

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

Um equipamento de deteção de incêndios tem dois sensores. Suponha que os dois sensores foram instalados na mesma altura, têm igual prazo de validade e operam de modo independente quer sabendo que estão ambos dentro do prazo de validade, quer sabendo que pelo menos um dos sensores está fora do prazo de validade. A probabilidade de cada sensor avariar sabendo que ambos estão dentro do prazo de validade é 0.01. A probabilidade de cada sensor avariar sabendo que pelo menos um dos sensores está fora do prazo de validade 0.5. De acordo com as inspeções efetuadas, a probabilidade de ambos os sensores estarem dentro do prazo de validade é de 0.7.

Numa inspeção em que os sensores são testados simultaneamente, a probabilidade de ambos os sensores estarem avariados é

A: 0.28505 **B:** 0.07507 **C:** 0.71495 **D:** 0.82493

• **Quadro de acontecimentos e probabilidades**

Acontecimento	Probabilidade
V = ambos os sensores estão dentro do prazo de validade	$P(V) = 0.7$
$\bar{V}$ = pelo menos um dos sensores está fora do prazo de validade	$P(\bar{V}) = 1 - 0.7 = 0.3$
A_i = sensor i avariar	$P(A_i) = ?$
	$P(A_i V) = 0.01, \quad i = 1, 2$
	$P(A_i \bar{V}) = 0.5, \quad i = 1, 2$

• **Prob. pedida,**

Invocando a lei da probabilidade total (LPT) e a independência condicional (IC), temos

$$\begin{aligned}
 P(A_1 \cap A_2) &\stackrel{LPT}{=} P(A_1 \cap A_2 | V) \times P(V) + P(A_1 \cap A_2 | \bar{V}) \times P(\bar{V}) \\
 &\stackrel{IC}{=} P(A_1 | V) \times P(A_2 | V) \times P(V) + P(A_1 | \bar{V}) \times P(A_2 | \bar{V}) \times P(\bar{V}) \\
 &= 0.01^2 \times 0.7 + 0.5^2 \times (1 - 0.7) = 0.07507
 \end{aligned}$$

Pergunta 2

2 valores

Considere que os aparelhos de radar produzidos por uma dada fábrica detetam corretamente o alvo com probabilidade p e operam de modo independente. Seja X a variável aleatória que representa o número de aparelhos de radar que detetam corretamente o alvo, em n utilizados na identificação de um alvo.

Qual é a moda de X , atendendo a que $E(X) = 9.6$ e $V(X) = 1.92$?

• **V.a.**

X = número de aparelhos de radar que detetam corretamente o alvo em n utilizados...

- **Distribuição**

$X \sim \text{binomial}(n, p)$, onde

$$(n, p) : \begin{cases} E(X) = np = 9.6 \\ V(X) = np(1-p) = 1.92 \end{cases} \Leftrightarrow \begin{cases} - \\ 9.6 \times (1-p) = 1.92 \Leftrightarrow p = \frac{9.6-1.92}{9.6} = 0.8 \end{cases}$$

$$\begin{cases} n = \frac{9.6}{p} = \frac{9.6}{0.8} = 12 \\ p = 0.8 \end{cases}$$

- **F.p. de X**

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \dots, n$$

- **Moda de X**

É sabido que $mo = mo(X) = \operatorname{argmax}_{x=0,1,\dots,n} P(X = x)$. Logo, $mo \in \{0, 1, \dots, n\}$ e é tal que

$$\begin{cases} P(X = mo) \geq P(X = mo - 1) \\ P(X = mo) \geq P(X = mo + 1) \end{cases} \Leftrightarrow \begin{cases} \frac{P(X=mo)}{P(X=mo-1)} \geq 1 \\ \frac{P(X=mo+1)}{P(X=mo)} \leq 1 \end{cases} \Leftrightarrow \begin{cases} \frac{\binom{n}{mo} p^{mo} (1-p)^{n-mo}}{\binom{n}{mo-1} p^{mo-1} (1-p)^{n-(mo-1)}} \geq 1 \\ \frac{\binom{n}{mo+1} p^{mo+1} (1-p)^{n-(mo+1)}}{\binom{n}{mo} p^{mo} (1-p)^{n-mo}} \leq 1 \end{cases}$$

