

Duração: 120 minutos

Exame Época Recurso - (b)

**Justifique convenientemente todas as respostas**

**Pergunta 1**

2 valores

Um teste para identificação de patógenos humanos em amostras de água conduz a resultado correto com probabilidade 0.99 em amostras com presença de patógenos humanos e 0.95 em amostras sem presença de patógenos humanos.

Considerando que o teste é aplicado a uma amostra de água escolhida ao acaso de uma população em que 1% das amostras exibem presença de patógenos humanos e o teste indica a presença de patógenos humanos na amostra, calcule a probabilidade de essa amostra conter efetivamente patógenos humanos.

• **Acontecimentos e probabilidades para uma amostra de água escolhida ao acaso**

| Acontecimento  | Probabilidade                 |
|--|-------------------------------|
| $A = \text{"a amostra de água contém patógenos humanos"}$                        | $P(A) = 0.01$                 |
| $I = \text{"o teste indica a presença de patógenos humanos na amostra de água"}$ | $P(I) = ?$                    |
|  | $P(I   A) = 0.99$             |
|  | $P(\bar{I}   \bar{A}) = 0.95$ |

• **Cálculo da probabilidade pedida**

$$\begin{aligned}
 P(A | I) &= \frac{P(A) \times P(I | A)}{P(A) \times P(I | A) + P(\bar{A}) \times P(I | \bar{A})} && \text{(teorema de Bayes)} \\
 &= \frac{P(A) \times P(I | A)}{P(A) \times P(I | A) + [1 - P(\bar{A})] \times [1 - P(\bar{I} | \bar{A})]} \\
 &= \frac{0.01 \times 0.99}{0.01 \times 0.99 + [1 - 0.01] \times [1 - 0.95]} \\
 &= \frac{1}{6} = 0.1(6).
 \end{aligned}$$

**Pergunta 2**

2 valores

Numa determinada população, após contacto com um vírus, cada indivíduo fica infetado com probabilidade 0.7. Suponha que as infeções ocorrem de forma independente.

- (a) Qual é a probabilidade de se observarem mais de 7 infeções num total de 15 contactos, quando ocorre pelo menos uma infeção?

- **V.a. de interesse**

$X = \text{"número de infecções num total de 15 contactos"}$ .

$$X \sim \text{binomial}(n = 15, p = 0.7).$$

- **Probabilidade pedida**

$$\begin{aligned} P(X > 7 | X \geq 1) &= \frac{1 - F_X(7)}{1 - 0.3^{15}} \\ &= \frac{F_{Bin(15,0.3)}(7)}{1 - 0.3^{15}} \\ &= \frac{0.95}{1 - 0.3^{15}} \\ &= 0.95. \end{aligned}$$

(b) Logo após ter ocorrido a primeira infecção, qual a probabilidade de serem necessários mais de 3 contactos para a segunda infecção?

- **Probabilidade pedida**

O número de contactos até infecção é uma variável aleatória geométrica(0.7).

Dada a independência, a probabilidade de serem necessários mais de 3 contactos para a segunda infecção logo após ter ocorrido a primeira infecção é igual a

$$\begin{aligned} P(\text{serem nec. mais de 3 contactos para a 1ª infecção}) &= P(\text{geométrica}(0.7) > 3) \\ &= 1 - 0.7(1 + 0.3 + 0.3^2) \\ &= 0.027. \end{aligned}$$

**Pergunta 3**

2 valores

A proporção de impurezas num produto químico é uma variável  $Y$  com função de densidade

$$f(y) = \begin{cases} 12y^2(1-y), & 0 \leq y \leq 1, \\ 0, & \text{outros casos.} \end{cases}$$

Determine a probabilidade de a proporção de impurezas ser superior a três vezes o seu desvio padrão.

- **V.a. de interesse**

$Y = \text{"proporção impurezas em um produto químico"}$

- **Variância de  $Y$**

$$\begin{aligned} V(Y) &= E(Y^2) - E^2(Y) \\ &= \int_0^1 y^2 12y^2(1-y) dy - \left( \int_0^1 y 12y^2(1-y) dy \right)^2 \\ &= 12 \int_0^1 (y^4 - y^5) dy - \left( 12 \int_0^1 (y^3 - y^4) dy \right)^2 \\ &= \frac{12}{30} - \left( \frac{12}{20} \right)^2 \\ &= \frac{1}{25} \\ &= 0.04. \end{aligned}$$

$$\text{Desvio padrão: } \sqrt{V(Y)} = \sqrt{\frac{1}{25}} = \frac{1}{5}.$$

- **Probabilidade pedida**

$$\begin{aligned} P\left(Y > 3 \times \frac{1}{5}\right) &= \int_{\frac{3}{5}}^1 12y^2(1-y) dy \\ &= 0.5248. \end{aligned}$$

**Pergunta 4**

2 valores

Os números de falhas em duas máquinas ( $A$  e  $B$ ) a trabalhar em simultâneo distribuem-se uniformemente no conjunto:  $\{(x, y) \in \{0, 1, 2\} \times \{0, 1, 2\} : x \leq y\}$  (recorde que  $A \times B = \{(x, y) : x \in A \text{ e } y \in B\}$ ).

