DM
DEPARTAMENTO DE MATEMÁTICA
TÉCNICO LISBOA

Probabilidade e Estatística

TODOS OS CURSOS

2º Semestre – 2023/2024
24/07/2024
10:30–12:30

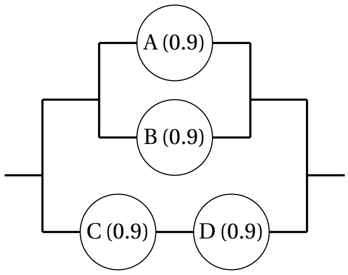
Duração: **120** minutos

Exame Época Especial

- As questões de escolha múltipla têm uma e apenas uma resposta correta. Caso assinale uma resposta errada ou mais de uma resposta, serão descontados 0.5 valores.
- As respostas às questões de escolha múltiplas têm de ser assinaladas na folha de instruções previamente distribuída.
- Para as restantes questões, terá de justificar convenientemente as suas respostas.
- Só pode sair da sala uma hora após o início do exame, devendo nesse caso entregar a sua prova ou desistir da mesma.

Pergunta 1 — 2 valores

Considere um circuito constituído pelas componentes A, B, C e D, que funcionam de forma mutuamente independente e estão dispostas conforme o diagrama ao lado.



A probabilidade de que cada componente funcione consta deste mesmo diagrama (e está entre parêntesis). O circuito funciona se e só se houver um caminho da esquerda para a direita que passe por componentes funcionais.

Qual é a probabilidade de o circuito funcionar?

- **Quadro de acontecimentos e probabilidades**

Acontecimento	Probabilidade
Z = componente Z está a funcionar	$P(Z) = 0.9,\ Z = A, B, C, D$

- **Prob. pedida**

Ao admitirmos que os eventos A, B, C e D são mutuamente independentes e que $P(A) = P(B) = P(C) = P(D) = 0.9$, temos

$$\begin{aligned}
&P(\text{“circuito funciona”})\\
&= P[(A\cup B)\cup(C\cap D)]\\
&= P(A\cup B) + P(C\cap D) - P[(A\cup B)\cap(C\cap D)]\\
&= P(A\cup B) + P(C\cap D) - P[(A\cap C\cap D)\cup(B\cap C\cap D)]\\
&= [P(A)+P(B)-P(A\cap B)] + P(C\cap D) - [P(A\cap C\cap D)+P(B\cap C\cap D)-P(A\cap B\cap C\cap D)]\\
&= [P(A)+P(B)-P(A)\times P(B)] + P(C)\times P(D)\\
&\quad -[P(A)\times P(C)\times P(D)+P(B)\times P(C)\times P(D)-P(A)\times P(B)\times P(C)\times P(D)]\\
&= (0.9+0.9-0.9^2)+0.9^2-(0.9^3+0.9^3-0.9^4)\\
&= 0.9981.
\end{aligned}$$

Pergunta 2 — 2 valores

O número mensal de ataques informáticos a uma empresa é uma variável aleatória com distribuição de Poisson com valor esperado igual a dois ataques.

Calcule a moda do número mensal de ataques informáticos a esta empresa.

- **V.a.**

  X = número mensal de ataques informáticos

- **Distribuição**

  $X \sim \text{Poisson}(\lambda)$, onde λ: $E(X) = 2 \Leftrightarrow \lambda = 2$

- **F.p. de X**

  $P(X = x) = e^{-\lambda}\frac{\lambda^x}{x!}$, $x \in \mathbb{N}_0$

- **Moda de X**

  É sabido que $mo = mo(X) = \text{argmax}_{x\in\mathbb{N}_0} P(X = x)$. Logo,

$$mo = mo(X) \in \mathbb{N}_0 \quad : \quad \begin{cases} P(X = mo) \geq P(X = mo - 1) \\ P(X = mo) \geq P(X = mo + 1) \end{cases} \quad \begin{cases} \frac{P(X=mo)}{P(X=mo-1)} \geq 1 \\ \frac{P(X=mo+1)}{P(X=mo)} \leq 1 \end{cases} \quad \begin{cases} \frac{\frac{e^{-\lambda}\lambda^{mo}}{mo!}}{\frac{e^{-\lambda}\lambda^{mo-1}}{(mo-1)!}} \geq 1 \\ \frac{\frac{e^{-\lambda}\lambda^{mo+1}}{(mo+1)!}}{\frac{e^{-\lambda}\lambda^{mo}}{mo!}} \leq 1 \end{cases}$$

