

Duração: 120 minutos

Exame de Época Normal - Versão A

- As questões de escolha única 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 única têm de ser assinaladas na folha de instruções distribuída.
- Justificar convenientemente as respostas a todas as restantes questões.

Pergunta 1 (corrigida)

2 valores

Para avaliar a eficácia do uso de máscara como proteção a uma doença, durante determinado período observou-se que na população em estudo: 45% usou máscara, 30% ficou infetada com a doença, 15% usou máscara e ficou infetada com a doença e 40% não usou máscara nem ficou infetada.

Considere um indivíduo escolhido ao acaso da população. Calcule as probabilidades, de que um indivíduo que tenha usado máscara fique doente, e de que um indivíduo que não tenha usado máscara fique doente.

Acontecimentos e probabilidades:

M = indivíduo usa máscara $P(M) = 0.45$ $P(M \cap D) = 0.15$
 B = indivíduo é infetado com a doença $P(D) = 0.3$ $P(\bar{M} \cap \bar{D}) = 0.4$

- $P(D|M) = \frac{P(D \cap M)}{P(M)} = \frac{0.15}{0.45} = \frac{1}{3}$
- $P(D|\bar{M}) = \frac{P(\bar{M} \cap D)}{P(\bar{M})} = \frac{P(\bar{M}) - P(\bar{M} \cap \bar{D})}{P(\bar{M})} = 1 - \frac{0.4}{1-0.45} = \frac{3}{11}$.

Pergunta 2

2 valores

Num estudo suficientemente abrangente sobre o consumo de um medicamento por indivíduos que desejam emagrecer, os resultados foram codificados do seguinte modo: consumo = 0 (nunca), 1 (ocasional), 2 (frequente). Admite-se que a função de distribuição do tipo de consumo (X) é definida por

$$F_X(x) = 0(x < 0), \quad F_X(x) = 0.45(0 \leq x < 1), \quad F_X(x) = 0.8(1 \leq x < 2), \quad F_X(x) = 1(x \geq 2).$$

Encontre a função (massa) de probabilidade de X e averigue se o desvio padrão do tipo de consumo é maior do que o seu valor esperado.

Função (massa) de probabilidade de X :

$$f_X(x) \equiv P(X = x) = P(X \leq x) - P(X < x) = F_X(x) - F_X(x^-) = \begin{cases} 0.45, & x = 0, \\ 0.35, & x = 1, \\ 0.2, & x = 2, \\ 0, & \text{caso contrário.} \end{cases}$$

Valor esperado, $E(X)$, e desvio padrão, $DP(X)$, de X :

$$E(X) = \sum_x x f_X(x) = 0 \times 0.45 + 1 \times 0.35 + 2 \times 0.2 = 0.75$$

$$E(X^2) = \sum_x x^2 f_X(x) = 0 \times 0.45 + 1 \times 0.35 + 4 \times 0.2 = 1.15$$

$$Var(X) = E(X^2) - E(X)^2 = 1.15 - 0.75^2 \approx 0.5875$$

$$DP(X) = \sqrt{Var(X)} = \sqrt{0.5875} \approx 0.7665$$

$$\Rightarrow DP(X) > E(X)$$

Pergunta 3

2 valores

Os autocarros da carreira 799 chegam a uma paragem específica com intervalos de 20 minutos, a partir das 07:00, ou seja, chegam às 07:00, 07:20, 07:40 e assim sucessivamente.

Se um passageiro chega à paragem num instante uniformemente distribuído entre as 07:00 e as 07:20, determine a probabilidade condicional de que ele espere mais do que 10 minutos por um autocarro, sabendo que ele vai esperar menos de 15 minutos.

A: 1 B: 3/4 C: 1/3 D: 1/4

X : Chegada do passageiro à paragem do autocarro 799.

$X \sim \text{Uniforme}(0, 20)$, i.e., $f_X(x) = \frac{1}{20}$, $0 < x < 20$, e 0, no caso contrário.

$$P(X \leq 10 | X > 5) = \frac{P(5 < X \leq 10)}{P(X > 5)} = \frac{\int_5^{10} \frac{1}{20} dx}{\int_5^{20} \frac{1}{20} dx} = \frac{(10-5)/20}{(20-5)/20} = \frac{1}{3}$$

Pergunta 4

2 valores

Considere uma variável aleatória discreta X tal que $P(X = 1) = 0.2$, $P(X = 2) = 0.5$ e $P(X = 3) = 0.3$.