(a) Calcule a correlação entre o número de falhas que ocorrem nas duas máquinas e comente o seu valor.

- **Distribuição conjunta**

$P(X = x, Y = y) = 1/6$ ,  $(x, y) \in D$ , com  $D = \{(x, y) \in \{0, 1, 2\} \times \{0, 1, 2\} : x \leq y\}$  e 0 outros casos.

- **Distribuições marginais de  $X$  e  $Y$**

$$P(X = x) = \begin{cases} 1/2, & x = 0, \\ 1/3, & x = 1, \\ 1/6, & x = 2, \\ 0, & \text{outros casos.} \end{cases}$$

$$P(Y = y) = \begin{cases} 1/6, & y = 0, \\ 1/3, & y = 1, \\ 1/2, & y = 2, \\ 0, & \text{outros casos.} \end{cases}$$

- **Valor médio e variância de  $X$**

$$\begin{aligned} E(X) &= 0 * \frac{1}{2} + 1 * \frac{1}{3} + 2 * \frac{1}{6} \\ &= \frac{2}{3}. \end{aligned}$$

$$\begin{aligned} E(X^2) &= 0 * \frac{1}{2} + 1 * \frac{1}{3} + 2^2 * \frac{1}{6} \\ &= 1. \end{aligned}$$

$$\begin{aligned} V(X) &= 1 - (2/3)^2 \\ &= 5/9. \end{aligned}$$

• **Valor médio e variância de  $Y$**

$$\begin{aligned} E(Y) &= 0 * \frac{1}{6} + 1 * \frac{1}{3} + 2 * \frac{1}{2} \\ &= \frac{4}{3} \end{aligned}$$

$$\begin{aligned} E(Y^2) &= 0 * \frac{1}{6} + 1 * \frac{1}{3} + 2^2 * \frac{1}{2} \\ &= \frac{7}{3} \end{aligned}$$

$$\begin{aligned} V(Y) &= \frac{7}{3} - (4/3)^2 \\ &= \frac{5}{9} \end{aligned}$$

• **Cálculo do valor esperado de  $XY$**

$$\begin{aligned} E(XY) &= \frac{1+2+4}{6} \\ &= \frac{7}{6} \end{aligned}$$

• **Cálculo da correlação ente  $X$  e  $Y$**

$$\begin{aligned} Cov(X, Y) &= 7/6 - (4/3) \times (2/3) \\ &= 15/54 \\ &= 5/18. \end{aligned}$$

$$\begin{aligned} Corr(X, Y) &= \frac{5/18}{\sqrt{5/9}\sqrt{5/9}} \\ &= \frac{1}{2}. \end{aligned}$$

Correlação moderada positiva indicando a existência de alguma dependência linear entre o número de falhas das duas máquinas.

- (b) Obtenha o número médio de falhas de  $A$  quando em  $B$  ocorre uma falha, e o número médio total de falhas quando em  $B$  ocorre uma falha.

$$\text{Sendo } P(X = x|Y = 1) = \begin{cases} 1/2, & x = 0, \\ 1/2, & x = 1, \\ 0, & \text{outros casos,} \end{cases}$$

logo

$$\begin{aligned} E[X|Y = 1] &= 1 * \frac{1}{2} \\ &= 0.5. \end{aligned}$$

$$\begin{aligned} E(X + Y|Y = 1) &= E(X|Y = 1) + E(Y|Y = 1) \\ &= 0.5 + 1 \\ &= 1.5. \end{aligned}$$

- (c) São as variáveis aleatórias independentes?

Não, por exemplo  $P(X = 0, Y = 0) = 1/6 \neq P(X = 0) \times P(Y = 0) = 1/2 \times 1/6$ .

Um jogador lança um dado equilibrado, este ganha um ponto se sair um número par e perde um ponto se sair um número ímpar. Ao fim de 50 jogadas qual é a probabilidade aproximada do número total de pontos obtidos pelo jogador ser superior a 2.

