DM DEPARTAMENTO DE MATEMÁTICA TÉCNICO LISBOA	**Probabilidade e Estatística** **LEAmb, LEAN, LEEC, LEGM, LEMat**	2º Semestre – 2023/2024 02/07/2024 **18:00–20:00**

Duração: **120** minutos **Exame Época Recurso – C**

- 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 múltipla 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

Em determinada população: 12% das pessoas adquirem a revista A; $p \times 100\%$ ($p \in (0,1)$) das pessoas adquire a revista B; e 5% não adquire nenhuma destas duas revistas. Sejam A e B os acontecimentos *pessoa selecionada ao acaso adquire a revista* A e *pessoa selecionada ao acaso adquire a revista* B, respetivamente.

Determine o valor de p, de modo a que $P(A \cap B) = 0$.

- **Quadro de acontecimentos e probabilidades**

Acontecimento	Probabilidade
A = pessoa selecionada ao acaso adquire a revista A	$P(A) = 0.12$
B = pessoa selecionada ao acaso adquire a revista B	$P(B) = p$
C = pessoa selecionada ao acaso não adquire nenhumas das duas revistas	$P(\bar{A} \cap \bar{B}) = 0.05$

- **Valor pedido**

$$\begin{aligned} p \quad : \quad & P(A \cap B) = 0 \\ & P(A \cup B) = P(A) + P(B) - P(A \cap B) \\ & 1 - P(\bar{A} \cap \bar{B}) = P(A) + P(B) - P(A \cap B) \\ & 1 - 0.05 = 0.12 + p - P(A \cap B) \\ & 0.95 - 0.12 = p - P(A \cap B) \\ & 0.83 = p - P(A \cap B). \end{aligned}$$

Para que $P(A \cap B) = 0$, devemos ter $p = 0.83$.

Pergunta 2 2 valores

A variável aleatória X representa o número de peixes a retirar com reposição de um lago até se obter uma amostra com 10 peixes de uma determinada espécie, que se estima constituir 40% da população de peixes do lago. A função de distribuição de X satisfaz a igualdade $F_X(x) = 1 - F_Y(9)$, em que $Y \sim \text{binomial}(x, 0.4)$, para $x = 10, 11, \ldots$

Sabendo que já foram retirados 10 peixes do lago sem ainda se ter obtido uma amostra com 10 peixes daquela espécie, calcule a probabilidade de não ser necessário capturar adicionalmente mais do que 10 peixes para se obter a amostra pretendida. **A**: 0.7637 **B**: 0.2446 **C**: 0.7554 **D**: 0.2529

- **V.a. de interesse**

X = # de peixes a retirar com reposição de um lago até obter amostra de 10 peixes de certa espécie

- **F.d. de X**

  $F_X(x) = P(X \leq x) = 1 - F_Y(9), \quad x = 10, 11, \ldots, \quad$ onde $Y \sim \text{binomial}(x, 0.4)$.

- **Probabilidade pedida**

$$\begin{aligned}
P(X \leq 20 \mid X > 10) &= \frac{P(10 < X \leq 20)}{P(X > 10)} = \frac{F_X(20) - F_X(10)}{1 - F_X(10)} \\
&= \frac{[1 - F_{binomial(20,0.4)}(9)] - [1 - F_{binomial(10,0.4)}(9)]}{1 - [1 - F_{binomial(10,0.4)}(9)]} \\
&= \frac{F_{binomial(10,0.4)}(9) - F_{binomial(20,0.4)}(9)}{1 - [1 - F_{binomial(10,0.4)}(9)]} \\
&\overset{tabelas/calc.}{=} \frac{0.9999 - 0.7553}{0.9999} \\
&\simeq 0.2446.
\end{aligned}$$

Pergunta 3	2 valores

Admita que determinado indicador financeiro é uma variável aleatória X com função de distribuição $F_X(x) = \frac{1}{\pi}\arctan(x) + \frac{1}{2}$, para $x \in \mathbb{R}$.

Determine o primeiro, segundo e terceiro quartis de X. Comente os resultados.

