DM
DEPARTAMENTO DE MATEMÁTICA
TÉCNICO LISBOA

Probabilidades e Estatística

— TODOS OS CURSOS

1º semestre – 2020/2021
04/02/2021 – **15:00**

Duração: 60+15 minutos **Teste 2C**

Nº : ________ **Nome:** ______________________ **Curso:** ________ **Sala:** ________

1. Admita que a distribui??o da vida útil das baterias do flash de câmaras de determinado tipo é representada pela variável aleatória X com função de densidade de probabilidade

$$f_X(x) = \begin{cases} \theta^2\, x e^{-\theta\, x}, & x > 0 \\ 0, & \text{caso contrário,} \end{cases}$$

onde θ é um parâmetro positivo desconhecido. Determine a estimativa de máxima verosimilhança de θ, baseada na amostra $(x_1, \ldots, x_5)$ proveniente da população X.

Preencha a caixa abaixo com o resultado obtido com, pelo menos, quatro casas decimais.

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

  $X =$ tempo de vida da bateria do flash de câmaras de determinado tipo

  $$f_X(x) = \begin{cases} \theta^2\, x e^{-\theta\, x}, & x > 0 \\ 0, & \text{caso contrário,} \end{cases}$$

- **Parâmetro desconhecido**

  $\theta, \quad \theta > 0$

- **Obtenção da estimativa de MV de θ**

  Passo 1 — Função de verosimilhança

  $$\begin{aligned} L(\theta \mid \underline{x}) &= f_{\underline{X}}(\underline{x}) \\ &\overset{X_i\, indep}{=} \prod_{i=1}^{n} f_{X_i}(x_i) \\ &\overset{X_i \sim X}{=} \prod_{i=1}^{n} f_X(x_i) \\ &= \prod_{i=1}^{n} \left[\theta^2\, x_i\, e^{-\theta x_i}\right] \\ &= \theta^{2n}\, e^{-\theta \sum_{i=1}^{n} x_i} \prod_{i=1}^{n} x_i, \quad \theta > 0, \quad x_{(1)} = \min x_i > 0 \end{aligned}$$

  Passo 2 — Função de log-verosimilhança

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

  Passo 3 — Maximização

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

  $$\begin{cases} \frac{2n}{\hat{\theta}} - \sum_{i=1}^{n} x_i = 0 \\ -\frac{2n}{\hat{\theta}^2} < 0 \quad \text{(prop. verdadeira)} \end{cases}$$

Passo 4 — Estimativa de MV de θ

$$\hat{\theta} = \frac{2\,n}{\sum_{i=1}^{n} x_i} = \frac{2}{\bar{x}}, \quad \text{onde } n = 5$$

2. Pretende-se comparar a esperança de vida (em anos) em dois países diferentes (A e B), μ_A e μ_B. Para tal, foram recolhidas observações casuais de n indivíduos de cada um dos paíes. A amostra de indivíuos do país A conduziu à média e ao desvio padrão amostrais $\bar{x}_A = x_A$ e $s_A = s_A$. Relativamente ao país B, a média e o desvio padrão amostrais são iguais a $\bar{x}_B = x_B$ e $s_B = s_B$.

Determine um intervalo de confiança a $(1-\alpha)\times 100\%$ para a diferença de esperanças de vida $\mu_A - \mu_B$.

Assinale a sua resposta com uma cruz.

- ⦿ [,]
- ⦿ [,]
- ⦿ [,]
- ⦿ [,]

- **V.a. de interesse**

  X_A = duração de vida no país A

  X_B = duração de vida no país B

- **Situação**

  X_A e X_B v.a. independentes com distribuições arbitrárias.

  $(\mu_A - \mu_B)$ DESCONHECIDO

  σ_H^2 e σ_M^2 desconhecidas

- **Obtenção do IC para $\mu_A - \mu_B$**

  Passo 1 — Seleção da v.a. fulcral para $(\mu_H - \mu_M)$

$$Z \;=\; \frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)}{\sqrt{\frac{S_A^2}{n} + \frac{S_B^2}{n}}} \overset{a}{\sim} \text{normal}(0,1)$$

  Passo 2 — Obtenção dos quantis de probabilidade

$$\begin{cases} a_\alpha = \Phi^{-1}(\alpha/2) = -\Phi^{-1}(1-\alpha/2) \\ b_\alpha = \Phi(1-\alpha/2) \end{cases}$$

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

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

$$P\left[a_\alpha \le \frac{(\bar{X}_A - \bar{X}_B) - (\mu_H - \mu_M)}{\sqrt{\frac{S_A^2}{n} + \frac{S_B^2}{n}}} \le b_\alpha\right] = 1-\alpha$$

$$P\left[(\bar{X}_H - \bar{X}_M) - b_\alpha \times \sqrt{\frac{S_A^2}{n} + \frac{S_B^2}{n}} \le \mu \le (\bar{X}_H - \bar{X}_M) - a_\alpha \times \sqrt{\frac{S_A^2}{n} + \frac{S_B^2}{n}}\right] = 1-\alpha$$

