



Duração: 90 minutos

2º teste B

Justifique convenientemente todas as respostas!

Grupo I

10 valores

1. Seja (X_1, X_2, \dots, X_n) uma amostra aleatória de uma população X com parâmetro θ , $\theta > 0$, possuindo função densidade de probabilidade

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

pelo que $E[X] = \frac{\theta}{\theta+1}$ e $E[X^2] = \frac{\theta}{\theta+2}$.

- (a) Determine o estimador de máxima verosimilhança do parâmetro θ .

(3.0)

• **Ed.p. de X**

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

• **Parâmetro desconhecido**

$$\theta, \quad \theta > 0$$

• **Amostra**

$\underline{x} = (x_1, \dots, x_n)$ amostra de dimensão n proveniente da população X

• **Obtenção do estimador de MV de θ**

Passo 1 — Função de verosimilhança

$$\begin{aligned} L(\theta|\underline{x}) &= f_{\underline{X}}(\underline{x}) \stackrel{X_i \text{ indep}}{=} \prod_{i=1}^n f_{X_i}(x_i) \stackrel{X_i \sim X}{=} \prod_{i=1}^n f_X(x_i) \\ &= \prod_{i=1}^n (\theta x_i^{\theta-1}) \\ &= \theta^n \times \left(\prod_{i=1}^n x_i \right)^{\theta-1}, \quad \theta > 0 \end{aligned}$$

Passo 2 — Função de log-verosimilhança

$$\ln L(\theta|\underline{x}) = n \ln(\theta) + (\theta - 1) \sum_{i=1}^n \ln(x_i) \quad [\text{onde } \sum_{i=1}^n \ln(x_i) < 0 \text{ já que } x_i \in (0, 1), i = 1, \dots, n.]$$

Passo 3 — Maximização

A estimativa de MV de θ é doravante representada por $\hat{\theta}$ e atente-se que $\hat{\theta} = \operatorname{argmax}_{\theta} L(\theta|\underline{x}) = \operatorname{argmax}_{\theta} \ln L(\theta|\underline{x})$. Neste caso em particular:

$$\hat{\theta} : \begin{cases} \frac{d \ln L(\theta|\underline{x})}{d\theta} \Big|_{\theta=\hat{\theta}} = 0 & (\text{ponto de estacionaridade}) \\ \frac{d^2 \ln L(\theta|\underline{x})}{d\theta^2} \Big|_{\theta=\hat{\theta}} < 0 & (\text{ponto de máximo}) \\ \frac{n}{\hat{\theta}} + \sum_{i=1}^n \ln(x_i) = 0 \\ -\frac{n}{\hat{\theta}^2} < 0 \\ \hat{\theta} = -\frac{n}{\sum_{i=1}^n \ln(x_i)} \\ -\frac{n}{\hat{\theta}^2} = -\frac{[\sum_{i=1}^n \ln(x_i)]^2}{n} < 0 \quad \text{proposição verdadeira pois } n \in \mathbb{N} \end{cases}$$

• **Passo 4 — Estimador de MV de θ**

$$\hat{\lambda} = -\frac{n}{\sum_{i=1}^n \ln(X_i)}$$

- (b) Tendo em vista a estimação do valor esperado de X , compare a eficiência do estimador \bar{X} (1.5) relativamente ao estimador $T = 2X_1 - X_n$.

• **Parâmetro desconhecido**

$$\mu = E(X)$$

• **Estimador de $\mu = E(X)$**

$$\bar{X}$$

• **Erro quadrático médio de \bar{X}**

$$\begin{aligned} EQM_{\mu}(\bar{X}) &= V(\bar{X}) + [bias_{\mu}(\bar{X})]^2 \\ &= V(\bar{X}) + [E(\bar{X}) - \mu]^2 \\ &\stackrel{X_i \sim i.i.d.X}{=} \frac{V(X)}{n} + [E(X) - E(X)]^2 \\ &= \frac{V(X)}{n} \end{aligned}$$

$$[\text{onde } V(X) = \frac{\theta}{\theta+2} - \left(\frac{\theta}{\theta+1}\right)^2 = \frac{\theta}{(\theta+2)(\theta+1)^2}.]$$