Para cada valor de X , $x = 1, 2, 3$: $P(Y = 2 | X = x) = q_x$ e $P(Y = -2 | X = x) = 1 - q_x$, onde $q_1 = 0.1$, $q_2 = 0.5$, e $q_3 = 0.7$.

Calcule o valor esperado de $X | Y = -2$.

- Função de probabilidade conjunta de (X, Y) :

$$P(X = x, Y = y) = P(X = x)P(Y = y | X = x) = \begin{cases} P(X = x) q_x, & y = 2, \\ P(X = x) (1 - q_x), & y = -2. \end{cases}$$

Ou seja,

x	$P(X = x, Y = 2)$	$P(X = x, Y = -2)$
1	$0.2 \times 0.1 = 0.02$	$0.2 \times 0.9 = 0.18$
2	$0.5 \times 0.5 = 0.25$	$0.5 \times 0.5 = 0.25$
3	$0.3 \times 0.7 = 0.21$	$0.3 \times 0.3 = 0.09$

- Função de probabilidade condicional de X dado $Y = -2$:

$$P(X = x | Y = -2) = \frac{P(X = x, Y = -2)}{P(Y = -2)} = \begin{cases} \frac{0.18}{0.52} \approx 0.3462, & x = 1, \\ \frac{0.25}{0.52} \approx 0.4808, & x = 2, \\ \frac{0.09}{0.52} \approx 0.1731, & x = 3. \end{cases}$$

onde a função de probabilidade marginal de $Y = -2$ é dada por

$$P(Y = -2) = \sum_x P(X = x, Y = -2) = 0.18 + 0.25 + 0.09 = 0.52$$

- Valor esperado de $X | Y = -2$:

$$\begin{aligned} E[X | Y = -2] &= \sum_x x P(X = x | Y = -2) \\ &= 1 \times 0.3462 + 2 \times 0.4808 + 3 \times 0.1731 \\ &= 1.827. \end{aligned}$$

Pergunta 5

2 valores

Na produção de componentes eletrônicas a tensão (em volts) de saída depende de duas etapas: uma primeira de ajuste inicial de tensão, representada pela variável aleatória $X \sim N(\mu = 10, \sigma^2 = 4)$, e numa segunda fase um ajuste fino, representada pela variável aleatória $Y \sim N(\mu = 5, \sigma^2 = 9)$. Considere X e Y variáveis aleatórias independentes. A tensão final da componente é dada por $W = 3X - 2Y$. A componente só é aprovada se $W > 0$.

Calcule a probabilidade de uma componente ser aprovada nesta linha de produção.

A: 0.991 B: 0.900 C: 0.832 D: 0.055

- $E(W) = 3E(X) - 2E(Y) = 3 \times 10 - 2 \times 5 = 20$
 - $Var(W) \stackrel{i.i.d.}{=} Var(3X) + Var(2Y) = 9Var(X) + 4Var(Y) = 9 \times 4 + 4 \times 9 = 72$
 - Sendo W uma combinação linear de variáveis aleatórias normais independentes, $W \sim N(20, 72)$.
- $$\therefore P(W > 0) = 1 - P\left(\frac{W - E(W)}{\sqrt{Var(W)}} \leq \frac{0 - 20}{\sqrt{72}}\right) = 1 - F_{N(0,1)}(-2.357) \equiv 1 - \Phi(-2.357) = \Phi(2.357) \approx 0.991.$$

Pergunta 6

2 valores

Num estudo prévio à instalação de uma central eólica em determinada região foram colocadas algumas questões a um meteorologista acerca da velocidade mínima (X) a que o vento sopra diariamente nessa região. O meteorologista suspeita que a função de densidade desta variável aleatória, em km/h, é:

$$f_X(x) = \frac{1}{\delta} \exp\left(-\frac{x-2}{\delta}\right), \quad x \geq 2, \quad \delta > 0.$$

Deduz a estimativa de máxima verosimilhança de δ , tendo em conta que 50 observações diárias de X conduziram a $\sum_{i=1}^{50} x_i = 150$.

