



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

Data: 15/06/2004

Duração: 3 horas

Nome: _____ N.º: _____

Curso: _____

Declaro que desisto _____ Classificação: _____

As cotações deste exame encontram-se na seguinte tabela:

Questão	1	2 a)	2 b)	2 c)	3 a)	3 b)	3 c)	4 a)	4 b)	4 c)	4 d)	4 e)	5 a)	5 b)	6 a)	6 b)	6 c)	6 d)	6 e)	7 a)	7 b)	7 c)
Cotação	1.5	0.5	0.5	0.5	0.5	1	1	1	1	1.5	1.5	1.5	1	1	0.5	0.5	0.5	0.5	0.5	1.5	1	1

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

Agora que o João terminou a licenciatura é hora de procurar o primeiro emprego.

1. Depois de analisar cuidadosamente várias ofertas de emprego, o João decidiu enviar o Curriculum Vitae (CV) para várias entidades (diferentes) das quais se destacam: um hospital, uma indústria metalomecânica e uma indústria química, todas na região de Aveiro. O João definiu os seguintes acontecimentos:

$A = \{\text{ser chamado a uma entrevista no hospital}\}$

$B = \{\text{ser chamado a uma entrevista na indústria metalomecânica}\}$

$C = \{\text{ser chamado a uma entrevista na indústria química}\}$

O João calculou mentalmente as seguintes probabilidades:

$$P(A) = P(B) = P(C) = 1/3$$

$$P(A \cap B) = P(A \cap C) = P(B \cap C) = 1/3^2 \text{ e } P(A \cap B \cap C) = 1/3^3 \text{ (porque as decisões são independentes)}$$

Após ter feito algumas contas, calculou que a probabilidade de ser chamado pelo menos para uma entrevista era de $19/27$. Verifique se o resultado a que chegou o João está correcto.

Passado uma semana, o João foi chamado a uma entrevista no hospital. O director de Recursos Humanos após ter analisado o seu CV, fez-lhe a seguinte proposta: “Dado ter frequentado Métodos Estatísticos na UA, tenho um lugar para si no departamento de Bioestatística deste hospital. Aceita?” Ao qual o João respondeu sem hesitar: “Com certeza”.

2. A primeira missão do João foi rever uns cálculos que um dos seus colegas fizera acerca dos dois únicos fornecedores de um determinado produto hospitalar. O fornecedor 1 entrega diariamente 10 unidades desse produto enquanto que o fornecedor 2 entrega 20. Sabendo que a probabilidade de cada unidade entregue por qualquer dos fornecedores ser defeituosa é de 0.06,
- a) Calcule a probabilidade de num dia o fornecedor 1 entregar exactamente 2 unidades defeituosas.

b) Calcule o número médio de unidades defeituosas fornecidas diariamente ao hospital (considerando os 2 fornecedores).

c) Indique a distribuição e os parâmetro(s) associado(s) do número total de unidades defeituosas fornecidas diariamente ao hospital.

3. O director do Serviço de Urgências está preocupado com a situação de catástrofe durante o EURO2004. As últimas estatísticas deste serviço, dizem que o nº médio de utentes que chegam à urgência em cada 5 minutos é de 4.5 e que se considera situação de catástrofe quando chegam 32 ou mais utentes em 30 minutos.

a) Qual a distribuição da variável X = “número de utentes que chega às urgências durante 5 minutos”, especificando o(s) parâmetro(s)?

b) Qual a distribuição da variável Y = “número de utentes que chega às urgências durante 30 minutos”, especificando o(s) parâmetro(s)?

c) Qual a probabilidade de ocorrer uma catástrofe?

Nota: para responder a esta questão o João foi aconselhado a utilizar o teorema do limite central.

Após impressionar o director do Departamento de Bioestatística com a sua resolução dos problemas anteriores, uma nova tarefa o espera.

4. Um investigador do Departamento de Farmacologia quer publicar o seu trabalho numa revista científica muito importante, pelo que pediu ajuda ao Departamento de Bioestatística para a análise dos dados e tomada de decisões. Ao rever as notas deste investigador, o João encontrou o seguinte quadro sobre dados experimentais de um novo fármaco que provoca uma diminuição da pressão arterial (PA em mmHg): (nota: as observações correspondem aos valores da diminuição)

Nome: _____ N° mecanográfico: _____

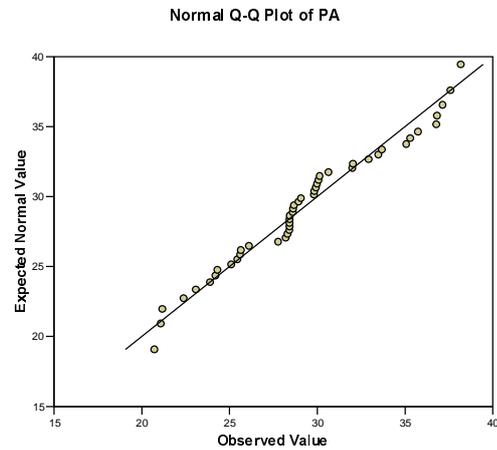
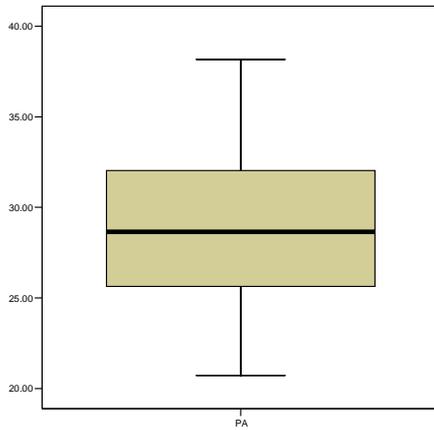
One-Sample Kolmogorov-Smirnov Test