• **Outro estimador de $\mu = E(X)$**

$$T = 2X_1 - X_n$$

• **Erro quadrático médio de T**

$$\begin{aligned} EQM_{\mu}(T) &= V(T) + [bias_{\mu}(T)]^2 \\ &= V(T) + [E(T) - \mu]^2 \\ &= V(2X_1 - X_n) + [E(2X_1 - X_n) - E(X)]^2 \\ &\stackrel{X_i \sim i.i.d.X}{=} 5V(X) + [E(X) - E(X)]^2 \\ &= 5V(X) \end{aligned}$$

• **Eficiência do estimador \bar{X} relativamente ao estimador $T = \frac{X_1 + X_{10}}{2}$**

$$\begin{aligned} e_{\mu}(\bar{X}, T) &= \frac{EQM_{\mu}(T)}{EQM_{\mu}(\bar{X})} \\ &= \frac{5V(X)}{\frac{V(X)}{n}} \\ &= 5n \end{aligned}$$

• **Comentário**

Tendo em conta que $n \in \mathbb{N}$, temos $e_{\mu}(\bar{X}, T) = 5n > 1$ (i.e., $EQM_{\mu}(T) > EQM_{\mu}(\bar{X})$), pelo que pode afirmar-se que \bar{X} é um estimador mais eficiente que $T = 2X_1 - X_n$ no que respeita à estimação de $\mu = E(X)$.

2. O limite de resistência à tração de cabos individuais de determinado tipo produzidos numa fábrica por cada um de dois métodos, A e B , possui distribuição normal em ambos os casos. Da experiência acumulada, sabe-se que o valor esperado do limite de resistência à tração de cabos produzidos pelo método tradicional A é igual a 1800 lb. Para avaliar a viabilidade de passar a usar o método B no fabrico desse tipo de cabos, um engenheiro de materiais realizou ensaios de tração que envolveram 25 cabos produzidos pelo método B , selecionados ao acaso, tendo obtido valores com média e variância amostrais de 1806.3 lb e 71.24 lb², respetivamente.

- (a) Obtenha um intervalo de confiança a 95% para o desvio padrão do limite de resistência à tração de (2.5)

cabos individuais produzidos pelo método B.

- **V.a. de interesse**

X = limite de resistência à tração (LRT) de cabo produzido de acordo com o método B

- **Situação**

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

μ desconhecido

σ^2 DESCONHECIDO

- **Obtenção do IC para σ**

Passo 1 — Selecção da v.a. fulcral para σ

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

uma vez que é suposto determinar um IC para a variância de uma população normal, com valor esperado desconhecido.

Passo 2 — Obtenção dos quantis de probabilidade

Ao ter-se em consideração que $n = 25$ e $(1 - \alpha) \times 100\% = 95\%$, far-se-á uso dos quantis

$$(a_\alpha, b_\alpha) : \begin{cases} P(a_\alpha \leq Z \leq b_\alpha) = 1 - \alpha \\ P(Z < a_\alpha) = P(Z > b_\alpha) = \alpha/2. \end{cases}$$

$$\begin{cases} a_\alpha = F_{\chi^2_{(n-1)}}^{-1}(\alpha/2) = F_{\chi^2_{(24)}}^{-1}(0.025) \stackrel{\text{tabela/calc.}}{=} 12.40 \\ b_\alpha = F_{\chi^2_{(n-1)}}^{-1}(1 - \alpha/2) = F_{\chi^2_{(24)}}^{-1}(0.975) \stackrel{\text{tabela/calc.}}{=} 39.36. \end{cases}$$

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

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

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

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

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

Passo 4 — Concretização

Atendendo ao par de quantis acima e ao facto de

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

$$s^2 = 71.24,$$

segue-se:

$$\begin{aligned} IC_{90\%}(\sigma) &= \left[\sqrt{\frac{(25-1) \times 71.24}{39.36}}, \sqrt{\frac{(25-1) \times 71.24}{12.40}} \right] \\ &\simeq [\sqrt{43.439}, \sqrt{137.884}] \\ &\simeq [6.590, 11.742]. \end{aligned}$$

- (b) A gestora da fábrica afirma que o valor esperado do limite de resistência à tração de cabos produzidos pelo método B é igual ao dos produzidos pelo método A. Considera que a opinião da gestora é suportada pelos dados ao nível de significância de 5%? (3.0)

- **Hipóteses**

$$H_0 : \mu = \mu_0 = 1800$$

$$H_1 : \mu \neq \mu_0$$

- **N.s.**

$$\alpha_0 = 5\%$$

- **Estatística de teste**

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

pois pretendemos efectuar um teste sobre o valor esperado de uma população normal, com variância desconhecida.