$$\begin{cases} \frac{\lambda}{mo} \geq 1 \\ \frac{\lambda}{mo+1} \leq 1 \end{cases} \quad \begin{cases} mo \leq \lambda \\ mo + 1 \geq \lambda, \end{cases}$$

  i.e., a v.a. X tem duas modas: 1 e 2. [Note-se que 1 e 2 estão, efectivamente, associados ao mesmo valor máximo da f.p., $P(X = 1) = P(X = 2) = e^{-2}$. X diz-se uma v.a. bimodal.]

Pergunta 3	2 valores

Suponha que o tempo de tromboplastina parcialmente ativada (TTPa) em segundos associado a um paciente saudável segue uma distribuição normal de valor esperado 35 segundos e desvio padrão 2 segundos.

Calcule a probabilidade de o TTPa medido num paciente saudável se encontrar entre 33 a 37 segundos, sabendo que tal TTPa se encontra entre os valores de referência 30 e 40 segundos.

- **V.a.**

  X = tempo de tromboplastina parcialmente ativada (TTPa), em segundos

- **Distribuição**

  $X \sim \text{normal}(\mu = 35, \sigma^2 = 2^2)$

- **Probabilidade pedida**

$$\begin{aligned}
P(33 \leq X \leq 37 \,|\, 30 \leq X \leq 40) &= \frac{P(33 \leq X \leq 37, 30 \leq X \leq 40)}{P(30 \leq X \leq 40)} = \frac{P(33 \leq X \leq 37)}{P(30 \leq X \leq 40)} \\
&= \frac{P\left(\frac{33-\mu}{\sigma} \leq Z = \frac{X-\mu}{\sigma} \leq \frac{37-\mu}{\sigma}\right)}{P\left(\frac{30-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{40-\mu}{\sigma}\right)} = \frac{\Phi\left(\frac{37-\mu}{\sigma}\right) - \Phi\left(\frac{33-\mu}{\sigma}\right)}{\Phi\left(\frac{40-\mu}{\sigma}\right) - \Phi\left(\frac{35-\mu}{\sigma}\right)} \\
&= \frac{\Phi\left(\frac{37-35}{2}\right) - \Phi\left(\frac{33-35}{2}\right)}{\Phi\left(\frac{40-35}{2}\right) - \Phi\left(\frac{30-35}{2}\right)} = \frac{\Phi(1) - \Phi(-1)}{\Phi(2.5) - \Phi(-2.5)} \\
&= \frac{\Phi(1) - [1 - \Phi(1)]}{\Phi(2.5) - [1 - \Phi(2.5)]} \\
&\overset{tabelas/calc.}{=} \frac{0.8413 - (1 - 0.8413)}{0.9938 - (1 - 0.9938)} \\
&= \frac{0.6826}{0.9876} \\
&\simeq 0.6912.
\end{aligned}$$

Pergunta 4	2 valores

Sejam X e Y variáveis aleatórias que representam, respetivamente, a largura (em dm) e o comprimento (em dm) de uma peça retangular. Admita que: para $0 < y < 1$,

$$f_{X|Y=y}(x) = \begin{cases} \frac{1}{y}, & 0 < x < y \\ 0, & \text{caso contrário;} \end{cases} \qquad f_Y(y) = \begin{cases} 2y, & 0 < y < 1 \\ 0, & \text{caso contrário.} \end{cases}$$

Obtenha $E[2(X+Y)]$, o valor esperado do perímetro desta peça.

- **Par aleatório**

  (X, Y)

  X = largura da peça retangular (em dm)

  Y = comprimento da peça retangular (dm)

- **Valor esperado de Y**

$$E(Y) = \int_{-\infty}^{+\infty} y \times f_Y(y)\, dy = \int_0^1 y \times 2y\, dy = \left.\frac{2y^3}{3}\right|_0^1 = \frac{2}{3}$$

