

Duração: 120 minutos

Exame de Época Normal — B

- 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

Uma empresa de material informático vende dois planos opcionais de assistência: A_1 (*software*) e A_2 (*hardware*). Sabe-se, da análise de dados históricos, que: 60% dos clientes não adquire nenhum plano; 25% dos clientes adquirem o plano A_1 ; 30% dos clientes adquirem o plano A_2 .

Qual das afirmações seguintes é verdadeira?

- A:** A_1 e A_2 são acontecimentos independentes
B: A_1 e A_2 são acontecimentos mutuamente exclusivos
C: A_1 e A_2 não são acontecimentos independentes nem são acontecimentos mutuamente exclusivos
D: A_1 e A_2 são acontecimentos independentes e também são acontecimentos mutuamente exclusivos

• **Quadro de acontecimentos e probabilidades**

Acontecimento	Probabilidade
A_1 = cliente adquire plano A_1	$P(A_1) = 0.25$
A_2 = cliente adquire plano A_2	$P(A_2) = 0.30$
$\bar{A}_1 \cap \bar{A}_2$ = cliente não adquire nenhum plano	$P(\bar{A}_1 \cap \bar{A}_2) = 0.6$
$A_1 \cup A_2$ = cliente adquire pelo menos um plano	$P(A_1 \cup A_2) = ?$
$A_1 \cap A_2$ = cliente adquire os dois planos	$P(A_1 \cap A_2) = ?$

• **Cálculos auxiliares**

Para determinar $P(A_1 \cap A_2)$ é necessário ter em conta que:

$$P(\bar{A}_1 \cap \bar{A}_2) = P(\overline{A_1 \cup A_2}) = 1 - P(A_1 \cup A_2) = 0.6 \Leftrightarrow P(A_1 \cup A_2) = 1 - 0.6 = 0.4;$$

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$

$$0.4 = 0.25 + 0.3 - P(A_1 \cap A_2)$$

$$P(A_1 \cap A_2) = 0.25 + 0.3 - 0.4 = 0.15.$$

• **Conclusão**

Os acontecimentos A_1 e A_2 são independentes se e só se $P(A_1 \cap A_2) = P(A_1) \times P(A_2)$, o que é falso, pois $0.15 \neq 0.25 \times 0.3$. Logo, as afirmações **A** e **D** são falsas.

Os acontecimentos A_1 e A_2 são mutuamente exclusivos se e só se $P(A_1 \cap A_2) = 0$, o que é falso, pois $P(A_1 \cap A_2) = 0.15 \neq 0$. Donde as afirmações **B** e **D** sejam falsas.

Com efeito, A_1 e A_2 não são acontecimentos independentes nem são acontecimentos mutuamente exclusivos. Assim sendo, a resposta correcta é **C**.

Em dias de mau tempo (aviso amarelo), o número de ocorrências reportadas em certa zona ribeirinha é uma variável aleatória X que segue uma distribuição de Poisson com valor esperado igual a 9.

Para um certo dia nas condições indicadas, calcule a probabilidade de se reportarem mais de 12 ocorrências quando o número reportado já excedeu o quantil de probabilidade 0.8 de X .

- **V.a.**

X = no. de ocorrências reportadas...

$X \sim \text{Poisson}(\lambda)$, onde $\lambda : E(X) = 9 \Leftrightarrow \lambda = 9$

- **F.p. e f.d. de X**

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

$$F_X(x) = P(X \leq x) = \sum_{n=0}^x P(X = n), \quad x \in \mathbb{N}_0 \quad [\text{tabelada...}]$$

- **Quantil pedido**

O quantil de probabilidade 0.8 de X , $q_{0.8}$, define-se do modo seguinte:

$$q_{0.8} \in \mathbb{N} : P(X \leq q_{0.8}) \geq 0.8 \quad \text{e} \quad P(X \geq q_{0.8}) \geq 1 - 0.8 \quad \Leftrightarrow \quad 0.8 \leq F_X(q_{0.8}) \leq 0.8 + P(X = q_{0.8})$$

$$F_X[(q_{0.8})^-] \leq 0.8 \leq F_X(q_{0.8}).$$

Ao recorrermos às tabelas ou a uma calculadora, temos

$$F_X(10) = F_X(11^-) = 0.7060 \leq 0.80 \leq F_X(12) = 0.8030,$$

pelo que $q_{0.8} = 11$.

