

Duração: 120 minutos

Exame de Época de Recurso - Versão B

- 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

2 valores

Um teste para avaliar o uso de esteroides em atletas profissionais é tal que, a probabilidade do teste resultar positivo quando administrado a um indivíduo que tenha tomado esteroides é igual a 0.87, e quando administrado a um indivíduo que não tenha tomado esteroides é igual a 0.02. Considere que, em cada 40 atletas profissionais, um destes toma esteroides.

Qual a probabilidade de um atleta profissional ao acaso ter tomado esteroides, se o seu resultado do teste for negativo?

A: 0.518 **B:** 0.003 **C:** 0.058 **D:** 0.126

Acontecimentos e probabilidades:

E = atleta toma esteroides T = teste positivo $P(T|E) = 0.87$ $P(T|\bar{E}) = 0.02$ $P(E) = \frac{1}{40}$

$$\begin{aligned} P(E|\bar{T}) &= \frac{P(E \cap \bar{T})}{P(\bar{T})} = \frac{P(E \cap \bar{T})}{P(E \cap \bar{T}) + P(\bar{E} \cap \bar{T})} = \frac{P(E) \times P(\bar{T}|E)}{P(E) \times (1 - P(T|E)) + (1 - P(E)) \times (1 - P(T|\bar{E}))} \\ &= \frac{\frac{1}{40} \times (1 - 0.87)}{\frac{1}{40} \times (1 - 0.87) + \frac{39}{40} \times (1 - 0.02)} = \frac{0.00325}{0.00325 + 0.9555} = \frac{0.00325}{0.95875} \approx 0.003 \end{aligned}$$

Pergunta 2

3 valores

Uma empresa de prospeção petrolífera constatou que o número de novas jazidas comercialmente viáveis descobertas por ano tem distribuição Poisson com parâmetro λ e que num ano é tão provável encontrar duas como três jazidas nas condições indicadas.

- (a) Determine λ e calcule a probabilidade de a empresa não encontrar qualquer jazida num ano. (1.5)

X : número de novas jazidas comercialmente viáveis descobertas num ano.

$X \sim \text{Poisson}(\lambda)$, com $P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}$, $x = 0, 1, \dots$, $\lambda > 0$.

$$P(X = 2) = P(X = 3) \Leftrightarrow \frac{e^{-\lambda} \lambda^2}{2!} = \frac{e^{-\lambda} \lambda^3}{3!} \Leftrightarrow \lambda = 3$$

$$P(X = 0) = \frac{e^{-3} 3^0}{0!} \approx 0.0498$$

[Se $\lambda = 4$, $P(X = 0) = e^{-4} \approx 0.0183$]

- (b) Em 8 anos de prospeção, qual é a probabilidade de haver menos de três anos em que não foi encontrada nenhuma jazida comercialmente viável? Suponha independência de ano para ano. (1.5)

W : número de anos em que não foi encontrados nenhuma jazida comercialmente viável em $n = 8$ anos.

Considerando $W = \sum_{i=1}^8 W_i$, onde W_i são variáveis aleatórias independentes com distribuição Bernoulli($p = 0.0498$), então $W \sim \text{Binomial}(n = 8, p = 0.0498)$, com $P(W = w) = \binom{n}{w} p^w (1 - p)^{n-w}$, $w = 0, 1, \dots, n, 0 < p < 1$.

$$P(W < 3) = F_W(2) = \sum_{w=0}^2 \binom{8}{w} 0.0498^w (1 - 0.0498)^{8-w} = 0.6646 + 0.2786 + 0.0511 = 0.9943$$

[Se $\lambda = 4$, $p = P(X = 0) = 0.0183$ e $P(W < 3) = F_W(2) = 0.9993$]

Nota: Se não determinar λ , use $\lambda = 4$.

Pergunta 3

3 valores

Num laboratório de engenharia física, ao analisar uma placa sujeita a forças em dois eixos, considere que X e Y representam as forças medidas ao longo dos eixos x e y , respectivamente.