$$\begin{cases} \frac{(n-mo+1)p}{mo(1-p)} \geq 1 \\ \frac{(n-mo)p}{(mo+1)(1-p)} \leq 1 \end{cases} \Leftrightarrow \begin{cases} mo \leq (n+1)p = (12+1) \times 0.8 = 10.4 \\ mo \geq np - (1-p) = 12 \times 0.8 - (1-0.8) = 9.4, \end{cases}$$

i.e., o número modal pedido é $mo = mo(X) = 10$.

Pergunta 3

2 valores

Suponha que o tempo de vida útil (em meses) de certo tipo de bateria de automóvel se distribui segundo uma lei normal com valor esperado e desvio padrão iguais a 40 e 4, respetivamente.

Determine a probabilidade de o tempo de vida útil de uma bateria selecionada ao acaso durar mais de 45 meses, sabendo que o seu tempo de vida útil excedeu 38 meses.

- **V.a.**

X = tempo de vida útil (em meses) de certo tipo de bateria de automóvel

- **Distribuição**

$$X \sim \text{normal}(\mu = 40, \sigma^2 = 4^2)$$

- **Probabilidade pedida**

$$\begin{aligned} P(X > 45 \mid X > 38) &= \frac{P(X > 45, X > 38)}{P(X > 38)} = \frac{P(X > 45)}{P(X > 38)} \\ &= \frac{1 - P\left(Z = \frac{X-\mu}{\sigma} \leq \frac{45-\mu}{\sigma}\right)}{1 - P\left(Z = \frac{X-\mu}{\sigma} \leq \frac{38-\mu}{\sigma}\right)} = \frac{1 - \Phi\left(\frac{45-\mu}{\sigma}\right)}{1 - \Phi\left(\frac{38-\mu}{\sigma}\right)} \\ &= \frac{1 - \Phi(1.25)}{1 - \Phi(-0.5)} = \frac{1 - \Phi(1.25)}{\Phi(0.5)} \stackrel{\text{tabelas/calcul.}}{=} \frac{1 - 0.8944}{0.6915} \approx 0.1527. \end{aligned}$$

Pergunta 4

2 valores

No desenvolvimento de um sistema de gestão de tráfego de rede, sabe-se que o tempo de resposta do sistema (X , normalizado em relação ao tempo máximo de resposta aceitável) e a quantidade de dados processados (Y , normalizada em relação à capacidade máxima de processamento de dados do sistema) são variáveis aleatórias contínuas que afetam o desempenho geral do sistema. As funções de densidade de

probabilidade conjunta de (X, Y) e marginais de X e de Y são dadas por

$$f_{X,Y}(x, y) = \begin{cases} \frac{4}{5}(1 + xy), & 0 \leq x \leq 1 \text{ e } 0 \leq y \leq 1 \\ 0, & \text{caso contrário} \end{cases} \quad \text{e} \quad f_X(x) \equiv f_Y(x) = \begin{cases} \frac{2}{5}(2 + x), & 0 \leq x \leq 1 \\ 0, & \text{caso contrário.} \end{cases}$$

Sabendo que $E(XY) = \frac{13}{45}$, calcule a variância de $(2X + 3Y)$.

- **Par aleatório** (X, Y)

X = tempo de resposta do sistema (normalizado)

Y = quantidade de dados processados (normalizado)

- **V.a. de interesse**

$2X + 3Y$ = medida do desempenho do sistema

- **Variância pedida**

Uma vez que $V(2X + 3Y) = 2^2 V(X) + 3^2 V(Y) + 2 \times 2 \times 3 \text{cov}(X, Y)$ são necessários alguns cálculos auxiliares que envolverão a f.d.p. conjunta de (X, Y) e as f.d.p. marginais de X e Y .

