

**Estatística Computacional**  
(Licenciatura em Matemática)

Duração: 2h

Exame

14/06/10

NOME: \_\_\_\_\_

**Observação:** A resolução completa das perguntas inclui a justificação do raciocínio utilizado.

O gerente de uma agência imobiliária decidiu efectuar um estudo estatístico sobre os preços dos apartamentos usados que a agência tem para venda em duas cidades do país. Para tal, seleccionou aleatoriamente, dos arquivos da agência, 95 apartamentos daquele tipo e registou, para cada um, a localização (cidade A ou B), a distância a que se encontra do centro da cidade (em km), a tipologia (T1, T2, T3 ou T4), a idade (em anos), o preço (em milhares de euros) e a área (em metros quadrados).

Apresentam-se seguidamente alguns excertos desse estudo.

1. A amostra relativa aos preços dos apartamentos foi resumida do seguinte modo:

Preço de Venda (em milhares de euros)

Frequency	Stem &	Leaf
2,00	0 .	89
12,00	1 .	122223334444
15,00	1 .	555667788899999
20,00	2 .	00001111122233334444
22,00	2 .	555566667777778888899
12,00	3 .	000111222344
7,00	3 .	6677789
3,00	4 .	001
1,00	4 .	5
1,00	Extremes	(>=600)

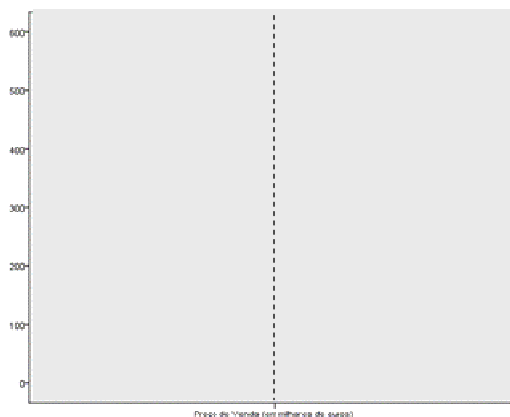
Stem width: 100  
Each leaf: 1 case(s)

**Statistics**

Preço de Venda (em milhares de euros)

N	Valid	95
	Missing	0
Mean		249,35
Variance		7818,021
Skewness		,773
Range		520
Percentiles	25	187,00
	50	243,00
	75	300,00

- a) Indique a proporção de apartamentos que custam mais de 300 mil euros.
- b) Trace o correspondente diagrama de extremos-e-quartis.



- c) Interprete o valor do 1º quartil.

- d) Das medidas descritivas apresentadas no quadro da página anterior, indique as que são fortemente influenciadas pela presença de *outliers*.
- e) Supondo que os preços dos apartamentos são convertidos em dólares (1 Euro=1.21 Dólares), converta os valores das referidas medidas, completando o quadro que se segue:

Preço de Venda (em milhares de dólares)

N	Valid	95
	Missing	0
Mean		
Variance		
Skewness		
Range		
Percentiles	25	

2. Por forma a encontrar um modelo para a distribuição da idade de um apartamento naquelas condições, foi efectuado o teste de Kolmogorov-Smirnov, cujo resultado figura no próximo quadro.

**One-Sample Kolmogorov-Smirnov Test**

		Idade
N		95
Poisson Parameter <sup>a, b</sup>	Mean	7,71
Most Extreme Differences	Absolute	,098
	Positive	,058
	Negative	-,098
Kolmogorov-Smirnov Z		,956
Asymp. Sig. (2-tailed)		,321

a. Test distribution is Poisson.

b. Calculated from data.

- a) Formule as hipóteses em confronto. Que conclusão pode tirar, aos níveis de significância usuais?
- b) Tendo em conta a decisão a que é conduzido, determine uma estimativa para a probabilidade de ser superior a 2 anos a idade de um apartamento usado que se encontra à venda naquela agência.
3. Outro aspecto analisado diz respeito à tipologia dos apartamentos.
- a) A fim de inferir sobre a proporção de apartamentos de cada um dos tipos, realizou-se o teste do qui-quadrado de ajustamento, que forneceu o seguinte *output*:

Tipologia			
	Observed N	Expected N	Residual
T1	18	23,8	-5,8
T2	42	23,8	18,3
T3	24	23,8	,3
T4	11	23,8	-12,8
Total	95		

Test Statistics	
	Tipologia
Chi-square	22,263 <sup>a</sup>
df	3
Asymp. Sig.	,000

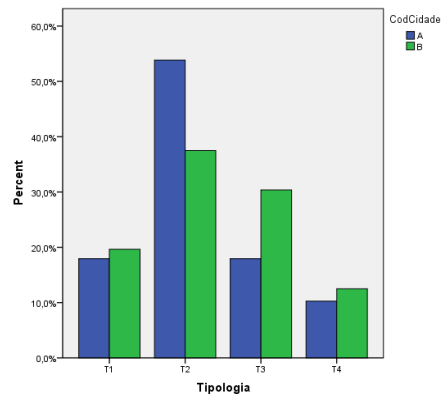
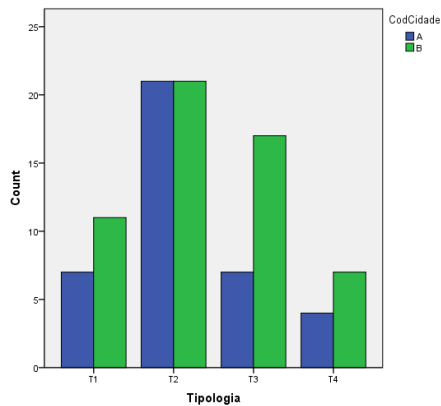
a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 23,8.