		PA
N		46
Normal Parameters(a,b)	Mean	29,2604
	Std. Deviation	4,60602
Most Extreme Differences	Absolute	,121
	Positive	,121
	Negative	-,104
Kolmogorov-Smirnov Z		,818
Asymp. Sig. (2-tailed)		,515

a Test distribution is Normal. b Calculated from data.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
PA	46	29,2604	4,60602	,67912



a) Com base nos gráficos e/ou na tabela que achar conveniente, indique uma possível distribuição para modelar os dados. Justifique.

b) O investigador pretende calcular um intervalo de confiança a 99% para a variância da população com base nesta amostra de dados. Para tal foi pedido ao João que determinasse o referido intervalo.

c) O investigador pretende também testar se a diminuição média da pressão arterial é superior a 25 mmHg. Para tal foi pedido ao João que efectuasse um teste de hipóteses com nível de significância a 5% com base na região crítica.

d) Designe a média e a variância da população de onde foram retirados os dados por μ e σ^2 , respectivamente. Como sabe, a variância amostral $S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ é um estimador de σ^2 . Verifique que, em média, a variância amostral não aponta para o verdadeiro valor de σ^2 , ou seja, verifique que S^2 não é um estimador centrado para σ^2 .

Nota: Utilize a representação $S^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2$ e recorde que $E[\bar{X}^2] = Var[\bar{X}] + E^2[\bar{X}]$

e) De seguida, o investigador forneceu ao João mais alguns dados sobre o mesmo fármaco (PA₁) comparado com um outro fármaco existente (PA₂). O investigador garantiu que o novo fármaco tinha um efeito na redução da pressão arterial consideravelmente melhor do que o existente, provocando maiores diminuições na pressão arterial. A afirmação feita pelo investigador é sustentada pelos dados? Especifique as hipóteses H₀ e H₁ e diga para que níveis de significância é que rejeita H₀.

Group Statistics

	Index1	N	Mean	Std. Deviation	Std. Error Mean
PA	1	46	29,2604	4,60602	,67912
	2	46	24,6694	3,43723	,50679

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
PA	Equal variances assumed	2,031	,158	2,468	90	,015	2,09094	,84737
	Equal variances not assumed			2,468	83,256	,016	2,09094	,84737

5. O director do Departamento de Bioestatística desenvolveu um trabalho sobre a qualidade de atendimento na urgência. Para tal elaborou um questionário, numa escala de 1 a 10, onde os utentes à saída do hospital classificavam o modo como tinham sido tratados (10-Excelente; 0-Muito mau). Resolveu experimentar este questionário em duas situações: no início do turno (Index1 = 1) e próximo do final do turno (Index1 = 2). O objectivo era avaliar se as classificações obtidas eram **piores** no final turno, isto é, se o cansaço influenciava a qualidade do atendimento. O teste escolhido pelo director foi o teste U de Mann-Whitney e os resultados que obteve no SPSS foram os seguintes:

amostra			
1	N		28
	Mean Rank		30,66
	Sum of Ranks		858,50
2	N		24
	Mean Rank		21,65
	Sum of Ranks		519,50
Total	N		52

	amostra
Mann-Whitney U	219,500
Wilcoxon W	519,500
Z	-2,158
Asymp. Sig. (2-tailed)	,031

a. Grouping Variable: Index1

A conclusão que o director chegou foi que ao nível de significância $\alpha = 0.02$ não rejeitava H_0 e a sua conclusão era que o cansaço de um turno não influenciava significativamente a qualidade do atendimento. Apostou com o João, o seu salário, que a sua conclusão estava correcta.

a) Comente a escolha do teste.

b) Será que o director perdeu a aposta? Justifique.

6. O tempo de um determinado acto cirúrgico foi medido durante um mês e existem 4 técnicas diferentes à escolha do cirurgião para o realizar. O departamento de cirurgia quer saber se o tempo médio medido é significativamente diferente de técnica para técnica. Lamentavelmente um dos colegas do João esqueceu-se de preencher a seguinte tabela:

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups					,000
Within Groups	7878,868	244			
Total	11395,413				

a) Preencha a tabela.

b) Indique a expressão que permite obter o valor do p-value fornecido na tabela. Não se esqueça de especificar a distribuição subjacente à estatística do teste.

c) Conclua, aos níveis usuais de confiança, se as técnicas usadas no acto de cirurgia têm o mesmo valor médio temporal.

d) Usando a seguinte tabela de comparações múltiplas indique, com base nos intervalos de confiança, quais os pares de médias que acusam diferenças significativas ao nível de 0.05, colocando * na coluna sig da tabela.