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

Tratando-se de um teste bilateral ($H_1 : \mu \neq \mu_0$), a região de rejeição de H_0 (para valores da estatística de teste) é do tipo $W = (-\infty, -c) \cup (c, +\infty)$, onde $c : P(\text{Rejeitar } H_0 | H_0) = \alpha_0$, i.e.,

$$c : P(T \in W | H_0) = \alpha_0$$

$$2 \times [1 - F_{t_{(n-1)}}(c)] = \alpha_0$$

$$c = F_{t_{(n-1)}}^{-1}(1 - \alpha_0/2)$$

$$c = F_{t_{(24)}}^{-1}(0.975)$$

$$c \stackrel{\text{tabela/ calc.}}{=} 2.064.$$

- **Decisão**

Uma vez que

$$n = 25$$

$$\bar{x} = 1806.3$$

$$s = \sqrt{71.24}$$

o valor observado da estatística de teste é igual a

$$\begin{aligned} t &= \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} \\ &= \frac{1806.3 - 1800}{\sqrt{\frac{71.24}{25}}} \\ &\approx 3.732. \end{aligned}$$

Como $t \approx 3.732 \in W = (-\infty, -2.064) \cup (2.064, +\infty)$, devemos rejeitar H_0 ao n.s. $\alpha_0 = 5\%$ [ou a qualquer n.s. superior a $\alpha_0 = 5\%$].

Grupo II	10 valores
-----------------	------------

1. Um modelo genérico especifica que as plantas de certa espécie se distribuem entre quatro categorias (1,2,3,4) de acordo com as seguintes proporções: $p_1 = 0.656$, $p_2 = p_3 = 0.093$ e $p_4 = 0.158$. Seleccionadas ao acaso 197 plantas dessa espécie, obtiveram-se as seguintes frequências observadas: $o_1 = 125$, $o_2 = 18$, $o_3 = 20$ e $o_4 = 34$. Averigüe, aplicando um teste apropriado, se tal modelo genérico é consistente com este conjunto de dados. Decida com base no valor-p. (4.0)

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

X = categoria da planta

$$p_i = \begin{cases} P(X = i), & i = 1, 2, 3, 4 \\ 0, & \text{caso contrário} \end{cases}$$

- **Hipóteses**

$$H_0: p_i = p_i^0, \quad \text{onde} \quad p_1^0 = 0.656, \quad p_2^0 = 0.093, \quad p_3^0 = 0.093, \quad p_4^0 = 0.158$$

$$H_1: p_i \neq p_i^0, \text{ para algum } i$$

- **Estatística de teste**

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

onde:

k = No. de classes = 4 (categorias)

O_i = Frequência absoluta observável da classe i

E_i = Frequência absoluta esperada, sob H_0 , da classe i

β = No. de parâmetros a estimar = 0 [dado que a distribuição conjecturada em H_0 está completamente especificada, i.e., H_0 é uma hipótese simples.]

- **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 $W = (c, +\infty)$.

- **Cálculo das frequências absolutas esperadas sob H_0**

As frequências absolutas esperadas sob H_0 são dadas por $E_i = n \times p_i^0$ ($i = 1, 2, 3, 4$) e iguais a

$$E_1 = 197 \times 0.656 = 129.232$$

$$E_2 = 197 \times 0.093 = 18.321$$

$$E_3 = 197 \times 0.093 = 18.321$$

$$E_4 = 197 \times 0.158 = 31.126.$$

[Importa notar que não é necessário fazer qualquer agrupamento de classes uma vez que em pelo menos 80% das classes se verifica $E_i \geq 5$ e $E_i \geq 1$ para todo o i .]

