



Exame de Recurso de Métodos Estatísticos
Departamento de Matemática – Universidade de Aveiro

Data: 10/07/2006

Duração: 3 horas

Nome: _____ N.º: _____

Curso: _____ Regime: _____

Declaro que desisto _____

Classificação: _____

As cotações deste exame encontram-se na seguinte tabela:

	1				2			3			4						5				
	a	b	c	d	a(i)	a(ii)	b	a	b	c	a	b	c	d	e	f	a	b	c	d	e
Cotação	,8	,8	,8	1,3	1,2	,8	1,6	,8	,8	1,3	,8	,5	,8	1	1,6	,8	,5	,5	1,5	,8	1
Classif.																					

Responda às questões utilizando o espaço reservado para o efeito. Tenha em atenção a clareza e apresentação das suas respostas.

Num dado município em que a pluviosidade é baixa, pretende-se tirar o máximo proveito da pouca água existente. As situações que se descrevem a seguir, visam conhecer melhor a quantidade de água que se ganha, trata, perde e consome.

1. Foi pedido um levantamento relativo à pluviosidade do local. Concluíram que a quantidade de chuva que cai durante um ano na área do município, em litros/m², pode ser considerada uma variável aleatória com a seguinte f.d.p.:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & , x \geq 0 \\ 0 & , x < 0 \end{cases} \quad \text{com } \lambda > 0 .$$

(a) Determine o valor da constante λ sabendo que a probabilidade de chover mais de 6 litros/m² é igual a 0.3

(Note: a função distribuição é dada por $F(x) = 1 - e^{-\lambda x}$, $x \geq 0$)

(b) Em média, qual a quantidade de chuva que cai por ano nesta cidade? Com que desvio padrão? Indique os cálculos que efectuar. (Sugestão: identifique a distribuição.) Se não resolveu a alínea anterior considere $\lambda=0.2$.

$\mu =$ _____

$\sigma =$ _____

- (c) Utilizando a propriedade da perda de memória da exponencial determinou-se a probabilidade da quantidade de chuva ser superior 11 litros/m² dado que já choveu 5 litros/m², $P(X > 11 | X > 5)$. Assinale a expressão correcta.
- $P(X > 11 | X > 5) = P(X > 11)P(X > 5) / P(X > 5) = P(X > 11)$
- $P(X > 11 | X > 5) = P(X > 11) / P(X > 5) = P(X > 6)$
- $P(X > 11 | X > 5) = P(X > 11)P(X > 5) / P(X > 11) = P(X > 5)$
- $P(X > 11 | X > 5) = P(5 < X < 11)$
- (d) Assumindo que a quantidade de chuva que cai durante um ano não depende dos outros anos, qual a probabilidade de em 30 anos obter uma quantidade de água superior a 200 litros/m² (Sugestão: use o Teorema do limite central e caso não tenha resolvido 1(b) considere $E[X] = \sqrt{Var[X]} = 5$).

2. O reservatório para recolha das águas residuais deste pequeno município tem uma capacidade de 18500 litros. Foram recolhidas de modo aleatório 100 observações relativas ao volume diário de águas residuais (podendo exceder 18500 litros). A amostra apresenta as seguintes características:

```

Volume diário de águas residuais Stem-and-Leaf Plot
Frequency      Stem & Leaf
  2,00 Extremes      (= < 17500)
  1,00      178 . 9
   ,00      179 .
  1,00      179 . 9
  4,00      180 . 0000
  2,00      180 . 55
 11,00      181 . 00000334444
 17,00      181 . 5555555555555559
  3,00      182 . 000
  9,00      182 . 555555556
  4,00      183 . 0001
 29,00      183 . 555555555555555555555568899999
  3,00      184 . 000
 12,00      184 . 555555699999
  1,00      185 . 1
   ,00      185 .
  1,00      186 . 0
Stem width:      100,00
Each leaf:       1 case(s)

```

Descriptive Statistics

	N	Mean	Std. Deviation
Volume diário de águas residuais	100	18250	200
Valid N (listwise)	100		

Admita a normalidade dos dados.

