



Departamento de Matemática

Exame Final de Métodos Estatísticos

Data: 1 de Junho de 2006

Duração: 3h00.

Nome: _____ N^o. Mec. _____

Curso: _____ Regime: _____

Declaro que desisto _____

As cotações deste exame encontram-se na seguinte tabela. Responda às questões utilizando o espaço reservado para o efeito. Tenha em atenção a clareza e apresentação das suas respostas.

1				2			3	4	5	6				7			8			9							
1	2	3	4	1	2	3				1	2	3	4	1	2	3	1	2	3	1	2	3	4	5			
1.2	.5	.5	.9	.6	.3	2				.5	.5	.5	.3	1.5	1	.5	0.5	1	1.5	.3	.6	.9	.4	.6	1.8	.6	1

O gestor da agência de viagens *Antistress* pretende compreender melhor as opções dos seus clientes. Com esse objectivo procedeu à realização de vários estudos.

1. Com base numa amostra de tempos (em dias) de duração de viagens a Londres obteve, usando o SPSS, o seguinte *output*.

Descriptives

		Statistic	Std. Error
tempos	Mean	11,78	1,685
	95% Confidence Interval for Mean	8,37	
	Lower Bound		15,19
	Upper Bound		
	5% Trimmed Mean	10,67	
	Median	8,96	
	Variance	113,552	
	Std. Deviation	10,656	
	Minimum	1	
	Maximum	48	
	Range	47	
	Interquartile Range	15	
	Skewness	1,490	,374
	Kurtosis	2,366	,733

Statistics

tempos

	Valid	Missing
N	40	0
Mean	11,78	
Median	8,96	
Mode	1 ^a	
Variance	113,552	
Range	47	
Minimum	1	
Maximum	48	
Percentiles		
25	3,17	
50	8,96	
75	18,42	

a. Multiple modes exist. The smallest value is shown

duração das viagens a Londres Stem-and-Leaf Plot

Frequency	Stem &	Leaf
13,00	0	. 1112222233334
8,00	0	. 56777889
8,00	1	. 00001123
3,00	1	. 889
3,00	2	. 013
3,00	2	. 569
,00	3	.
1,00	3	. 7
1,00	Extremes	(>=48)

Stem width: 10
Each leaf: 1 case(s)

1.1. Construa a caixa de bigodes correspondente a esta amostra.

Vertical line for drawing the box plot.

1.2. Sugira, justificando, uma distribuição (com base na informação dada) para a população subjacente a esta amostra.

Vertical line for justification.

1.3. Indique, utilizando o método dos momentos, uma estimativa pontual para o(s) parâmetro(s) da distribuição seleccionada na alínea anterior.

Vertical line for estimation.

1.4. Comente a validade do intervalo de confiança apresentado para o valor médio da duração da viagem a Londres (**Nota:** lembre-se que o SPSS construiu o intervalo com base na lei t-student).

Vertical line for comment.

Nome: _____ N^o. Mec. _____

2. A *Antistress* organiza regularmente viagens de autocarro a Santiago de Compostela. A experiência das viagens realizadas no passado mostrou que 10% das pessoas que se inscrevem na viagem acabam por desistir.

2.1. Sabendo que são admitidas 50 inscrições na próxima viagem a Santiago de Compostela, então a v.a. que indica o número de desistências nas 50 inscrições tem distribuição _____ com parâmetro(s) _____.

2.2. O número esperado de desistências nas 50 inscrições é de _____.

2.3. Sabe-se que o peso (em kg.) da bagagem de cada passageiro segue uma distribuição normal de média 3 e desvio padrão 0.8.

(i) Determine a probabilidade de um passageiro trazer uma bagagem com mais de 5kg.

(ii) Indique a distribuição do peso total das bagagens admitindo que a viagem se realiza com as 50 pessoas.

3. Dos seguintes métodos gráficos seleccione, com uma cruz, aqueles que podem ser úteis para avaliar a normalidade da população subjacente à amostra:

- histograma;
- diagrama de dispersão;
- PP-plot;
- QQ-plot.

4. Dos seguintes testes seleccione, com uma cruz, aqueles que podem ser úteis para avaliar a normalidade da população subjacente à amostra:

- Mann-Whitney;
- Kolmogorov-Smirnov;
- Kruskal-Wallis;
- Shapiro-Wilk.

5. Indique, com uma cruz, quais dos seguintes factores influencia a amplitude do intervalo de confiança para a média de uma população normal com variância desconhecida.

- diminuição da média amostral;
- aumento da dimensão da amostra;
- diminuição do grau de confiança;
- aumento da variância amostral.

