

Estatística Computacional
(Licenciatura em Matemática)

Duração: 2h

Exame

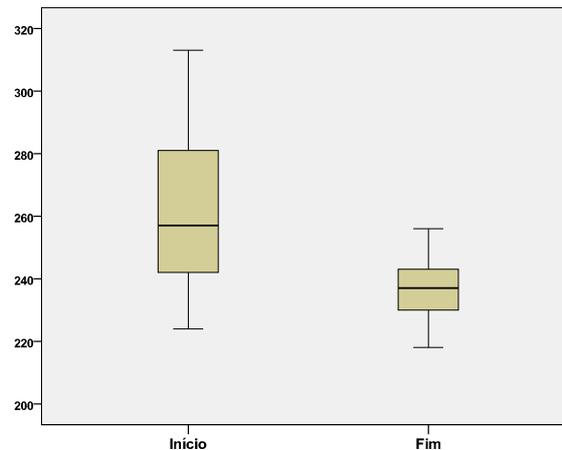
28-06-2011

NOME: _____

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

1. Para analisar a eficácia de um novo medicamento para a redução do nível de colesterol total, um laboratório seleccionou aleatoriamente 15 doentes de determinada população com esta patologia. Foi medido o nível de colesterol de cada um deles no início e no fim do tratamento com o referido medicamento.

a) Na figura seguinte estão representados os *Box-Plot* correspondentes aos registos do nível de colesterol no início e no fim do tratamento com o referido medicamento. Compare os gráficos relativamente à localização e à dispersão e tire conclusões sobre a eficácia do medicamento.



b) Pretende-se agora efectuar um teste de Student para averiguar se houve, de facto, redução do nível médio de colesterol total com aquele tratamento.

(i) Sob que condições pode usar este teste?

(ii) Admita que tais condições são verificadas. O resultado do teste efectuado no SPSS é apresentado no quadro que se segue. Que pode concluir?

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	Fim - Início	-25,000	21,037	5,432	-36,650	-13,350	-4,603	14	,000

(iii) Relativamente ao quadro apresentado na alínea anterior, o que significam os valores da coluna *95% Confidence Interval of the Difference*?

2. Num estudo sobre técnicas utilizadas para medir o volume dos pulmões, foram recolhidos dados fisiológicos de 10 indivíduos adultos, escolhidos ao acaso. Para cada indivíduo, observaram-se as capacidades residuais funcionais (expressas em litros) tanto na posição sentada como na posição deitada.

a) Uma vez que as amostras são de dimensão reduzida, optou-se pelo teste dos sinais para averiguar se existem diferenças significativas entre as medidas registadas nas duas posições. Formule as hipóteses em teste.

b) O *output* deste teste, obtido no SPSS, é o seguinte:

		N
Deitado - Sentado	Negative Differences ^a	10
	Positive Differences ^b	0
	Ties ^c	0
	Total	10

a. Deitado < Sentado
 b. Deitado > Sentado
 c. Deitado = Sentado

	Deitado - Sentado
Exact Sig. (2-tailed)	,002 ^a
Exact Sig. (1-tailed)	,001
Point Probability	,001

a. Binomial distribution used.
 b. Sign Test

Que pode concluir aos níveis de significância usuais?

3. No âmbito dos ensaios clínicos de um novo medicamento, dois investigadores pretendem averiguar se existe alguma relação de dependência entre a concentração plasmática de determinada proteína e o tempo que decorre até o medicamento fazer efeito. Com este objectivo recolheram uma amostra de 60 pacientes e registaram a concentração da proteína (em µg/ml) em cada um deles. Administraram então o medicamento a cada um dos pacientes e observaram o tempo (em minutos) até o medicamento fazer efeito.

a) Um dos investigadores optou por classificar as duas amostras e testar a independência entre os atributos a partir de uma tabela de contingência. A tabela que obteve foi a seguinte:

Count		Tempo			Total
]47,77]]77,107]]107,137]	
Concentração]9,16]	11	4	5	20
]16,23]	3	12	7	22
]23,30]	6	4	8	18
Total		20	20	20	60

(i) Dos pacientes com uma concentração plasmática da proteína superior a 23 µg/ml, qual é a percentagem dos que sentiram o efeito do medicamento passados mais de 107 minutos?