Suponha que as forças estão sujeitas a variação devido a fatores ambientais e imperfeições do material, sendo tratadas como variáveis aleatórias contínuas com função densidade de probabilidade conjunta dada por:

$$f(x, y) = \begin{cases} 4xy, & \text{se } 0 \leq x \leq 1, 0 \leq y \leq 1, \\ 0, & \text{caso contrário} \end{cases}$$

Determine a covariância de X e Y e comente a afirmação: “as variáveis X e Y são independentes”.

- Funções densidade de probabilidade marginais de X e Y :

$$f_X(x) = \int_0^1 f(x, y) dy = \int_0^1 4xy dy = 4x \int_0^1 y dy = 4x \cdot \frac{1}{2} = 2x, \quad 0 \leq x \leq 1$$

$$f_Y(y) = \int_0^1 f(x, y) dx = \int_0^1 4xy dx = 4y \int_0^1 x dx = 4y \cdot \frac{1}{2} = 2y, \quad 0 \leq y \leq 1$$

- Valores esperados de X , Y e XY :

$$E[X] = \int_0^1 x f_X(x) dx = 2 \int_0^1 x^2 dx = 2 \cdot \frac{x^3}{3} \Big|_0^1 = \frac{2}{3}$$

$$E[Y] = \int_0^1 y f_Y(y) dy = \dots = \frac{2}{3} \quad (\text{f.d.p. de } Y \text{ é igual à de } X)$$

$$\begin{aligned} E[XY] &= \int_0^1 \int_0^1 xy 4xy dy dx = \int_0^1 4y^2 \int_0^1 x^2 dx dy = 4 \int_0^1 y^2 (x^3/3) \Big|_0^1 dy \\ &= (4/3)(y^3/3) \Big|_0^1 = 4/9 \end{aligned}$$

- Covariância de X e Y :

$$\text{Cov}(X, Y) = E[XY] - E[X] E[Y] = \frac{4}{9} - \left(\frac{2}{3}\right)^2 = \frac{4}{9} - \frac{4}{9} = 0$$

- Independência:

Usando o valor da covariância, nada se pode concluir sobre a independência, mas, como

$$f(x, y) = f_X(x) f_Y(y), \forall (x, y) \in \mathbb{R}^2,$$

as variáveis aleatórias X e Y são independentes. Afirmação correta!

Pergunta 4

3 valores

O número de acidentes ligeiros de trabalho por mês numa grande fábrica é considerada uma variável aleatória com valor esperado 3 e desvio padrão igual a 4.

Considerando as ocorrências desses acidentes de trabalho independentes entre si, calcule a probabilidade aproximada de se observarem mais de 150 acidentes ligeiros de trabalho em 4 anos.

A: 0.14 B: 0.33 C: 0.41 D: 0.59

- X_i : número de acidentes ligeiros de trabalho no mês i , $i = 1, \dots, 48$.
- $W = \sum_{i=1}^{48} X_i$ é total de acidentes ligeiros de trabalho em 4 anos.
- Considerando $E(X_i) = 3$ e $DP(X_i) \equiv \sqrt{\text{Var}(X_i)} = 4$, $\forall i$ (i.d.) e $X_1, \dots, X_{48}$ independentes (ind.),

$$\begin{aligned} E(W) &= E\left(\sum_{i=1}^{48} X_i\right) = \sum_{i=1}^{48} E(X_i) \stackrel{i.d.}{=} 48 \times 3 = 144 \\ \text{Var}(W) &= \text{Var}\left(\sum_{i=1}^{48} X_i\right) \stackrel{ind.}{=} \sum_{i=1}^{48} \text{Var}(X_i) \stackrel{i.d.}{=} 48 \times 16 = 768 \end{aligned}$$

- Para grandes amostras ($n = 48$), pelo T.L.C., $W \stackrel{d}{\sim} N(144, 768)$, i.e., $\frac{W-144}{\sqrt{768}} \stackrel{d}{\sim} N(0, 1)$.

$$P(W > 150) \approx P\left(\frac{W - E(W)}{\sqrt{\text{Var}(W)}} > \frac{150 - 144}{\sqrt{768}}\right) = 1 - F_{N(0,1)}(0.2165) \equiv 1 - \Phi(0.2165) = 1 - 0.5857 \approx 0.41$$

Pergunta 5

3 valores

Num hospital, a administração deseja monitorizar a probabilidade de um paciente não comparecer à consulta marcada ("falta" sem aviso prévio). Num determinado mês, foram selecionadas aleatoriamente 150 consultas agendadas e constatou-se que 12 pacientes faltaram à consulta sem aviso prévio.

