

Grupo I

10 valores

1. O tempo de processamento, em segundos, de programas que chegam a um determinado servidor é uma variável aleatória, X , com distribuição normal de valor esperado μ e desvio padrão 1.5 segundos. Com base numa concretização $(x_1, x_2, \dots, x_{20})$ de uma amostra aleatória (de dimensão 20) de X da qual resultou $\sum_{i=1}^{20} x_i = 410$, deduzia:

- (a) As estimativas de máxima verosimilhança de μ e da probabilidade de a média de uma amostra aleatória (de X) de dimensão 10 ser inferior a 20 segundos. (3.5)

$$\mathcal{L}(\mu|x_1, \dots, x_{20}) = \prod_{i=1}^{20} f_X(x_i) \quad (\text{porque } X_i \text{ são iid})$$

$$= \prod_{i=1}^{20} (1.5\sqrt{2\pi})^{-1} \exp\left\{-\frac{2}{9}(x_i - \mu)^2\right\} = (1.5\sqrt{2\pi})^{-20} \exp\left\{-\frac{2}{9}\sum_{i=1}^{20}(x_i - \mu)^2\right\}, \text{ com } \mu \in \mathbb{R}$$

$$\log(\mathcal{L}(\mu|x_1, \dots, x_{20})) = -20\log(1.5\sqrt{2\pi}) - \frac{2}{9}\sum_{i=1}^{20}(x_i - \mu)^2 \quad (\text{diferenciável em ordem a } \mu)$$

Procura-se o valor de μ , que se denomina por $\hat{\mu}$ (a estimativa de máxima verosimilhança de μ), que maximiza $\log(\mathcal{L}(\mu|x_1, \dots, x_{20}))$, i.e.

$$\frac{d\log(\mathcal{L}(\mu|x_1, \dots, x_{20}))}{d\mu} = 0 \Leftrightarrow \frac{4}{9}\sum_{i=1}^{20}(x_i - \mu) = 0 \Leftrightarrow \sum_{i=1}^{20}x_i - 20\mu = 0 \Leftrightarrow \mu = \frac{\sum_{i=1}^{20}x_i}{20} = \bar{x}, \text{ e}$$

$$\left. \frac{d^2\log(\mathcal{L}(\mu|x_1, \dots, x_{20}))}{d\mu^2} \right|_{\mu=\bar{x}} = -\frac{80}{9} < 0, \forall \mu \in \mathbb{R}, \text{ concluiu-se que } \hat{\mu} = \bar{x} \text{ é a estimativa de máxima verosimilhança de } \mu.$$

Para a amostra observada obtém-se a estimativa $\bar{x} = 410/20 = 20.5$.

Seja \bar{X}_{10} a média de uma amostra aleatória de dimensão 10. Temos que $\bar{X}_{10} \sim N(\mu, 1.5^2/10)$.

Sejam ψ a probabilidade da média de uma amostra aleatória (de X) de dimensão 10 ser inferior a 20 segundos e ψ_{MV} ($\hat{\psi}_{MV}$) o seu estimador (estimativa) de máxima verosimilhança.

Como $\psi = P(\bar{X}_{10} < 20) = \Phi\left(\sqrt{10}\frac{20-\mu}{1.5}\right)$, pela invariância dos estimadores de máxima verosimilhança temos que $\psi_{MV} = \Phi\left(\sqrt{10}\frac{20-\bar{X}}{1.5}\right)$, logo

$$\hat{\psi}_{MV} = \Phi\left(\sqrt{10}\frac{20-20.5}{1.5}\right) = \Phi(-1.05) = 0.1469.$$

(b) Um intervalo de confiança a 95% para o parâmetro μ .

(2.5)

Sejam $Z = \frac{\bar{X} - \mu}{1.5/\sqrt{20}} \sim N(0, 1)$ e $a = \Phi^{-1}(0.975) = 1.96$.

$$P(-a \leq Z \leq a) = 0.95 \Leftrightarrow P\left(\bar{X} - a \frac{1.5}{\sqrt{20}} \leq \mu \leq \bar{X} + a \frac{1.5}{\sqrt{20}}\right) = 0.95.$$

$$IAC_{0.95}(\mu) = \left[\bar{X} - 1.96 \frac{1.5}{\sqrt{20}}, \bar{X} + 1.96 \frac{1.5}{\sqrt{20}} \right].$$

Para a amostra observada temos $\bar{x} = 20.5$ e

$$IC_{0.95}(\mu) = [19.843, 21.157].$$

2. Um operador de uma rede móvel está a equacionar lançar um novo pacote de preços, especialmente vocacionado para assinantes que fazem chamadas de longa duração. O departamento comercial deste operador alega não valer a pena lançar esse pacote, pois considera que a proporção de aderentes a uma eventual campanha de lançamento desse pacote seria 15%. Para avaliar a suposição do departamento comercial, foi efectuado um inquérito a 200 assinantes, escolhidos ao acaso, tendo 34 deles manifestado interesse em usufruir do dito pacote de preços.