- **Valor esperado de X**

  Como $f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$, temos

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x \times f_X(x)\, dx = \int_{-\infty}^{+\infty} x \times \left[\int_{-\infty}^{+\infty} f_{X,Y}(x,y)\, dy\right] dx \\ &= \int_{-\infty}^{+\infty} x \times \left[\int_{-\infty}^{+\infty} f_{X|Y=y}(x,y) \times f_Y(y)\, dy\right] dx = \int_0^1 x \times \left[\int_x^1 \frac{1}{y} \times 2y\, dy\right] dx \\ &= \int_0^1 x \times \left[\int_x^1 2\, dy\right] dx = \int_0^1 x \times 2y\Big|_x^1\, dx = \int_0^1 x \times 2(1-x)\, dx = \left.\left(x^2 - \frac{2x^3}{3}\right)\right|_0^1 = \left(1 - \frac{2}{3}\right) = \frac{1}{3}. \end{aligned}$$

  [Alternativamente, $X \mid Y = y \sim \text{uniforme}(0, y)$, $E(X \mid Y = y) = y/2$, $E(X \mid Y) = Y/2$, $E(X) = E[E(X \mid Y)] = E(Y/2) = 1/2 \times 2/3 = 1/3$.]

- **Valor esperado pedido**

$$E[2(X+Y)] = 2\,[E(X) + E(Y)] = 2 \times \left(\frac{1}{3} + \frac{2}{3}\right) = 2.$$

Pergunta 5	2 valores

Considere que os tempos entre chegadas consecutivas de aeronaves a um aeroporto são variáveis aleatórias independentes e identicamente distribuídas a X com distribuição exponencial e valor esperado igual a μ.

Indique o valor aproximado da probabilidade de a média de 100 de tais tempos pertencer ao intervalo $[(1-0.075)\,\mu,\ (1+0.075)\,\mu]$? **A**: 0.4533 **B**: 0.5468 **C**: 0.0801 **D**: 0.9199

- **V.a.; valor esperado e variância comuns**

  X_i = i-ésimo tempo entre chegadas consecutivas de aeronaves, $i = 1, \ldots, n$

  $n = 100$

  $X_i \overset{i.i.d.}{\sim} X \sim \text{exponencial}(1/\mu)$

  $E(X_i) = E(X) = \mu$

  $V(X_i) = V(X) = \sigma^2 \overset{form.}{=} \mu^2$

- **V.a. de interesse**

  $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ = média de n tempos entre chegadas consecutivas de aeronaves a um aeroporto

  $E(\bar{X}) = \cdots = \mu$

  $V(\bar{X}) = \cdots = \frac{\sigma^2}{n} = \frac{\mu^2}{100}$

- **Distribuição aproximada de** $\bar{X}$

  De acordo com o TLC,

$$Z = \frac{\bar{X} - E(\bar{X})}{\sqrt{V(\bar{X})}} = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{\bar{X} - \mu}{\sqrt{\frac{\mu^2}{n}}} \stackrel{a}{\sim} \text{normal}(0,1).$$

- **Prob. pedida** (valor aproximado)

$$\begin{aligned}
P[(1-0.075)\,\mu \le \bar{X} \le (1+0.075)\,\mu] &= P\left[Z \le \frac{(1+0.075)\,\mu - \mu}{\sqrt{\frac{\mu^2}{n}}}\right] - P\left[Z \le \frac{(1-0.075)\,\mu - \mu}{\sqrt{\frac{\mu^2}{n}}}\right] \\
&\stackrel{TLC}{\simeq} \Phi\left[\frac{(1+0.075)\,\mu - \mu}{\sqrt{\frac{\mu^2}{n}}}\right] - \Phi\left[\frac{(1-0.075)\,\mu - \mu}{\sqrt{\frac{\mu^2}{n}}}\right] \\
&= \Phi(\sqrt{n} \times 0.075) - \Phi(-\sqrt{n} \times 0.075) \\
&= \Phi(\sqrt{n} \times 0.075) - [1 - \Phi(\sqrt{n} \times 0.075)] \\
&= 2 \times \Phi(\sqrt{n} \times 0.075) - 1 \\
&= 2 \times \Phi(0.75) - 1 \\
&\stackrel{tab./calc.}{\simeq} 2 \times 0.7734 - 1 \\
&= 0.5468.
\end{aligned}$$

Pergunta 6 2 valores

Seja X o volume de combustível armazenado num reservatório de capacidade unitária. Admita que X é uma variável aleatória com função de densidade de probabilidade

$$f_X(x) = \begin{cases} 4\,\beta\,x^3\,(1-x^4)^{\beta-1}, & 0 < x < 1 \\ 0, & \text{caso contrário,} \end{cases}$$

onde β é um parâmetro positivo desconhecido. Deduza os estimadores de máxima verosimilhança de β e da mediana de X, $(1-2^{-1/\beta})^{1/4}$, com base numa amostra aleatória de dimensão n proveniente de X.