- **Valores esperados e variâncias de X e de Y**

Como X e Y são v.a. identicamente distribuídas ($X \sim Y$), temos

$$\begin{aligned} E(X) \equiv E(Y) &= \int_{-\infty}^{+\infty} x \times f_X(x) dx \\ &= \int_0^1 x \times \frac{2}{5}(2 + x) dx = \frac{2}{5} \int_0^1 (2x + x^2) dx = \frac{2}{5} \left(x^2 + \frac{x^3}{3} \right) \Big|_0^1 = \frac{2}{5} \left(1 + \frac{1}{3} \right) = \frac{8}{15} = 0.5(3) \end{aligned}$$

$$\begin{aligned} E(X^2) \equiv E(Y^2) &= \int_{-\infty}^{+\infty} x^2 \times f_X(x) dx \\ &= \int_0^1 x^2 \times \frac{2}{5}(2 + x) dx = \frac{2}{5} \int_0^1 (2x^2 + x^3) dx = \frac{2}{5} \left(\frac{2x^3}{3} + \frac{x^4}{4} \right) \Big|_0^1 = \frac{2}{5} \left(\frac{2}{3} + \frac{1}{4} \right) = \frac{11}{30} \\ &= 0.3(6) \end{aligned}$$

$$\begin{aligned} V(X) \equiv V(Y) &= E(X^2) - E^2(X) \\ &= \frac{11}{30} - \left(\frac{8}{15} \right)^2 = \frac{11}{30} - \frac{64}{225} = \frac{165}{450} - \frac{128}{450} = \frac{37}{450} = 0.08(2). \end{aligned}$$

- **[Valor esperado de XY]**

$$\begin{aligned} E(XY) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy \times f_{X,Y}(x, y) dx dy \\ &= \int_0^1 \int_0^1 xy \times \frac{4}{5}(1 + xy) dx dy = \frac{4}{5} \int_0^1 \int_0^1 (xy + x^2 y^2) dx dy = \frac{4}{5} \int_0^1 \left(\frac{x^2 y}{2} + \frac{x^3 y^2}{3} \right) \Big|_0^1 dy \\ &= \frac{4}{5} \int_0^1 \left(\frac{y}{2} + \frac{y^2}{3} \right) dy = \frac{4}{5} \times \left(\frac{y^2}{4} + \frac{y^3}{9} \right) \Big|_0^1 = \frac{4}{5} \times \left(\frac{1}{4} + \frac{1}{9} \right) = \frac{1}{5} + \frac{4}{45} = \frac{13}{45} = 0.2(8) \end{aligned}$$

- **Covariância entre X e Y**

$$\begin{aligned} \text{cov}(X, Y) &= E(XY) - E(X) \times E(Y) \\ &= \frac{13}{45} - \frac{8}{15} \times \frac{8}{15} = \frac{13}{45} - \frac{64}{225} \\ &= \frac{1}{225} \end{aligned}$$

- **Variância pedida (cont.)**

$$\begin{aligned} V(2X + 3Y) &= 4 \times V(X) + 9 \times V(Y) + 12 \times \text{cov}(X, Y) \\ &= 4 \times \frac{37}{450} + 9 \times \frac{37}{450} + 12 \times \frac{1}{225} = \frac{101}{90} \\ &= 1.1(2). \end{aligned}$$

O tempo de processamento (X , medido em segundos) de uma certa tarefa num servidor remoto é uma variável aleatória com distribuição uniforme no intervalo $[0, 1]$. Admita que se realizam sequencialmente n tarefas com tempos de processamento ($X_i, i = 1, \dots, n$) independentes e identicamente distribuídas a U .