- (a) A estação de tratamento garante o escoamento das águas desde que o desvio padrão não seja superior a 175 litros. Ao nível de significância 5% e com base na amostra recolhida é de crer que não haja problemas no escoamento das águas residuais? Para responder à questão formularam-se as seguintes hipóteses:

$$H_0 : \sigma = 175 \quad \text{vs} \quad H_1 : \sigma > 175$$

(i) Realize o teste com base na região crítica.

(ii) Calcule o *p-value* do teste.

(b) Teste ao nível de significância 5% se há razões para acreditar que a proporção de dias em que o volume diário de águas excede 18500 litros é inferior 0.025.

3. Recolheram-se aleatoriamente 25 observações relativas à pluviosidade diária e o respectivo valor previsto pelo serviço municipal de protecção civil. Alguns dos resultados obtidos a partir dos dados foram:

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
previsão em l/m ² de precipitação	,279	25	,000	,736	25	,000
valor registado em l/m ² de precipitação	,295	25	,000	,803	25	,000

a. Lilliefors Significance Correction

Test Statistics^b

	valor registado em l/m ² de precipitação - previsão em l/m ² de precipitação
Z	-2,354 ^a
Asymp. Sig. (2-tailed)	-----

a. Based on positive ranks.

b. Wilcoxon Signed Ranks Test

(a) As amostras relativas à pluviosidade e previsão são emparelhadas ou independentes? Justifique.

(b) O que pode concluir com base nos resultados da tabela onde se apresentam os resultados de testes da normalidade.

(c) Calcule o *p-value* assintótico do teste de Wilcoxon, para a diferença de médias e conclua se existem diferenças significativas ao nível de significância 5%. (Sugestão: $Z \underset{\text{sob } H_0}{\overset{\circ}{\cap}} N(0, 1)$)

4. Pretende-se estudar o consumo de água de um município de acordo com as temperaturas máximas diárias. Escolheram-se aleatoriamente 200 indivíduos e separaram-se em 4 grupos equilibrados e para cada grupo foi atribuído um dia escolhido aleatoriamente. Registaram-se as máximas diárias dos dias atribuídos aos grupos e no dia atribuído os indivíduos do grupo registaram a água por eles consumida. As máximas diárias atingidas nos dias escolhidos (aleatoriamente) foram 26°C, 27°C, 30°C e 35°C.

Tests of Normality

	Index1	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
consumo	26°C	,110	50	,177	,970	50	,241
	27°C	,094	50	,200*	,982	50	,622
	30°C	,099	50	,200*	,962	50	,109
	35°C	,087	50	,200*	,975	50	,362

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Test of Homogeneity of Variances

consumo			
Levene Statistic	df1	df2	Sig.
1,709	3	196	,167

ANOVA

consumo					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	30566,393	-----	-----	-----	,000
Within Groups	83394,875	-----	-----		
Total	-----	-----			

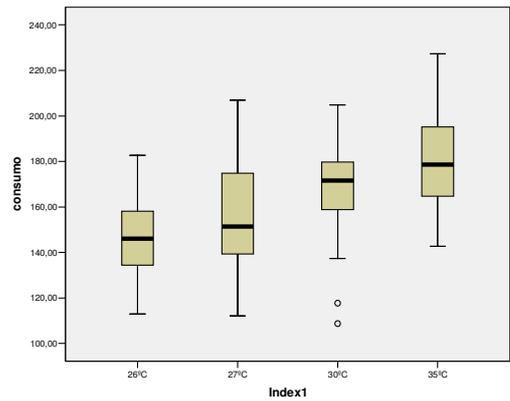
Descriptives

consumo						
		Std. Deviation	Std. Error	95% Confidence Interval for Mean		Between-Component Variance
				Lower Bound	Upper Bound	
Model	Fixed Effects	20,62726	1,45857	160,2006	165,9536	
	Random Effects		7,13751	140,3624	185,7918	-----

