

**Exame de Estatística**

**Duração:** 2h 30m

12-06-2007

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

(9.0) **I** - Para estudar o efeito da qualidade do ar na água de um lago fizeram-se, em diferentes períodos, medições do pH da água, descrito por uma variável aleatória real (v.a.r.)  $Y$ , e do índice de qualidade do ar, descrito por uma v.a.r.  $X$  (valores crescentes de  $X$  indicam aumento de poluição). Os valores  $Y$  das medidas do pH obtidas nesses períodos foram registados para cada valor  $x_n$  do índice  $X$  ( $n = 1, 2, \dots, 202$ ).

1. Começou por se efectuar a análise descritiva da variável estatística marginal  $Y$  a partir de uma classificação da amostra, em 4 intervalos de igual amplitude, no intervalo de observação  $]0, 8]$ . Apresenta-se abaixo a tabela de frequências fornecida pelo *software* estatístico SPSS.

**PhAguaCL**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1,00	3	1,5	1,5	1,5
3,00	44	21,8	21,8	23,3
5,00	121	59,9	59,9	83,2
7,00	34	16,8	16,8	100,0
Total	202	100,0	100,0	

- a) Obtenha a função cumulativa desta distribuição estatística e deduzo o correspondente intervalo interquartis.
- b) Obtenha a média e variância da amostra classificada.

2. Foi analisada, por meio do *software* estatístico SPSS, a existência de uma relação linear da forma  $Y_n = ax_n + b + U_n$  ( $n = 1, 2, \dots, 202$ ), com  $(U_n)$  variáveis aleatórias independentes e identicamente distribuídas com uma lei normal centrada de desvio padrão  $\sigma$  ( $\sigma > 0$ , desconhecido). A análise descritiva dos dados conduziu aos seguintes resultados:

**Descriptive Statistics**

	Mean	Std. Deviation	N
PhAgua	4,8118	1,15890	202
indice	39,9752	25,85598	202

**Correlations**

	PhAgua	indice
Pearson Correlation	1,000	-,898
	indice	-,898
Sig. (1-tailed)	PhAgua	,000
	indice	,000

- a) Indique a média e o coeficiente de correlação da amostra observada. Interprete os valores obtidos.

Nos quadros abaixo encontram-se os resultados correspondentes à regressão linear efectuada sobre os dados.

**Coefficients<sup>a</sup>**

Model		Unstandardized Coefficients		95% Confidence Interval for B	
		B	Std. Error	Lower Bound	Upper Bound
1	(Constant)	6,421	,066	6,290	6,551
	indice	-,040	,001	-,043	-,037

a. Dependent Variable: PhAgua

**One-Sample Kolmogorov-Smirnov Test**

	Unstandardized Residual
N	202
Normal Parameters <sup>a,b</sup>	Mean
	Std. Deviation
Kolmogorov-Smirnov Z	,368
Asymp. Sig. (2-tailed)	,999

a. Test distribution is Normal.

b. Calculated from data.

Além disso, a soma dos quadrados dos resíduos da estimação é igual a 52.304.

- b) Indique a relação linear estimada e tire conclusões sobre a adequação, aos dados observados, do modelo estocástico considerado.
- c) Indique intervalos de confiança para os coeficientes do modelo, precisando o grau de confiança.

- d) Mostre que a amostra observada permite concluir, ao nível de significância 0.01, que o coeficiente  $a$  é significativamente diferente de zero.
- e) Indique uma estimativa pontual para o pH da água do referido lago se for de 25 o índice de qualidade do ar.

(4.0) **II** - No quadro abaixo estão resumidos, por meio de uma amostra classificada, os tempos (em minutos) compreendidos entre duas chegadas consecutivas de doentes ao Serviço de Urgências de certo Hospital:

Tempo	]0, 1.5]	]1.5, 3]	]3, 4.5]	]4.5, 6]
Nº de ocorrências	59	28	8	7

Poderá concluir, ao nível de significância 0.02, que os dados observados são compatíveis com uma lei exponencial?

**Nota:** A função de distribuição de uma lei exponencial de parâmetro  $\lambda$  ( $\lambda > 0$ ) é dada por  $F(x) = (1 - e^{-\lambda x}) \mathbb{I}_{]0, +\infty[}(x)$ .

(7.0) **III** - Seja  $X$  uma variável aleatória real (v.a.r) uniformemente distribuída sobre  $[1, \theta]$ , com  $\theta > 1$ , arbitrariamente fixo e desconhecido. Considere uma amostra aleatória de dimensão  $n$  de  $X$ ,  $(X_1, \dots, X_n)$ .

- a) Determine, usando o método dos momentos, um estimador  $T_n$  de  $\theta$ .
- b) Prove que  $T_n$  é cêntrico e convergente, precisando os modos de convergência.
- c) Mostre que  $U_n = \max_{1 \leq i \leq n} X_i$  é um estimador da máxima verosimilhança de  $\theta$ .
- d) Conclua que a variável aleatória real  $R_n = \frac{U_n - 1}{\theta - 1}$  segue a lei de função de distribuição

$$G(x) = x^n \mathbf{I}_{]0, 1]}(x) + \mathbb{I}_{]1, +\infty[}(x).$$

- e) Determinada sequência de 50 números foi obtida com um gerador de números aleatórios de acordo com uma lei uniforme num certo intervalo real. Sabe-se que foi 1 o valor escolhido na geração para o extremo inferior do intervalo. Obtenha uma estimativa intervalar, com 95% de confiança e caudas igualmente repartidas, para o extremo superior do intervalo, sabendo que o maior valor gerado foi 4.5.