



Duração: 90 minutos

2º teste A

Justifique convenientemente todas as respostas!

Grupo I

10 valores

1. Da análise da sua carteira de empréstimos a particulares com algum incumprimento de pagamento, uma instituição bancária concluiu que o número de meses que decorre até ao primeiro incumprimento de pagamento é modelado pela variável aleatória X com distribuição geométrica de parâmetro p , $0 < p \leq 1$. Considere que (X_1, \dots, X_n) , $n \geq 3$, é uma amostra aleatória de X .

- (a) Determine o erro quadrático médio do estimador $T = \frac{\sum_{i=1}^3 iX_i}{6}$ do valor esperado do número de meses até ao primeiro incumprimento de pagamento. (2.0)

$$\begin{aligned} E[T] &= E\left[\frac{\sum_{i=1}^3 iX_i}{6}\right] = E\left[\frac{\sum_{i=1}^3 iE[X_i]}{6}\right] = \frac{\sum_{i=1}^3 i}{6} E[X] = E[X] = \frac{1}{p} \\ \text{Var}[T] &= \text{Var}\left[\frac{\sum_{i=1}^3 iX_i}{6}\right] = \frac{\sum_{i=1}^3 i^2 \text{Var}[X_i]}{36} = \frac{\sum_{i=1}^3 i^2}{36} \text{Var}[X] = \frac{7}{18} \text{Var}[X] = \frac{7}{18} \frac{(1-p)}{p^2} \\ EQM[T] &= \text{Var}[T] + (E[T] - 1/p)^2 = \frac{7}{18} \frac{(1-p)}{p^2} \end{aligned}$$

- (b) Determine o estimador de máxima verosimilhança do parâmetro p . (2.5)

$$\begin{aligned} \mathcal{L}(p, x_1, \dots, x_n) &= \prod_{i=1}^n f_X(x_i) = \prod_{i=1}^n p(1-p)^{x_i-1} = p^n (1-p)^{\sum_{i=1}^n x_i - n} \\ \text{Para } p \in]0, 1[, \log(\mathcal{L}(p, x_1, \dots, x_n)) &= n \log p + (\sum_{i=1}^n x_i - n) \log(1-p) \text{ (diferenciável em ordem a } p) \\ \text{Para } \sum_{i=1}^n x_i > n, \frac{d\mathcal{L}}{dp} = 0 &\iff \frac{n}{p} - \frac{(\sum_{i=1}^n x_i - n)}{1-p} = 0 \iff p = \frac{1}{\bar{x}} \text{ e} \\ \frac{d^2\mathcal{L}}{dp^2} &= -\frac{n}{p^2} - \frac{(\sum_{i=1}^n x_i - n)}{(1-p)^2} < 0, \forall 0 < p < 1 \text{ uma vez que } \sum_{i=1}^n x_i > n \\ \therefore \hat{p}_{MV} &= \frac{1}{\bar{x}} \\ \text{Nota: se } \sum_{i=1}^n x_i = n \text{ então } \mathcal{L}(p, x_1, \dots, x_n) &= p^n \text{ e } \hat{p}_{MV} = 1 = \frac{1}{\bar{x}} \end{aligned}$$

- (c) Com base numa realização da amostra de tamanho 100 em que $\sum_{i=1}^{100} x_i = 4965$, obtenha a estimativa de máxima verosimilhança da probabilidade de o primeiro incumprimento de pagamento não ocorrer durante o primeiro trimestre de um empréstimo. (1.0)

$$\begin{aligned} \text{Pretende-se estimar } g(p) = P(X > 3) &= 1 - F_{Geo(p)}(3). \text{ Pela invariância dos estimadores de invariância} \\ \text{tem-se que } \hat{g}_{MV}(p) = g(\hat{p}_{MV}) &= 1 - F_{Geo(\frac{1}{\bar{x}})}(3). \\ \text{Com } \bar{x} = 49.65 \text{ tem-se } \hat{P}(X > 3) &\approx 1 - F_{Geo(0.02)}(3) = 0.9412. \end{aligned}$$

