

Duração: 120 minutos

Exame de Época de Recurso - 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

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 0.9, e quando administrado a um indivíduo que não tenha tomado esteroides é igual a 0.01. Considere que, em cada 50 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: 0.002 B: 0.124 C: 0.647 D: 0.972

Acontecimentos e probabilidades:

E = atleta toma esteroides T = teste positivo $P(T|E) = 0.9$ $P(T|\bar{E}) = 0.01$ $P(E) = \frac{1}{50}$

$$\begin{aligned} P(E|\bar{T}) &= \frac{P(E \cap \bar{T})}{P(\bar{T})} = \frac{P(E \cap \bar{T})}{P(E \cap \bar{T}) + P(\bar{E} \cap \bar{T})} = \frac{P(E) \times P(\bar{T}|E)}{P(E) \times (1 - P(T|E)) + (1 - P(E)) \times (1 - P(T|\bar{E}))} \\ &= \frac{\frac{1}{50} \times (1 - 0.9)}{\frac{1}{50} \times (1 - 0.9) + \frac{49}{50} \times (1 - 0.01)} = \frac{0.002}{0.002 + 0.9702} = \frac{0.002}{0.9722} \approx 0.002 \end{aligned}$$

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 três como quatro 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 $P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}$, $x = 0, 1, \dots$, $\lambda > 0$.

$$P(X = 3) = P(X = 4) \Leftrightarrow \frac{e^{-\lambda} \lambda^3}{3!} = \frac{e^{-\lambda} \lambda^4}{4!} \Rightarrow \lambda = 4$$

$$P(X = 0) = \frac{e^{-4} 4^0}{0!} \approx 0.0183$$

[Se $\lambda = 5$, $P(X = 0) = e^{-5} \approx 0.0067$]

- (b) Em 10 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 $n = 10$ anos.

Considerando $W = \sum_{i=1}^{10} W_i$, onde W_i são variáveis aleatórias independentes com distribuição Bernoulli($p = 0.0183$), então $W \sim \text{Binomial}(n = 10, p = 0.0183)$, com $P(W = w) = \binom{n}{w} p^w (1 - p)^{n-w}$, $w = 0, 1, \dots, n, 0 < p < 1$.

$$P(W < 3) = F_W(2) = \sum_{w=0}^2 \binom{10}{w} 0.0183^w (1 - 0.0183)^{10-w} = 0.8312 + 0.1551 + 0.0130 = 0.9993$$

[Se $\lambda = 5$, $p = P(X = 0) = 0.0067$ e $P(W < 3) = F_W(2) = 0.9999$]

Nota: Se não determinar λ , use $\lambda = 5$.

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:

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

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

- Funções densidade de probabilidade (f.d.p.) marginais de X e Y :

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

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

- Valores esperados de X , Y e XY :

$$E[X] = \int_0^1 x f_X(x) dx = \int_0^1 x \left(x + \frac{1}{2} \right) dx = \int_0^1 x^2 dx + \int_0^1 \frac{x}{2} dx = \frac{1}{3} + \frac{1}{4} = \frac{4}{12} + \frac{3}{12} = \frac{7}{12}$$

$$E[Y] = \int_0^1 y f_Y(y) dy = \dots = \frac{7}{12} \quad (\text{f.d.p. de } Y \text{ é igual à de } X)$$

$$\begin{aligned} E[XY] &= \int_0^1 \int_0^1 xy(x + y) dy dx = \int_0^1 y \int_0^1 (x^2 + xy) dx dy = \int_0^1 y \left[\frac{x^3}{3} + \frac{x^2 y}{2} \right]_0^1 dy \\ &= \int_0^1 (y/3 + y^2/2) dy = (y^2/6 + y^3/6) \Big|_0^1 = (1/6 + 1/6) = 1/3 \end{aligned}$$

- Covariância de X e Y :

$$\text{Cov}(X, Y) = E[XY] - E[X] E[Y] = \frac{1}{3} - \left(\frac{7}{12} \right)^2 = \frac{1}{3} - \frac{49}{144} = \frac{48 - 49}{144} = -\frac{1}{144}$$

- Independência:

Como $\text{Cov}(X, Y) = -\frac{1}{144} \neq 0$, então X e Y não são variáveis aleatórias (v.a.) independentes. Afirmação incorreta!

[Em alternativa] Se $\exists (x, y) \in \mathbb{R}^2$: $f(x, y) \neq f_X(x) f_Y(y)$, então X e Y não são v.a. independentes, e.g., $f(\frac{1}{3}, \frac{1}{3}) = \frac{2}{3} \neq \frac{5}{6} \times \frac{5}{6} = f_X(\frac{1}{3}) \times f_Y(\frac{1}{3})$.

O número de acidentes ligeiros de trabalho por mês numa grande fábrica é considerada uma variável aleatória com valor esperado 2.5 e desvio padrão igual a 3.

Considerando as ocorrências desses acidentes de trabalho independentes entre si, calcule a probabilidade aproximada de se observarem mais de 100 acidentes ligeiros de trabalho em 3 anos.