6. A *Antistress* decidiu analisar a variabilidade dos preços (em Euros) praticados por uma companhia aérea nos voos de Lisboa para Paris. Assim recolheu uma amostra aleatória de 16 preços, praticados durante a época baixa, e para a tarifa mais económica. Com base nessa amostra obteve $\bar{x} = 220$ e $s_c = 12.2$. Assuma que se pode considerar que as observações foram retiradas de uma distribuição normal.

6.1. Indique uma estimativa pontual para a respectiva variância dos preços σ^2 , $\hat{\sigma}^2 =$ _____

6.2. Teste com base na construção da região crítica, e a um nível de significância de 5%, as hipóteses $H_0 : \sigma^2 = 100$ vs $H_1 : \sigma^2 > 100$.

Vertical line for answer to 6.2.

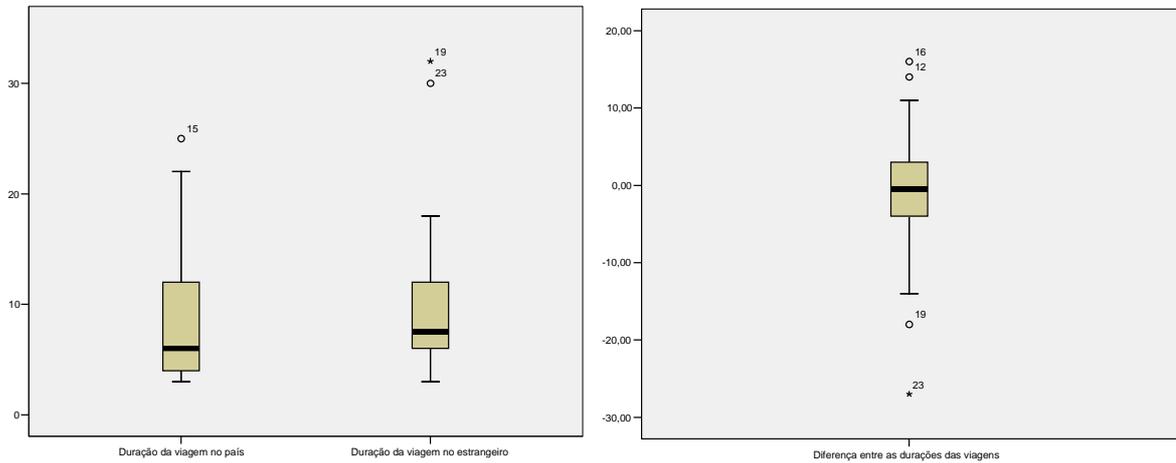
6.3. Calcule o p - *value* do teste realizado na alínea anterior.

Vertical line for answer to 6.3.

6.4. Calcule o p - *value* do teste bilateral $H_0 : \sigma^2 = 100$ vs $H_1 : \sigma^2 \neq 100$.

Vertical line for answer to 6.4.

7. De modo a averiguar se existem diferenças significativas entre a duração das viagens ao estrangeiro e a duração das viagens no país, a *Antistress* seleccionou uma amostra de durações (em dias) de viagens dos dois tipos. Para tal seleccionou aleatoriamente 30 dos seus clientes que efectuaram ambos os tipos de viagens e a partir da ficha de cada cliente seleccionou aleatoriamente uma viagem ao estrangeiro e um viagem no país. Com base nessa amostra obteve o seguinte *output*.

