Departamento de Matemática da Universidade de Coimbra

Exame de Estatística

Duração: 2h 30min — 12-01-2011

Observação: A resolução completa das questões apresentadas inclui a justificação do raciocínio utilizado e a apresentação dos cálculos efectuados.

---

1. Com o objectivo de relacionar o nível de colesterol total (Y, em miligramas por decilitro) com o consumo diário médio de gordura saturada (X, em gramas) em indivíduos considerados saudáveis, uma equipa médica registou os dados correspondentes a 50 pessoas em consultas de rotina.

a) Começou por se efectuar a análise descritiva da variável estatística marginal X a partir de uma classificação da amostra, em 4 intervalos de igual amplitude, no intervalo de observação $]35, 75[$. Apresenta-se a seguir a tabela de frequências fornecida pelo *software* estatístico SPSS.

Consumo de gordura

		Frequency	Percent	Cumulative Percent
Valid	40,00	13	26,0	26,0
	50,00	12	24,0	50,0
	60,00	13	26,0	76,0
	70,00	12	24,0	100,0
	Total	50	100,0	

(i) Obtenha a função cumulativa associada a esta distribuição estatística e calcule o respectivo primeiro quartil.

(ii) Determine a ordem do percentil correspondente a um indivíduo que ingere diariamente, em média, 58 gramas de gordura saturada. Interprete o resultado obtido.

b) Foi analisada, por meio do software estatístico SPSS, a existência de uma relação linear que exprima Y como função de X. Apresentam-se a seguir alguns dos resultados obtidos.

Correlations

	Nível de colesterol	Consumo de gordura
Nível de colesterol	1,000	,951
Consumo de gordura	,951	1,000

Coefficients[a]

Model	Unstandardized Coefficients B
(Constant)	29,066
Consumo de gordura	3,417

a. Dependent Variable: Nível de colesterol

(i) Indique o coeficiente de correlação da amostra observada e interprete-o.

(ii) Obtenha uma previsão do nível máximo de colesterol de um indivíduo cujo consumo diário médio de gordura saturada não exceda 49 gramas.

v.p.f.

2. Seja X uma variável aleatória real absolutamente contínua cuja função densidade é dada por

$$f_\theta(x) = \frac{1}{2\theta}\, e^{-\frac{|x|}{\theta}}, \quad x \in \mathbb{R},$$

onde θ é um parâmetro real positivo desconhecido. Nestas condições, a v.a.r. $Y = |X|$ segue a lei exponencial de parâmetro θ^{-1}.

Com o objectivo de estimar θ, considere uma amostra aleatória de dimensão n de X, $(X_1, \ldots, X_n)$.

a) Utilizando o método da máxima verosimilhança, obtenha um estimador do parâmetro θ.

b) Deduza um estimador da máxima verosimilhança da probabilidade do acontecimento $\{X > 1\}$.

c) Verifique que $Z_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} |X_i|$ é um estimador de θ cêntrico e convergente em probabilidade.

d) Mostre que a v.a.r. $\dfrac{2n}{\theta}\, Z_n$ segue a lei do qui-quadrado com $2n$ graus de liberdade.

e) Foi observada a seguinte amostra de X, de dimensão 14:

$$0.7,\ -3.9,\ 0.8,\ 1.0,\ -2.1,\ -1.9,\ 4.0,\ -1.0,\ 2.2,\ 0.9,\ -1.0,\ -0.5,\ -0.9,\ 1.8.$$

Com base nesta amostra, determine um intervalo de confiança, de caudas igualmente ponderadas, que inclua o verdadeiro valor de θ com grau de confiança 0.98.

3. Determinada máquina de produção de componentes electrónicas fabrica, em cada período de funcionamento contínuo, uma série de s unidades. As séries são produzidas independentemente umas das outras e quando a máquina se avaria a série considera-se perdida. Além disso, quando há uma avaria a máquina é reparada, retomando a produção nas mesmas condições. Seja p, $p \in]0,1[$, a probabilidade de que a máquina se avarie durante um período de funcionamento contínuo. A v.a.r. X que representa o número de séries completas produzidas pela referida máquina até que ocorra uma avaria na mesma segue a lei geométrica de parâmetro p. Sejam $(X_1, \ldots, X_n)$ uma amostra aleatória de dimensão n de X e $\overline{X}_n$ a correspondente média empírica.

a) Verifique que, para n suficientemente grande, a lei da v.a.r. $U_n = \sqrt{\frac{n}{1-p}}\,\big(p\overline{X}_n - 1 + p\big)$ é bem aproximada pela lei normal centrada e reduzida.

b) Obtenha a região crítica de um teste que seja o mais potente de nível de significância α, $\alpha \in]0,1[$, para testar $H_0 : p = 0.05$ contra $H_1 : p = 0.07$ com base numa amostra de X de dimensão n $(n > 30)$.

c) Nas condições da alínea anterior, se se pretender que a probabilidade de erro de segunda espécie seja também igual a α, qual a decisão a tomar perante uma amostra observada com média 18?