(a) Teste, ao nível de significância de 5%, a validade da alegada suposição do departamento comercial contra a alternativa de a proporção de aderentes à campanha ser superior a 15%. (3.0)

A amostra observada é uma concretização de uma amostra aleatória de $X \sim Ber(p)$.

Hipóteses: $H_0 : p = 0.15$ contra $H_1 : p > 0.15$.

Uma vez que o tamanho da amostra é suficientemente grande temos, pelo TLC,

$$Z = \frac{\bar{X} - E[X]}{\sqrt{\frac{Var[X]}{200}}} = \frac{\bar{X} - p}{\sqrt{\frac{p(1-p)}{200}}} \stackrel{a}{\sim} N(0, 1).$$

$$\text{Sob } H_0, \text{ obtemos a estatística do teste: } Z_0 = \frac{\bar{X} - 0.15}{\sqrt{\frac{0.15 \times 0.85}{200}}} \stackrel{a}{\sim} N(0, 1).$$

Para $\alpha = 0.05$ deve rejeitar-se H_0 se $Z_0 > \Phi^{-1}(0.95) = 1.645$.

Para a amostra observada temos $\bar{x} = 34/200 = 0.17$ e $z_0 = \frac{0.02}{0.02525} = 0.792$.

Como z_0 não pertence à região de rejeição então H_0 não é rejeitada para $\alpha = 0.05$.

Alternativa: $\text{valor} - p \approx 1 - \Phi(0.79) = 0.215 > 0.05$, pelo que H_0 não é rejeitada para $\alpha = 0.05$.

(b) Calcule a probabilidade (aproximada) de o teste anterior rejeitar correctamente a hipótese nula caso a verdadeira proporção (populacional) de aderentes à campanha de lançamento desse pacote fosse 20%. (1.0)

$$P(Z_0 > 1.645 | p = 0.2) = P(\bar{X} > 0.192 | p = 0.2).$$

$$\text{Com } p = 0.2 \text{ tem-se } Z^* = \frac{\bar{X} - 0.2}{\sqrt{\frac{0.2 \times 0.8}{200}}} \stackrel{a}{\sim} N(0, 1).$$

$$\text{Então, } P(\bar{X} > 0.192 | p = 0.2) = P(Z^* > -0.283) \approx 1 - \Phi(-0.283) = 0.611.$$

1. Uma companhia de seguros registou para um conjunto de 240 dias, escolhidos ao acaso, o número diário de participações relativas a um certo tipo de acidentes (X), tendo obtido os seguintes resultados:

Nº diário de participações	0	1	2	≥ 3
Nº de dias	72	87	52	29

- (a) Os dados recolhidos são consistentes com a hipótese da variável aleatória X seguir uma distribuição de Poisson com variância igual a 1.2, ao nível de significância de 1%? (3.5)

$X \sim Poi(\lambda)$ com $\text{Var}[X] = \lambda = 1.2$, logo $X \sim Poi(1.2)$.

Hipóteses: $H_0 : X \sim Poi(1.2)$ contra $H_1 : X \not\sim Poi(1.2)$.

Estatística de teste: $Q_0 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \underset{H_0}{\sim} \chi_{(k-\beta-1)}^2$, onde β é o número de parâmetros a estimar.

Sejam $p_i^0 = P(X = i - 1 | H_0) = e^{-1.2} \frac{1.2^{i-1}}{(i-1)!}$, $i = 1, 2, 3$ e $p_4^0 = P(X \geq 3 | H_0) = 1 - \sum_{i=1}^3 p_i^0 = 0.1205$.

	o_i	p_i^0	$E_i = np_i^0$
0	72	0.3012	72.29
1	87	0.3614	86.74
2	52	0.2169	52.06
≥ 3	29	0.1205	28.92
	$n = 240$		

Neste caso, não é necessário agrupar classes ($k = 4$) e não há qualquer parâmetro estimado ($\beta = 0$).

Para $\alpha = 0.01$, deve rejeitar-se H_0 se $Q_0 > F_{\chi_{(3)}^2}^{-1}(0.99) = 11.345$.

Como $q_0 = 0.002$ não pertence à região de rejeição então não se rejeita H_0 para $\alpha = 0.01$.

- (b) Calcule, justificando, o valor- p do teste anterior e decida com base no valor obtido. (1.0)

Uma vez que a região crítica é unilateral, rejeitando-se H_0 para valores elevados de Q_0 , tem-se valor- $p = P(Q_0 > 0.002 | H_0 \text{ verdadeira}) = 1 - F_{\chi_{(3)}^2}(0.002) = 0.99998$.

Como o valor- p é tão elevado, pode concluir-se que não há evidência para contestar que o número diário de participações relativas ao tipo de acidentes considerado segue uma distribuição de Poisson com variância igual a 1.2.