Mostre que o estimador de máxima verosimilhança da probabilidade de um paciente faltar a consulta sem aviso prévio é a média amostral e averigue se esse estimador é centrado.

- Seja $(X_1, \dots, X_n)$ uma amostra aleatória de $X \sim \text{Bernoulli}(p)$, onde p é a probabilidade de um paciente faltar a consulta sem aviso prévio, e a sua concretização $n = 150$ e $\sum_{i=1}^{150} x_i = 12$.
- Método da máxima verosimilhança (estimação de p):

1. Funções de verosimilhança e de log-verosimilhança:

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

$$\ln L(p|\mathbf{x}) = \left(\sum_{i=1}^n x_i\right) \ln p + \left(n - \sum_{i=1}^n x_i\right) \ln(1-p)$$

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

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

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

3. Estimador de máxima verosimilhança (EMV) de p :

$$\text{EMV}(p) = \bar{X} \quad (\text{média amostral}).$$

- Como $E(\bar{X}) = E\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{\sum_{i=1}^n E(X_i)}{n} \stackrel{i.d.}{=} \frac{nE(X)}{n} = E(X) = p$, então $EMV(p) = \bar{X}$ é um estimador centrado de p .

Pergunta 6

3 valores

Trinta e seis amostras de aço com uma certa percentagem de carbono foram testadas com o objetivo de determinar o valor da respetiva tensão de rotura, X , expressa em certas unidades. Os dados obtidos foram tais que

$$\sum_{i=1}^{36} x_i = 1450 \quad \text{e} \quad \sum_{i=1}^{36} x_i^2 = 59018.$$

O engenheiro responsável pelo delineamento dos ensaios e pela análise estatística dos resultados crê, de estudos anteriores, que a tensão de rotura segue uma distribuição Normal de média μ e variância σ^2 desconhecidos. Considere que $(X_1, \dots, X_{36})$ é uma amostra aleatória correspondente aos dados em causa, com média amostral denotada por $\bar{X}$.

- (a) Construa um intervalo de confiança a 95% para o desvio padrão da tensão de rotura.

(2.0)

- Seja $(X_1, \dots, X_n)$ uma amostra aleatória de $X \sim N(\mu, \sigma^2)$ e a sua concretização $n = 36$, $\sum_{i=1}^{36} x_i = 1450$ e $\sum_{i=1}^{36} x_i^2 = 59018$, implicando $s^2 = \frac{1}{36-1} \left(\sum_{i=1}^{36} x_i^2 - \frac{(\sum_{i=1}^{36} x_i)^2}{36} \right) = 17.5778$.
- Método da variável fulcral (estimação intervalar de σ):
 1. Variável fulcral: $W = \frac{(n-1)S^2}{\sigma^2} \sim \chi_{(n-1)}^2 = \chi_{(35)}^2$.
 2. Fixado um grau de confiança $1 - \alpha = 0.95$, obtêm-se a e b tais que

$$P(a < W < b) = 1 - \alpha = 0.95 \quad \Rightarrow \quad a = F_{\chi_{(35)}^2}^{-1}(0.025) = 20.57 \quad \text{e} \quad b = F_{\chi_{(35)}^2}^{-1}(0.975) = 53.20$$

3. Desigualdade equivalente:

$$P\left(a < \frac{(n-1)S^2}{\sigma^2} < b\right) = 0.95 \quad \Leftrightarrow \quad P\left(\sqrt{\frac{(n-1)S^2}{b}} < \sigma < \sqrt{\frac{(n-1)S^2}{a}}\right) = 0.95$$