- **V.a. de interesse; f.d.p.**

  X = volume de combustível armazenado num reservatório de capacidade unitária

$$f_X(x) = \begin{cases} 4\,\beta\,x^3\,(1-x^4)^{\beta-1}, & 0 < x < 1 \\ 0, & \text{caso contrário} \end{cases}$$

- **Parâmetro desconhecido**

  $\beta, \quad \beta > 0$

- **Amostra; amostra aleatória**

  $\underline{x} = (x_1, \ldots, x_n)$, amostra de dimensão n proveniente de X.

  $\underline{X} = (X_1, \ldots, X_n)$ amostra aleatória de dimensão n proveniente de X.

- **Obtenção do estimador de MV de β**

  Passo 1 — Função de verosimilhança

$$\begin{aligned} L(\beta \mid \underline{x}) &= f_{\underline{X}}(\underline{x}) \stackrel{X_i\, indep}{=} \prod_{i=1}^{n} f_{X_i}(x_i) \stackrel{X_i \sim X}{=} \prod_{i=1}^{n} f_X(x_i) = \prod_{i=1}^{n} \left[4\,\beta\, x_i^3\,(1-x_i^4)^{\beta-1}\right] \\ &= (4\,\beta)^n \left(\prod_{i=1}^{n} x_i\right)^3 \left[\prod_{i=1}^{n}(1-x_i^4)\right]^{\beta-1} \end{aligned}$$

  Passo 2 — Função de log-verosimilhança

$$\ln L(\beta \mid \underline{x}) = n\ln 4 + n\ln(\beta) + 3\sum_{i=1}^{n}\ln(x_i) + (\beta-1)\sum_{i=1}^{n}\ln(1-x_i^4)$$

  Passo 3 — Maximização

  A estimativa de MV de β é representada por $\hat{\beta}$ e

$$\hat{\beta} \;:\; \begin{cases} \left.\frac{d\ln L(\beta|\underline{x})}{d\beta}\right|_{\beta=\hat{\beta}} = 0 & \text{(ponto de estacionaridade)} \\ \left.\frac{d^2\ln L(\beta|\underline{x})}{d\beta^2}\right|_{\beta=\hat{\beta}} < 0 & \text{(ponto de máximo)} \end{cases}$$

$$\begin{cases} \frac{n}{\hat{\beta}} + \sum_{i=1}^{n}\ln(1-x_i^4) = 0 \\ -\frac{n}{\hat{\beta}^2} < 0 \quad \text{(prop. verdadeira desde que } \hat{\beta} \neq 0) \end{cases}$$

$$\begin{cases} \hat{\beta} = -\frac{n}{\sum_{i=1}^{n}\ln(1-x_i^4)} \\ - \end{cases}$$

  Passo 4 — Estimador de MV de β

$$EMV(\beta) = -\frac{n}{\sum_{i=1}^{n}\ln(1-X_i^4)}$$

- **Outro parâmetro desconhecido**

  $h(\beta) = (1-2^{-1/\beta})^{1/4}$ [mediana de X]

- **Estimador de MV de $h(\beta)$**

  Ao invocar a propriedade de invariância dos EMV, concluímos que o estimador de MV de $h(\beta)$ é

$$EMV[h(\beta)] = h[EMV(\beta)] = \left(1-2^{-\frac{1}{EMV(\beta)}}\right)^{1/4} = \left[1-2^{\frac{\sum_{i=1}^{n}\ln(1-X_i^4)}{n}}\right]^{1/4}.$$

Pergunta 7	2 valores

Seja p a probabilidade de uma peça, escolhida ao acaso da produção diária de uma unidade fabril, ser defeituosa. Para estimar p, o gabinete de controlo de qualidade selecionou aleatoriamente uma amostra de 100 peças, sendo que 3 das peças foram classificadas como defeituosas.

Determine um intervalo aproximado de confiança a 90% para p.