- **V.a.  $X_i$**

Seja  $X_i$  uma va que assume o valor 1 se sair face par e -1 se sair face ímpar no  $i$ -ésimo lançamento do dado equilibrado,  $i = 1, \dots, n$  com  $n = 50$ .

- **Valor esperado de  $X_i$**

$$E(X_i) = 0.$$

- **Variância de  $X_i$**

$$\text{Var}(X_i) = E(X_i)^2 = (-1)^2/2 + 1^2/2 = 1.$$

- **V.a. de interesse**

$S_n = \sum_{i=1}^n X_i$ , número total de pontos em  $n$  jogadas.

- **Valor esperado e variância de  $S_n$**

$E(S_n) = E(\sum_{i=1}^n X_i) = \sum_{i=1}^n E(X_i) = n \times 0 = 0$ , porque as va são identicamente distribuídas.

$\text{Var}(S_n) = \text{Var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \text{Var}(X_i) = n \times 1 = 50$ , porque as variáveis aleatórias iid.

- **Distribuição aproximada de  $S_n$**

$$\frac{S_n - E(S_n)}{\sqrt{\text{Var}(S_n)}} = \frac{S_n}{\sqrt{n}} \stackrel{a}{\approx} \text{normal}(0, 1).$$

- **Valor aproximado da probabilidade pedida**

De acordo com o teorema do limite central, temos

$$\begin{aligned} P\left(\sum_{i=1}^{50} X_i \geq 2\right) &= 1 - P\left(\sum_{i=1}^{50} X_i \leq 1\right) \\ &\stackrel{TLC}{\approx} 1 - \Phi\left(\frac{1}{\sqrt{50}}\right) \\ &\approx 1 - \Phi(0.14) \\ &= 1 - 0.5557 \\ &= 0.4443. \end{aligned}$$

Admita que  $X$  é uma variável aleatória com função densidade de probabilidade

$$f_X(x; \theta) = \begin{cases} \frac{3}{\theta} x^2 e^{-x^3/\theta}, & x > 0 \\ 0, & \text{outros casos} \end{cases}$$

com  $\theta > 0$  desconhecido. A concretização de uma amostra aleatória de dimensão 4 proveniente de  $X$ , conduziu a  $x_1 = 1.5$ ,  $x_2 = 3.5$ ,  $x_3 = 2.9$ ,  $x_4 = 4.7$ . Obtenha a estimativa de máxima verosimilhança do parâmetro  $\theta$ .

- Seja  $\underline{x} = (x_1, x_2, x_3, x_4)$  uma amostra aleatória de dimensão 3 proveniente da população  $X$ .

- **Obtenção do estimador da máxima verosimilhança de  $\theta$**

Passo 1 - Função de verosimilhança

$$\begin{aligned} L(\theta|\underline{x}) &= f_{\underline{X}}(\underline{x}) \\ &\stackrel{X_i \text{ indep}}{=} \prod_{i=1}^4 f_{X_i}(x_i) \\ &\stackrel{X_i \sim X}{=} \prod_{i=1}^4 f_X(x_i) \\ &= \prod_{i=1}^4 \frac{3}{\theta} x_i^2 e^{-x_i^3/\theta} \\ &= \frac{3^4}{\theta^4} \left( \prod_{i=1}^4 x_i^2 \right) e^{-\frac{1}{\theta} \sum_{i=1}^4 x_i^3} \end{aligned}$$

Passo 2 - Função de log-verosimilhança

$$\ln L(\theta|\underline{x}) = 4 \ln 3 - 4 \ln \theta + \ln \left( \prod_{i=1}^4 x_i^2 \right) - \frac{1}{\theta} \sum_{i=1}^4 x_i^3$$

Passo 3 - Maximização

A estimativa de MV de  $\theta$  é doravante representada por  $\hat{\theta}$  e

$$\hat{\theta} : \begin{cases} \frac{d \ln L(\theta|\underline{x})}{d\theta} \Big|_{\theta=\hat{\theta}} = 0 & \text{(ponto de estacionariedade)} \\ \frac{d^2 \ln L(\theta|\underline{x})}{d\theta^2} \Big|_{\theta=\hat{\theta}} < 0 & \text{(ponto de máximo)} \end{cases}$$

$$\frac{d \ln L(\theta|\underline{x})}{d\theta} = 0 \Leftrightarrow -\frac{4}{\theta} + \frac{\sum_{i=1}^4 x_i^3}{\theta^2} = 0 \Leftrightarrow \frac{\sum_{i=1}^4 x_i^3}{\theta^2} = \frac{4}{\theta} \Rightarrow \hat{\theta} = \frac{\sum_{i=1}^4 x_i^3}{4}$$

$$\frac{d^2 \ln L(\theta|\underline{x})}{d\theta^2} \Big|_{\theta=\hat{\theta}} = -0.0021 < 0, \text{ (proposição verdadeira)}$$