- **Prob. pedida**

$$P(X > 12 | X > 11) = \frac{P(X > 12, X > 11)}{P(X > 11)} = \frac{P(X > 12)}{P(X > 11)} = \frac{1 - P(X \leq 12)}{1 - P(X \leq 11)} = \frac{1 - F_X(12)}{1 - F_X(11)}$$

$$\stackrel{\text{tabelas/calc.}}{=} \frac{1 - 0.8758}{1 - 0.8030} \approx 0.630457$$

O tempo entre avistamentos sucessivos de cetáceos numa estação oceanográfica, em horas, é uma variável aleatória X que se admite variar de acordo com uma distribuição exponencial de mediana igual a 4.5 horas.

Calcule a variância de X .

- **V.a. e f.d.p.**

X = tempo entre avistamentos sucessivos de cetáceos...

$X \sim \text{exponencial}(\lambda)$

$f_X(x) = \lambda e^{-\lambda x}$, $x > 0$, onde

$$\lambda : me = me(X) = F_X^{-1}(0.5) = 4.5$$

$$F_X(4.5) = \int_{-\infty}^{4.5} f_X(x) dx = \int_0^{4.5} \lambda e^{-\lambda x} dx = 1 - e^{-\lambda \times 4.5} = 0.5$$

$$\lambda = -\frac{1}{4.5} \ln(1 - 0.5) \approx 0.154033$$

- **Variância pedida**

$$V(X) \stackrel{\text{form.}}{=} \frac{1}{\lambda^2}$$

$$\approx \frac{1}{0.154033^2} \approx 42.147561.$$

Sejam X o número de vezes que um técnico é chamado de emergência para reparar uma avaria de uma máquina de controlo numérico e Y o número diário de avarias dessa máquina. A função de probabilidade conjunta do par aleatório (X, Y) é dada pela tabela à direita.

Calcule $V(2Y | X = 3)$.

X	Y		
	1	3	5
1	0.05	0.05	0
2	0.05	0.10	0.20
3	0.10	0.35	0.10

- **Par aleatório**

(X, Y)

X = número de vezes que um técnico é chamado de emergência para consertar uma avaria

Y = número de avarias de uma máquina de controlo numérico num dia

- **V.a. de interesse**

$(2Y + | X = 3)$

- **Variância pedida**

Aplicando as propriedades do operador variância,

$$V(2Y | X = 3) = 2^2 \times V(Y | X = 3) = 4 \times [E(Y^2 | X = 3) - E^2(Y | X = 3)].$$

Ora,

$$P(X = 3) = \sum_y P(X = 3, Y = y) = 0.10 + 0.35 + 0.10 = 0.55$$

$$P(Y = y | X = 3) = \frac{P(X = 3, Y = y)}{P(X = 3)} = \begin{cases} \frac{0.10}{0.55} = \frac{10}{55} = \frac{2}{11}, & y = 1, 5 \\ \frac{0.35}{0.55} = \frac{35}{55} = \frac{7}{11}, & y = 3 \\ 0, & \text{outros valores de } y \end{cases}$$

$$E(Y | X = 3) = \sum_{y=1}^5 y P(Y = y | X = 3) = 1 \times \frac{2}{11} + 3 \times \frac{7}{11} + 5 \times \frac{2}{11} = 3$$

$$E(Y^2 | X = 3) = \sum_{y=1}^5 y^2 P(Y = y | X = 3) = 1^2 \times \frac{2}{11} + 3^2 \times \frac{7}{11} + 5^2 \times \frac{2}{11} = \frac{115}{11}$$

$$V(Y | X = 3) = E(Y^2 | X = 3) - E^2(Y | X = 3) = \frac{115}{11} - 3^2 = \frac{115}{11} - \frac{99}{11} = \frac{16}{11}.$$