A: 1 B: 2 C: 3 D: 4

- Seja $(X_1, \dots, X_n)$ uma amostra aleatória de X com função de densidade $f_X(x)$ e sua concretização $n = 50$ e $\sum_{i=1}^{50} x_i = 150$.
- Método da máxima verosimilhança (estimação de δ):

1. Funções de verosimilhança e de log-verosimilhança:

$$L(\delta|\mathbf{x}) = f_{\mathbf{X}}(\mathbf{x}|\delta) \stackrel{i.i.d.}{=} \prod_{i=1}^n f_X(x_i|\delta) = \prod_{i=1}^n \frac{1}{\delta} \exp\left(-\frac{x_i-2}{\delta}\right) = \left(\frac{1}{\delta}\right)^n \exp\left(-\frac{\sum_{i=1}^n x_i - 2n}{\delta}\right)$$

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

2. Maximização de $\ln L(\delta|\mathbf{x})$, com $\hat{\delta}$ a denotar a estimativa de MV de δ :

$$\left. \frac{d}{d\delta} \ln L(\delta|\mathbf{x}) \right|_{\delta=\hat{\delta}} = 0 \quad \Leftrightarrow \quad -\frac{n}{\hat{\delta}} + \frac{\sum_{i=1}^n x_i - 2n}{\hat{\delta}^2} = 0 \quad \Rightarrow \quad \hat{\delta} = \frac{1}{n} \left(\sum_{i=1}^n x_i - 2n \right) = \bar{x} - 2$$

$$\left. \frac{d^2}{d\delta^2} \ln L(\delta|\mathbf{x}) \right|_{\delta=\hat{\delta}} < 0 \quad \Rightarrow \quad -\frac{n}{\hat{\delta}^2} \left(-1 + \frac{2(\sum_{i=1}^n x_i - 2n)}{n\hat{\delta}} \right) = -\frac{n}{\hat{\delta}^2} < 0 \quad (\hat{\delta} \text{ é ponto de máximo})$$

3. Estimador de máxima verosimilhança (EMV) de δ e sua estimativa (emv):

$$\text{EMV}(\delta) = \bar{X} - 2 \quad \text{e} \quad \text{emv}(\delta) = \bar{x} - 2 = \frac{150}{50} - 2 = 1.$$

Considere os seguintes resultados de 8 medições da temperatura, X em graus Celsius, à qual um certo líquido entra em ebulição.

$$74.1 \quad 74.9 \quad 75.0 \quad 75.7 \quad 75.8 \quad 75.0 \quad 75.6 \quad 75.9 \Rightarrow \sum_{i=1}^8 x_i = 602$$

Admitindo que aquela temperatura segue uma distribuição Normal ($\mu, \sigma^2 = 0.4$), construa um intervalo de confiança a 98% para o valor esperado da temperatura e diga que dimensão deveria ter a amostra para que a amplitude do correspondente intervalo fosse metade da do anterior.

- $(X_1, \dots, X_n)$ uma amostra aleatória de $N(\mu, \sigma^2 = 0.4)$, com concretização $\bar{x} = \frac{\sum_{i=1}^8 x_i}{n} = \frac{602}{8} = 75.25$.
- Método da variável fulcral (estimação intervalar de μ):

1. Variável fulcral: $W = \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \sim N(0, 1)$.

2. Fixado um grau de confiança $1 - \alpha = 0.98$, obtêm-se a e b tais que

$$P(a < W < b) = 1 - \alpha = 0.98 \Rightarrow b = -a = F_{N(0,1)}^{-1}(1 - \alpha/2) = F_{N(0,1)}^{-1}(0.99) = 2.3263$$

3. Desigualdade equivalente:

$$P\left(a < \frac{\bar{X} - \mu}{\sqrt{0.4/8}} < b\right) = 0.98 \Leftrightarrow P\left(\bar{X} - 2.3263 \sqrt{\frac{0.4}{8}} < \mu < \bar{X} + 2.3263 \sqrt{\frac{0.4}{8}}\right) = 0.98$$

4. Intervalo aleatório de confiança (IAC) e sua concretização (IC) a 98% para μ :

$$IAC(\mu; 0.98) : \bar{X} \pm 2.3263 \sqrt{\frac{0.4}{8}} \quad IC(\mu; 0.98) : 75.25 \pm 2.3263 \sqrt{\frac{0.4}{8}} = 75.25 \pm 0.5202 = (74.7298, 75.7702)$$

