \sum_{i=1}^{16} x_i^2 = 60409, \quad \sum_{i=1}^{16} y_i^2 = 121\,887$$

$$\sum_{i=1}^{16} x_i y_i = 81331.$$

Admite que o modelo RLS ($Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$) é adequado.

a) Calcule:

i) uma estimativa pontual da diferença entre as massas musculares médias de mulheres cujas idades diferem de um ano;

ii) uma estimativa pontual da massa muscular média para as mulheres de 60 anos;

iii) o valor do resíduo para a 8ª observação;

(iv) uma estimativa pontual de $Var(\varepsilon_i)=\sigma^2$;

(v) o coeficiente de determinação e interprete o valor obtido.

b) O nutricionista pensa que, na gama de idades considerada, a massa muscular é significativam/ influenciada pela idade. Acha que as observações confirmam esta hipótese? Use um nível de significância de 5% e indique as hipóteses de trabalho de que necessita para efectuar o teste.

c) Calcule o intervalo de confiança a 99% para o valor esperado da massa muscular para uma mulher de 45 anos. Acha legítimo usar o mesmo procedimento tratando-se de uma mulher com 20 anos em vez de 45?

Resolução: cálculos auxiliares $(n=16)$

$$\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} = 81331 - 16\left(\frac{967}{16}\right)\left(\frac{1379}{16}\right) = -2012.3125$$

$$\sum_{i=1}^{16} x_i^2 - n\bar{x}^2 = 60409 - 16\left(\frac{967}{16}\right)^2 = 1965.9375$$

$$\sum_{i=1}^{16} y_i^2 - n\bar{y}^2 = 121887 - 16\left(\frac{1379}{16}\right)^2 = 3034.4375$$

a) i) $E(Y|(x_j+1)) - E(Y|x_j) = \beta_0+\beta_1(x_j+1) - (\beta_0+\beta_1 x_j)$

$= \beta_1$ e $\hat{\beta}_1 = \dfrac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} = \dfrac{-2012.3125}{1965.9375} \simeq \underline{\underline{-1.0236}}$

(ii) $\widehat{E(Y|x=60)} = \hat{\beta}_0 + \hat{\beta}_1 \times 60$

$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x} = \frac{1379}{16} + 1.0236 \times \frac{967}{16} \simeq 148.0513$

$\Rightarrow \widehat{E(Y|x=60)} \simeq 148.0513 - 1.0236 \times 60 \simeq 86.6353$

(iii) $e_8 = Y_8 - \hat{Y}_8$ com $\hat{Y}_8 = \widehat{E(Y|x_8=56)} =$

$= \hat{\beta}_0 + \hat{\beta}_1 x_8 = 148.0513 - 1.0236 \times 56 = 90.7297$

$\Rightarrow e_8 = 80 - 90.7297 = -10.7297$

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

$= \frac{1}{14}\left[3034.4375 - (-1.0236)^2 \times 1965.9375\right]$

$\simeq 69.6152$

(v) $r^2 = \frac{\left(\sum x_i y_i - n\bar{x}\bar{y}\right)^2}{\left(\sum x_i^2 - n\bar{x}^2\right)\left(\sum y_i^2 - n\bar{y}^2\right)} = \frac{(-2012.3125)^2}{(1965.9375)(3034.4375)}$

$\simeq 0.6788 \Rightarrow$ cerca de 68% da variabilidade deste índice de massa muscular é explicada pela idade da mulher.

b) $H_0: \beta_1 = 0$ vs $H_1: \beta_1 \neq 0$

Hipótese de trabalho necessária:

$\varepsilon_i \underset{i.i.d.}{\sim} N(0, \sigma^2), \quad \forall i = 1, \dots, 16.$

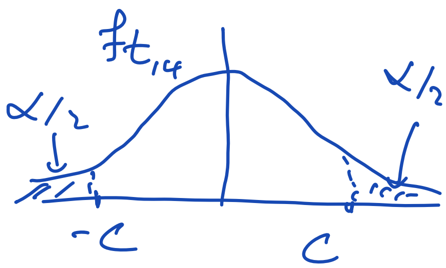
Estatística de teste:

$$T_{H_0} = \frac{\hat{\beta}_1 - \overbrace{\beta_{1,0}}^{=0}}{\sqrt{\hat{\sigma}^2 / \left(\sum x_i^2 - n\bar{x}^2\right)}} \sim t_{(n-2)}$$