- **V.a.**

  $X =$ razão entre o comprimento e o diâmetro de determinada peça metálica

- **F.d. de X**

  $F_X(x) = \frac{1}{\pi}\arctan(x) + \frac{1}{2}, \quad x \in \mathbb{R}$

- **Quartis pedidos**

$$\begin{aligned}
\chi_p \quad : \quad & F_X(\chi_p) = p, \quad p = 0.25, 0.5, 0.75 \\
& \frac{1}{\pi}\arctan(x) + \frac{1}{2} = p \\
& x = \tan[\pi(p - 1/2)] = \begin{cases} \tan[\pi(0.25 - 1/2)] = \tan(-\pi/4) = -1, & p = 0.25 \\ \tan[\pi(0.5 - 1/2)] = \tan(0) = 0, & p = 0.5 \\ \tan[\pi(0.75 - 1/2)] = \tan(\pi/4) = +1, & p = 0.75. \end{cases}
\end{aligned}$$

- **Comentário**

  O primeiro e terceiro quartis são simétricos, sugerindo que os valores da v.a. X se distribuem simetricamente em torno da origem. [Efectivamente, a f.d.p. de X, $f_X(x) = \frac{dF_X(x)}{dx} = [\pi(1 + x^2)]^{-1}$ é simétrica em torno da origem. Mais, $F_X^{-1}(1-p) = \tan\{\pi[(1-p) - 1/2)]\} = -\tan[\pi(p - 1/2)] = -F_X^{-1}(p)$, para $p \in (0,1)$, à semelhança do que acontece com os quantis da normal padrão ($\Phi^{-1}(1-p) = -\Phi^{-1}(p)$, $p \in (0,1)$.]

Pergunta 4	2 valores

Um sistema elétrico é constituído por duas peças, A e B, que podem estar: em funcionamento (estado 0); em funcionamento mas com deficiência (estado 1); ou em falha (estado 2). Sejam X e Y as variáveis aleatórias que representam o estado da peça A e B, respetivamente. Suponha que a função de probabilidade conjunta de X e Y é dada pela tabela à direita.

	Y		
X	0	1	2
0	0	1/4	0
1	1/4	0	1/4
2	0	1/4	0

Determine e comente o coeficiente de correlação entre X e Y. Serão X e Y variáveis aleatórias independentes?

- **Par aleatório**

  (X,Y)

  $X =$ estado da peça A

  $Y =$ estado da peça B

- **Correlação pedida**

$$corr(X,Y) = \frac{cov(X,Y)}{\sqrt{V(X)\,V(Y)}} = \frac{E(XY)-E(X)\times E(Y)}{\sqrt{V(X)\,V(Y)}}.$$

  Logo são necessários alguns cálculos auxiliares que envolverão as f.p. conjunta de (X,Y) e marginais de X e Y dadas por $P(X=x)=\sum_y P(X=x,Y=y)$ e $P(Y=y)=\sum_x P(X=x,Y=y)$, respectivamente.

	Y			
X	0	1	2	$P(X=x)$
0	0	1/4	0	1/4
1	1/4	0	1/4	1/2
2	0	1/4	0	1/4
$P(Y=y)$	1/4	1/2	1/4	1

- **Valor esperado e variância de X e Y**

  Dado que as variáveis aleatórias X e Y são identicamente distribuídas

$$E(X) = \sum_{x=0}^{2} x \times P(X=x) = 1\times 1/2 + 2\times 1/4 = 1 = E(Y)$$

$$V(X) = E(X^2)-E^2(X) = \sum_{x=0}^{2} x^2\,P(X=x) - E^2(X) = 1^2\times 1/2 + 2^2\times 1/4 - 1^2 = 0.5 = V(Y).$$

- **Valor esperado de XY e covariância entre X e Y**

$$E(XY) = \sum_{x=0}^{2}\sum_{y=0}^{2} xy \times P(X=x,Y=y) = 1\times 2\times 1/4 + 2\times 1\times 1/4 = 1$$

$$cov(X,Y) = E(XY)-E(X)\times E(Y) = 1 - 1\times 1 = 0$$

- **Correlação pedida** (cont.)

$$corr(X,Y) = \frac{cov(X,Y)}{\sqrt{V(X)\,V(Y)}} = 0$$

- **Comentários**

  Como $corr(X,Y)=0$, X e Y são v.a. não correlacionadas.

  Pelo valor da correlação nada podemos afirmar quanto à independência das v.a.

- **Averiguação de (in)dependência**

  X e Y são v.a. INDEPENDENTES sse $P(X=y,Y=y) = P(X=x)\times P(Y=y)$, $\forall (x,y)\in\mathbb{R}^2$. Como

$$P(X=-1,Y=-1)=0 \quad \neq \quad P(X=-1)\times P(Y=-1) = \frac{1}{4}\times\frac{1}{4} = \frac{1}{16},$$

  podemos afirmar que X e Y são v.a. DEPENDENTES embora não correlacionadas.