4. Intervalo aleatório de confiança (IAC) e sua concretização (IC) a 95% para σ :

$$IAC(\sigma; 0.95): \left(\sqrt{\frac{(n-1)S^2}{b}}, \sqrt{\frac{(n-1)S^2}{a}} \right), \quad IC(\sigma; 0.95): \left(\sqrt{\frac{35 \times 17.5778}{53.20}}, \sqrt{\frac{35 \times 17.5778}{20.57}} \right) = (3.400, 5.469)$$

- (b) No teste das hipóteses $H_0 : \mu = 48$ contra $H_1 : \mu > 48$, a regra de decisão para rejeitar H_0 é $\bar{X} \geq 49.8$. (1.0)
Qual a probabilidade de rejeitar incorretamente H_0 ? Considere agora que a variância populacional é dada por $\sigma^2 = 16$.

- Seja $(X_1, \dots, X_n)$ uma amostra aleatória de $X \sim N(\mu, \sigma^2 = 16)$ com $n = 36$.
- Região crítica (rejeição de H_0): $RC = \{\bar{X} \geq 49.8\}$. Região de aceitação de H_0 : $RA = \{\bar{X} < 49.8\}$.

$$\begin{aligned} P(\text{Rejeitar } H_0 | H_0 \text{ verdadeiro}) &= P(\bar{X} \geq 49.8 | X \sim N(48, 16)) = 1 - F_{N(0,1)}\left(\frac{49.8 - 48}{\sqrt{16/36}}\right) \\ &= 1 - \Phi(2.7) = 1 - 0.997 = 0.003 \end{aligned}$$

Pergunta 7

2 valores

Na gestão de um estaleiro naval interessa considerar a diferença entre a data efetiva de chegada e a respetiva data planeada dos navios (X , em dias). Foi recolhida uma amostra com a diferença temporal (em dias) destes tempos,

Diferença entre datas de chegada e planeada	]0, 0.4]	]0.4, 0.8]	]0.8, +∞[
Frequência absoluta observada	18	14	8

Averigue se a variável aleatória representativa desta diferença segue uma distribuição Exponencial de valor médio 0.4, recorrendo ao valor-p do teste para essa hipótese.

A: 0.36 B: 0.13 C: 0.06 D: 0.01

- Se $X \sim \text{Exponencial}(\lambda)$, $F_X(x) = P(X \leq x) = \int_0^x \lambda e^{-\lambda u} du = -e^{-\lambda u} \Big|_0^x = 1 - e^{-\lambda x}$, para $x > 0$ e $\lambda > 0$, com $E(X) = 1/\lambda = 0.4$, i.e., $\lambda = 2.5$.
- $(X_1, \dots, X_n)$ uma amostra aleatória de X , com $n = 40$, agrupada em $k = 3$ classes.
- Teste de hipóteses (Teste de ajustamento do Qui-quadrado de Pearson):

1. Hipóteses: $H_0 : X \sim \text{Exponencial}(\lambda = 2.5)$ versus $H_1 : X \not\sim \text{Exponencial}(\lambda = 2.5)$

2. Estatística do teste:

$$T = \sum_{i=1}^3 \frac{(O_i - E_i)^2}{E_i} \underset{H_0}{\sim} \chi_{(k-1)}^2 = \chi_{(2)}^2$$

cujo valor observado é $t = 5.6723$, 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 $\lambda = 2.5$ 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, 0.4]	18	25.2848	2.0571
2	]0.4, 0.8]	14	9.3018	2.3753
3	]0.8, +∞[	8	5.4134	1.2399
total		40	40	5.6723

3. Valor-p: $P = P(T \geq t | H_0) = (1 - F_{\chi_{(2)}^2}(5.6723)) = 1 - 0.9414 \approx 0.06$. Note-se que $F_{\chi_{(2)}^2}(5.181) = 0.925 < F_{\chi_{(2)}^2}(5.6723) = 0.9414 < F_{\chi_{(2)}^2}(5.991) = 0.95$.