Obtenha o valor exacto de $P(X_1 + X_2 > 1.5)$, tirando partido do facto de

$$F_{X_1+X_2}(x) = \begin{cases} 0 & x < 0 \\ \frac{x^2}{2} & 0 \leq x < 1 \\ 1 - \frac{(2-x)^2}{2} & 1 \leq x \leq 2 \\ 1 & x > 2. \end{cases}$$

Calcule um valor aproximado para $P(X_1 + X_2 > 1.5)$, recorrendo ao teorema do limite central, apesar de o número de parcelas ser muito inferior a 30. Por último, indique a diferença (arredondada às centésimas) entre os valores exacto e aproximado.

A: 0.11 B: 0.12 C: 0.01 D: 0.76

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

X_i = tempo de processamento da tarefa i , $i = 1, \dots, n$

$n = 2$

$X_i \stackrel{i.i.d.}{\sim} X \sim \text{uniforme}(0, 1)$

$$E(X_i) = E(X) = \mu \stackrel{\text{form.}}{=} \frac{0+1}{2} = \frac{1}{2}$$

$$V(X_i) = V(X) = \sigma^2 \stackrel{\text{form.}}{=} \frac{(1-0)^2}{12} = \frac{1}{12}$$

• **V.a. de interesse**

$S_n = \sum_{i=1}^n X_i$ = tempo total de processamento das n tarefas

• **Prob. pedida** (valor exacto)

$$P(X_1 + X_2 > 1.5) = 1 - F_{X_1+X_2}(1.5) = 1 - \left[1 - \frac{(2-x)^2}{2} \right] = \frac{(2-1.5)^2}{2} = 0.125.$$

• **Valor esperado e variância de S_n ; distribuição aproximada de S_n**

$$E(S_n) = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) \stackrel{X_i \sim X}{=} n E(X) = n\mu$$

$$V(S_n) = V\left(\sum_{i=1}^n X_i\right) \stackrel{X_i \text{ indep.}}{=} \sum_{i=1}^n V(X_i) \stackrel{X_i \sim X}{=} n V(X) = n\sigma^2$$

De acordo com o teorema do limite central (TLC),

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

• **Prob. pedida** (valor aproximado)

Apesar de o número de parcelas ser inferior a 30, podemos adiantar que

$$\begin{aligned} P(S_n > 1.5) &= 1 - P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq \frac{1.5 - n\mu}{\sqrt{n\sigma^2}}\right) \stackrel{TLC}{\approx} 1 - \Phi\left(\frac{1.5 - n\mu}{\sqrt{n\sigma^2}}\right) \\ &= 1 - \Phi\left(\frac{1.5 - 2 \times 0.5}{\sqrt{2 \times \frac{1}{12}}}\right) = 1 - \Phi(1.22) \stackrel{\text{tabelas}}{\approx} 1 - 0.8888 = 0.1112 \quad [\stackrel{\text{calc.}}{\approx} 0.1103]. \end{aligned}$$

• **Valor pedido** (arredondado às centésimas)

$$0.125 - 0.1112 \approx 0.01.$$

O número de defeitos por chip (produzidos numa linha de uma fábrica), X , segue uma distribuição de Poisson(λ), com λ podendo assumir os valores 1 (processo de fabrico sob controlo) e 3 (processo de fabrico fora de controlo).

Sabendo que uma amostra casual de 30 chips, produzidos sem mudança do estado de controlo do processo de fabrico, conduziu a um total de 48 defeitos, determine a estimativa de máxima verosimilhança de $E(X^2)$.

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

X = número de defeitos por chip

$$f_X(x) = P(X = x) = \begin{cases} e^{-\lambda} \frac{\lambda^x}{x!}, & x = 0, 1, 2, \dots \\ 0, & \text{outros valores de } x \end{cases}$$

- **Parâmetro desconhecido e espaço paramétrico**

λ

$\Theta = \{1, 3\}$

- **Amostra; amostra aleatória**

$\underline{x} = (x_1, \dots, x_n)$, amostra de dimensão n proveniente de X , onde $n = 30$ e $\sum_{i=1}^n x_i = 48$.