  Passo 4 — Concretização

$$IC_{(1-\alpha)\times 100\%}(\mu_A - \mu_B) = \left[(\bar{x}_A - \bar{x}_B) - \Phi^{-1}(1-\alpha/2)\sqrt{\frac{s_A{}^2}{n} + \frac{s_B{}^2}{n}},\ (\bar{x}_A - \bar{x}_B) + \Phi^{-1}(1-\alpha/2)\sqrt{\frac{s_A{}^2}{n} + \frac{s_B{}^2}{n}}\right]$$

3. Um delegado de saúde pública está convicto que pelo menos 80% dos adultos portugueses consideram que as crianças saudáveis devem ser vacinadas. Para avaliar a hipótese do delegado de saúde, foram inquiridos n adultos portugueses dos quais tot manifestaram-se favoráveis à vacinação de crianças saudáveis.

Rejeita-se ou não a hipótese sustentada pelo delegado de saúde pública? Decida com base no valor-p aproximado.

Assinale a sua resposta com uma cruz.

- ⦿ Rejeita-se H_0 a 1%, 5% e 10%.
- ⦿ Rejeita-se H_0 a 5% e 10% e não se rejeita H_0 a 1%.
- ⦿ Rejeita-se H_0 a 10% e não se rejeita H_0 a 1% e 5%.
- ⦿ Não se rejeita H_0 a 1%, 5% e 10%.

- **V.a. de interesse**

  $X = \begin{cases} 1, & \text{se adulto se manifesta favoravelmente à vacinação de crianças saudáveis} \\ 0, & \text{caso contrário} \end{cases}$

- **Situação**

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

  p DESCONHECIDO

- **Hipóteses**

  $H_0 : p \geq p_0 = 0.80$

  $H_1 : p < p_0$

- **Estatística de teste**

  $T = \frac{\bar{X}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \overset{a}{\sim}_{p=p_0} \text{normal}(0,1)$

- **Região de rejeição de H_0**

  Teste unilateral inferior, r.r. de H_0 é um intervalo à esquerda, $W = (-\infty, c)$.

- **Decisão (com base no valor-p)**

$$valor-p \quad = \quad P\left[T < t = \frac{\bar{x}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{tot/n - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \mid p = p_0\right] \simeq \Phi\left(\frac{tot/n - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}\right)$$

  Devemos escolher a resposta, das quatro apresentadas acima, que se coaduna com:

  - a não rejeição de H_0 a qualquer n.s. $\alpha_0 \leq valor-p$;
  - a rejeição de H_0 a qualquer n.s. $\alpha_0 > valor-p$.

4. Segundo a ANSR, no ano de 2019 ocorreram 135 058 acidentes rodoviários em Portugal Continental. Na tabela abaixo indica-se o número de acidentes por dia da semana.

Dia da semana	segunda	terça	quarta	quinta	sexta	sábado	domingo
Número de acidentes ocorridos	19294	19306	19320	o_4	o_5	o_6	o_7

Averigue se os dados permitem afirmar que os acidentes rodoviários em Portugal Continental se distribuem uniformemente ao longo dos sete dias da semana. Decida com base no valor-p aproximado.

Assinale a sua resposta com uma cruz.

- ⦿ Rejeita-se H_0 a 1%, 5% e 10%.

⦿ Rejeita-se H_0 a 5% e 10% e não se rejeita H_0 a 1%.

⦿ Rejeita-se H_0 a 10% e não se rejeita H_0 a 1% e 5%.

⦿ Não se rejeita H_0 a 1%, 5% e 10%.

- **V.a. de interesse**
  X = dia da semana em que ocorre acidente (1 = segunda, ..., 7 = domingo)

- **Hipóteses**
  $H_0 : X \sim$ uniforme discreta$(\{1,\ldots,7\})$
  $H_1 : X \not\sim$ uniforme discreta$(\{1,\ldots,7\})$

- **Estatística de teste**
  $T = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \overset{a}{\sim}_{H_0} \chi_{(k-\beta-1)}$,
  onde:

  k = No. de classes = 7
  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 em H_0 se conjectura uma distribuição específica.]

- **Frequências absolutas esperadas sob** H_0
  Atendendo à dimensão da amostra n e ao facto de a f.p. conjecturada ser dada por $P(X = x \mid H_0) = \frac{1}{7}$, $x = 1,\ldots,7$, as frequências absolutas esperadas sob H_0 são, para $i = 1,\ldots,7$, iguais a:

  $$E_i \quad = \quad n \times p_i^0 = n \times \frac{1}{7}.$$

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

- **Região de rejeição de** H_0 (para valores de T)
  Tratando-se de um teste de ajustamento, a região de rejeição de H_0 escrita para valores de T é o intervalo à direita $W = (c, +\infty)$.