(i) Quais são as hipóteses em confronto?

(ii) Como é definida a estatística de teste?

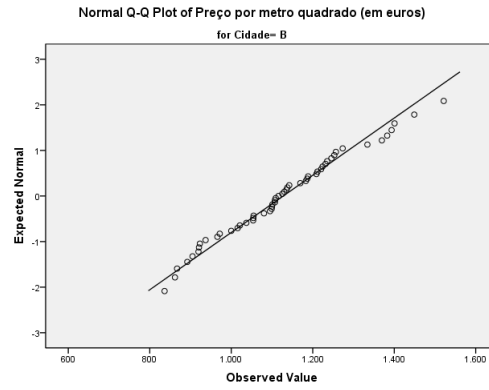
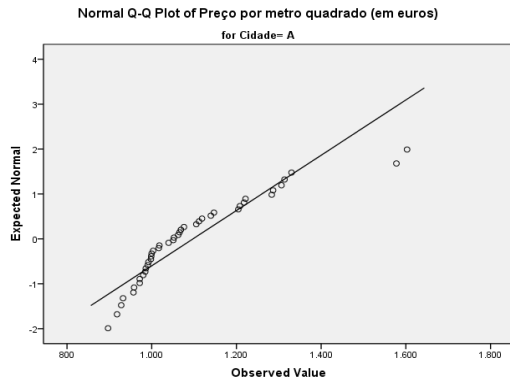
(iii) Que pode concluir?

b) Pretende-se também elaborar um gráfico de barras que permita comparar o mercado de vendas nas duas cidades segundo a tipologia do apartamento. Qual dos gráficos apresentados lhe parece mais adequado para o efeito? Justifique.



4. O preço por metro quadrado de um apartamento nas condições referidas foi outra das características estudadas. Após a análise descritiva da amostra global, procedeu-se à comparação dos preços nas duas cidades.

a) Para verificar os pressupostos do teste paramétrico utilizado para comparar as localizações de duas distribuições, foram traçados, com base em cada uma das amostras anteriores, os seguintes papéis de probabilidade:



(i) Que informação fornecem estes gráficos?

(ii) Qual é a ideia que preside à sua construção?

b) Optou-se por efectuar tal comparação através de um teste não paramétrico, que conduziu aos seguintes resultados:

Ranks

	CodCidade	N	Mean Rank	Sum of Ranks
Preço por metro quadrado (em euros)	A	39	40,59	1583,00
	B	56	53,16	2977,00
	Total	95		

Test Statistics<sup>a</sup>

	Preço por metro quadrado (em euros)
Mann-Whitney U	803,000
Wilcoxon W	1583,000
Z	-2,186
Asymp. Sig. (2-tailed)	,029
Exact Sig. (2-tailed)	,029
Exact Sig. (1-tailed)	,014

a. Grouping Variable: CodCidade

(i) Como são calculados os valores da coluna *Sum of Ranks*?

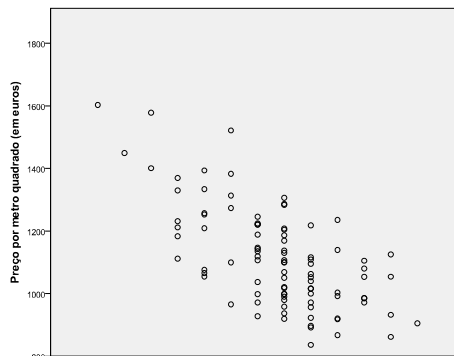
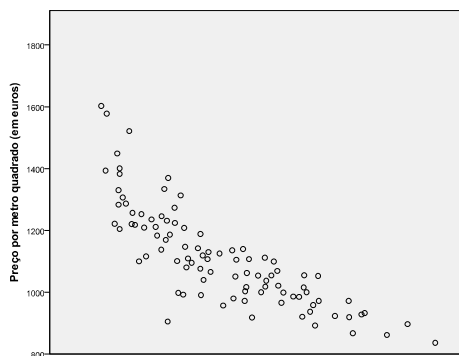
(ii) Indique uma expressão que permita calcular o valor de prova assintótico.

(iii) Que conclusão pode tirar?

(iv) Refira uma vantagem e uma desvantagem deste teste face ao teste paramétrico alternativo.

5. Pensa-se que dois dos factores que mais influenciam o preço por metro quadrado de um apartamento usado são a idade e a distância a que se encontra do centro da cidade.

a) Com o objectivo de detectar uma eventual relação entre estas características, foram traçados diagramas de dispersão e calculados os coeficientes de correlação que se apresentam. Associe a cada diagrama o quadro respectivo.