2. Suspeita-se que a quantidade de frango vendida por dia num supermercado (Y , em kg) está relacionada directamente com o preço (x , em euros/kg) a que os frangos são comercializados. Para investigar esta suspeita, registaram-se os resultados referentes a 10 dias, escolhidos ao acaso:

x_i	3.40	2.80	3.10	4.00	2.10	3.60	1.99	2.85	3.50	2.90
y_i	34.0	41.3	38.3	23.9	48.1	34.7	51.5	43.1	33.5	42.2

$$\sum_{i=1}^{10} x_i = 30.24, \quad \sum_{i=1}^{10} y_i = 390.6, \quad \sum_{i=1}^{10} x_i^2 = 95.1226, \quad \sum_{i=1}^{10} y_i^2 = 15830.44 \quad \text{e} \quad \sum_{i=1}^{10} x_i y_i = 1136.45.$$

Considerando o modelo de regressão linear simples de Y em x , com as hipóteses de trabalho habituais:

- (a) Determine a recta de regressão de mínimos quadrados e obtenha uma estimativa da variância da quantidade de frango (kg) vendida num dia em que o preço dos frangos é conhecido (fixo). (2.0)

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} = \frac{1136.45 - 30.24 \times 390.6/10}{95.1226 - 30.24^2/10} = -\frac{44.7244}{3.6768} = -12.164,$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 390.6/10 + 12.164 \times 30.24/10 = 75.843,$$

$$\hat{E}[Y|x] = 75.843 - 12.164x.$$

$$\text{Var}[Y|x] = \sigma^2 \text{ e}$$

$$\hat{\sigma}^2 = \frac{1}{n-2} \left[\left(\sum_{i=1}^n y_i^2 - n\bar{y}^2 \right) - (\hat{\beta}_1)^2 \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right) \right] = \\ = \frac{1}{8} [(15830.44 - 390.6^2/10) - (-12.164)^2 \times 3.6768] = 3.6981.$$

- (b) Construa um intervalo de confiança a 90% para a quantidade média de frango (kg) vendida num dia em que o preço dos frangos seja 2.50 euros/kg. Seria adequado seguir um procedimento análogo para construir um intervalo de confiança a 90% para a quantidade de frango vendida num dia em que o preço dos frangos fosse 5.00 euros/kg? Justifique. (3.5)

$$E[Y|x = 2.50] = \beta_0 + 2.5\beta_1.$$

$$\text{Sejam } T = \frac{(\hat{\beta}_0 + 2.5\hat{\beta}_1) - (\beta_0 + 2.5\beta_1)}{\sqrt{\left(\frac{1}{10} + \frac{(\bar{x} - 2.5)^2}{\sum x_i^2 - 10\bar{x}^2} \right) \hat{\sigma}^2}} \sim t_{(8)} \text{ e } a = F_{t(8)}^{-1}(0.95) = 1.860.$$

$$P(-a \leq T \leq a) = 0.9 \Leftrightarrow$$

$$P\left(\hat{\beta}_0 + 2.5\hat{\beta}_1 - a \sqrt{\left(\frac{1}{10} + \frac{(\bar{x} - 2.5)^2}{\sum x_i^2 - 10\bar{x}^2} \right) \hat{\sigma}^2} \leq \beta_0 + 2.5\beta_1 \leq \hat{\beta}_0 + 2.5\hat{\beta}_1 + a \sqrt{\left(\frac{1}{10} + \frac{(\bar{x} - 2.5)^2}{\sum x_i^2 - 10\bar{x}^2} \right) \hat{\sigma}^2} \right) = 0.9. \text{ Logo,}$$

$$\text{IAC}_{0.9}(\beta_0 + 2.5\beta_1) =$$

$$\left[\hat{\beta}_0 + 2.5\hat{\beta}_1 - 1.860 \sqrt{\left(\frac{1}{10} + \frac{(\bar{x} - 2.5)^2}{\sum x_i^2 - 10\bar{x}^2} \right) \hat{\sigma}^2}, \hat{\beta}_0 + 2.5\hat{\beta}_1 + 1.860 \sqrt{\left(\frac{1}{10} + \frac{(\bar{x} - 2.5)^2}{\sum x_i^2 - 10\bar{x}^2} \right) \hat{\sigma}^2} \right].$$

$$\text{Como } \hat{\beta}_0 + 2.5\hat{\beta}_1 = 45.434 \text{ e } \sqrt{\left(\frac{1}{10} + \frac{(\bar{x} - 2.5)^2}{\sum x_i^2 - 10\bar{x}^2} \right) \hat{\sigma}^2} = 0.804, \text{ obtem-se}$$

$$\text{IC}_{0.9}(\beta_0 + 2.5\beta_1) = [43.94, 46.93].$$

Para $x_0 = 5$ não é adequado aplicar o procedimento anterior pois esse valor de x encontra-se fora da gama de valores utilizada na experiência [1.99, 4.00] e, por isso, não se pode garantir que o modelo ajustado continua a ser adequado para $x_0 = 5$.