A: 0.17 B: 0.29 C: 0.52 D: 0.71

- X_i : número de acidentes ligeiros de trabalho no mês i , $i = 1, \dots, 36$.
- $W = \sum_{i=1}^{36} X_i$ é total de acidentes ligeiros de trabalho em 3 anos.
- Considerando $E(X_i) = 2.5$ e $DP(X_i) \equiv \sqrt{\text{Var}(X_i)} = 3$, $\forall i$ (i.d.) e $X_1, \dots, X_{36}$ independentes (ind.),

$$\begin{aligned} E(W) &= E\left(\sum_{i=1}^{36} X_i\right) = \sum_{i=1}^{36} E(X_i) \stackrel{i.d.}{=} 36 \times 2.5 = 90 \\ \text{Var}(W) &= \text{Var}\left(\sum_{i=1}^{36} X_i\right) \stackrel{ind.}{=} \sum_{i=1}^{36} \text{Var}(X_i) \stackrel{i.d.}{=} 36 \times 9 = 324 \end{aligned}$$

- Para grandes amostras ($n = 36$), pelo T.L.C., $W \stackrel{d}{\sim} N(90, 324)$, i.e., $\frac{W-90}{\sqrt{324}} \stackrel{d}{\sim} N(0, 1)$.

$$P(W > 100) \approx P\left(\frac{W - E(W)}{\sqrt{\text{Var}(W)}} > \frac{100 - 90}{\sqrt{324}}\right) = 1 - F_{N(0,1)}(0.5556) \equiv 1 - \Phi(0.5556) = 1 - 0.7107 \approx 0.29$$

Pergunta 5

3 valores

O gerente de uma agência de reparação de um dado equipamento eletrónico investiga a probabilidade de reparação devida à má utilização do equipamento. Para esse efeito, 200 processos de reparação foram selecionados aleatoriamente, dos quais 15 revelaram que a avaria foi consequência direta de má utilização.

Mostre que o estimador de máxima verosimilhança da probabilidade de reparação devida à má utilização do equipamento é a média amostral e averigüe 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 reparação devida à má utilização do equipamento, e a sua concretização $n = 200$ e $\sum_{i=1}^{200} x_i = 15$.
- 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

Vinte 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

$$\sum_{i=1}^{26} x_i = 1560 \quad \text{e} \quad \sum_{i=1}^{26} x_i^2 = 93850.$$

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 $(X_1, \dots, X_{26})$ é uma amostra aleatória correspondente aos dados em causa, com média amostral denotada por $\bar{X}$.

- (a) Construa um intervalo de confiança a 98% 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 $n = 26$, $\sum_{i=1}^{26} x_i = 1560$ e $\sum_{i=1}^{26} x_i^2 = 93850$, implicando $s^2 = \frac{1}{26-1} \left(\sum_{i=1}^{26} x_i^2 - \frac{(\sum_{i=1}^{26} x_i)^2}{26} \right) = 10$.

- Método da variável fulcral (estimação intervalar de σ):

1. Variável fulcral: $W = \frac{(n-1)S^2}{\sigma^2} \sim \chi_{(n-1)}^2 = \chi_{(25)}^2$.

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

$$P(a < W < b) = 1 - \alpha = 0.98 \quad \Rightarrow \quad a = F_{\chi_{(25)}^2}^{-1}(0.01) = 11.52 \quad \text{e} \quad b = F_{\chi_{(25)}^2}^{-1}(0.99) = 44.31$$

3. Desigualdade equivalente:

$$P\left(a < \frac{(n-1)S^2}{\sigma^2} < b\right) = 0.98 \quad \Leftrightarrow \quad P\left(\sqrt{\frac{(n-1)S^2}{b}} < \sigma < \sqrt{\frac{(n-1)S^2}{a}}\right) = 0.98$$

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

$$IAC(\sigma; 0.98) : \left(\sqrt{\frac{(n-1)S^2}{b}}, \sqrt{\frac{(n-1)S^2}{a}} \right)$$

$$IC(\sigma; 0.98) : \left(\sqrt{\frac{25 \times 10}{44.31}}, \sqrt{\frac{25 \times 10}{11.52}} \right) = (2.375, 4.658)$$

- (b) No teste das hipóteses $H_0 : \mu = 50$ contra $H_1 : \mu > 50$, a regra de decisão para rejeitar H_0 é $\bar{X} \geq 50.8$. Qual a probabilidade de rejeitar incorretamente H_0 ? Considere agora que a variância populacional é dada por $\sigma^2 = 9$. (1.0)

- Seja $(X_1, \dots, X_n)$ uma amostra aleatória de $X \sim N(\mu, \sigma^2 = 9)$ com $n = 26$.
- Região crítica (rejeição de H_0): $RC = \{\bar{X} \geq 50.8\}$. Região de aceitação de H_0 : $RA = \{\bar{X} < 50.8\}$.

$$\begin{aligned} P(\text{Rejeitar } H_0 | H_0 \text{ verdadeiro}) &= P(\bar{X} \geq 50.8 | X \sim N(50, 9)) = 1 - F_{N(0,1)}\left(\frac{50.8-50}{\sqrt{9/26}}\right) \\ &= 1 - \Phi(1.35) = 1 - 0.913 = 0.087 \end{aligned}$$

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	9	3

Averigue se a variável aleatória representativa desta diferença segue uma distribuição Exponencial de valor médio 0.5, recorrendo ao valor-p do teste para essa hipótese.

