

Estatística Computacional
(Licenciatura em Matemática)

Duração: 2h

Frequência

24-05-2011

NOME: _____

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

I

Uma equipa médica desenvolveu um estudo envolvendo 271 doentes com perturbações gastrointestinais, de ambos os sexos e de várias faixas etárias. Para cada doente registou-se o sexo, a idade (em anos), o facto de ser ou não diabético e o grau de gravidade de alguns sintomas (Ausente, Ligeiro, Moderado, Grave, Muito grave, Excessivo).

1. Os valores observados das idades conduziram aos seguintes resultados:

Idade Stem-and-Leaf Plot

```

Frequency      Stem & Leaf

    5,00 Extremes      (<=27)
    1,00          3 .  3
    7,00          3 .  6777899
   16,00          4 .  0111122333444444
   24,00          4 .  555556667777788889999999
   42,00          5 .  0000000011112222222233333333444444444444
   53,00          5 .  55555555556666666677777777777778888888889999999999
   40,00          6 .  0000000011122222222222333333334444444444
   53,00          6 .  55555555556666666666666677777777788888888999999999
   24,00          7 .  000000000011111112222334
    4,00          7 .  5579
    2,00          8 .  11

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

Statistics

Idade		
N	Valid	271
	Missing	0
Mean		57,75
Median		58,00
Mode		66
Std. Deviation		10,387
Skewness		-,668
Std. Error of Skewness		,148

Percentiles

		Percentiles						
		5	10	25	50	75	90	95
Weighted Average (Definition 1)	Idade	39,60	44,00	52,00	58,00	66,00	70,00	71,40

a) Que idade tinha o doente mais velho?

b) Qual a percentagem de doentes com menos de 40 anos?

c) Interprete o valor 44 que figura no quadro *Percentiles*.

d) Descreva a amostra no que diz respeito à localização central e à dispersão.

e) Que pode afirmar sobre a assimetria da amostra?

2. No quadro seguinte figuram os resultados do teste de Kolmogorov-Smirnov, com a correcção de Lilliefors, efectuado com base na amostra observada das idades.

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Idade	,072	271	,002	,970	271	,000

a. Lilliefors Significance Correction

a) Quais são as hipóteses em teste?

b) Qual a decisão a tomar relativamente às hipóteses referidas na alínea anterior?

c) Considera legítima a aplicação do teste de Student para verificar se a média das idades é inferior a 59 anos?

d) Apresenta-se a seguir o *output* do teste de Student.

One-Sample Test

	Test Value = 59					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Idade	-1,977	270	,049	-1,247	-2,49	,00

(i) A que parâmetro corresponde o intervalo de confiança apresentado no quadro acima?

(ii) Qual é o p-valor do teste referido na alínea c)? A que conclusão conduz este valor?

3. Neste estudo investigaram-se possíveis relações entre a presença dos referidos sintomas e a presença da diabetes. Em particular, no que diz respeito a um determinado sintoma, designado *sintoma A*, obteve-se a seguinte tabela de contingência e o correspondente teste do qui-quadrado.

Diabetes * Sintoma A Crosstabulation

Count

		Sintoma A		Total
		Ausente	Presente	
Diabetes	Não diabético	94	37	131
	Diabético	89	51	140
Total		183	88	271

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)
Pearson Chi-Square	2,067 ^a	1	,150	,156
Continuity Correction ^b	1,711	1	,191	
N of Valid Cases	271			

a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 42,54.
 b. Computed only for a 2x2 table

a) Indique

(i) a percentagem de doentes que apresentam o sintoma A;

(ii) a percentagem de doentes diabéticos que apresentam o sintoma A.

b) O teste efectuado permite concluir que há associação entre a presença do sintoma A e a presença da diabetes?

4. A análise da relação entre o grau de gravidade do sintoma A e o grau de gravidade de outro sintoma, designado *sintoma B*, nos doentes diabéticos, conduziu ao seguinte *output*:

Sintoma A * Sintoma B Crosstabulation^a

			Sintoma B						Total
			Ausente	Ligeiro	Moderado	Grave	Muito grave	Excessivo	
Sintoma A	Ausente	Count	62	9	7	6	2	3	89
		Expected Count	55,9	8,3	6,4	7,0	5,1	6,4	89,0
	Ligeiro	Count	7	1	2	1	3	0	14
		Expected Count	8,8	1,3	1,0	1,1	,8	1,0	14,0
	Moderado	Count	12	1	0	2	0	2	17
		Expected Count	10,7	1,6	1,2	1,3	1,0	1,2	17,0
	Grave	Count	4	1	0	1	3	2	11
		Expected Count	6,9	1,0	,8	,9	,6	,8	11,0
	Muito grave	Count	3	1	1	1	0	3	9
		Expected Count	5,7	,8	,6	,7	,5	,6	9,0
Total		Count	88	13	10	11	8	10	140
		Expected Count	88,0	13,0	10,0	11,0	8,0	10,0	140,0

a. Diabetes = Diabético

Chi-Square Tests^b

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	40,083 ^a	20	,005