Pergunta 5	2 valores

Admita que as massas (em mg) de dois chips são variáveis aleatórias independentes e identicamente distribuídas normais com valor esperado 5.7 mg e desvio padrão 0.5 mg.

Determine a probabilidade exacta de a massa média desses dois chips ser não inferior a 5 mg e não exceder 7 mg.

- **V.a.**

  $X_i =$ massa (em mg) do chip i, $\quad i = 1, \ldots, n$

  $n = 2$

  $X_i \overset{i.i.d.}{\sim} X \sim \text{normal}(\mu = 5.7, \sigma^2 = 0.5^2)$

- **V.a. de interesse**

  $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i =$ média das massas de n chips

- **Valor esperado, variância e distribuição de $\bar{X}$**

  $E(\bar{X}) = \frac{1}{n}\sum_{i=1}^n E(X_i) \overset{X_i \sim X}{=} E(X) = \mu$

  $V(\bar{X}) = \frac{1}{n^2}\sum_{i=1}^n V(X_i) \overset{X_i \sim_{i.i.d.} X}{=} \frac{V(X)}{n} = \frac{\sigma^2}{n}$

  Pelo facto de $\bar{X}$ ser uma combinação linear de v.a. com distribuição normal, também tem distribuição normal: $\bar{X} \sim \text{normal}(\mu, \sigma^2/n)$.

- **Probabilidade pedida**

$$
\begin{aligned}
P(5 \le \bar{X} \le 7) &= P\left(\frac{5 - E(\bar{X})}{\sqrt{V(\bar{X})}} \le Z = \frac{\bar{X} - E(\bar{X})}{\sqrt{V(\bar{X})}} \le \frac{7 - E(\bar{X})}{\sqrt{V(\bar{X})}}\right) = P\left(\frac{5 - 5.7}{0.5/\sqrt{2}} \le Z \le \frac{7 - 5.7}{0.5/\sqrt{2}}\right) \\
&\simeq P(-1.98 \le Z \le 3.68) = \Phi(3.68) - \Phi(-1.98) = \Phi(3.68) - [1 - \Phi(1.98)] \\
&\overset{tabelas/calc.}{=} 0.999883 - (1 - 0.9761) \\
&= 0.975983 \simeq 0.9760.
\end{aligned}
$$

Pergunta 6	2 valores

Admita que a velocidade do vento em certo local é representada pela variável aleatória X com função de densidade de probabilidade dada por $f_X(x) = \frac{3}{\lambda}\left(\frac{x}{\lambda}\right)^2 e^{-(x/\lambda)^3}$, para $x > 0$, onde λ é um parâmetro positivo desconhecido.

Deduza o estimador de máxima verosimilhança de λ com base numa amostra aleatória $(X_1, \ldots, X_n)$ de X.

- **V.a. de interesse**

  $X =$ velocidade do vento em certo local

  $f_X(x) = \frac{3}{\lambda}\left(\frac{x}{\lambda}\right)^2 e^{-(x/\lambda)^3}, \quad x > 0$

- **Parâmetro desconhecido**

  $\lambda \ (\lambda > 0)$

- **Amostra aleatória; amostra**

  $\underline{X} = (X_1, \ldots, X_n), \quad X_i \overset{i.i.d}{\sim} X, \quad i = 1, \ldots, n;$

  $\underline{x} = (x_1, \ldots, x_n)$

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

  Passo 1 — Função de verosimilhança

$$
\begin{aligned}
L(\lambda \mid \underline{x}) &= f_{\underline{X}}(\underline{x}) \overset{X_i\, indep}{=} \prod_{i=1}^n f_{X_i}(x_i) \overset{X_i \sim X}{=} \prod_{i=1}^n f_X(x_i) \\
&= \prod_{i=1}^n \frac{3}{\lambda^3} x_i^2 e^{-(x_i/\lambda)^3} \\
&= \frac{3^n}{\lambda^{3n}} \left(\prod_{i=1}^n x_i^2\right) e^{-\sum_{i=1}^n (x_i/\lambda)^3}
\end{aligned}
$$

Passo 2 — Função de log-verosimilhança

$$\ln L(\lambda \mid \underline{x}) \quad = \quad n\ln(3) - 3n\ln(\lambda) + 2\sum_{i=1}^{n}\ln(x_i) - \frac{1}{\lambda^3}\sum_{i=1}^{n}x_i^3$$