2. Um vendedor de uma marca de baterias, usadas em *tablets*, mediu os tempos (em horas) de autonomia de 16 dessas baterias, $x_i, i = 1, 2, \dots, 16$, tendo obtido: $\sum_{i=1}^{16} x_i = 79.5$ e $\sum_{i=1}^{16} x_i^2 = 395.65$. Admitindo que o tempo de autonomia dessas baterias tem distribuição normal:

- (a) Obtenha um intervalo de confiança a 95% para o desvio padrão do tempo de autonomia das baterias em questão. (3.0)

$$\begin{aligned} \text{Sejam } Q = \frac{15S^2}{\sigma^2} &\sim \chi_{(15)}^2, a = F_{\chi_{(15)}^2}^{-1}(0.025) \approx 6.262 \text{ e } b = F_{\chi_{(15)}^2}^{-1}(0.975) \approx 27.49. \\ P(a \leq Q \leq b) = 0.95 &\iff P\left(\frac{15S^2}{b} \leq \sigma^2 \leq \frac{15S^2}{a}\right) = 0.95 \iff P\left(\sqrt{\frac{15S^2}{b}} \leq \sigma \leq \sqrt{\frac{15S^2}{a}}\right) = 0.95 \\ \text{IAC}_{0.95}(\sigma) &= \left[\sqrt{\frac{15S^2}{27.49}}, \sqrt{\frac{15S^2}{6.262}}\right] \\ \text{Para a amostra observada tem-se } \bar{x} &= 79.5/16 \approx 4.97 \text{ e } 15s^2 = \sum_{i=1}^{16} x_i^2 - 16\bar{x}^2 \approx 0.6344 \text{ (} s^2 \approx 0.0423) \text{ e} \\ \text{IC}_{0.95}(\sigma) &= [0.1519, 0.3183]. \end{aligned}$$

- (b) Usando o resultado da alínea anterior, teste ao nível de significância de 5% se o desvio padrão do tempo de autonomia das baterias em questão é igual a 0.25 horas. (1.5)

Pretende-se testar $H_0 : \sigma = 0.25$ contra $H_1 : \sigma \neq 0.25$. Uma vez que $0.25 \in IC_{0.95}(\sigma)$ não se rejeita H_0 a um n. s. de 0.05 recorrendo à relação entre intervalos de confiança e testes de hipóteses bilaterais.

Grupo II

10 valores

1. O registo, efetuado ao longo de um conjunto de várias semanas, de faltas ao trabalho dos funcionários de uma empresa conduziu aos seguintes resultados: (4.0)

Dia	Segunda	Terça	Quarta	Quinta	Sexta
Nº Faltas	23	17	14	20	26

Teste a hipótese de as faltas ao trabalho dos funcionários da empresa se distribuírem uniformemente pelos 5 dias úteis da semana. Decida com base no valor-p.

Numerando os dias da semana de 1 a 5, seja $X =$ “dia em que ocorre uma falta ao trabalho”. Pretende-se testar $H_0 : X \sim U(\{1, \dots, 5\})$ contra $H_1 : X \neq U(\{1, \dots, 5\})$.

Seja $p_i^0 = P(X = i | H_0) = 0.2, i = 1, \dots, 5$.

i	o_i	p_i^0	$e_i = np_i^0$
1	23	0.2	20
2	17	0.2	20
3	14	0.2	20
4	20	0.2	20
5	26	0.2	20
	$n = 100$		

Como todas as classes têm uma frequência esperada superior a 5 não é necessário agrupar classes ($k = 5$) e, não havendo qualquer parâmetro estimado ($\beta = 0$), a estatística de teste é $Q_0 = \sum_{i=1}^5 \frac{(O_i - E_i)^2}{E_i} \underset{H_0}{\sim} \chi_{(4)}^2$.

Tem-se $q_0 = 4.5$ e valor-p = $P(Q_0 > q_0 | H_0) = 1 - F_{\chi_{(4)}^2}(4.5) = 0.3426$. Deve-se rejeitar H_0 para níveis de significância ≥ 0.3426 e não rejeitar no caso contrário. Para os níveis de significância usuais, $\alpha \in [0.01, 0.1]$, não há evidência suficiente para rejeitar H_0 .