A: 0.06 **B:** 0.17 **C:** 0.23 **D:** 0.36

- 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 $E(X) = 1/\lambda = 0.5$, i.e., $\lambda = 2$.
- $(X_1, \dots, X_n)$ uma amostra aleatória de X , com $n = 30$, agrupada em $k = 3$ classes.
- Teste de hipóteses (Teste de ajustamento do Qui-quadrado de Pearson):

1. Hipóteses: $H_0 : X \sim \text{Exponencial}(\lambda = 2)$ versus $H_1 : X \not\sim \text{Exponencial}(\lambda = 2)$

2. Estatística do teste:

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

cujo valor observado é $t = 2.0141$, 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 $\lambda = 2$ 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	16.5201	0.1326
2	]0.4,0.8]	9	7.4230	0.3364
3	]0.8,+∞[	7	6.0569	1.5451
total		30	30	2.0141

3. Valor-p: $P = P(T \geq t | H_0) = (1 - F_{\chi_{(2)}^2}(2.0141)) = 1 - 0.6347 \approx 0.36$. Note-se que $F_{\chi_{(2)}^2}(1.833) = 0.6 < F_{\chi_{(2)}^2}(2.0141) = 0.6347 < F_{\chi_{(2)}^2}(2.408) = 0.7$.

4. Conclusão: Rejeita-se H_0 para o nível de significância $\alpha > 0.365$ e não se rejeita H_0 no caso contrário ($\alpha = 1\%, 5\%$ e 10%). Ou seja, há forte evidência de que a diferença entre datas de chegada e planeada não segue uma distribuição exponencial de média 0.5.

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 20 execuções dessa tarefa foram os seguintes:

$$n = 20 \quad \sum_{i=1}^n x_i = 640 \quad \sum_{i=1}^n y_i = 3400 \quad \sum_{i=1}^n x_i^2 = 27480 \quad \sum_{i=1}^n x_i y_i = 126300$$

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 90% para o consumo energético esperado quando o tempo de processamento é de $x_0 = 30$ segundos.

Para os cálculos que efetuar considere a seguinte estimativa da variância do erro aleatório (ε) na regressão linear: $\hat{\sigma}^2 = 1000$.

- 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 $E(Y|x = x_0 = 30)$):

1. Variável fulcral:

$$T = \frac{(\hat{\beta}_0 + \hat{\beta}_1 x_0) - (\beta_0 + \beta_1 x_0)}{\sqrt{\left(\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}\right) \hat{\sigma}^2}} \sim t_{(n-2)} = t_{(18)}$$

2. Fixado um grau de confiança $1 - \alpha = 0.90$, obtêm-se a e b tais que

$$P(a < W < b) = 1 - \alpha = 0.9 \quad \Rightarrow \quad a = -b \text{ e } b = F_{t_{(18)}}^{-1}(0.95) = 1.734$$

3. Desigualdade equivalente:

$$P\left(-b < \frac{(\hat{\beta}_0 + \hat{\beta}_1 x_0) - (\beta_0 + \beta_1 x_0)}{\hat{k}} < b\right) = 0.9$$

$$\Leftrightarrow P((\hat{\beta}_0 + \hat{\beta}_1 x_0) - b\hat{k} < E(Y|x = x_0)) \equiv \beta_0 + \beta_1 x_0 < (\hat{\beta}_0 + \hat{\beta}_1 x_0) + b\hat{k} = 0.9,$$

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 $\hat{\sigma}^2 = 1000$.

4. Intervalo aleatório de confiança (IAC) e sua concretização (IC) a 90% para $E(Y|x = 30)$:

$$IAC(E(Y|x = 30); 0.9) : (\hat{\beta}_0 + \hat{\beta}_1 \times 30) \pm b \hat{k}$$

$$IC(E(Y|x = 30); 0.9) : (90 + 2.5 \times 30) \pm 1.734 \times 7.1114 \approx (152.7, 177.3),$$

tendo em conta os cálculos auxiliares abaixo.

$$\hat{\beta}_1 = \frac{126300 - 20 \times 32 \times 170}{27480 - 20 \times (32)^2} = 2.5, \quad \hat{\beta}_0 = 170 - 2.5 \times 32 = 90,$$

$$\hat{E}(Y|x = 30) = 90 + 2.5 \times 30 = 165,$$

$$\hat{k} = \sqrt{[(1/20) + (32 - 30)^2] / (27480 - 20 \times 32^2) \times 1000} = 7.11145.$$