Logo,

$$V(2Y | X = 3) = 2^2 \times V(Y | X = 3) = 4 \times \frac{16}{11} = \frac{64}{11} \approx 5.81.$$

Admita que a massa do comprimido (X , em grama) de um tipo de vitamina segue uma distribuição normal com valor esperado 0.6 e desvio padrão 0.1. Sabe-se que determinada terapia consiste na toma de uma caixa de vitaminas com 30 destes comprimidos e que esta terapia não é eficaz se a massa destes 30 comprimidos for inferior a 17.5 gramas.

Considerando as massas dos diferentes comprimidos variáveis aleatórias independentes e identicamente distribuídas a X , calcule a probabilidade exata de a terapia com uma caixa de vitaminas não ser eficaz e indique o valor obtido arredondado às milésimas.

A: 0.433 B: 0.000 C: 0.181 D: 0.386

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

X_i = massa (em grama) do comprimido i , $i = 1, \dots, n$

$$n = 30, \quad E(X_i) = E(X) = \mu = 0.6, \quad V(X_i) = V(X) = \sigma^2 = 0.1^2 = 0.01$$

- **V.a. de interesse**

$S_n = \sum_{i=1}^n X_i$ = massa total de n comprimidos

- **Valor esperado e variância de S_n**

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

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

- **Distribuição de S_n**

Atendendo a que a v.a. de interesse é uma combinação linear de v.a. normais independentes, ela também possui distribuição normal:

$$S_n \sim \text{normal}(n\mu, n\sigma^2).$$

- **Prob. pedida**

$$\begin{aligned} P(S_n \leq 17.5) &= P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq \frac{17.5 - n\mu}{\sqrt{n\sigma^2}}\right) = \Phi\left(\frac{17.5 - 30 \times 0.6}{\sqrt{30 \times 0.01}}\right) = \Phi\left(\frac{17.5 - 18}{\sqrt{0.3}}\right) = \Phi(-0.91) \\ &= 1 - \Phi(0.91) = 1 - 0.8186 = 0.1814. \end{aligned}$$

Pergunta 6

2 valores

No desenvolvimento de um sistema de navegação para veículos autónomos sabe-se que o desvio (em metros) entre a localização real de um veículo e a localização dada por este sistema é uma variável aleatória X com função de densidade de probabilidade

$$f_X(x) = \frac{1}{2\beta} \exp\left(-\frac{|x|}{\beta}\right), \quad x \in \mathbb{R},$$

onde β é um parâmetro desconhecido positivo.

Para melhorar a precisão deste sistema, pretende-se estimar a variância de X , $V(X) = 2\beta^2$, com base em $(x_1, \dots, x_n)$, uma amostra de dimensão $n = 10$ proveniente de X tal que $\sum_{i=1}^n |x_i| = 25$. A estimativa de máxima verosimilhança de $V(X)$ com base nessa amostra é

A: 6.25 **B:** 5 **C:** 50 **D:** 12.5

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

X = desvio entre localização real e dada pelo sistema de navegação

$$f_X(x) = \frac{1}{2\beta} \exp\left(-\frac{|x|}{\beta}\right), \quad x \in \mathbb{R}$$

- **Parâmetro desconhecido**

$$\beta \quad (\beta > 0)$$

- **Amostra**

$$(x_1, \dots, x_n), \quad n = 10$$

- **Obtenção da estimativa de MV de β**

Passo 1 — Função de verosimilhança

$$L(\beta | \underline{x}) = f_{\underline{X}}(\underline{x}) \stackrel{X_i \text{ indep.}}{=} \prod_{i=1}^n f_{X_i}(x_i) \stackrel{X_i \sim X}{=} \prod_{i=1}^n f_X(x_i) = \prod_{i=1}^n \frac{1}{2\beta} \exp\left(-\frac{|x_i|}{\beta}\right) = (2\beta)^{-n} \prod_{i=1}^n \exp\left(-\frac{|x_i|}{\beta}\right), \beta > 0$$