4. Conclusão: Rejeita-se H_0 para o nível de significância $\alpha > 0.059$ ($\alpha = 10\%$) e não se rejeita H_0 no caso contrário ($\alpha = 1\%$ e 5%). Ou seja, há alguma evidência de que a diferença entre datas de chegada e planeada não segue uma distribuição exponencial de média 0.4.

Pergunta 8

2 valores

Numa equipa de engenharia informática pretende-se estudar o impacto do tempo de processamento de uma tarefa (x , em segundos) no consumo energético médio (Y , em Watts) de um servidor. Os dados recolhidos ao acaso em 42 execuções dessa tarefa foram os seguintes:

$$n = 42 \quad \sum_{i=1}^n x_i = 1280 \quad \sum_{i=1}^n y_i = 6800 \quad \sum_{i=1}^n x_i^2 = 54960 \quad \sum_{i=1}^n x_i y_i = 252600$$

Admita a validade do modelo de regressão linear simples, $Y = \beta_0 + \beta_1 x + \varepsilon$, e as hipóteses de trabalho convenientes.

Indique o intervalo de confiança de 95% para o consumo energético esperado quando o tempo de processamento é de $x_0 = 15$ segundos.

Para os cálculos que efetuar considere a seguinte estimativa da variância do erro aleatório (ε) na regressão linear: $\hat{\sigma}^2 = 1238.548$.

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

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

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

- Método da variável fulcral (estimação intervalar de $E(Y|x = x_0 = 15)$):

1. Variável fulcral:

$$T = \frac{(\hat{\beta}_0 + \hat{\beta}_1 x_0) - (\beta_0 + \beta_1 x_0)}{\sqrt{\left(\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}\right) \hat{\sigma}^2}} \sim t_{(n-2)} = t_{(40)}$$

2. Fixado um grau de confiança $1 - \alpha = 0.95$, obtêm-se a e b tais que

$$P(a < W < b) = 1 - \alpha = 0.95 \quad \Rightarrow \quad a = -b \text{ e } b = F_{t_{(40)}}^{-1}(0.975) = 2.021$$

3. Desigualdade equivalente:

$$P\left(-b < \frac{(\hat{\beta}_0 + \hat{\beta}_1 x_0) - (\beta_0 + \beta_1 x_0)}{\hat{k}} < b\right) = 0.95$$

$$\Leftrightarrow P((\hat{\beta}_0 + \hat{\beta}_1 x_0) - b\hat{k} < E(Y|x = x_0)) \equiv \beta_0 + \beta_1 x_0 < (\hat{\beta}_0 + \hat{\beta}_1 x_0) + b\hat{k} = 0.95,$$

onde $\hat{k} = \sqrt{\left(\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}\right) \hat{\sigma}^2}$, com a estimativa $\hat{\sigma}^2 = 1238.548$.

4. Intervalo aleatório de confiança (IAC) e sua concretização (IC) a 95% para $E(Y|x = 15)$:

$$IAC(E(Y|x = 15); 0.95) : (\hat{\beta}_0 + \hat{\beta}_1 \times 15) \pm b \hat{k}$$

$$IC(E(Y|x = 15); 0.95) : (75.233 + 2.844 \times 15) \pm 2.021 \times 6.9345 \approx (103.9, 131.9),$$

tendo em conta os cálculos auxiliares abaixo.

$$\hat{\beta}_1 = \frac{252600 - 42 \times 30.476 \times 161.905}{54960 - 42 \times (30.476)^2} = 2.844, \quad \hat{\beta}_0 = 161.905 - 2.844 \times 30.476 = 75.233,$$

$$\hat{E}(Y|x = 15) = 75.233 + 2.844 \times 15 = 117.892,$$

$$\hat{k} = \sqrt{[(1/42) + (30.476 - 15)^2] / (54960 - 42 \times (30.476)^2) \times 1238.548} = 6.9345.$$