$(X_1, X_2, \dots, X_n)$, amostra aleatória de dimensão n proveniente de X .

- **Obtenção da estimativa de MV de λ**

Será representada por $\hat{\lambda}$ e é tal que: $L(\hat{\lambda} | \underline{x}) = \max_{\lambda \in \Theta} L(\lambda | \underline{x})$ ou $\ln L(\hat{\lambda} | \underline{x}) = \max_{\lambda \in \Theta} \ln L(\lambda | \underline{x})$, onde $L(\lambda | \underline{x})$ e $\ln L(\lambda | \underline{x})$ representam as f. de verosimilhança e de log-verosimilhança.

Funções de verosimilhança e log-verosimilhança

$$L(\lambda | \underline{x}) = f_{\underline{X}}(\underline{x}) \stackrel{X_i \text{ indep}}{=} \prod_{i=1}^n f_{X_i}(x_i) \stackrel{X_i \sim X}{=} \prod_{i=1}^n f_X(x_i) = \prod_{i=1}^n \left[e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} \right] = e^{-n\lambda} \frac{\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}$$

$$\ln L(\lambda | \underline{x}) = -n\lambda + \ln(\lambda) \sum_{i=1}^n x_i - \ln \left(\prod_{i=1}^n x_i! \right)$$

Maximização e concretização

[Como Θ é um conjunto discreto, a estimativa de MV de λ obtém-se calculando os vários valores de $\ln L(\lambda | \underline{x})$, para $\lambda \in \Theta$, e identificando o ponto de máximo.]

λ	$\ln L(\lambda \underline{x})$
1	$L(1 \underline{x}) = -30 \times 1 - \ln(1) \times 48 - \ln \left(\prod_{i=1}^n x_i! \right) = -30 - \ln \left(\prod_{i=1}^n x_i! \right)$
3	$L(3 \underline{x}) = -30 \times 3 - \ln(3) \times 48 - \ln \left(\prod_{i=1}^n x_i! \right) \approx -142.733 - \ln \left(\prod_{i=1}^n x_i! \right)$

A inspeção da tabela anterior leva a concluir que a função de log-verosimilhança atinge o máximo quando $\lambda = 1$, pelo que $\hat{\lambda} = 1$.

- **Outro parâmetro desconhecido**

Como $V(X) = E(X) = \lambda$, decorre que $h(\lambda) = E(X^2) = V(X) + [E(X)]^2 = \lambda + \lambda^2$.

- **Estimativa de MV de $h(\lambda)$**

Usando o princípio de invariância dos estimadores (e, bem assim, das estimativas) de máxima verosimilhança, decorre então que a estimativa de máxima verosimilhança de $E(X^2)$ é:

$$\widehat{h(\lambda)} = h(\hat{\lambda}) = \hat{\lambda} + (\hat{\lambda})^2 = 1 + 1^2 = 2.$$

Pergunta 7

2 valores

Uma empresa de engenharia está a desenvolver um novo material compósito para uso em estruturas aeroespaciais. Para garantir a qualidade do material, é importante estimar a variância da resistência à tensão do material (em MPa). Assuma que a resistência à tensão segue uma distribuição normal e que foram realizados 20 testes de resistência à tensão, tendo-se obtido uma amostra $(x_1, \dots, x_{20})$ tal que $\sum_{i=1}^{20} x_i = 5003.5$ e $\sum_{i=1}^{20} x_i^2 = 4062131.25$.