Multiple Comparisons

Dependent Variable: consumo
Tukey HSD

(I) Index1	(J) Index1	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
26°C	27°C	-8,15981	4,12545		-18,8497	2,5301
	30°C	-20,38972	4,12545	,000	-31,0796	-9,6998
	35°C	-32,63062	4,12545	,000	-43,3205	-21,9407
27°C	26°C	8,15981	4,12545		-2,5301	18,8497
	30°C	-12,22990	4,12545	,018	-22,9198	-1,5400
	35°C	-24,47081	4,12545	,000	-35,1607	-13,7809
30°C	26°C	20,38972	4,12545	,000	9,6998	31,0796
	27°C	12,22990	4,12545	,018	1,5400	22,9198
	35°C	-12,24091	4,12545	,018	-22,9308	-1,5510
35°C	26°C	32,63062	4,12545	,000	21,9407	43,3205
	27°C	24,47081	4,12545	,000	13,7809	35,1607
	30°C	12,24091	4,12545	,018	1,5510	22,9308



Descriptives

consumo					
	N	Mean	Std. Deviation	Minimum	Maximum
26°C	50	147,7821	17,92617	112,88	182,68
27°C	50	155,9419	22,67381	112,15	206,93
30°C	50	168,1718	20,39663	108,74	204,84
35°C	50	180,4127	21,22415	142,62	227,22
Total	200	163,0771	23,93052	108,74	227,22

- (a) Complete os espaços a tracejado das tabelas.
- (b) Com base nas tabelas anteriores que achar conveniente, discuta se são válidos os pressupostos da ANOVA paramétrica.
- (c) Com base nas amostras e assumindo válidos os pressupostos da ANOVA paramétrica, utilize o método dos momentos para apresentar uma estimativa para o(s) parâmetro(s) da distribuição do volume de água consumida a 30°C.
- (d) Com base na tabela de ANOVA diga se os consumos são significativamente diferentes para diferentes máximas diárias ($\alpha = 1\%$). Justifique.
- (e) Com base nos resultados disponíveis na tabela resultante do método de Tuckey – comparações múltiplas, indique ao nível de significância 5% quais os pares de médias significativamente diferentes. Justifique. (Note que os asteriscos foram retirados da tabela)
- (f) Comente a validade da análise de comparações múltiplas efectuada neste problema.

5. No sentido de estudar o total de água consumida por família mensalmente registou-se:

- O tempo semanal de duchas e/ou banhos (em horas),
- O tempo semanal que a casa está habitada (em horas),
- O número de habitantes da casa,
- A área do jardim (em m²)

de 60 indivíduos escolhidos aleatoriamente. Optou-se por ajustar um modelo de regressão linear, o qual permitiu obter os resultados seguintes:

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,999 ^a	,998	,998	,98619

a. Predictors: (Constant), área do jardim (x4), tempo semanal de banhos e/ou duchas (x1), número de habitantes da casa (x3), tempo semanal médio que a casa está habitada (x2)

b. Dependent Variable: total de água consumida (y)

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	23721,804	4	5930,451	6097,663	,000 ^a
	Residual	53,492	55	,973		
	Total	23775,296	59			

a. Predictors: (Constant), área do jardim (x4), tempo semanal de banhos e/ou duchas (x1), número de habitantes da casa (x3), tempo semanal médio que a casa está habitada (x2)

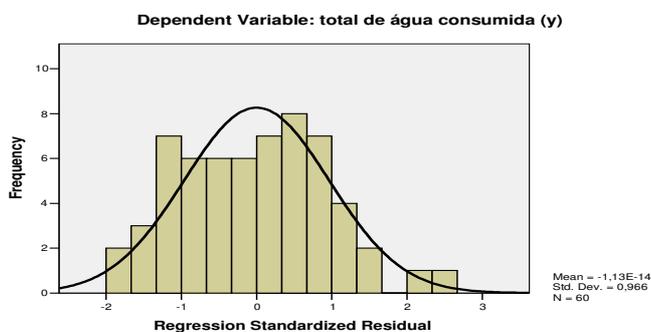
b. Dependent Variable: total de água consumida (y)