Passo 2 — Função de log-verosimilhança

$$\ln L(\sigma | \underline{x}) = -n \ln(2\beta) - \frac{1}{\beta} \sum_{i=1}^n |x_i|$$

Passo 3 — Maximização

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

$$\hat{\beta} : \begin{cases} \left. \frac{d \ln L(\beta|x)}{d\beta} \right|_{\beta=\hat{\beta}} = 0 & \text{(ponto de estacionaridade)} \\ \left. \frac{d^2 \ln L(\beta|x)}{d\beta^2} \right|_{\beta=\hat{\beta}} < 0 & \text{(ponto de máximo)} \end{cases}$$
$$\begin{cases} -\frac{n}{\hat{\beta}} + \frac{1}{\hat{\beta}^2} \sum_{i=1}^n |x_i| = 0 \\ -\frac{n}{\hat{\beta}^2} - \frac{2}{\hat{\beta}^3} \sum_{i=1}^n |x_i| < 0 \end{cases} \Leftrightarrow \begin{cases} \hat{\beta} = \frac{1}{n} \sum_{i=1}^n |x_i| \\ -\frac{n}{\hat{\beta}^2} - \frac{2n}{\hat{\beta}^2} < 0 \quad (\text{prop. verdadeira}) \end{cases}$$

Passo 4 — Estimativa de MV de β

$$\hat{\beta} = \frac{1}{n} \sum_{i=1}^n |x_i| = \frac{1}{10} \times (|-5| + |-3| + |0| + |2| + |-1| + |4| + |-2| + |1| + |3| + |-4|) = \frac{25}{10} = 2.5$$

- **Outro parâmetro**

$$h(\beta) = V(X) = 2\beta^2$$

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

Pela propriedade de invariância dos EMV, temos

$$\widehat{h(\beta)} = h(\hat{\beta}) = 2(\hat{\beta})^2 = 2 \times 2.5^2 = 12.5.$$

Pergunta 7

2 valores

Suponha que a variável aleatória X representa o desvio em relação a um valor nominal de uma medição efectuada usando um espectómetro de massa. Admita que X possui distribuição normal com valor esperado desconhecido μ e variância conhecida e igual a $\sigma^2 = 0.04$.

Após ter obtido a expressão geral para o intervalo aleatório de confiança a 98% para μ baseado numa amostra aleatória $(X_1, X_2, \dots, X_n)$ de X , determine o valor mínimo de n para que a amplitude de tal intervalo não seja superior a 0.1.

- **V.a. de interesse**

X = desvio em relação a um valor nominal de uma medição...

- **Situação**

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

μ DESCONHECIDO

$$\sigma^2 = 0.04$$

- **Obtenção de IAC para μ**

Passo 1 — Seleção da v.a. fulcral para μ

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \text{normal}(0, 1)$$

Passo 2 — Obtenção dos quantis de probabilidade

Dado que $(1 - \alpha) \times 100\% = 98\%$ (i. e., $\alpha = 0.02$), faremos uso dos quantis simétricos

$$\begin{cases} a_\alpha = -b_\alpha \\ b_\alpha = \Phi^{-1}(1 - \alpha/2) = \Phi^{-1}(0.99) \stackrel{\text{tabela/calcul.}}{=} 2.3263. \end{cases}$$

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

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

$$P\left[-b_\alpha \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq b_\alpha\right] = 1 - \alpha$$

$$P\left(\bar{X} - b_\alpha \times \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + b_\alpha \times \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Passo 4 — IAC a $(1 - \alpha) \times 100\%$ para μ