Ao utilizar-se quantis aproximados às centésimas, o intervalo de confiança a 90% para a variância da

resistência à tensão do material é dado por

A: [89 474.1, 259 021.3] B: [88 582.0, 263 820.3] C: [93 244.2, 277 705.6] D: [103 322.8, 241 234.4]

- **V.a. de interesse**

X = resistência à tensão do novo material

- **Situação**

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

μ desconhecido

σ^2 DESCONHECIDO

- **Obtenção de IC para σ^2**

Passo 1 - Selecção da v.a. fulcral para σ^2

$$Z = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

Passo 2 — Obtenção dos quantis de probabilidade

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

$$\begin{cases} a_\alpha = F_{\chi^2_{(n-1)}}^{-1}(\alpha/2) = F_{\chi^2_{(19)}}^{-1}(0.05) \stackrel{\text{tabelas/calc.}}{=} 10.12 \\ b_\alpha = F_{\chi^2_{(n-1)}}^{-1}(1 - \alpha/2) = F_{\chi^2_{(19)}}^{-1}(0.95) \stackrel{\text{tabelas/calc.}}{=} 30.14 \end{cases}$$

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

$$P(a_\alpha \leq Z \leq b_\alpha) = 1 - \alpha$$

$$P\left[a_\alpha \leq \frac{(n-1)S^2}{\sigma^2} \leq b_\alpha\right] = 1 - \alpha$$

$$P\left[\frac{(n-1)S^2}{b_\alpha} \leq \sigma^2 \leq \frac{(n-1)S^2}{a_\alpha}\right] = 1 - \alpha$$

Passo 4 — Concretização

A expressão geral do intervalo de confiança a $(1 - \alpha) \times 100\%$ para σ^2 é

$$IC_{(1-\alpha) \times 100\%}(\sigma^2) = \left[\frac{(n-1)S^2}{b_\alpha}, \frac{(n-1)S^2}{a_\alpha} \right]$$

Uma vez que $n = 20$, $\alpha = 0.10$, $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{5003.5}{20} = 250.175$ e $(n-1)s^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2 = 4062131.25 - 20 \times 250.175^2 = 2810380.6375$, segue-se

$$IC_{90\%}(\sigma^2) = \left[\frac{2810380.6375}{30.14}, \frac{2810380.6375}{10.12} \right] \simeq [93244.2, 277705.6]$$

Pergunta 8

2 valores

Deseja-se testar a resistência de um novo tipo de vidro utilizado em janelas de edifícios altos. As especificações exigem que o valor esperado da resistência do vidro a impactos seja no mínimo de 1500 Joule, sendo o vidro considerado inadequado caso a sua resistência esperada seja inferior a esse valor. Tendo sido recolhidas 36 amostras do vidro, observou-se uma média de resistência ao impacto de 1 492.05 Joule.

Teste a hipótese $H_0 : \mu = 1500$ contra $H_1 : \mu < 1500$, assumindo que a resistência do vidro ao impacto possui desvio padrão conhecido e igual a 9 Joule e considerando um nível de significância de 5%.

- **V.a. de interesse**

X = resistência do vidro ao impacto

- **Situação**

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

μ DESCONHECIDO

$\sigma^2 = 9^2 = 81$

- **Nível de significância**

$\alpha_0 = 5\%$

- **Hipóteses**

$H_0 : \mu = \mu_0 = 1500$

$H_1 : \mu < \mu_0 = 1500$

- **Estatística de teste**

$$T = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \underset{H_0}{\sim} \text{normal}(0, 1)$$

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

Como a hipótese alternativa é $H_1 : \mu < 1500$, estamos realizando um teste unilateral à esquerda, logo a região de rejeição de H_0 é do tipo $W = (-\infty, c)$, onde

$$c : P(\text{Rejeitar } H_0 | H_0) \simeq \alpha_0 \Leftrightarrow c = \Phi^{-1}(\alpha_0) = -\Phi^{-1}(1 - \alpha_0) = -\Phi^{-1}(0.95) \stackrel{\text{tabelas/calc.}}{=} -1.6449.$$