Passo 3 — Maximização

A estimativa de MV de λ é aqui representada por $\hat{\lambda}$ e

$$\hat{\lambda} \;:\; \begin{cases} \left.\frac{d\ln L(\lambda|\underline{x})}{d\lambda}\right|_{\lambda=\hat{\lambda}} = 0 & \text{(ponto de estacionaridade)} \\ \left.\frac{d^2\ln L(\lambda|\underline{x})}{d\lambda^2}\right|_{\lambda=\hat{\lambda}} < 0 & \text{(ponto de máximo)} \end{cases} \Leftrightarrow \begin{cases} -\frac{3n}{\hat{\lambda}} + \frac{3}{\hat{\lambda}^4}\sum_{i=1}^{n}x_i^3 = 0 \\ -\frac{12}{\hat{\lambda}^5} < 0 \quad \text{(prop. verdadeira)} \end{cases}$$

$$\begin{cases} \frac{1}{\hat{\lambda}^3}\sum_{i=1}^{n}x_i^3 = n \\ - \end{cases}$$

Passo 4 — Estimativa de MV de λ

$$\hat{\lambda} \quad = \quad \left(\frac{1}{n}\sum_{i=1}^{n}x_i^3\right)^{1/3}$$

- **Estimador de MV de λ**

$$EMV(\lambda) = \left(\frac{1}{n}\sum_{i=1}^{n}X_i^3\right)^{1/3}.$$

Pergunta 7	2 valores

Um engenheiro assume que a despesa diária de manutenção por aeronave é uma variável aleatória (X, em euros) com distribuição normal, com valor esperado e desvio padrão desconhecidos. Uma amostra casual de $n = 11$ registos de tal despesa conduziu a $\bar{x} = 5100$ e $s^2 = 1020^2$.

Determine um intervalo de confiança a 95% para o desvio padrão de X.

A: [753.8, 1625.0] **B**: [727.1, 1508.0] **C**: [712.7, 1790.0] **D**: [688.9, 1651.2]

- **V.a. de interesse**

  X = despesa diária de manutenção por aeronave (em euros)

- **Situação**

  $X \sim \text{normal}(\mu, \sigma^2)$, $\mu = E(X)$ desconhecido, $\sigma = \sqrt{V(X)}$ DESCONHECIDO

- **Obtenção do IC para σ**

Passo 1 — Variável aleatória fulcral para σ

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

Passo 2 — Quantis de probabilidade

Dado que $n = 11$ e $(1-\alpha)\times 100\% = 95\%$, usaremos os quantis

$$(a_\alpha, b_\alpha): \begin{cases} P(a_\alpha \le Z \le b_\alpha) = 1-\alpha \\ P(Z < a_\alpha) = P(Z > b_\alpha) = \alpha/2 \end{cases} \Leftrightarrow \begin{cases} a_\alpha = F^{-1}_{\chi^2_{(n-1)}}(\alpha/2) = F^{-1}_{\chi^2_{(10)}}(0.025) \overset{tabela/calc.}{=} 3.247 \\ b_\alpha = F^{-1}_{\chi^2_{(n-1)}}(1-\alpha/2) = F^{-1}_{\chi^2_{(10)}}(0.975) \overset{tabela/calc.}{=} 20.48. \end{cases}$$

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

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

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

Passo 4 — Concretização

Atendendo à expressão geral do IC para σ,

$$IC_{(1-\alpha)\times 100\%}(\sigma) \quad = \quad \left[\sqrt{\frac{(n-1)\,s^2}{F^{-1}_{\chi^2_{(n-1)}}(1-\alpha/2)}},\ \sqrt{\frac{(n-1)\,s^2}{F^{-1}_{\chi^2_{(n-1)}}(\alpha/2)}}\right],$$

ao par de quantis acima e ao facto de $s^2 = 2.5$, temos:

$$IC_{95\%}(\sigma) \quad = \quad \left[\sqrt{\frac{(11-1)\times 1020^2}{20.48}},\ \sqrt{\frac{(11-1)\times 1020^2}{3.247}}\right] \simeq [712.8,\, 1790.0].$$

Pergunta 8	2 valores

Um fabricante de pendurais para camas articuladas supõe que a massa (em kg) dos mesmos é uma variável aleatória X com valor esperado igual a 3 kg e desvio padrão $\sigma = 0.9$ kg. Foram selecionados aleatoriamente 36 pendurais, tendo-se obtido uma média amostral de 3.5 kg.

Confronte as hipóteses $H_0 : \mu = 3$ e $H_1 : \mu > 3$, ao nível de significância de 2%.