(ii) Qual é a percentagem de pacientes nos quais a concentração plasmática da proteína não excede 23 µg/ml?

(iii) Qual é o teste adequado para testar a independência acima referida? Em que condições pode ser usado o p-valor assintótico para este teste?

(iv) Sabendo que o p-valor do teste é aproximadamente igual a 0.025, que decisão deve tomar ao nível de significância 0.05? (Nota: as condições referidas na alínea (iii) são verificadas neste caso.)

b) O outro investigador usou os dados originais e procedeu à sua análise através do menu do SPSS *Analyze* → *Regression* → *Linear*, tendo obtido, entre outros, o seguinte *output*:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,055 ^a	,003	-,014	26,340

a. Predictors: (Constant), Concentração

(i) Como interpreta o valor 0.055 que figura neste quadro?

(ii) Parece-lhe que os resultados obtidos pelos dois investigadores são contraditórios?

c) O *output* seguinte permite tirar conclusões sobre o 1º quartil da variável aleatória real que representa o tempo que o medicamento leva a fazer efeito. Comente esta afirmação, apresentando as suas conclusões.

Binomial Test

	Category	N	Observed Prop.	Test Prop.	Exact Sig. (1-tailed)	
Tempo	Group 1	<= 62	17	,28	,25	,320
	Group 2	> 62	43	,72		
Total			60	1,00		

4. Determinada fábrica possui uma máquina que embala pacotes com 1 kg de açúcar. A secção de controlo de qualidade analisa regularmente o estado da máquina, recolhendo amostras de 100 pacotes e registando os respectivos pesos (em gramas).

a) A amostra recolhida num dos controlos de qualidade forneceu o seguinte *output* do teste de *runs*:

	Peso
Test Value ^a	1000,0000
Total Cases	100
Number of Runs	54
Z	,759
Asymp. Sig. (2-tailed)	,448

a. User-specified.

(i) Como são definidos os runs neste caso concreto?

(ii) Quais são as hipóteses em teste? Perante a amostra observada, qual delas deve aceitar?

b) A opção *Normality plots with tests* do menu *Analyze* → *Descriptive Statistics* → *Explore* do SPSS forneceu o seguinte quadro:

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Peso	,047	100	,200*	,993	100	,909

a. Lilliefors Significance Correction

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

Indique as hipóteses em confronto. Que pode concluir?

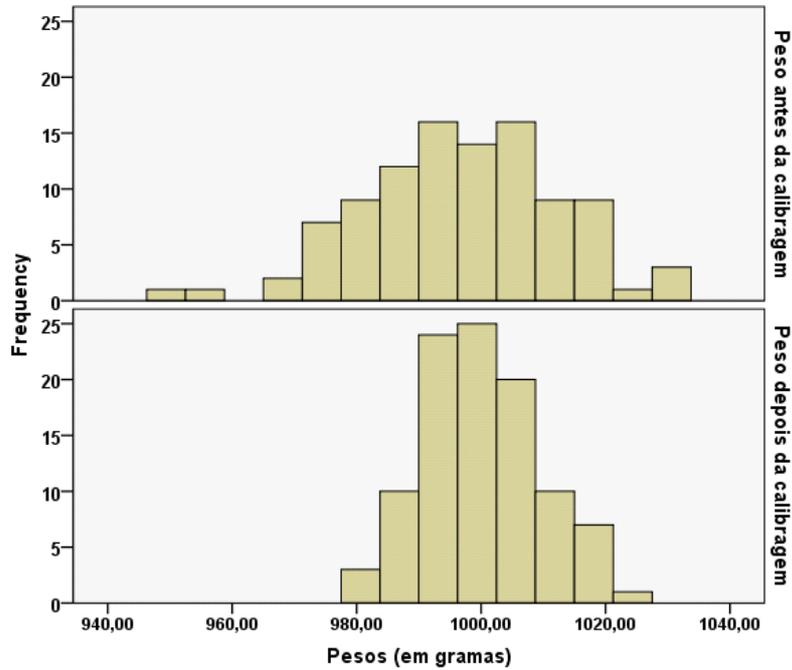
c) Suspeitando-se que os pacotes apresentam, em média, um peso inferior ao estipulado, foi efectuado um teste ao valor médio do peso por pacote, cujos resultados figuram a seguir.