Atendendo aos extremos aleatórios acima, temos

$$IAC_{98\%}(\mu) = \left[\bar{X} - b_\alpha \times \frac{\sigma}{\sqrt{n}}, \quad \bar{X} + b_\alpha \times \frac{\sigma}{\sqrt{n}} \right].$$

• **Dimensão mínima pedida**

Pretendemos a dimensão mínima da a.a. de modo a que amplitude do IAC para μ não seja superior a $amp = 0.1$, i.e., o menor dos inteiros positivos que satisfaça

$$n : \left(\bar{X} + b_\alpha \times \frac{\sigma}{\sqrt{n}} \right) - \left(\bar{X} - b_\alpha \times \frac{\sigma}{\sqrt{n}} \right) \leq amp$$

$$2 b_\alpha \times \frac{\sigma}{\sqrt{n}} \leq amp$$

$$n \geq \left(\frac{2 b_\alpha \times \sigma}{amp} \right)^2 = \left(\frac{2 \times 2.3263 \times 0.2}{0.1} \right)^2 = 86.5867.$$

Logo, a dimensão mínima pedida é $n_{min} = 87$.

Pergunta 8

2 valores

Uma engenheira de qualidade de uma fábrica de sensores de temperatura suspeita que a probabilidade de um novo sensor ser defeituoso, p , poderá ser superior ao valor histórico de 2% de sensores defeituosos produzidos na fábrica.

Sabendo que a engenheira encontrou 15 sensores defeituosos num total de 500 novos sensores, escolhidos aleatoriamente, teste $H_0 : p = 0.02$ versus $H_1 : p > 0.02$ ao nível de significância de 7.5%.

• **V.a. de interesse**

$$X = \begin{cases} 1, & \text{se novo sensor é defeituoso} \\ 0, & \text{c.c.} \end{cases}$$

• **Situação**

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

$$p = P(\text{novo sensor ser defeituoso}) \quad \text{DESCONHECIDA}$$

$$n = 500 \quad (n \geq 30)$$

• **Hipóteses**

$$H_0 : p = p_0 = 0.02$$

$$H_1 : p > p_0$$

• **N.s.**

$$\alpha_0 = 7.5\%$$

• **Estatística de teste**

$$T = \frac{\bar{X} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \underset{H_0}{\sim} \text{normal}(0, 1)$$

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

Lidamos com um teste unilateral superior ($H_1 : p > p_0$), logo a região de rejeição de H_0 é do tipo $W = (c, +\infty)$, sendo

$$c = \Phi^{-1}(1 - \alpha_0) = \Phi^{-1}(1 - 0.075) = \Phi^{-1}(0.925) \stackrel{\text{tabela/calcul.}}{=} 1.4395.$$

- **Decisão com base no valor observado da estatística de teste**

O valor observado da estatística de teste é

$$t = \frac{\bar{x} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{\frac{15}{500} - 0.02}{\sqrt{\frac{0.02 \times (1-0.02)}{500}}} \approx 1.5972.$$

Como $t \approx 1.5972 > 1.4395 = c$, rejeita-se $H_0 : p = 2\%$ em favor de $H_1 : p > 2\%$ ao n.s. de $\alpha_0 = 7.5\%$ [ou a qualquer outro n.s. superior a α_0].

Pergunta 9

2 valores

Uma engenheira conjectura que o tempo de incubação de uma doença viral (X , em dias) possui função de distribuição dada por $F_0(x) = \Phi[(\ln x - 3.36)/0.28]$, para $x > 0$ (hipótese H_0). Os dados de uma amostra casual de 120 pacientes foram divididos em 6 classes equiprováveis sob H_0 e as frequências absolutas observadas são as facultadas abaixo.