Likelihood Ratio	34,179	20	,025
Linear-by-Linear Association	13,956	1	,000
N of Valid Cases	140		

a. 20 cells (66,7%) have expected count less than 5. The minimum expected count is ,51.

b. Diabetes = Diabético

a) Qual é o significado dos valores *Count* e *Expected Count* que figuram no primeiro quadro?

b) Será recomendável tomar uma decisão com base neste *output*?

5. Pretendeu-se também tirar conclusões sobre a percentagem de diabéticos que não apresentam o sintoma B. Para tal, foi efectuado um teste binomial, cujos resultados figuram a seguir.

Binomial Test^a

	Category	N	Observed Prop.	Test Prop.	Exact Sig. (1-tailed)
Sintoma B	Group 1	Presente	52	,37	,326
	Group 2	Ausente	88	,63	
	Total		140	1,00	

a. Diabetes = Diabético

a) Quais são as hipóteses em teste subjacentes a este *output*?

b) A percentagem de diabéticos que não apresentam o sintoma B pode ser considerada superior a 65%?

II

O gestor de uma agência de viagens fez vários estudos para melhor compreender as opções dos seus clientes relativamente a viagens turísticas.

1. No que diz respeito a uma amostra de tempos de duração das viagens ao estrangeiro (em dias), obteve o seguinte *output*:

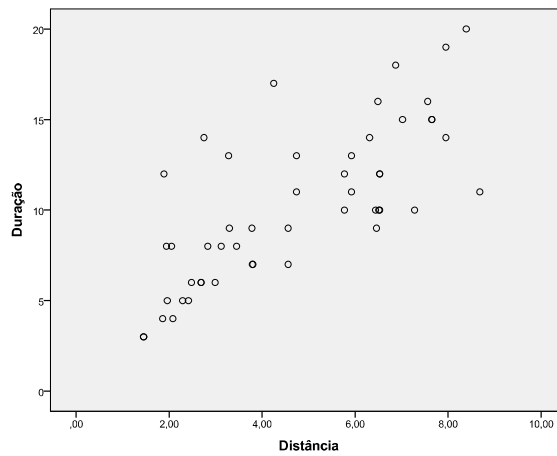
	Duração
Test Value ^a	10
Cases < Test Value	23
Cases >= Test Value	27
Total Cases	50
Number of Runs	24
Z	-,529
Asymp. Sig. (2-tailed)	,597

a. Median

a) Como são definidos os *Runs*?

b) Quais as hipóteses em confronto? Que pode concluir aos níveis de significância usuais?

2. Para cada viagem, foi também registada a região de destino. Para relacionar a duração da viagem com a distância ao seu destino (em milhares de quilómetros), foram construídos o diagrama de dispersão e o quadro que se segue.



Dos quadros seguintes, contendo valores dos coeficientes de correlação de Pearson e de Spearman, qual poderá corresponder a este diagrama de dispersão?

Symmetric Measures			Symmetric Measures		
		Value			Value
Interval by Interval	Pearson's R	,022	Interval by Interval	Pearson's R	,751
Ordinal by Ordinal	Spearman Correlation	,036	Ordinal by Ordinal	Spearman Correlation	,766
N of Valid Cases		50	N of Valid Cases		50

Quadro 1

Quadro 2

Symmetric Measures			Symmetric Measures		
		Value			Value
Interval by Interval	Pearson's R	-,679	Interval by Interval	Pearson's R	,679
Ordinal by Ordinal	Spearman Correlation	-,634	Ordinal by Ordinal	Spearman Correlation	,898
N of Valid Cases		50	N of Valid Cases		50

Quadro 3

Quadro 4

3. Com o objectivo de verificar se as regiões de destino das viagens eram igualmente preferidas pelos clientes da agência, foi efectuado um teste de ajustamento do qui-quadrado, que forneceu os seguintes resultados:

Destino			
	Observed N	Expected N	Residual
África	14	12,5	1,5
Brasil	11	12,5	-1,5
Caraíbas	9	12,5	-3,5
Europa Central	16	12,5	3,5
Total	50		

Test Statistics	
	Destino
Chi-Square	2,320
df	3
Asymp. Sig.	,509

Relativamente ao objectivo referido, que pode concluir?

III

Sejam X e Y duas variáveis aleatórias absolutamente contínuas e independentes. Encontram-se abaixo alguns resultados da análise descritiva de duas amostras, uma da variável X e outra da variável Y , bem como o resultado do teste de Mann-Whitney aplicado a estas amostras.

Descriptives			
Variável		Statistic	Std. Error
X	Mean	,9081	,05565
	Median	,6630	
	Std. Deviation	,87997	
	Minimum	,00	
	Maximum	4,77	
	Range	4,76	
	Interquartile Range	1,03	
	Skewness	1,569	,154
	Kurtosis	2,574	,307
Y	Mean	,4618	,05008
	Median	,6630	
	Std. Deviation	,77581	
	Minimum	-2,14	
	Maximum	2,00	
	Range	4,14	
	Interquartile Range	,98	
	Skewness	-1,200	,157
	Kurtosis	1,056	,313

Test Statistics ^a	
Mann-Whitney U	24974,500
Z	-3,207
Asymp. Sig. (2-tailed)	,001

a. Grouping Variable: Variável

1. Como explica o resultado do teste de Mann-Whitney, face à igualdade das medianas das amostras observadas?

2. Pretende-se agora testar a hipótese $H_0: \mu_Y = 0.75$, onde μ_Y representa a mediana de Y .

a) Qual é a hipótese alternativa que está de acordo com a tendência da amostra?

b) Que teste usaria neste caso?

c) O SPSS fornece o valor 0.107 para o p-valor assintótico do teste bilateral adequado para esta situação. Para que valores de significância aceitaria a hipótese alternativa a que se refere a alínea a)?

3. Seja Z uma outra variável aleatória real absolutamente contínua da qual se dispõe uma amostra tal que o coeficiente de correlação amostral entre X e Z é igual a 0.00004. Este valor permite concluir que X e Z não estão relacionadas?