One-Sample Test

	Test Value = 1000					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Peso	-1,948	99	,054	-3,03081	-6,1172	,0556

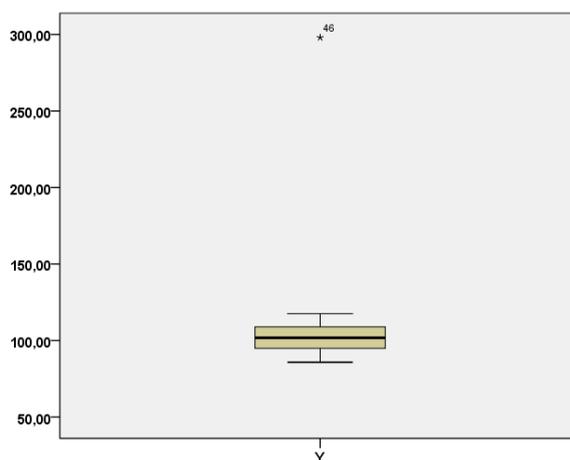
Poderá afirmar que a máquina está efectivamente a prejudicar os consumidores relativamente ao peso médio de um pacote de açúcar?

d) Depois de se proceder à calibragem da máquina, foi recolhida uma nova amostra de dimensão 100 e pesados os pacotes da amostra. Na figura que se segue apresentam-se os histogramas associados às duas amostras. Compare as distribuições dos pesos das duas amostras relativamente à localização, à dispersão e à forma e pronuncie-se sobre a eficácia do procedimento de calibragem.



5. Com o objectivo de explicar a relação de dependência entre uma variável aleatória, Y , e uma variável não aleatória, x , observaram-se 46 concretizações do par (x, Y) .

a) Começou por se efectuar a análise descritiva da variável estatística marginal Y , na qual foi obtido o *Box-Plot* que figura a seguir. Como procederia perante este resultado?



b) Depois de efectuado o devido procedimento, uma nova análise descritiva da variável Y conduziu aos seguintes resultados:

Descriptives			Statistic	Std. Error
Y	Mean		102,2400	1,27256
	95% Confidence Interval for Mean	Lower Bound	99,6753	
		Upper Bound	104,8046	
	5% Trimmed Mean		102,2277	
	Median		101,4420	
	Variance		72,873	
	Std. Deviation		8,53658	
	Minimum		85,86	
	Maximum		117,54	
	Interquartile Range		13,92	
	Skewness		,172	,354

(i) Como interpreta o facto da média da amostra ser aproximadamente igual à média aparada a 5%?

(ii) Que pode afirmar relativamente à assimetria da amostra?

c) Procedeu-se seguidamente ao ajustamento de um modelo de regressão linear, tendo sido obtidos os resultados que figuram nos quadros seguintes.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,993 ^a	,987	,986	1,00046

a. Predictors: (Constant), x

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	,440	1,817		,242	,810
	x	1,988	,035	,993	56,218	,000

a. Dependent Variable: Y

(i) Parece-lhe que um modelo de regressão linear é adequado para relacionar Y com x ?

(ii) Admitindo que estão asseguradas as hipóteses do modelo gaussiano, será legítimo considerar que a recta de regressão correspondente contém a origem das coordenadas?