Symmetric Measures

		Value
Interval by Interval	Pearson's R	-.817
Ordinal by Ordinal	Spearman Correlation	-.854
N of Valid Cases		95

Symmetric Measures

		Value
Interval by Interval	Pearson's R	-.670
Ordinal by Ordinal	Spearman Correlation	-.618
N of Valid Cases		95

b) Na sequência das conclusões obtidas na alínea anterior, foi ajustado um modelo de regressão linear simples por forma a exprimir preço por metro quadrado de um apartamento em função da sua distância ao centro da cidade. Os resultados fornecidos pelo SPSS foram os seguintes:

ANOVA<sup>b</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1614707,777	1	1614707,777	186,078	,000 <sup>a</sup>
	Residual	807015,354	93	8677,584		
	Total	2421723,131	94			

a. Predictors: (Constant), Distância ao centro da cidade (em km)

b. Dependent Variable: Preço por metro quadrado (em euros)

Coefficients<sup>a</sup>

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1365,561	20,755		65,795	,000
	Distância ao centro da cidade (em km)	-103,604	7,595	-.817	-13,641	,000

a. Dependent Variable: Preço por metro quadrado (em euros)

(i) Obtenha a equação da recta de regressão estimada e interprete o valor do coeficiente associado ao regressor.

(ii) Com base na 2ª tabela, escreva a equação da análise da variância associada ao modelo.

(iii) Relacione os resultados do teste F que figuram na tabela ANOVA com os do teste à nulidade do coeficiente associado ao regressor.

(iv) Porque razão a utilização do coeficiente de determinação se deve revestir de algum cuidado?

6. Tendo constatado que, passado um ano, muitos dos apartamentos publicitados ainda não tinham sido vendidos, o director da agência procurou aprofundar o estudo destes casos.

a) O teste binomial, cujos resultados figuram no próximo quadro, permite inferir sobre a proporção de apartamentos vendidos naquele período.

	Category	N	Observed Prop.	Test Prop.	Asymp. Sig. (1-tailed)
Vendido	Group 1	Sim	56	,6	,456 <sup>a b</sup>
	Group 2	Não	39	,4	
	Total		95	1,0	

a. Alternative hypothesis states that the proportion of cases in the first group < .6.

b. Based on Z Approximation.

(i) Qual é o critério utilizado pelo SPSS para definir os grupos?

(ii) Indique as hipóteses do teste subjacente a este *output*. Que conclusão pode tirar, aos níveis de significância usuais?

(iii) Em que condições o SPSS fornece o valor de prova de um teste bilateral?

b) A fim de analisar a influência da tipologia do apartamento no sucesso das vendas, foi elaborada a tabela de contingência seguidamente apresentada.

Tipologia \* Vendido Crosstabulation

			Vendido		Total
			Não	Sim	
Tipologia	T1	Count	4	14	18
		Expected Count			
		% within Tipologia	a%		
		% within Vendido			
	% of Total				
	T2	Count	8	34	42
		Expected Count		b	
		% within Tipologia			
		% within Vendido	c%		
		% of Total			
	T3	Count	18	6	24
		Expected Count			
		% within Tipologia			
		% within Vendido			e%
		% of Total			
	T4	Count	9	2	11
		Expected Count			d
		% within Tipologia			
		% within Vendido			
		% of Total	g%		
Total		Count	39	56	95
		Expected Count			
		% within Tipologia			
		% within Vendido		f%	
		% of Total			

(i) Indique os valores de a, b, c, d, e, f, g.

(ii) Qual o significado das frequências designadas por *Expected Count*?

(iii) Dos seguintes coeficientes, assinale aqueles que podem ser utilizados para medir a associação entre os atributos anteriores:

Coeficiente gama

Coeficiente de contingência de Pearson

Coeficiente V de Cramer

Coeficiente tau-b de Kendall

(iv) Que teste poderia utilizar para inferir sobre uma eventual relação entre os atributos em causa ?

c) Por fim, decidi averiguar se o insucesso das vendas provocou alguma redução nos preços de tais apartamentos naquele período. Para tal, efectuei um teste t de comparação de médias, que conduziu aos seguintes resultados:

**Paired Samples Statistics**

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Preço do apartamento (em milhares de euros)	227,50	39	24,824	3,975
	Novo Preço (em milhares de euros)	226,67	39	24,771	3,967

**Paired Samples Test**

		Paired Differences				t	df	Sig. (2-tailed)	
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower				Upper
Pair 1	Preço do apartamento (em milhares de euros) - Novo Preço (em milhares de euros)	,837	2,200	,352	,124	1,550	2,377	38	,023

(i) O intervalo de confiança apresentado diz respeito a que parâmetro?

(ii) Para que níveis de significância poderá afirmar que, em média, ocorreu uma redução no preço de um apartamento não vendido?

(iii) Indique o(s) teste(s) não paramétrico(s) alternativo(s) ao teste anterior.