- **V.a. de interesse**

  X = massa (em kg) de pendural para camas articulada

- **Situação**

  X v.a. com distribuição arbitrária

  $\mu = E(X)$ DESCONHECIDO

  $\sigma^2 = 0.9^2$

  $n = 36 \geq 30$

- **Hipóteses**

  $H_0 : \mu = \mu_0 = 3 \quad$ vs. $\quad H_1 : \mu > \mu_0$

- **N.s.**

  $\alpha_0 = 2\%$

- **Estatística de teste**

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

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

  Tratando-se de um teste unilateral superior ($H_1 : \mu > \mu_0$), a região de rejeição de H_0 é o intervalo $W = (c, +\infty)$, onde

  $$c \quad = \quad \Phi^{-1}(1-\alpha_0) = \Phi^{-1}(1-0.02) = \Phi^{-1}(0.98) \stackrel{tabelas/calc.}{=} 2.0537.$$

  [De notar que $P(\text{Rejeitar } H_0 \mid H_0) \simeq \alpha_0$.]

- **Decisão**

  O valor observado da estatística de teste é igual a

  $$t \quad = \quad \frac{3.5-3}{0.9/\sqrt{36}} = \frac{10}{3} \simeq 3.33,$$

  Como $t \simeq 3.33 \in W = (2.0537, +\infty)$, devemos rejeitar H_0 ao n.s. de $\alpha_0 = 2\%$ [ou a qualquer outro n.s. superior a α_0].

Pergunta 9	2 valores

Um engenheiro do ambiente suspeita que o desvio entre a concentração observada de um certo poluente nas águas de um lago e o limite legal (em dg/l) é uma variável aleatória contínua (X) com função de densidade de probabilidade,

$$f_0(x) = \begin{cases} 1+x, & -1 < x \le 0 \\ 1-x, & 0 < x < 1 \\ 0, & \text{caso contrário} \end{cases}$$

(hipótese H_0). Para testar a sua suspeita, o engenheiro realizou, de forma independente, 100 medições de X e classificou-as de acordo com a tabela apresentada a seguir, tendo incluído também as frequências absolutas esperadas sob H_0 (arredondadas à unidade) para cada uma das 5 classes em que foi dividido o contradomínio de X.

Classe	$[-1.0,-0.6[$	$[-0.6,-0.2[$	$[-0.2,0.2[$	$[0.2,0.6[$	$[0.6,1.0[$
Freq. absoluta observada	9	24	40	24	3
Freq. absoluta esperada sob H_0	8	24	36	24	8

Obtenha o valor-p aproximado (ou um intervalo para o valor-p aproximado) do teste de ajustamento do qui-quadrado. Qual deverá ser a conclusão do engenheiro ao nível de significância de 2.5%?

- **V.a. de interesse**

  X = desvio entre a concentração observada e o limite legal (em dg/l)

- **Hipóteses**

  $H_0 : f_X(x) = f_0(x), \forall x \in \mathbb{R}$ vs. $H_1 : f_X(x) \neq f_0(x)$, para algum x

- **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)

  No cálculo do valor observado da estatística de teste convém recorrer à seguinte tabela auxiliar:

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	$[-1.0,-0.6[$	9	8	$\frac{(9-8)^2}{8} = 0.125$
2	$[-0.6,-0.2[$	24	24	0
3	$[-0.2,0.2[$	40	36	0.444
4	$[0.2,0.6[$	24	24	0
5	$[0.6,1.0[$	3	8	3.125
		$\sum_{i=1}^{k} o_i = n = 100$	$\sum_{i=1}^{k} E_i = n = 100$	$t = \sum_{i=1}^{k} \frac{(o_i-E_i)^2}{E_i} \simeq 3.694$

  Dado que a região de rejeição de H_0 é um intervalo à direita, temos:

  $$valor-p \;=\; P(T > t \mid H_0) \simeq 1 - F_{\chi^2_{(k-\beta-1)}}(t) = 1 - F_{\chi^2_{(4)}}(3.694) \overset{calc.}{\simeq} 0.4490.$$

  Por consequência, é suposto:

  - não rejeitarmos H_0 a qualquer n.s. $\alpha_0 \le 44.90\%$, designadamente a qualquer dos n.u.s. $(1\%, 5\%, 10\%)$ e ao n.s. de 2.5%[;

– rejeitar H_0 a qualquer n.s. $\alpha_0 > 44.90\%$].