Valor observado da estatística de teste:

$$t_{obs} = \frac{-1.0236}{\sqrt{69.6152/1965.9375}} \simeq -5.4396$$

Região crítica para T_{H_0} ao nível de significância $\alpha = 0.05$:



$$RC_{0.05} = (-\infty, -c) \cup (c, +\infty)$$

com $c = F^{-1}_{t_{14}}(1-\alpha/2) = F^{-1}_{t_{14}}(0.975) \underset{\substack{\uparrow \\ \text{tabela}}}{=} 2.145$

Decisão: $t_{obs} = -5.4396 \in RC_{0.05}$

$\Rightarrow$ rejeitar H_0 ao nível de significância de 5%, i.e. parece haver evidência de que a idade da mulher influencia a sua massa muscular.

Nota: valor-p $= 2P(T_{H_0} < t_{obs} \simeq -5.44) =$

$$= 2\left(1 - P(T_{H_0} < 5.44)\right) = 2\Big(1 - \underbrace{F_{t_{14}}(5.44)}_{> 0.9995 \text{ (tabela)}}\Big) < 0.001$$

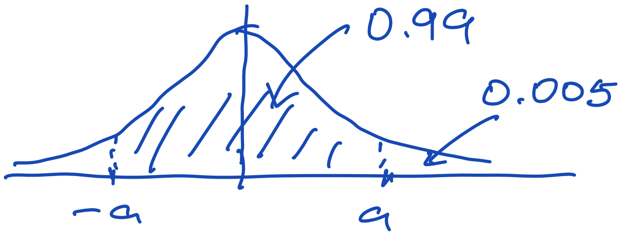
$\Rightarrow$ rejeitar H_0 até com nível de significância 0.1%.

c) $IC_{99\%}$ para $E(Y\,|\,x=45) = \beta_0 + \beta_1 \times 45$

<u>Nota</u>: $x = 45 \in [\min(x_i), \max(x_i)] = [43, 78]$ ✓

$x = 20 \notin [43,78] \Rightarrow$ não é legítimo fazer o mesmo para uma mulher com 20 anos.

Variável aleatória fulcral: $$T = \frac{(\hat\beta_0 + \hat\beta_1 \times 45) - (\beta_0 + \beta_1 \times 45)}{\sqrt{\left(\frac{1}{16} + \frac{(\bar x - x_0)^2}{(\sum x_i^2 - 16\bar x^2)}\right)\hat\sigma^2}} \sim t_{14}$$



$$\Rightarrow a = F^{-1}_{t_{14}}(0.995) \overset{\text{tabela}}{\doteq} 2.977$$

$$\Rightarrow IC_{99\%}(\beta_0 + \beta_1 \times 45) = \left((\hat\beta_0 + \hat\beta_1 \times 45) \pm a \sqrt{(\ \cdot\ )\hat\sigma^2} \right)$$

Concretização: $\hat\beta_0 + \hat\beta_1 \times 45 \simeq 148.0513 - 1.0236 \times 45 \simeq 101.99$

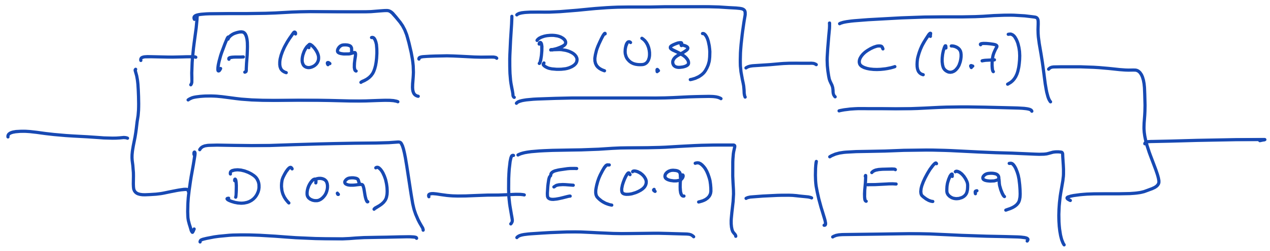
$$\sqrt{\left(\frac{1}{16} + \frac{\left(\frac{967}{16} - 45\right)^2}{1965.9375}\right) \times 69.6152} \simeq 3.5763$$

$$\Rightarrow IC_{99\%}(\beta_0 + \beta_1 \times 45) = \left[101.99 - 2.977 \times 3.5763,\ 101.99 + \underbrace{2.977 \times 3.5763}_{\simeq 10.6466}\right]$$

$$\simeq [91.34,\ 112.64]$$

Resolução do 1º Exame do 1º semestre de 2021/22

(1)