- n tal que a amplitude do correspondente intervalo fosse metade da do $IC(\mu; 0.98)$:

$$\text{amplitude} = 2 \times 0.5202 \Rightarrow \frac{2 \times 2.3263 \times \sqrt{0.4}}{\sqrt{n}} = \frac{2 \times 0.5202}{2} \Rightarrow n = \left(\frac{2 \times 2.3263 \times \sqrt{0.4}}{0.5202} \right)^2 \approx 32$$

Uma publicação recente numa rede social afirma que metade dos engenheiros recém-licenciados no país pretendem seguir estudos no mestrado da sua licenciatura. Numa sondagem encomendada por uma escola de engenharia foram contactados aleatoriamente 400 engenheiros recém-licenciados, tendo 228 deles indicado que pretendiam seguir estudos no mestrado da sua licenciatura.

Construa um teste de hipótese para investigar se é razoável, à luz dos resultados observados na sondagem, a afirmação dessa publicação e indique o respetivo valor-p.

- X : Indica se o engenheiro recém-licenciado pretende ou não seguir estudos no mestrado da sua licenciatura ($x = 1, 0$, respetivamente).
- $X \sim \text{Bernoulli}(p)$, onde $p = P(X = 1) = 1 - P(X = 0)$, com $E(X) = p$ e $\text{Var}(X) = p(1 - p)$, $0 < p < 1$.
- $(X_1, \dots, X_n)$ amostra aleatória de X com concretização $n = 400$ e $\sum_{i=1}^{400} x_i = 228$.
- Para grandes amostras ($n = 400$), pelo T.L.C., $\bar{X} \stackrel{a}{\sim} N\left(p, \frac{p(1-p)}{n}\right)$, i.e., $\frac{\bar{X} - p}{\sqrt{p(1-p)/n}} \stackrel{a}{\sim} N(0, 1)$.

- Teste de hipóteses:

1. Hipóteses: $H_0 : p = 0.5$ versus $H_1 : p \neq 0.5$

2. Estatística do teste: $T = \frac{\bar{X} - 0.5}{\sqrt{(0.5 \times (1 - 0.5))/400}} \stackrel{a}{\sim}_{H_0} N(0, 1)$, com valor observado da estatística do teste $t = \frac{228/400 - 0.5}{\sqrt{(0.5 \times (1 - 0.5))/400}} = 2.8$

3. Valor-p: $P = P(|T| \geq |t| | H_0) = 2 \times (1 - F_{N(0,1)}(2.8)) = 0.0051$.

4. Conclusão: Rejeita-se H_0 para o nível de significância $\alpha > 0.0051$ ($\alpha = 1\%$, 5% e 10%) e não se rejeita H_0 no caso contrário. Ou seja, há forte evidência contra a afirmação de que metade dos engenheiros recém-licenciados no país pretendem seguir estudos no mestrado da sua licenciatura.

Pergunta 9

2 valores

Num posto de inspeção final de uma fábrica, o número de defeitos (ligeiros) por aparelho submetido à inspeção supostamente segue uma distribuição de Poisson de valor médio 1.5.

Controlos regulares são necessários para avaliar se tal padrão se mantém. Recolhida uma amostra (observações ao acaso e independentes) de dimensão 40 obtiveram-se as frequências absolutas:

Número de defeitos por aparelho	0	1	2	3	4 ou mais
Frequência absoluta observada	4	12	9	8	7

Avalie a plausibilidade do modelo padrão perante a amostra obtida, com um nível de significância de 5%.

- X : Número de defeitos por aparelho. Se $X \sim \text{Poisson}(\lambda)$, $f_X(x) = \frac{e^{-\lambda} \lambda^x}{x!}$, $x = 0, 1, 2, \dots$, $\lambda > 0$.
- $(X_1, \dots, X_n)$ uma amostra aleatória de X , com $n = 40$, agrupada em $k = 5$ classes.
- Teste de hipóteses (Teste de ajustamento do Qui-quadrado de Pearson):