2. O gestor de marketing de um supermercado quer avaliar o efeito do espaço usado de prateleiras nas vendas diárias de comida para animais. Para o efeito, o gestor recolheu os dados da tabela seguinte, em que $x(m)$ representa o comprimento usado de prateleira com comida para animais e $Y(10^3\text{€})$ o respetivo volume diário de vendas de comida para animais, tendo os dias em que foram obtidas as observações sido escolhidos ao acaso.

x_i	0.5	0.5	0.5	1.0	1.0	1.0	1.5	1.5	1.5	2.0	2.0	2.0
y_i	1.6	2.2	1.4	1.9	2.4	2.6	2.3	2.7	2.8	2.6	2.9	3.1

$$\sum_{i=1}^{12} x_i = 15.0, \quad \sum_{i=1}^{12} x_i^2 = 22.50, \quad \sum_{i=1}^{12} y_i = 28.5, \quad \sum_{i=1}^{12} y_i^2 = 70.69, \quad \sum_{i=1}^{12} x_i y_i = 38.40$$

- (a) Admitindo que é válido o modelo de regressão linear simples com as suposições de trabalho habituais, determine a respetiva estimativa da reta de regressão de mínimos quadrados e interprete o valor de $\hat{\beta}_1$. (2.0)

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} = 0.74$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 1.45$$

$$\hat{E}[Y|x] = 1.45 + 0.74x \text{ para } x \in [0.5, 2.0].$$

Estima-se que um aumento de 1.0m no comprimento da prateleira produza um aumento de 0.74 milhares de euros no volume médio diário de vendas.

- (b) Determine um intervalo de confiança a 96% para o valor esperado do volume de vendas de comida para animais no supermercado num dia em que seja usada uma prateleira de 1.0 m para expor esses produtos. É correto utilizar o mesmo procedimento aplicado a uma prateleira de 10.0 m? Justifique. (4.0)

$$E[Y|x = 1.0] = \beta_0 + \beta_1$$

$$\text{Sejam } T = \frac{(\hat{\beta}_0 + \hat{\beta}_1) - (\beta_0 + \beta_1)}{\sqrt{\left(\frac{1}{12} + \frac{(\bar{x}-1.0)^2}{\sum x_i^2 - 12\bar{x}^2}\right) \hat{\sigma}^2}} \sim t_{(10)} \text{ e } a = F_{t(10)}^{-1}(0.98) = 2.359$$

$$P(-a \leq T \leq a) = 0.96 \iff$$

$$P\left(\hat{\beta}_0 + \hat{\beta}_1 - a \sqrt{\left(\frac{1}{12} + \frac{(\bar{x}-1.0)^2}{\sum x_i^2 - 12\bar{x}^2}\right) \hat{\sigma}^2} \leq \beta_0 + \beta_1 \leq \hat{\beta}_0 + \hat{\beta}_1 + a \sqrt{\left(\frac{1}{12} + \frac{(\bar{x}-1.0)^2}{\sum x_i^2 - 12\bar{x}^2}\right) \hat{\sigma}^2}\right) = 0.96$$

$$\text{IAC}_{0.96}(\beta_0 + \beta_1) = \left[\hat{\beta}_0 + \hat{\beta}_1 - 2.359 \sqrt{\left(\frac{1}{12} + \frac{(\bar{x}-1.0)^2}{\sum x_i^2 - 12\bar{x}^2}\right) \hat{\sigma}^2}, \hat{\beta}_0 + \hat{\beta}_1 + 2.359 \sqrt{\left(\frac{1}{12} + \frac{(\bar{x}-1.0)^2}{\sum x_i^2 - 12\bar{x}^2}\right) \hat{\sigma}^2} \right]$$

$$\hat{\beta}_0 + \hat{\beta}_1 = 2.19$$

$$\sqrt{\left(\frac{1}{12} + \frac{(\bar{x}-1.0)^2}{\sum x_i^2 - 12\bar{x}^2}\right) \hat{\sigma}^2} = 0.0974$$

$$\text{IC}_{0.96}(\beta_0 + \beta_1) = [1.9602, 2.4198]$$

Para $x_0 = 10.0\text{m}$ não é adequado aplicar o procedimento anterior pois esse valor de x encontra-se fora da gama de valores utilizada na experiência $[0.5, 2.0]$ e, por isso, não se pode garantir que o modelo ajustado continua a ser adequado para $x_0 = 10.0$.