- **V.a. de interesse**

$$X = \begin{cases} 1, & \text{peça defeituosa} \\ 0, & \text{caso contrário} \end{cases}$$

- **Situação**

  $X \sim \text{Bernoulli}(p)$

  p DESCONHECIDO

- **Obtenção de IC aproximado para p**

  Passo 1 - Seleção da v.a. fulcral para p

  $$Z = \frac{\bar{X} - p}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \overset{a}{\sim} \text{normal}(0,1)$$

  Passo 2 — Obtenção dos quantis de probabilidade

  Como $(1-\alpha)\times 100\% = 90\% \Leftrightarrow \alpha = 0.10$, lidaremos com os quantis seguintes:

  $$\begin{cases} a_\alpha = \Phi^{-1}(\alpha/2) = -\Phi^{-1}(1-\alpha/2) = -\Phi^{-1}(0.95) \overset{tabela/calc.}{=} 1.6449 \\ b_\alpha = \Phi^{-1}(1-\alpha/2) = \Phi^{-1}(0.95) = 1.6449. \end{cases}$$

  Passo 3 — Inversão da desigualdade $a_\alpha \le Z \le b_\alpha$

  $$P(a_\alpha \le Z \le b_\alpha) \simeq 1-\alpha$$

  $$P\left[a_\alpha \le (\bar{X}-p)/\sqrt{\bar{X}(1-\bar{X})/n} \le b_\alpha\right] \simeq 1-\alpha$$

  $$P\left[\bar{X} - b_\alpha \times \sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \le p \le \bar{X} - a_\alpha \times \sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right] \simeq 1-\alpha$$

  Passo 4 — Concretização

  A expressão geral do intervalo aproximado de confiança a $(1-\alpha)\times 100\%$ para p é

  $$IC_{(1-\alpha)\times 100\%}(p) \simeq \left[\bar{x} \pm \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{\bar{x}(1-\bar{x})}{n}}\right].$$

  Ao termos em conta que $n = 100$, $\alpha = 0.10$, $\Phi^{-1}(1-\alpha/2) = 1.6449$ e $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{3}{100} = 0.03$,

  $$\begin{aligned} IC_{90\%}(p) &\simeq \left[0.03 - 1.6449 \times \sqrt{\frac{0.03\times(1-0.03)}{100}}, 0.03 + 1.6449 \times \sqrt{\frac{0.03\times(1-0.03)}{100}}\right] \\ &\simeq [0.0019\ ,\ 0.0581]. \end{aligned}$$

Pergunta 8	2 valores

Admita que o TTPa (tempo de tromboplastina parcialmente ativada) em segundos associado a um adulto saudável é uma variável aleatória com distribuição normal com valor esperado μ e desvio padrão σ desconhecidos. Uma equipa médica mediu o TTPa em 30 adultos saudáveis, tendo obtido uma média e um desvio padrão amostrais iguais a $\bar{x} = 33.8$ e $s = 3.30$.

Teste a hipótese $H_0 : \mu = 35$ contra $H_1 : \mu \neq 35$, ao nível de significância de 5%.

- **V.a. de interesse**

  X = tempo de tromboplastina parcial ativada (TTPa), em segundos

- **Situação**

  $X \sim \text{normal}(\mu, \sigma^2)$

  μ DESCONHECIDO

  σ desconhecido

- **Hipóteses**

  $H_0 : \mu = \mu_0 = 35$ vs. $H_1 : \mu \neq 35$

- **N.s.**

  $\alpha_0 = 5\%$

- **Estatística de teste**

  $$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim_{H_0} t_{n-1}$$

- **Região de rejeição de H_0 (para valores da estatística de teste)**

  O teste é bilateral ($H_1 : \mu \neq \mu_0$), logo a região de rejeição de H_0 é do tipo $W = (-\infty, -c) \cup (c, +\infty)$, onde

  $$c \quad : \quad P(\text{Rejeitar } H_0 \mid H_0) = \alpha_0 \Leftrightarrow c = F_{t_{(n-1)}}(1 - \alpha_0/2) = F^{-1}_{t_{(29)}}(0.975) \overset{tabelas/calc.}{=} 2.045.$$

- **Decisão**

  Atendendo a que o valor observado da estatística de teste é igual a

  $$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{33.8 - 35}{3.30/\sqrt{30}} = -1.992$$

  e $t \notin W = (-\infty, -2.045) \cup (2.045, +\infty)$, é suposto não rejeitarmos H_0 ao n.s. $\alpha_0 = 5\%$ [ou a qualquer outro n.s. inferior a 5%].