[Em alternativa, o recurso às tabelas de quantis da distribuição do qui-quadrado permitiria adiantar um intervalo aproximado para o valor-p:

$$\begin{array}{rcl} F^{-1}_{\chi^2_{(4)}}(0.50) = & < & t = 3.694 < 4.045 = F^{-1}_{\chi^2_{(3)}}(0.60) \\ 0.40 = 1 - 0.60 & < & 1 - F_{\chi^2_{(3)}}(3.694) < 1 - 0.50 = 0.50. \end{array}$$

Uma vez que o intervalo aproximado para o valor-p é (0.40, 0.50) podemos adiantar que:

– não rejeitarmos H_0 a qualquer n.s. $\alpha_0 \geq 40\%$, em particular aos n.u.s. $(1\%, 5\%, 10\%)$ ou ao n.s. de 2.5%[;

– rejeitarmos H_0 a qualquer n.s. $\alpha_0 \geq 50\%$].]

Pergunta 10 — 2 valores

Uma engenheira está a investigar a relação entre o diâmetro das rodas motrizes (x, em cm) de uma cadeira de rodas e a distância de propulsão (Y, em cm). Obtiveram-se os seguintes resultados para uma amostra casual de dimensão $n = 10$:

$$\sum_{i=1}^{10} x_i = 545.9, \quad \sum_{i=1}^{10} x_i^2 = 29\,810.68, \quad \sum_{i=1}^{10} y_i = 871.3, \quad \sum_{i=1}^{10} x_i y_i = 47\,500.85, \quad \hat{\sigma}^2 \simeq 54.802402,$$

onde $[\min_{i=1,\ldots,10} x_i, \max_{i=1,\ldots,10} x_i] = [53.0, 56.6]$. Admita que x e Y estão relacionadas de acordo com o modelo de regressão linear simples, $Y = \beta_0 + \beta_1 x + \epsilon$.

Tendo em conta os valores acima e as hipóteses de trabalho convenientes, obtenha um intervalo de confiança a 95% para $E(Y \mid x = 54)$.

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

  $Y =$ distância de propulsão (v.a. resposta)

  $x =$ diâmetro das rodas motrizes (variável explicativa)

  $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \ldots, n$

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

- **Estimativas de β_0 e β_1**

  Temos

  - $n = 10$
  - $\sum_{i=1}^n x_i = 545.9$
    $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i = \frac{545.9}{10} = 54.59$
    $\sum_{i=1}^n x_i^2 = 29\,810.68$
    $\sum_{i=1}^n x_i^2 - n\,\bar{x}^2 = 29\,810.68 - 10 \times 54.59^2 = 9.9990$
  - $\sum_{i=1}^n y_i = 871.3$
    $\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i = \frac{871.3}{10} = 87.13$
  - $\sum_{i=1}^n x_i y_i = 47\,500.85$
    $\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y} = 47\,500.85 - 10 \times 54.59 \times 87.13 = -63.417.$

  Consequentemente,

  $$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} \simeq \frac{-63.417}{9.9990} \simeq -6.342334$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \times \bar{x} \simeq 87.13 - (-6.342334) \times 54.59 \simeq 433.358013$$

- **Obtenção do IC a** 95% **para** $E(Y \mid x_0) = \beta_0 + \beta_1\, x_0)$

- **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 $(1-\alpha) \times 100\% = 95\%$, temos $\alpha = 0.05$ e lidaremos com

$$\begin{cases} a_\alpha = -b_\alpha \\ b_\alpha = F^{-1}_{t_{(n-2)}}(1-\alpha/2) = F^{-1}_{t_{(10-2)}}(1-0.05/2) = F^{-1}_{t_{(8)}}(0.975) \stackrel{tabela/calc.}{=} 2.306. \end{cases}$$

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

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

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

Passo 4 — Concretização

A expressão geral do $IC_{(1-\alpha)\times 100\%}(\beta_0 + \beta_1\, x_0)$ é dada por $\left[(\hat{\beta}_0 + \hat{\beta}_1\, x_0) \pm b_\alpha \times se(\hat{\beta}_0 + \hat{\beta}_1\, x_0)\right]$. Logo,

$$\begin{aligned} IC_{95\%}(\beta_0 + \beta_1\, x_0) &= \left[(433.358013 - 6.342334 \times 54) \pm 2.306 \times \sqrt{\left[\frac{1}{10} + \frac{(54.59-54)^2}{9.9990}\right] \times \hat{\sigma}^2}\right] \\ &\simeq [90.871977 \pm 2.306 \times 2.718106] \\ &\simeq [84.604025, 97.139929]. \end{aligned}$$