

Duração: 90 minutos

2º teste A

Justifique convenientemente todas as respostas!

Grupo I

10 valores

1. Em cada uma de 500 operações stop, selecionadas ao acaso, foi registado o número X de condutores fiscalizados até ser encontrado o primeiro em infração, tendo-se observado que $\sum_{i=1}^{500} x_i = 6250$. Admitindo que X tem distribuição geométrica de parâmetro p , determine:

- (a) O estimador de máxima verosimilhança de p . (2.5)

$$\mathcal{L}(p, x_1, \dots, x_n) \equiv f_{\mathbf{X}}(x_1, \dots, x_n) \stackrel{iid}{=} \prod_{i=1}^n f_X(x_i) = \prod_{i=1}^n p(1-p)^{x_i-1} = p^n (1-p)^{\sum_{i=1}^n x_i - n}$$

Para $p \in]0, 1[$, $\log(\mathcal{L}(p, x_1, \dots, x_n)) = n \log p + (\sum_{i=1}^n x_i - n) \log(1-p)$ (diferenciável em ordem a p)

Para $\sum_{i=1}^n x_i > n$, $\frac{d\mathcal{L}}{dp} = 0 \iff \frac{n}{p} - \frac{(\sum_{i=1}^n x_i - n)}{1-p} = 0 \iff p = \frac{1}{\bar{x}}$ e

$$\frac{d^2\mathcal{L}}{dp^2} = -\frac{n}{p^2} - \frac{(\sum_{i=1}^n x_i - n)}{(1-p)^2} < 0, \forall 0 < p < 1 \text{ uma vez que } \sum_{i=1}^n x_i > n$$

$\therefore \hat{p}_{MV} = \frac{1}{\bar{x}}$

- (b) A estimativa de máxima verosimilhança da probabilidade de ser necessário fiscalizar mais de 3 condutores até encontrar a primeira infração. (1.5)

Pretende-se estimar $g(p) = P(X > 3) = 1 - F_{Geo(p)}(3)$. Pela invariância dos estimadores de invariância tem-se que $\hat{g}_{MV}(p) = g(\hat{p}_{MV}) = 1 - F_{Geo(\frac{1}{\bar{x}})}(3)$.

Com $\bar{x} = 12.5$ tem-se $\hat{P}(X > 3) \approx 1 - F_{Geo(0.08)}(3) = 0.7787$.

2. Com o objetivo de reduzir a poluição num troço da Ribeira dos Milagres foi desencadeada uma campanha de limpeza. Para avaliar a eficácia da medida adotada foram feitas medições em 10 pontos escolhidos ao acaso antes (X) e medições em outros 10 pontos um ano após a campanha de limpeza (Y), obtendo-se os seguintes resultados: $\bar{x} = 79.9$ e $\bar{y} = 74.9$, numa unidade tal que menores níveis de poluição conduzem a valores de X e Y mais baixos.

Admitindo que ambas as variáveis aleatórias têm distribuição normal com variância comum igual a 13:

- (a) A um nível de significância de 0.04, teste a hipótese de a campanha de limpeza não ter tido o efeito pretendido. (4.0)

Pretende-se testar $H_0 : \mu_X - \mu_Y \leq 0$ contra $H_1 : \mu_X - \mu_Y > 0$.

Seja $Z = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} \sim N(0, 1)$. Sob H_0 obtemos a estatística do teste, $Z_0 = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{26}{10}}} \stackrel{\mu_X - \mu_Y = 0}{\sim} N(0, 1)$.

Para $\alpha = 0.04$ deve rejeitar-se H_0 se $Z_0 > \Phi^{-1}(0.96) = 1.7506$.

Como $z_0 \approx 3.1$ pertence à região de rejeição então H_0 é rejeitada para $\alpha = 0.04$.

- (b) Calcule a probabilidade de o procedimento que aplicou na alínea anterior conduzir a uma decisão errada no caso de a campanha de limpeza ter contribuído para uma redução de 5 unidades no nível médio de poluição. (2.0)

$$P(Z_0 \leq 1.7506 \mid \mu_X - \mu_Y = 5) = P\left(Z_0 \leq 1.7506 \mid Z^* = \frac{\bar{X} - \bar{Y} - 5}{\sqrt{\frac{26}{10}}} \sim N(0, 1)\right) = P(Z^* \leq -1.35) = \Phi(-1.35) = 0.088.$$

1. Os dados relativos ao tempo de espera X , em dias úteis, até se conseguir uma consulta de certa especialidade num dado agrupamento de centros saúde de 100 doentes foram registados, tendo-se observado: (4.5)

Nº dias	≤ 5]5, 8]]8, 12]	>12
Nº doentes	52	28	16	4

Teste a hipótese do tempo de espera até se conseguir uma consulta ter distribuição exponencial de parâmetro λ . Decida com base no valor-p do teste, sabendo que a estimativa de máxima verosimilhança de λ é 0.2.

Pretende-se testar $H_0 : X \sim Exp(\lambda)$ contra $H_1 : X \not\sim Exp(\lambda)$.