Passo 4 - Estimativa de MV de  $\theta$ .

$$\begin{aligned} \hat{\theta} &= \frac{\sum_{i=1}^4 x_i^3}{4} \\ &= \frac{174.462}{4} \\ &= 43.61. \end{aligned}$$

### Pergunta 7

2 valores

Numa investigação sobre um novo processo de refinação de um certo minério está-se a analisar o teor de lítio no produto refinado, em percentagem. Numa amostra aleatória de tamanho 12 observou-se  $\sum_{i=1}^{12} x_i = 1140$  e  $\sum_{i=1}^{12} x_i^2 = 108900$ .

Admitindo que o teor de lítio após refinação tem distribuição normal, determine um intervalo de confiança para a variância do teor de lítio a um nível de confiança de 95%.

- **V.a. de interesse**

$X =$  "teor de lítio no produto refinado, em percentagem".

- **Situação**  $X \sim \text{normal}(\mu, \sigma^2)$ .

- **Seleção da variável aleatória fulcral para  $\sigma^2$**

$$Z = \frac{11S^2}{\sigma^2} \sim \chi^2_{(11)}$$

- **Obtenção dos quantis de probabilidade**

$$a_\alpha = F_{\chi^2_{(11)}}^{-1}(0.025) = 3.816;$$

$$b_\alpha = F_{\chi^2_{(11)}}^{-1}(0.975) = 21.92.$$

- **Inversão da desigualdade  $a_\alpha \leq Z \leq b_\alpha$**

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

$\Leftrightarrow$

$$P\left(3.816 \leq \frac{11S^2}{\sigma^2} \leq 21.92\right) = 0.95$$

$\Leftrightarrow$

$$P\left(\frac{11S^2}{21.92} \leq \sigma^2 \leq \frac{11S^2}{3.816}\right) = 0.95$$

- **Concretização**

$$\begin{aligned} IC_{95\%}(\sigma^2) &= \left[ \frac{11s^2}{21.92}, \frac{11s^2}{3.816} \right] \\ &= \left[ \frac{11 \times 54.55}{21.92}, \frac{11 \times 54.55}{3.816} \right] \\ &= [27.37, 157.24]. \end{aligned}$$

|                   |           |
|-------------------|-----------|
| <b>Pergunta 8</b> | 2 valores |
|-------------------|-----------|

A verminose (doença causada por vermes) tem impacto na produção de leite dos rebanhos. O veterinário de uma exploração agrícola afirma que a proporção de animais afetados é de 10%. Um exame a 100 cabeças do rebanho desta exploração agrícola, retiradas ao acaso, indicou 12 delas com verminose. Verifique se a afirmação do veterinário pode ser considerada válida ou se, pelo contrário, a proporção de animais afetados é superior a esse valor.

- **V.a. de interesse**

$$X = \text{"indicador relativo à presença da verminose"} = \begin{cases} 1 & \text{com verminose} \\ 0 & \text{outros casos} \end{cases}$$

- **Situação**  $X \sim \text{bernoulli}(p)$ .

- **Hipóteses**

$$H_0 : p = p_0 = 0.10$$

$$H_1 : p = p_0 > 0.10$$

- **Estatística de teste**

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

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

O teste é unilateral, logo a região de rejeição de  $H_0$  é do tipo  $W = (c, \infty)$ .

- **Decisão (com base no valor-p)** Atendendo a que o valor observado da estatística de teste:

$$t = \frac{0.12 - 0.1}{\sqrt{0.1 \times (1 - 0.1)/100}} = 0.66666$$

$$\text{valor-p} \approx 1 - \Phi(0.67) = (1 - 0.7486) = 0.2514$$

pelo que  $H_0$  não é rejeitada para níveis de significância usuais.

|                   |           |
|-------------------|-----------|
| <b>Pergunta 9</b> | 2 valores |
|-------------------|-----------|

Num estudo para avaliar as preferências entre os utilizadores de cinco tipos de passadeiras motorizadas, foram inquiridos 410 utilizadores, tendo sido obtidos os resultados seguintes: 30 mostraram preferência pela passadeira do tipo A; 21 pela de tipo B; 110 pela de tipo C; 35 pela de tipo D; e 214 pela de tipo E. Vários treinadores desportivos sustentam que, entre os utilizadores de passadeiras motorizadas, a probabilidade de escolha de uma passadeira do tipo A é metade da probabilidade de escolha de uma passadeira do tipo C; o triplo da de tipo B; o dobro da do tipo D; e um terço da do tipo E. Averigue se a opinião dos treinadores desportivos é consistente com os resultados do estudo ao nível de significância de 1%.