- **Decisão (com base no p-value)**

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

i	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	125	129.232	$\frac{[125 - 129.232]^2}{129.232} \approx 0.145$
2	18	18.321	$\frac{[18 - 18.321]^2}{18.321} \approx 0.006$
3	20	18.321	$\frac{[20 - 18.321]^2}{18.321} \approx 0.154$
4	34	31.126	$\frac{[34 - 31.126]^2}{31.126} \approx 0.265$
$\sum_{i=1}^4 o_i = n = 197$		$\sum_{i=1}^4 E_i = n = 197$	$t = \sum_{i=1}^4 \frac{(o_i - E_i)^2}{E_i} \approx 0.57$

Assim, temos

$$\begin{aligned} t &= \sum_{i=1}^4 \frac{(o_i - E_i)^2}{E_i} \\ &\approx 0.57. \end{aligned}$$

Uma vez que a região de rejeição de H_0 é para este teste um intervalo à direita temos:

$$\begin{aligned} \text{valor} - p &= P(T > t \mid H_0) \\ &= P[T > 0.57 \mid H_0] \\ &\approx 1 - F_{\chi^2_{(4-1-0)}}(0.57) \\ &\stackrel{calc.}{\approx} 0.903265. \end{aligned}$$

Consequentemente, é suposto:

- não rejeitar H_0 a qualquer n.s. $\alpha_0 \leq 90.3265\%$, pelo que o modelo genérico é consistente com os dados a qualquer dos níveis usuais de significância (1%, 5% e 10%);
- rejeitar H_0 a qualquer n.s. $\alpha_0 > 90.3265\%$.

[Em alternativa, poderíamos recorrer às tabelas de quantis da distribuição do qui-quadrado com 2 graus de liberdade e adiantar um intervalo para o *valor-p*:

$$\begin{aligned} F_{\chi^2_{(3)}}^{-1}(0.075) = 0.472 &< t = 0.57 < 0.584 = F_{\chi^2_{(3)}}^{-1}(0.10) \\ 0.075 &< F_{\chi^2_{(3)}}^2(0.57) < 0.10 \\ 1 - 0.10 &< 1 - F_{\chi^2_{(3)}}^2(0.57) < 1 - 0.075 \\ 0.90 &< \text{valor} - p < 0.925. \end{aligned}$$

Logo o intervalo para o *valor-p* é (0.90, 0.925) e é suposto:

- não rejeitar H_0 a qualquer n.s. $\alpha_0 \leq 90.0\%$, pelo que o modelo genérico é consistente com os dados a qualquer dos níveis usuais de significância (1%, 5% e 10%).
- rejeitar H_0 a qualquer n.s. $\alpha_0 \geq 92.5\%$.]

2. O *Archaeopteryx* é um animal extinto, tendo penas como um pássaro bem como dentes e uma cauda óssea como um réptil. As medições de comprimentos em centímetros do fémur (osso da perna), x , e do úmero (um osso do braço), Y , para dez espécimes fósseis que preservam os dois ossos, conduziram aos seguintes resultados:

$$\sum_{i=1}^{10} x_i = 580, \quad \sum_{i=1}^{10} x_i^2 = 35080, \quad \sum_{i=1}^{10} y_i = 657, \quad \sum_{i=1}^{10} y_i^2 = 45251, \quad \sum_{i=1}^{10} x_i y_i = 39829$$

- (a) Obtenha as estimativas de mínimos quadrados dos parâmetros da reta de regressão linear simples de Y em x e interprete a estimativa do parâmetro β_1 do modelo. (2.0)

• **Estimativas de β_0 e β_1**

Dado que

$$n = 10$$

$$\sum_{i=1}^n x_i = 580$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{580}{10} = 58$$

$$\sum_{i=1}^n x_i^2 = 35080$$

$$\sum_{i=1}^n x_i^2 - n(\bar{x})^2 = 35080 - 10 \times 58^2 = 1440$$

$$\sum_{i=1}^n y_i = 657$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = \frac{657}{10} = 65.7$$

$$\sum_{i=1}^n y_i^2 = 45251$$

$$\sum_{i=1}^n y_i^2 - n(\bar{y})^2 = 45251 - 10 \times 65.7^2 = 2086.1$$

$$\sum_{i=1}^n x_i y_i = 39829$$

$$\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y} = 39829 - 10 \times 58 \times 65.7 = 1723,$$

as estimativas de β_1 e β_0 são, para este modelo de RLS, iguais a:

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n(\bar{x})^2} \\ &= \frac{1723}{1440} \\ &\approx 1.196528 \end{aligned}$$

$$\begin{aligned}\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \times \bar{x} \\ &\simeq 65.7 - 1.196528 \times 58 \\ &= -3.698611\end{aligned}$$

- **Interpretação da estimativa de β_1**

$$\hat{\beta}_1 \simeq 1.396226$$

Estima-se que um aumento de um cm no comprimento do fémur esteja associado a um aumento no valor esperado do comprimento do úmero de aproximadamente 1.196528 cm.

- (b) Após ter enunciado as hipóteses de trabalho que entender por convenientes, obtenha um intervalo de confiança a 90% para o valor esperado do comprimento do úmero de um espécime fóssil cujo fémur tem 74 cm de comprimento. (4.0)

- **Hipóteses de trabalho**

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

β_0, β_1 e σ^2 DESCONHECIDOS

- **Obtenção do IC para $E(Y | x_0) = \beta_0 + \beta_1 x_0$**

Passo 1 — V.a. fulcral para $E(Y | x_0) = \beta_0 + \beta_1 x_0$

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

Passo 2 — Quantis de probabilidade

Já que $(1 - \alpha) \times 100\% = 90\%$, temos $\alpha = 0.10$ e lidaremos com os quantis

$$(a_\alpha, b_\alpha) : \begin{cases} P(a_\alpha \leq Z \leq b_\alpha) = 1 - \alpha \\ P(Z < a_\alpha) = P(Z > b_\alpha) = \alpha/2. \end{cases}$$

$$\begin{cases} a_\alpha = F_{t_{(n-2)}}^{-1}(\alpha/2) = -F_{t_{(10-2)}}^{-1}(1 - 0.10/2) \stackrel{\text{tabela/calcul.}}{=} -1.860 \\ b_\alpha = F_{t_{(10-2)}}^{-1}(1 - 0.10/2) \stackrel{\text{tabela/calcul.}}{=} 1.860. \end{cases}$$

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

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

$$P \left[a_\alpha \leq \frac{(\hat{\beta}_0 + \hat{\beta}_1 x_0) - (\beta_0 + \beta_1 x_0)}{\sqrt{\hat{\sigma}^2 \times \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} \right]}} \leq b_\alpha \right] = 1 - \alpha$$

$$\begin{aligned}P \left[(\hat{\beta}_0 + \hat{\beta}_1 x_0) - b_\alpha \times \sqrt{\hat{\sigma}^2 \times \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} \right]} \leq \beta_0 + \beta_1 x_0 \right. \\ \left. \leq (\hat{\beta}_0 + \hat{\beta}_1 x_0) - a_\alpha \times \sqrt{\hat{\sigma}^2 \times \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} \right]} \right] = 1 - \alpha\end{aligned}$$

- **Passo 4 — Concretização**

Uma vez que a estimativa de σ^2 é igual a

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{n-2} \left[\left(\sum_{i=1}^n y_i^2 - n \bar{y}^2 \right) - (\hat{\beta}_1)^2 \left(\sum_{i=1}^n x_i^2 - n \bar{x}^2 \right) \right] \\ &= \frac{1}{10-2} (2086.1 - 1.196528^2 \times 1440) \\ &\simeq 3.060234\end{aligned}$$

e a expressão geral do IC pretendido é

$$IC_{(1-\alpha) \times 100\%}(\beta_0 + \beta_1 x_0) \\ = \left[(\hat{\beta}_0 + \hat{\beta}_1 x_0) \pm F_{t(n-2)}^{-1} (1 - \alpha/2) \times \sqrt{\hat{\sigma}^2 \times \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} \right]} \right],$$

temos

$$IC_{90\%}(\beta_0 + \beta_1 \times 74) \\ = \left[(-3.698611 + 1.196528 \times 74) \pm 1.860 \times \sqrt{3.060234 \times \left[\frac{1}{10} + \frac{(74-58)^2}{1440} \right]} \right] \\ = [84.844461 \pm 1.860 \times 0.921990] \\ = [84.844461 \pm 1.71490] \\ = [83.129560, 86.559362].$$