- **Decisão**

O valor observado da estatística de teste é

$$t = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{1492.05 - 1500}{9 / \sqrt{36}} = -5.3.$$

Como $t = 5.3 \in W = (-\infty, -1.6449)$, devemos rejeitar a hipótese nula H_0 ao n.s. $\alpha_0 = 5\%$ [ou a qualquer outro n.s. superior a 5%. Com base na amostra recolhida e atendendo às hipóteses nula e alternativa, podemos afirmar que o valor esperado da resistência do vidro ao impacto é significativamente inferior a 1500 Joules, ao n.s. de 5%].

Pergunta 9

2 valores

A produção de grandes transformadores e condensadores elétricos requer a utilização de bifenilos policlorados (PCB), que são perigosos quando lançados no meio ambiente. Uma engenheira do ambiente conjectura a hipótese H_0 de que o nível de PCB (em partes por milhão) em peixes de determinado lago é uma variável aleatória com distribuição normal com valor esperado $\mu = 6$ e variância $\sigma^2 = 0.09$. Teste essa conjectura com base na seguinte tabela de frequências referentes a 200 registos casuais do nível de PCB.

Nível de PCB	]5.2, 5.5]	]5.5, 5.7]	]5.7, 5.85]	]5.85, 6.15]	]6.15, 6.3]	]6.3, 6.5]	]6.5, 6.8]
Freq. abs. observada	9	15	33	76	38	21	8
Freq. abs. esperada sob H_0	9.56	22.17	29.98	76.58	29.98	22.17	9.56

Indique o valor p aproximado (às centésimas).

A: 0.91 B: 0.40 C: 0.53 D: 0.60

- **V.a. de interesse**

X = nível de PCB em peixes do lago

- **Hipóteses**

$H_0 : X \sim \text{normal}(\mu = 6, \sigma^2 = 0.09)$ vs. $H_1 : X \not\sim \text{normal}(\mu = 6, \sigma^2 = 0.09)$

- **Estatística de teste**

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

onde: k = número de classes = 7; 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 $W = (c, +\infty)$.

- **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	$(-\infty, 5.5]$	9	9.56	$\frac{(9-9.56)^2}{9.56} \approx 0.033$
2	$(5.5, 5.7]$	15	22.17	2.319
3	$(5.7, 5.85]$	33	76.58	0.304
4	$(5.85, 6.15]$	76	76.58	0.004
5	$(6.15, 6.3]$	38	76.58	2.145
6	$(6.3, 6.5]$	21	22.17	0.062
7	$(6.5, +\infty)$	8	9.56	0.255
		$\sum_{i=1}^k o_i = n = 200$	$\sum_{i=1}^k E_i = n = 200$	$t = \sum_{i=1}^k \frac{(o_i - E_i)^2}{E_i} \approx 5.122$

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

$$valor - p = P(T > t | H_0) \approx 1 - F_{\chi^2_{(7-1)}}(5.122) \stackrel{calc.}{\approx} 0.528264 \approx 0.53.$$

Por consequência, a resposta, das quatro apresentadas, que se coaduna com o valor referido acima é **C: 0.53**.

[Mais, devemos decidir pela: rejeição de H_0 a qualquer n.s. $\alpha_0 > 52.8264\%$; não rejeição de H_0 a qualquer n.s. $\alpha_0 \leq 52.8264\%$, nomeadamente aos n.u.s. (1%, 5%, 10%).]