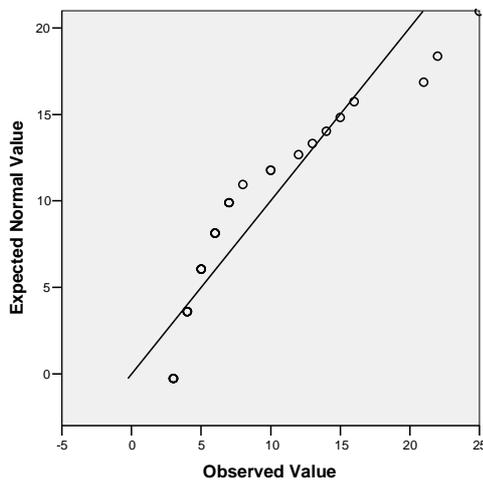


Estimated Distribution Parameters

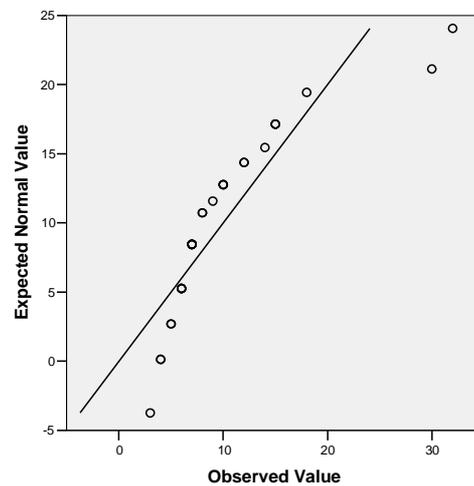
		Duração da viagem no país	Duração da viagem no estrangeiro
Normal Distribution	Location	8,6333	10,1667
	Scale	6,04285	6,80813

The cases are unweighted.

Normal Q-Q Plot of Duração da viagem no país



Normal Q-Q Plot of Duração da viagem no estrangeiro



7.1. Indique, justificando, se as amostras devem ser consideradas emparelhadas ou independentes.

|

7.2. Indique, justificando, que teste se deve realizar.

|

7.3. Suponha que o teste que indicou na alínea anterior forneceu um p -value bilateral no valor de 0.418.

(i) Identifique as hipóteses deste teste identificando os parâmetros em causa.

|

(ii) Calcule o p -value do teste unilateral que averigua se as viagens ao estrangeiro têm duração superior às no país.

|

(iii) Calcule o p -value do teste unilateral que averigua se as viagens ao estrangeiro têm duração inferior às no país.

|

8. Para seleccionar o destino duma nova campanha a *Antistress* decidiu comparar as receitas líquidas (em Euros) que no ano passado tinha obtido com quatro destinos: Brasil, Cuba, Cabo Verde e Macau. Para tal seleccionou aleatoriamente 10 fichas de clientes que efectuem viagens para cada um daqueles destinos.

Após verificados os devidos pressupostos procedeu-se à realização da seguinte análise de variância (ANOVA):

ANOVA

rendimento líquido

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	10897,965	3	3632,655	1,164	,337
Within Groups	112322,95	36	3120,082		
Total	123220,92	39			

8.1. Indique as hipóteses (nula e alternativa) do teste efectuado.

8.2. Aos níveis de significância usuais (1, 5 e 10%) que conclusões pode retirar do *output* apresentado?

8.3. Para esta situação em concreto acha que se deve proceder a uma análise de comparações múltiplas? Justifique.

9. Pretendendo avaliar a influência da distância do destino da viagem (em km) na duração das férias (em dias), procedeu-se à realização do seguinte estudo usando o SPSS.

Model Summary					ANOVA ^b					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Model	Sum of Squares	df	Mean Square	F	Sig.
1	,692 ^a	,478	,441	4,48473	1	Regression 258,171	1	258,171	12,836	,003 ^a
						Residual 281,579	14	20,113		
						Total 539,750	15			

a. Predictors: (Constant), Distância dos destinos

a. Predictors: (Constant), Distância dos destinos

b. Dependent Variable: Duração das férias

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	5,926	1,673		3,542	,003	2,338	---
	Distância dos destinos	,002	,001	,692	3,583	,003	---	---

a. Dependent Variable: Duração das férias

9.1. Indique a expressão da recta de regressão obtida pelo SPSS.

9.2. A um nível de significância $\alpha = 0.05$, diga, justificando, se ambos os coeficientes são significativamente diferentes de zero.

9.3. Preencha os espaços em falta na tabela de coeficientes. (**Sugestão:** para o cálculo do intervalo de confiança para o declive note que sempre que um parâmetro θ é estimado por um estimador $\hat{\theta}$ com distribuição $N(\theta, \sigma_{\hat{\theta}}^2)$ tem-se que $\frac{\hat{\theta} - \theta}{\hat{\sigma}_{\hat{\theta}}} \sim t_m$ e um IC para θ é dado por $\hat{\theta} \pm t_{1-\frac{\alpha}{2}, m} \hat{\sigma}_{\hat{\theta}}$, onde m representa o número de graus de liberdade associados à distribuição do estimador $\hat{\sigma}_{\hat{\theta}}^2$.)

Para o declive ($\theta = b_1$) especifique os seguintes valores:

$\hat{\theta} =$

$\hat{\sigma}_{\hat{\theta}} =$

$t_{1-\frac{\alpha}{2}, m} =$

9.4. Com base no intervalo de confiança obtido na alínea anterior teste se a ordenada na origem é significativamente diferente de 7.

9.5. O que pode afirmar relativamente à qualidade e significado do modelo de regressão obtido? Justifique.