Seja  $X$  tipo preferido de passadeira motorizada pelo utilizador inquirido

$$P(X = x) = \begin{cases} p_i, & i = 1, 2, 3, 4, 5, \\ 0, & \text{outros casos.} \end{cases}$$

**Hipóteses:**

$$H_0 : p_i = p_i^0, i = 1, \dots, 5, \text{ vs } H_1 : p_i \neq p_i^0, \text{ para algum } i, \text{ onde}$$

$$\begin{cases} p_1^0 = 3 \times p_2^0 \\ p_1^0 = \frac{1}{2} \times p_3^0 \\ p_1^0 = 2 \times p_4^0 \\ p_1^0 = \frac{1}{3} \times p_5^0 \\ p_1^0 + p_2^0 + p_3^0 + p_4^0 + p_5^0 = 1 \end{cases}$$

o que implica que  $p_1^0 = \frac{6}{41}$ ,  $p_2^0 = \frac{2}{41}$ ,  $p_3^0 = \frac{12}{41}$ ,  $p_4^0 = \frac{3}{41}$  e  $p_5^0 = \frac{18}{41}$ .



**Estatística de teste:**

$$T = \sum_{i=1}^5 \frac{(O_i - E_i)^2}{E_i} \underset{a}{\sim} \chi_4^2.$$

**Nível de significância:**  $\alpha_0 = 1\%$

**Região de rejeição de  $H_0$** 

Ao efetuarmos um teste de ajustamento do qui-quadrado a região de rejeição de  $H_0$  é um intervalo à direita  $W = (c, \infty)$ , onde

$$c = F_{\chi_4^2}^{-1}(1 - \alpha_0) = F_{\chi_4^2}^{-1}(0.99) = 13.28.$$

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

As frequências absolutas esperadas sob  $H_0$  são dadas por  $E_i = 345 \times p_i^0$ ,  $i = 1, \dots, 5$ , e iguais a

$$E_1 = 410 \times \frac{6}{41} = 60,$$

$$E_2 = 410 \times \frac{2}{41} = 20,$$

$$E_3 = 410 \times \frac{12}{41} = 120,$$

$$E_4 = 410 \times \frac{3}{41} = 30,$$

$$E_5 = 410 \times \frac{18}{41} = 180.$$

**Cálculo do valor observado da estatística de teste**

$$\begin{aligned} t &= \frac{(30 - 60)^2}{60} + \frac{(21 - 20)^2}{20} + \frac{(110 - 120)^2}{120} + \frac{(35 - 30)^2}{30} + \frac{(214 - 180)^2}{180} \\ &= 15.00 + 0.05 + 0.83 + 0.83 + 6.42 \\ &= 23.13. \end{aligned}$$

Como  $t \in (13.28, +\infty)$ , rejeita-se  $H_0$  ao nível de significância 5%.

**Pergunta 10**

2 valores

Considere uma experiência em que se analisa a octanagem da gasolina em função da adição de um novo aditivo. Para isso, foram realizados ensaios com diferentes % de quantidade de aditivo ( $x$ , g/L), e registado o índice de octanagem ( $y$ ), tendo-se obtido os seguintes valores:

| $x$ (em %)              | 1    | 2    | 3    | 4    | 5    | 6  |
|-------------------------|------|------|------|------|------|----|
| índice de octanagem $y$ | 80.5 | 81.6 | 82.1 | 83.7 | 83.9 | 85 |

$$\sum_{i=1}^6 x_i = 21, \quad \sum_{i=1}^6 x_i^2 = 91, \quad \sum_{i=1}^6 y_i = 496.8, \quad \sum_{i=1}^6 x_i y_i = 1754.30.$$

Admita que as variáveis estão relacionadas de acordo com um modelo de regressão linear simples  $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ , com as hipóteses de trabalho usuais. Estime os parâmetros da reta e interprete a estimativa de  $\beta_1$ .

### Estimativas de MQ dos parâmetros desconhecidos $\beta_0$ e $\beta_1$

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum_{i=1}^6 x_i y_i - 6 * \bar{x} \bar{y}}{\sum_{i=1}^6 x_i^2 - 6 * \bar{x}^2} \\ &= 0.886.\end{aligned}$$

$$\begin{aligned}\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 * \bar{x} \\ &= 80.55,\end{aligned}$$

pelos que se estima que o índice da octanagem aumente em média 0.886 por cada g/L extra de aditivo adicionado, para valores de aditivo adicionado entre 1 e 6.