Componentes funcionam de forma mutuam/ indep.

$$P(\text{"circuito funcionar"}) = P[\underbrace{(A\cap B\cap C)}_{ABC} \cup \underbrace{(D\cap E\cap F)}_{DEF}]$$

$$= P(ABC) + P(DEF) - P(ABC \cap DEF)$$

indep.

$$= P(A)P(B)P(C) + P(D)P(E)P(F) - P(A)P(B)P(C)P(D)P(E)P(F)$$

$$= 0.9\times0.8\times0.7 + 0.9\times0.9\times0.9 - 0.9\times0.8\times0.7\times0.9\times0.9\times0.9$$

$$= 0.504 + 0.729 - 0.504\times0.729$$

$$= 0.865584$$

(2) $X \sim \text{Poisson}(\lambda)$, $Var(X) = 10.5$

Mostre que $mo(X) = 10$ e obtenha $P(X > 10)$.

- $P(X=x) = \dfrac{e^{-\lambda}\lambda^{x}}{x!}$, $x = 0, 1, \dots$, $E(X) = Var(X) = \underline{\underline{\lambda = 10.5}}$

- $\dfrac{P(X=mo)}{P(X=mo-1)} \geq 1 \Leftrightarrow \dfrac{e^{-\lambda}\lambda^{mo}}{mo!}\cdot\dfrac{(mo-1)!}{e^{-\lambda}\lambda^{mo-1}} = \dfrac{\lambda}{mo} \geq 1$

$$\Downarrow$$

$$\frac{P(X=mo)}{P(X=mo+1)} \geq 1 \Leftrightarrow \frac{mo+1}{\lambda} \geq 1 \Longrightarrow$$

$$\lambda - 1 \leq mo \leq \lambda$$

$$\Rightarrow 9.5 \leq mo \leq 10.5 \Rightarrow mo = \underline{\underline{10}}\ \checkmark$$

- $P(X>10) = 1 - P(X\leq 10) = 1 - F_{\text{Poisson}(10.5)}(10)$

(tabela) $= 1 - 0.5207 = 0.4793$ //

(3) $X \sim \exp(\lambda)$, $E(X^2 + X + 1) = 2$

$\uparrow$ duração em <u>centenas de horas</u>

$\lambda = ?$ $\qquad P(X\leq 2 \mid X>0.5) = ?$

- $f_X(x) = \lambda e^{-\lambda x}$, $x \geq 0$; $E(X) = \frac{1}{\lambda}$; $Var(X) = \frac{1}{\lambda^2} =$
$= E(X^2) - E(X)^2$

- $E(X^2+X+1) = 2 \Rightarrow E(X^2) + E(X) + 1 = 2$

$\Rightarrow \left(\frac{1}{\lambda^2} + \left(\frac{1}{\lambda}\right)^2\right) + \frac{1}{\lambda} = 1 \Rightarrow \frac{2}{\lambda^2} + \frac{\lambda}{\lambda^2} = \frac{\lambda^2}{\lambda^2}$

$\Rightarrow \lambda^2 - \lambda - 2 = 0 \Rightarrow \lambda = \frac{1 \pm \sqrt{1+8}}{2} = \frac{1\pm 3}{2} \begin{cases} 2 \\ -1 \end{cases}$

$\Rightarrow \boxed{\lambda = 2}$

propriedade de falta de memória da exponencial.

- $P(X\leq 2 \mid X>0.5) = P(X\leq 1.5) =$

$$= \int_0^{1.5} \lambda e^{-\lambda x}\,dx = \left[-e^{-\lambda x}\right]_0^{1.5} = -e^{-1.5\times\lambda} + 1$$

$$= 1 - e^{-3} = 1 - 0.049787 = 0.950213 \text{ //}$$

[<u>Nota</u>: $P(X\leq 2 \mid X>0.5) = \frac{P(0.5<X\leq 2)}{P(X>0.5)} =$

$$= \frac{(1-e^{-2\lambda}) - (1-e^{-0.5\lambda})}{1-(1-e^{-0.5\lambda})} = \frac{e^{-0.5\lambda} - e^{-2\lambda}}{e^{-0.5\lambda}} =$$

$$= 1 - e^{-1.5\lambda} \text{ //}$$]