1. Hipóteses: $H_0 : X \sim \text{Poisson}(\lambda = 1.5)$ versus $H_1 : X \not\sim \text{Poisson}(\lambda = 1.5)$

2. Estatística do teste:

$$T = \sum_{i=1}^4 \frac{(O_i - E_i)^2}{E_i} \stackrel{a}{\sim}_{H_0} \chi^2_{(k-1)} = \chi^2_{(4)}$$

cujo valor observado é $t = 12.025$, onde $E_i = E(O_i | H_0) = n p_i$ é na prática calculado com $e_i = n \times p_i^0 = n \times P(X \in \text{classe}(x)_i)$, tendo em conta $\lambda = 1.5$ e a tabela auxiliar abaixo.

i	$\text{classe}(x)_i$	o_i	$e_i = n \times p_i^0$	$\frac{(o_i - e_i)^2}{e_i}$
1	0	4	8.925	2.7179
2	1	12	13.388	0.1439
3	2	9	10.041	0.1079
4	3	8	5.020	1.7683
5	≥ 4	7	2.625	7.2874
total		40	40	12.025

3. Região crítica (RC): Fixado $\alpha = 0.05$, $RC_{5\%} = (q, +\infty)$, onde $q = F_{\chi^2_{(4)}}^{-1}(0.95) = 9.488$.

4. Conclusão: Como $t = 12.025 \in RC_{5\%}$, rejeita-se H_0 ao nível de significância de 5%, i.e., o modelo padrão ($\text{Poisson}(\lambda = 1.5)$) não é plausível perante a amostra obtida a 5%.

Pergunta 10

2 valores

Numa investigação sobre erupções de um determinado géiser, pretende-se estudar a relação entre o tempo de espera desde a última erupção (x , em minutos) e a duração da erupção (Y , em minutos). Estes fenómenos são importantes para compreender o comportamento cíclico do géiser. Recolheram-se de forma aleatória dados sobre este fenómeno, obtendo-se os seguintes resultados:

$$n = 27 \quad \sum_{i=1}^n x_i = 1898.144 \quad \sum_{i=1}^n y_i = 93.768 \quad \sum_{i=1}^n x_i^2 = 134369.9 \quad \sum_{i=1}^n y_i^2 = 334.326 \quad \sum_{i=1}^n x_i y_i = 6634.803$$

Admitindo a validade do modelo de regressão linear simples e as hipóteses de trabalho convenientes, teste se a duração da erupção não é significativamente influenciada pelo tempo de espera desde a última erupção (H_0) e, com base nos seus resultados, indique o respetivo valor-p.

A: 0.006 **B:** 0.012 **C:** 0.053 **D:** 0.258

- Modelo de regressão linear simples (hipóteses de trabalho):

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

com $\varepsilon_i \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$, $i = 1, \dots, n$.

- Teste de hipóteses:

1. Hipóteses: $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$ (regressão significativa).

2. Estatística do teste:

$$T = \frac{\hat{\beta}_1 - 0}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}}} \underset{H_0}{\sim} t_{(n-2)} = t_{(25)}$$

cujo valor observado é $t = 2.711$, , tendo em conta os cálculos auxiliares abaixo.

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} = \frac{6634.80 - 27 \frac{1898.144}{27} \frac{93.768}{27}}{134369.9 - 27 \left(\frac{1898.14}{27} \right)^2} = \frac{42.7568}{927.8459} = 0.0461$$

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n-2} \left[\sum_{i=1}^n y_i^2 - n\bar{y}^2 - \hat{\beta}_1^2 \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right) \right] \\ &= \frac{1}{25} \left[334.326 - 27 \left(\frac{93.768}{27} \right)^2 - (0.0461)^2 (134369.9 - 27 \left(\frac{1898.14}{27} \right)^2) \right] \\ &= 0.2683 \end{aligned}$$

3. Valor-p: $P = P(|T| \geq |t| | H_0) = 2 \times (1 - F_{t_{(25)}}(2.711)) = 2 \times (1 - 0.994) = 0.012$. Note-se que $F_{t_{(25)}}(2.485) = 0.99 < F_{t_{(25)}}(2.711) = 0.994 < F_{t_{(25)}}(2.787) = 0.995$.

4. Conclusão: Rejeita-se H_0 para o nível de significância $\alpha > 0.012$ ($\alpha = 5\%$ e 10%) e não se rejeita H_0 no caso contrário. Ou seja, há alguma evidência de que a duração da erupção é significativamente influenciada pelo tempo de espera desde a última erupção (regressão significativa).