- **Decisão (com base no valor-p)**
  Como as classes são equiprováveis, o valor observado da estatí?tica de teste é igual a

  $$t \quad = \quad \sum_{i=1}^{k} \frac{(o_i - E_i)^2}{E_i} \equiv \frac{k}{n}\left(\sum_{i=1}^{k} {o_i}^2\right) - n.$$

  Além disso, importa referir que

  $$valor - p \quad = \quad P(T > t \mid H_0) \simeq 1 - F_{\chi^2_{(k-1)}}(t).$$

  Devemos escolher a resposta, das quatro apresentadas acima, que se coaduna com:
  - a não rejeição de H_0 a qualquer n.s. $\alpha_0 \leq valor - p$;
  - a rejeição de H_0 a qualquer n.s. $\alpha_0 > valor - p$.

5. Um estudo pretende avaliar a relação entre a percentagem de gordura corporal (Y) e aidade (x). Recolheu-se uma amostra casual de n indivíduos com idades compreendidas entre os 23 e 61 anos que conduziu a:

$$\textstyle\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i, \quad \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i^2, \quad \sum_{i=1}^{n} y_i = \sum_{i=1}^{n} y_i, \quad \sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{n} y_i^2, \quad \sum_{i=1}^{n} x_i y_i = \sum_{i=1}^{n} x_i\, y_i.$$

Confronte as hipóteses $H_0 : E(Y \mid x = 53) = a$ e $H_1 : E(Y \mid x = 53) \neq a$. Decida com base no valor-p.

Assinale a sua resposta com uma cruz.

- ⦿ Rejeita-se H_0 a 1%, 5% e 10%.
- ⦿ Rejeita-se H_0 a 5% e 10% e não se rejeita H_0 a 1%.
- ⦿ Rejeita-se H_0 a 10% e não se rejeita H_0 a 1% e 5%.
- ⦿ Não se rejeita H_0 a 1%, 5% e 10%.

- **Hipóteses de trabalho**

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

- **Hipóteses**

  $H_0 : E(Y \mid x = 53) = \beta_0 + \beta_1 \times 53 = a$

  $H_1 : E(Y \mid x = 53) = \beta_0 + \beta_1 \times 53 \neq a$

- **Estatística de teste**

$$T = \frac{(\hat{\beta}_0 + \hat{\beta}_1 \times 53) - a}{\sqrt{\hat{\sigma}^2 \times \left(\frac{1}{n} + \frac{(\bar{x}-53)^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}\right)}} \sim_{H_0} t_{(n-2)}$$

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

  Dado que o teste é bilateral, a região de rejeição de H_0 é a reunião de intervalos $W = (-\infty, -c) \cup (c, +\infty)$.

- **Valor observado da estatística de teste**

$$t = \frac{(\hat{\beta}_0 + \hat{\beta}_1 \times 53) - a}{\sqrt{\hat{\sigma}^2 \left(\frac{1}{n} + \frac{(\bar{x}-53)^2}{\sum_{i=1}^n x_i^2 - n\left(\frac{\sum_{i=1}^n x_i}{n}\right)^2}\right)}}$$

  onde

$$\begin{aligned}
\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \times \bar{x} \\
&= \bar{y} - \frac{\sum_{i=1}^n x_i y_i - n \times \frac{\sum_{i=1}^n x_i}{n} \times \frac{\sum_{i=1}^n y_i}{n}}{\sum_{i=1}^n x_i^2 - n \times \left(\frac{\sum_{i=1}^n x_i}{n}\right)^2} \times \bar{x} \\
\hat{\beta}_1 &= \frac{\sum_{i=1}^n x_i y_i - n\left(\frac{\sum_{i=1}^n x_i}{n}\right)\left(\frac{\sum_{i=1}^n y_i}{n}\right)}{\sum_{i=1}^n x_i^2 - n\left(\frac{\sum_{i=1}^n x_i}{n}\right)^2} \\
\hat{\sigma}^2 &= \frac{1}{n-2}\left[\left(\sum_{i=1}^n y_i^2 - n\,\bar{y}^2\right) - (\hat{\beta}_1)^2\left(\sum_{i=1}^n x_i^2 - n\,\bar{x}^2\right)\right] \\
&= \frac{1}{n-2}\left\{\left[\sum_{i=1}^n y_i^2 - n\left(\frac{\sum_{i=1}^n y_i}{n}\right)^2\right] - (\hat{\beta}_1)^2\left[\sum_{i=1}^n x_i^2 - n\left(\frac{\sum_{i=1}^n x_i}{n}\right)^2\right]\right\}.
\end{aligned}$$

- **Decisão (com base no valor-p)**

$$valor-p = 2 \times \left[1 - F_{t_{(n-2)}}(|t|)\right].$$

  Devemos escolher a resposta, das quatro apresentadas acima, que se coaduna com:

  - a não rejeição de H_0 a qualquer n.s. $\alpha_0 \leq valor-p$;
  - a rejeição de H_0 a qualquer n.s. $\alpha_0 > valor-p$.