Pergunta 9	2 valores

Um engenheiro de materiais conjectura a hipótese H_0 de que o tamanho de partículas em determinado pó metálico possui função de distribuição (f.d.) dada por $F_0(x) = \Phi[\ln(x)]$, para $x > 0$, onde $\Phi(z)$ representa a f.d. da normal padrão. O engenheiro organizou a seguinte tabela de frequências referente a 200 registos semanais obtidos casualmente para testar a adequação de tal distribuição.

Tamanho da partícula	$(0, 0.4310]$	$(0.4310, 0.7762]$	$(0.7762, 1.2883]$	$(1.2883, 2.3201]$	> 2.3201
Freq. abs. observada	37	38	55	40	30
Freq. abs. esperada sob H_0	40.0	40.0	40.0	40.0	40.0

Com base no teste de ajustamento do qui-quadrado, indique que decisão deverá tomar o engenheiro?

A: Rejeitar H_0 a 1%, 5% e 10%. **B**: Rejeitar H_0 a 5% e 10% e não rejeitar H_0 a 1%.
C: Rejeitar H_0 a 10% e não rejeitar H_0 a 1% e 5%. **D**: Não rejeitar H_0 a 1%, 5% e 10%.

- **V.a. de interesse**

  X = tamanho de partículas em determinado pó metálico

- **Hipóteses**

  $H_0 : F_X(x) = F_0(x) = \Phi[\ln(x)], \quad \forall x > 0 \quad$ vs. $\quad H_1 : F_X(x) \neq F_0(x), \quad$ para algum $x > 0$

- **Estatística de teste**

  $$T = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \overset{a}{\sim}_{H_0} \chi^2_{(k-1)}$$

  onde: k = número de classes = 5; O_i = frequência absoluta observável da classe i; E_i = frequência absoluta esperada, sob H_0, da classe i.

- **Região de rejeição de H_0** (para valores de T)

  Ao efectuarmos um teste de ajustamento do qui-quadrado, a região de rejeição de H_0 é um intervalo à direita $W = (c, +\infty)$.

- **Decisão (com base no valor-p aproximado)**

i	Classe	Freq. abs. obs. o_i	Freq. abs. esper. sob H_0 $E_i = n \times p_i^0$	Parcelas valor obs. estat. teste $\frac{(o_i-E_i)^2}{E_i}$
1	$(0, 0.4310]$	37	40.0	$\frac{(37-40.0)^2}{40.0} = 0.225$
2	$(0.4310, 0.7762]$	38	40.0	0.1
3	$(0.7762, 1.2883]$	55	40.0	5.625
4	$(1.2883, 2.3201]$	40	40.0	0
5	$(2.3201, +\infty)$	30	40.0	2.5
		$\sum_{i=1}^k o_i = n = 200$	$\sum_{i=1}^k E_i = n = 200$	$t = \sum_{i=1}^k \frac{(o_i-E_i)^2}{E_i} = 8.45$

Atendendo a que a região de rejeição de H_0 é um intervalo à direita num teste de ajustamento do qui-quadrado, temos

$$valor - p \quad = \quad P(T > t = | H_0) \simeq 1 - F_{\chi^2_{(5-1)}}(8.45) \overset{calc.}{\simeq} 0.076.$$

Assim sendo, devemos decidir pela:

- rejeição de H_0 a qualquer n.s. $\alpha_0 > 7.6\%$, designadamente ao n.u.s. de 10%;
- não rejeição de H_0 a qualquer n.s. $\alpha_0 \leq 7.6\%$, nomeadamente ao n.u.s. de 1% e 5%.

Por consequência, a resposta, das quatro apresentadas, que se coaduna o referido acima é:

- **C**: Rejeitar H_0 a 10% e não rejeitar H_0 a 1% e 5%.