[Alternativamente, a consulta das tabelas dos quantis de probabilidade da distribuição do qui-quadrado com $k - 1 = 7$ graus de liberdade permitem concluir que

$$\begin{aligned} F_{\chi^2_{(7-1)}}^{-1}(0.40) = 4.570 &< 5.122 < 5.348 = F_{\chi^2_{(7-1)}}^{-1}(0.50) \\ 0.50 = 1 - 0.50 &< 1 - F_{\chi^2_{(7-1)}}(5.122) < 1 - 0.40 = 0.60. \end{aligned}$$

Logo, devemos: rejeitar H_0 a qualquer n.s. $\alpha_0 \geq 60\%$; não rejeitar de H_0 a qualquer n.s. $\alpha_0 \leq 7.5\%$, nomeadamente aos n.u.s. (1%, 5%, 10%). Deste modo, a resposta correcta é **C: 0.53**.]

Pergunta 10

2 valores

A percentagem de matéria gorda, x , e a percentagem de resíduos sólidos, Y , foram medidas em 16 amostras de leite de vaca. Para analisar a relação entre as variáveis, foi considerado um modelo de regressão linear com $Y_i \sim \text{normal}(\beta_0 + \beta_1 x_i, \sigma^2)$, $i = 1, \dots, 16$, independentes. Alguns dos resultados obtidos nessa análise são: $\bar{x} = 3.3313$, $\bar{y} = 7.6875$, $\sum_{i=1}^{16} x_i^2 = 182.6500$, $\sum_{i=1}^{16} y_i^2 = 993.3400$, $\hat{\beta}_1 = 2.7042$, para $[\min_{i=1, \dots, 16} x_i, \max_{i=1, \dots, 16} x_i] = [2, 5]$.

Obtenha uma estimativa da percentagem esperada de resíduos sólidos para um leite com 4% de matéria gorda. Calcule e interprete o valor do coeficiente de determinação.

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

Y = percentagem de resíduos sólidos (v.a. resposta)

x = matéria gorda (variável explicativa)

$Y_i \stackrel{indep.}{\sim} \text{normal}(\beta_0 + \beta_1 x_i, \sigma^2), i = 1, \dots, n]$

- **Estimativa de β_0 , recta de MQ e $\hat{E}(Y | x = 4)$**

Temos

- $n = 16, \quad \bar{x} = 3.3313, \quad \sum_{i=1}^n x_i^2 = 182.65, \quad \sum_{i=1}^n x_i^2 - n \bar{x}^2 = 182.65 - 16 \times 3.3313^2 = 5.08904$
- $\bar{y} = 7.6875, \quad \sum_{i=1}^n y_i^2 = 993.3400, \quad \sum_{i=1}^n y_i^2 - n \bar{y}^2 = 993.3400 - 16 \times 7.6875^2 = 47.7775$
- $\hat{\beta}_1 = 2.7042.$

A estimativa de β_0 é

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \times \bar{x} = 7.6875 - 2.7042 \times 3.3313 \simeq -1.321;$$

a recta de MQ e estimativa de $E(Y | x = 4)$ são dadas por

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \simeq -1.321 + 2.7042 x, \quad x \in [\min_{i=1, \dots, 16} x_i, \max_{i=1, \dots, 16} x_i] = [2, 5];$$

$$\hat{E}(Y | x = 4) = \hat{\beta}_0 + 4 \hat{\beta}_1 = 9.4958$$

- **Coeficiente de determinação**

Como $\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2}$, temos

$$\begin{aligned} r^2 &= \frac{(\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y})^2}{(\sum_{i=1}^n x_i^2 - n \bar{x}^2) \times (\sum_{i=1}^n y_i^2 - n \bar{y}^2)} \\ &= (\hat{\beta}_1)^2 \times \frac{\sum_{i=1}^n x_i^2 - n \bar{x}^2}{\sum_{i=1}^n y_i^2 - n \bar{y}^2} \\ &= 2.7042^2 \times \frac{5.08904}{47.7775} \\ &\simeq 0.778916. \end{aligned}$$

- **Interpretação do coeficiente de determinação**

Cerca de 78% da variação total da variável resposta Y é explicada pela variável x , através do modelo de regressão linear simples ajustado, donde podemos afirmar que a recta estimada parece ajustar-se bem ao conjunto de dados.