Classe	C_1	C_2	C_3	C_4	C_5	C_6
Freq. abs. observada	14	12	18	17	29	30

Com base no teste de ajustamento do qui-quadrado, devemos

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 = tempo de incubação (em dias) de uma doença viral

- **Hipóteses**

$$H_0 : F_X(x) = F_0(x) = \Phi[(\ln x - 3.36)/0.28], \quad \forall x$$

$$H_1 : F_X(x) \neq F_0(x), \quad \text{para algum } x$$

- **Estatística de teste**

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

onde:

k = número de classes = 6;

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)

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

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

Por forma a obter o valor observado da estatística de teste, é útil considerar a tabela seguinte:

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	C_1	14	20	$\frac{(14-20)^2}{20} = 1.8$
2	C_2	12	20	3.2
3	C_3	18	20	0.2
4	C_4	17	20	0.45
5	C_5	29	20	4.05
6	C_6	30	20	5
		$\sum_{i=1}^k o_i = n = 120$	$\sum_{i=1}^k E_i = n = 100$	$t = \sum_{i=1}^k \frac{(o_i - E_i)^2}{E_i} \approx 14.7$

[Note-se que $t = \sum_{i=1}^k \frac{(o_i - E_i)^2}{E_i} \frac{k}{n} = (\sum_{i=1}^k o_i^2) - n = \frac{6}{120} (14^2 + 12^2 + 18^2 + 17^2 + 29^2 + 30^2) - 120 = 14.7$.]

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

$$valor - p = P(T > t | H_0) \approx 1 - F_{\chi_{(6-1)}^2}(14.7) \stackrel{calc.}{\approx} 0.0117.$$

Assim sendo, devemos decidir pela:

- rejeição de H_0 a qualquer n.s. $\alpha_0 > 1.17\%$, nomeadamente ao n.u.s. de 5% e 10%.
- não rejeição de H_0 a qualquer n.s. $\alpha_0 \leq 1.17\%$, nomeadamente ao n.u.s. de 1%.

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

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

[Alternativamente, a consulta das tabelas dos quantis de probabilidade da distribuição do qui-quadrado com $k - 1 = 5$ graus de liberdade permitem concluir que

$$\begin{aligned} F_{\chi_{(6-1)}^2}^{-1}(0.975) = 12.83 &< 14.7 < 15.09 = F_{\chi_{(6-1)}^2}^{-1}(0.990) \\ 0.01 = 1 - 0.990 &< 1 - F_{\chi_{(6-1)}^2}(14.7) < 1 - 0.975 = 0.025. \end{aligned}$$

Logo, devemos: rejeitar H_0 a qualquer n.s. $\alpha_0 \geq 2.5\%$, designadamente aos n.u.s. de 5% e 10%; não rejeitar de H_0 a qualquer n.s. $\alpha_0 \leq 1\%$, nomeadamente ao n.u.s. de 1%. Deste modo, a resposta correcta é **B**: Rejeitar H_0 a 5% e 10% e não rejeitar H_0 a 1%.]

Pergunta 10

2 valores

De a forma a estudar relação entre a temperatura (x , em grau Fahrenheit) à superfície de um pavimento e a deflexão do pavimento (Y , em polegada), uma engenheira de materiais recolheu uma amostra casual e obteve os seguintes resultados:

$$\sum_{i=1}^{20} x_i = 1478, \quad \sum_{i=1}^{20} x_i^2 = 143215.8, \quad \sum_{i=1}^{20} y_i = 12.8, \quad \sum_{i=1}^{20} x_i y_i = 1083.67, \quad \hat{\sigma}^2 \approx 0.006106,$$

onde $[\min_{i=1,\dots,20} x_i, \max_{i=1,\dots,20} x_i] = [70, 100]$. Admita que as variáveis x e Y estão relacionadas de acordo com o modelo de regressão linear simples, $Y = \beta_0 + \beta_1 x + \epsilon$.

Atendendo aos resultados acima e considerando as hipóteses de trabalho convenientes, obtenha um intervalo de confiança a 99% para $E(Y | x = 85)$.