Seja $\hat{p}_i^0 = P(X \in Classe_i | H_0)$, $i = 1, \dots, 4$. Como o valor de λ não é especificado apenas podemos calcular as estimativas $\hat{p}_i^0 = P(X \in Classe_i | H_0, \hat{\lambda} = 0.2)$.

Com $X \sim Exp(0.2)$ tem-se $F_X(x) = 1 - e^{-0.2x}$, para $x > 0$, e então $\hat{p}_1^0 = F_X(5) = 0.6321$, $\hat{p}_2^0 = F_X(8) - F_X(5) = 0.1660$, $\hat{p}_3^0 = F_X(12) - F_X(8) = 0.1112$ e $\hat{p}_4^0 = 1 - F_X(12) = 0.0907$.

i	o_i	\hat{p}_i^0	$e_i = n\hat{p}_i^0$
1	52	0.6321	63.21
2	28	0.1660	16.60
3	16	0.1112	11.12
4	4	0.0907	9.07
n = 100			

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

Tem-se $q_0 \approx 14.8$ e valor-p = $P(Q_0 > q_0 | H_0) = 1 - F_{\chi^2_{(2)}}(14.8) \approx 6 \times 10^{-4}$. Deve-se rejeitar H_0 para níveis de significância $\geq 6 \times 10^{-4}$ e não rejeitar no caso contrário. Para os níveis de significância usuais, $\alpha \in [0.01, 0.1]$, há evidência suficiente para rejeitar H_0 .

2. A densidade ótica (Y) de uma solução de certa substância química foi medida para várias concentrações ($x \in]0, 15]$), tendo-se obtido o seguinte conjunto de resultados sumariados:

$$\sum_{i=1}^8 x_i = 57, \quad \sum_{i=1}^8 y_i = 234, \quad \sum_{i=1}^8 x_i^2 = 579, \quad \sum_{i=1}^8 y_i^2 = 9536, \quad \sum_{i=1}^8 x_i y_i = 2348.$$

- (a) Após ter enunciado as hipóteses de trabalho que entender mais convenientes, obtenha e interprete as estimativas dos coeficientes da reta de regressão. (2.0)

Admite-se que as variáveis $Y_i = Y | x = x_i$ são não correlacionadas e que $Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$ para $i = 1, \dots, 8$.

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} = 3.9378 \text{ e } \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 1.1931$$

Estima-se que o aumento de uma unidade na concentração da substância química conduza a um aumento de 3.9378 unidades no valor esperado da densidade ótica da solução. Estima-se também que a densidade ótica média do solvente (ou seja, a "solução" sem a substância química em questão) é igual a 1.1931 unidades.

- (b) Obtenha uma estimativa pontual e intervalar a 99% do valor esperado da densidade ótica de uma solução com uma concentração de $x_0 = 7$. (3.5)

Pretende-se estimar $E[Y|x=7] = \beta_0 + 7\beta_1$.

$$\hat{E}[Y|x=7] = \hat{\beta}_0 + 7\hat{\beta}_1 = 28.76$$

$$\text{Sejam } T = \frac{(\hat{\beta}_0 + 7\hat{\beta}_1) - (\beta_0 + 7\beta_1)}{\sqrt{\left(\frac{1}{8} + \frac{(\bar{x}-7)^2}{\sum x_i^2 - 8\bar{x}^2}\right) \hat{\sigma}^2}} \sim t_{(6)} \text{ e } a = F_{t_{(6)}}^{-1}(0.995) = 3.707$$

$$P(-a \leq T \leq a) = 0.99 \iff$$

$$P\left(\hat{\beta}_0 + 7\hat{\beta}_1 - a\sqrt{\left(\frac{1}{8} + \frac{(\bar{x}-7)^2}{\sum x_i^2 - 8\bar{x}^2}\right) \hat{\sigma}^2} \leq \beta_0 + 7\beta_1 \leq \hat{\beta}_0 + 7\hat{\beta}_1 + a\sqrt{\left(\frac{1}{8} + \frac{(\bar{x}-7)^2}{\sum x_i^2 - 8\bar{x}^2}\right) \hat{\sigma}^2}\right) = 0.99$$

$$\text{IAC}_{0.99}(\beta_0 + 7\beta_1) = \left[\hat{\beta}_0 + 7\hat{\beta}_1 - 3.707\sqrt{\left(\frac{1}{8} + \frac{(\bar{x}-7)^2}{\sum x_i^2 - 8\bar{x}^2}\right) \hat{\sigma}^2}, \hat{\beta}_0 + 7\hat{\beta}_1 + 3.707\sqrt{\left(\frac{1}{8} + \frac{(\bar{x}-7)^2}{\sum x_i^2 - 8\bar{x}^2}\right) \hat{\sigma}^2} \right]$$

$$\hat{\sigma}^2 \approx 1.81$$

$$\sqrt{\left(\frac{1}{8} + \frac{(\bar{x}-7)^2}{\sum x_i^2 - 8\bar{x}^2}\right) \hat{\sigma}^2} = 0.4758$$

$$\text{IC}_{0.99}(\beta_0 + 7\beta_1) = [26.99, 30.52]$$