Multiple Comparisons

Dependent Variable: Met
Tukey HSD

(I) Index1	(J) Index1	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	-3,32072	1,02060		-5,9608	-,6806
	3	4,59126	1,02060		1,9511	7,2314
	4	-5,36725	1,02060		-8,0074	-2,7271
2	1	3,32072	1,02060		,6806	5,9608
	3	7,91198	1,02060		5,2719	10,5521
	4	-2,04653	1,02060		-4,6866	,5936
3	1	-4,59126	1,02060		-7,2314	-1,9511
	2	-7,91198	1,02060		-10,5521	-5,2719
	4	-9,95851	1,02060		-12,5986	-7,3184
4	1	5,36725	1,02060		2,7271	8,0074
	2	2,04653	1,02060		-,5936	4,6866
	3	9,95851	1,02060		7,3184	12,5986

The mean difference is significant at the .05 level.

e) O que pode concluir da tabela anterior, ao nível de significância 0.05?

Após ter recebido os parabéns do director do Departamento de Bioestatística pelo trabalho desenvolvido, uma última missão ainda separa o João das desejadas férias.

7. Através de um inquérito a 28 pessoas, o Departamento de Psicologia está a tentar relacionar o nível de depressão dos seus doentes com as seguintes variáveis: nível de stress, abuso de álcool, idade, abuso de drogas e condições climatéricas. A tabela de regressão obtida foi a seguinte:

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-4,802	1,545		-3,109	,005
	Nível de stress	,381		,399	3,686	,001
	Abuso de álcool	,581	,137	,474	4,247	
	Idade		,078	,210	1,916	,069
	Abuso de drogas	,504	,099	,550		,000
	Condições climatéricas	-,043	,121	-,041	-,358	,724

a Dependent Variable: Nível de depressão

a) Complete os espaços.

b) Identifique as variáveis cujos coeficientes são significativamente diferentes de zero, ao nível de significância de 5%. Justifique.

c) Proceda à validação dos pressupostos do modelo de regressão (análise dos resíduos) usando as **tabelas e/ou gráficos** que achar convenientes.

Model Summary(b)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,866(a)	,750	,694	1,10768

a Predictors: (Constant), Condições climatéricas, Abuso de drogas, Nível de stress, Idade, Abuso de álcool

b Dependent Variable: Nível de depressão

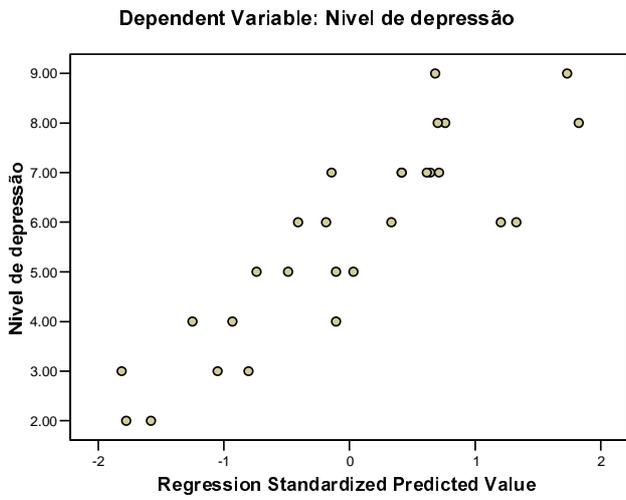
ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	81,114	5	16,223	13,222	,000(a)
	Residual	26,993	22	1,227		
	Total	108,107	27			

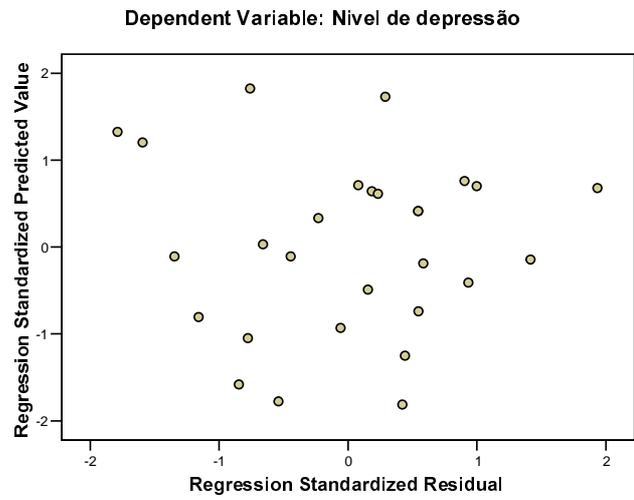
a Predictors: (Constant), Condições climatéricas, Abuso de drogas, Nível de stress, Idade, Abuso de álcool

b Dependent Variable: Nível de depressão

Scatterplot



Scatterplot



Normal P-P Plot of Regression Standardized Residual