• [Modelo de RLS e hipóteses de trabalho

Y = deflexão do pavimento (v.a. resposta)

x = temperatura à superfície do pavimento (variável explicativa)

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, \dots, n$$

$$\epsilon_i \stackrel{i.i.d.}{\sim} \text{normal}(0, \sigma^2), \quad i = 1, \dots, n]$$

- **Estimativas de β_0 e β_1**

Temos

- $n = 20$
- $\sum_{i=1}^n x_i = 1478$
 $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1478}{20} = 73.9$
 $\sum_{i=1}^n x_i^2 = 143215.8$
 $\sum_{i=1}^n x_i^2 - n \bar{x}^2 = 143215.8 - 20 \times 73.9^2 = 33991.6$
- $\sum_{i=1}^n y_i = 12.8$
 $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = \frac{12.8}{20} = 0.64$
- $\sum_{i=1}^n x_i y_i = 1083.67$
 $\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y} = 1083.67 - 20 \times 73.9 \times 0.64 = 137.75.$

Consequentemente,

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} \simeq \frac{137.75}{33991.6} \simeq 0.004052$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \times \bar{x} \simeq 0.64 - 0.004052 \times 73.9 \simeq 0.340557$$

- **Obtenção do IC a 99% para $E(Y | x_0) = \beta_0 + \beta_1 x_0$**
- **Passo 1 — V.a. fulcral para $(\beta_0 + \beta_1 x_0)$**

$$Z = \frac{(\hat{\beta}_0 + \hat{\beta}_1 x_0) - (\beta_0 + \beta_1 x_0)}{se(\hat{\beta}_0 + \hat{\beta}_1 x_0)} \sim t_{(n-2)}, \quad \text{onde } se(\hat{\beta}_0 + \hat{\beta}_1 x_0) = \sqrt{\left[\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} \right] \hat{\sigma}^2}$$

Passo 2 — Quantis de probabilidade

Já que $(1 - \alpha) \times 100\% = 99\%$, temos $\alpha = 0.01$ e lidaremos com

$$\begin{cases} a_\alpha = -b_\alpha \\ b_\alpha = F_{t_{(n-2)}}^{-1}(1 - \alpha/2) = F_{t_{(20-2)}}^{-1}(1 - 0.05/2) = F_{t_{(18)}}^{-1}(0.995) \stackrel{\text{tabela/calc.}}{=} 2.878. \end{cases}$$

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

$$P(a_\alpha \leq Z \leq b_\alpha) = 1 - \alpha \Leftrightarrow P\left[-b_\alpha \leq \frac{(\hat{\beta}_0 + \hat{\beta}_1 x_0) - (\beta_0 + \beta_1 x_0)}{se(\hat{\beta}_0 + \hat{\beta}_1 x_0)} \leq b_\alpha\right] = 1 - \alpha$$

$$P[(\hat{\beta}_0 + \hat{\beta}_1 x_0) - b_\alpha \times se(\hat{\beta}_0 + \hat{\beta}_1 x_0) \leq \beta_0 + \beta_1 x_0 \leq (\hat{\beta}_0 + \hat{\beta}_1 x_0) + b_\alpha \times se(\hat{\beta}_0 + \hat{\beta}_1 x_0)] = 1 - \alpha$$

Passo 4 — Concretização

A expressão geral do $IC_{(1-\alpha) \times 100\%}(\beta_0 + \beta_1 x_0)$ é dada por $[(\hat{\beta}_0 + \hat{\beta}_1 x_0) \pm b_\alpha \times se(\hat{\beta}_0 + \hat{\beta}_1 x_0)]$. Assim sendo, para $x_0 = 85$, temos

$$\begin{aligned} IC_{99\%}(\beta_0 + \beta_1 x_0) &= \left[(0.340557 + 0.004052 \times 85) \pm 2.878 \times \sqrt{\left[\frac{1}{20} + \frac{(73.9 - 85)^2}{33991.6} \right] \times \hat{\sigma}^2} \right] \\ &\simeq [0.684977 \pm 2.878 \times 0.0180946] \\ &\simeq [0.632901, 0.737053]. \end{aligned}$$