Coefficients^a

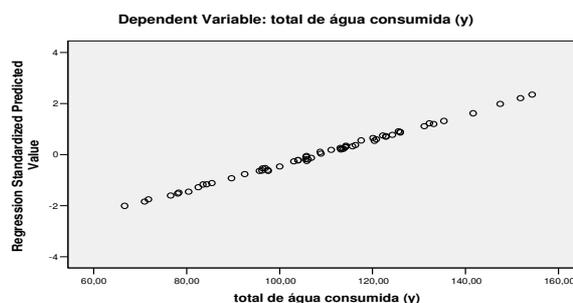
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	,092	,938		,098	,922
	tempo semanal de banhos e/ou duchas (x1)		,136	,024	3,699	,001
	tempo semanal médio que a casa está habitada (x2)	-,067	,042	-,011	-1,589	,118
	número de habitantes da casa (x3)	1,220		,073	11,316	,000
	área do jardim (x4)	,019	,001	,019		,000

a. Dependent Variable: total de água consumida (y)

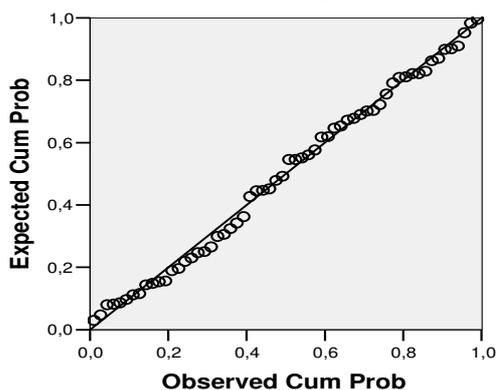
Histogram



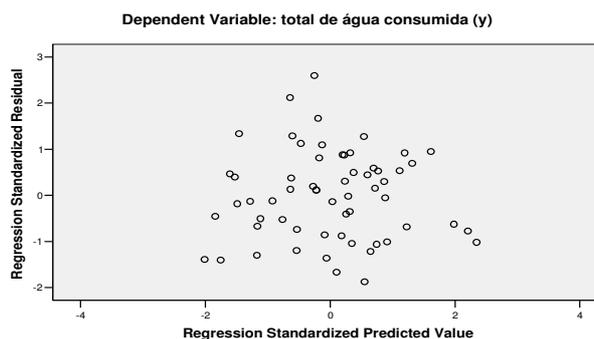
Scatterplot



Dependent Variable: total de água consumida (y)



Scatterplot



Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Unstandardized Residual	,058	60	,200*	,985	60	,661

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

- (a) Complete os espaços da tabela.
 (b) Identifique os coeficientes que não são significativamente diferentes de zero ao nível de significância 5%. Justifique.

- (c) Complete as seguintes frases e risque o que não interessa:
 Em média, estima-se que cada novo habitante de uma casa provoca um(a) acréscimo / diminuição de _____ m³ no consumo mensal de água.
 Em média, se uma família pavimentar 100 m² do seu jardim, irá ter um(a) acréscimo / diminuição de _____ m³ no consumo mensal de água.
 Em média, estima-se que cada hora semanal em que a casa está habitada provoca um(a) acréscimo / diminuição de _____ m³ no consumo mensal de água. Esta variação é / não é significativa.
 O coeficiente \hat{b}_0 é / não é significativo, porque o seu valor é próximo de zero / o p-value associado é próximo de 1.

- (d) Averigúe, aos níveis usuais de significância, se existem evidências para concluir que o coeficiente associado ao tempo semanal de banhos e/ou duches (x1) é negativo. (Indique o valor do p-value do teste unilateral.)

- (e) Com base nos gráficos e tabelas fornecidos (que achar conveniente) proceda à validação dos pressupostos do modelo de regressão.