[Alternativamente, a consulta das tabelas dos quantis de probabilidade da distribuição do qui-quadrado com $k-1=4$ graus de liberdade permitem concluir que

$$\begin{aligned} F^{-1}_{\chi^2_{(5-1)}}(0.9) = 7.779 \quad &< \quad 8.45 < 8.496 = F^{-1}_{\chi^2_{(5-1)}}(0.925) \\ 0.075 = 1 - 0.925 \quad &< \quad 1 - F_{\chi^2_{(5-1)}}(8.45) < 1 - 0.90 = 0.10. \end{aligned}$$

Logo, devemos:

- rejeitar H_0 a qualquer n.s. $\alpha_0 \geq 10\%$, designadamente ao n.u.s. de 10%;
- não rejeitar de H_0 a qualquer n.s. $\alpha_0 \leq 7.5\%$, nomeadamente aos n.u.s. de 1% e 5%.

Deste modo, a resposta correcta é **C**: Rejeitar H_0 a 10% e não rejeitar H_0 a 1% e 5%.]

Pergunta 10 2 valores

Foi realizada uma experiência para estudar o efeito da dose de determinado barbitúrico (x, em miligrama por kg de peso) no tempo de sono (Y, em horas). Obtiveram-se os seguintes resultados para uma amostra casual de 9 pacientes:

$$\sum_{i=1}^{9} x_i = 84, \quad \sum_{i=1}^{9} x_i^2 = 1\,002, \quad \sum_{i=1}^{9} y_i = 72, \quad \sum_{i=1}^{9} y_i^2 = 642, \quad \sum_{i=1}^{9} x_i y_i = 780,$$

onde $[\min_{i=1,\ldots,9} x_i, \max_{i=1,\ldots,9} x_i] = [3, 15]$. Admita que as variáveis x e Y estão relacionadas de acordo com o modelo de regressão linear simples, $Y = \beta_0 + \beta_1 x + \epsilon$.

Obtenha a reta de mínimos quadrados com base nos dados fornecidos. Calcule o coeficiente de determinação e interprete o seu valor.

- **[Modelo de RLS e hipóteses de trabalho**

  Y = horas de sono (v.a. resposta)

  x = dosagem do barbitúrico (variável explicativa)

$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \qquad E(\epsilon_i) = 0, \quad V(\epsilon_i) = \sigma^2, \quad i = 1,\ldots,n, \qquad cov(\epsilon_i, \epsilon_j) = 0,\ i \neq j,\ i,j = 1,\ldots,n]$

- **Estimativas de MQ dos parâmetros desconhecidos β_0 e β_1; recta de MQ**

  Temos

  - $n = 9$
  - $\sum_{i=1}^n x_i = 84$

    $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i = \frac{84}{9} = \frac{28}{3}$

    $\sum_{i=1}^n x_i^2 = 1\,002$

    $\sum_{i=1}^n x_i^2 - n\,\bar{x}^2 = 1\,002 - \frac{84^2}{9} = 218$
  - $\sum_{i=1}^n y_i = 72$

    $\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i = \frac{72}{9} = 8$

    $\sum_{i=1}^n y_i^2 = 642$

    $\sum_{i=1}^n y_i^2 - n\,\bar{y}^2 = 642 - \frac{72^2}{9} = 66$
  - $\sum_{i=1}^n x_i y_i = 780$

    $\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y} = 780 - 84 \times 8 = 108.$

  Logo, as estimativas de MQ dos parâmetros desconhecidos β_1 e β_0 são dadas por

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y}}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2} = \frac{108}{218} = \frac{54}{109} \simeq 0.495413$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \times \bar{x} = 8 - \frac{54}{109} \times \frac{28}{3} = \frac{368}{109} \simeq 3.376147.$$

  e a recta de MQ por

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1\,x \simeq 3.376147 + 0.495413\,x, \quad x \in [\min_{i=1,\ldots,9} x_i, \max_{i=1,\ldots,9} x_i] = [3, 15].$$

- **Coeficiente de determinação**

$$r^2 = \frac{\left(\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y}\right)^2}{\left(\sum_{i=1}^n x_i^2 - n\,\bar{x}^2\right) \times \left(\sum_{i=1}^n y_i^2 - n\,\bar{y}^2\right)} = \frac{108^2}{218 \times 66} = \frac{11\,664}{14\,388} \simeq 0.8106756.$$

- **Interpretação coeficiente de determinação**

  Cerca de 81% da variação total da variável resposta Y é explicada pela variável x, através do modelo de regressão linear simples ajustado, donde possamos afirmar que a recta estimada parece ajustar-se bem ao conjunto de dados.