Universidade do Minho
Escola de Engenharia
Departamento de Produção e Sistemas

ESTATÍSTICA APLICADA

Licenciatura em Engenharia Informática

18 de Janeiro de 2010 – 2ª Frequência

Duração 2 horas

Nº **Nome:**

Leia com atenção os enunciados e apresente todos os cálculos que tiver de efectuar na resolução dos exercícios. Justifique todas as respostas. Boa Sorte! RUBRIQUE TODAS AS FOLHAS.

1. Numa certa firma, duas máquinas produzem peças metálicas que, de acordo com uma norma estabelecida, devem ter um comprimento médio igual a 6 cm. Uma amostra aleatória de 73 peças da produção da máquina I e uma outra de 61 peças da máquina II conduziram aos resultados seguintes:

Máquina	n_i	$\overline{x}_i$	s_i^2
Máquina I	73	5.95	0.018
Máquina II	61	6.01	0.020

a) Será de admitir que o comprimento médio da amostra I é inferior ao estabelecido pela norma, para um nível de significância de 5%?

b) Haverá evidência suficiente nos resultados obtidos que nos permita concluir pela diferença significativa entre os comprimentos médios das duas amostras, ao nível de significância de 5%?

Resposta: a) Teste ao valor médio unilateral para grandes amostras

1. Formulação das hipóteses

$H_0 : \mu = 6.0$ (O comprimento médio da amostra I é igual ao estabelecido pela norma)

$H_1 : \mu < 6.0$ (O comprimento médio da amostra I é inferior ao estabelecido pela norma)

$RR : Z < -c$

2. Região crítica: $z < -z_{0.95} = -1.65$

3. Teste estatístico

$$z \approx \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{5.95 - 6}{\sqrt{0.018}/\sqrt{73}} = -3.18$$

4. Decisão

Como -3.18<-1.65, rejeita-se a Hipótese Nula, ou seja, o comprimento médio das peças produzidas pela máquina I é inferior ao estabelecido pela norma.

b) Teste à diferença dos valores médios para amostras independentes de grande dimensão

1. Formulação das hipóteses

$H_0 : \mu_1 - \mu_2 = 0$ (O comprimento médio das peças produzidas pelas duas máquinas é igual)

$H_1 : \mu_1 - \mu_2 \neq 0$ (O comprimento médio das peças produzidas pelas duas máquinas é diferente)

$RR : |Z| > c$

2. Região crítica: $z < -z_{0.975} = -1.96 \vee z > z_{0.975} = 1.96$

3. Teste estatístico

$$Z=\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}=\frac{(5.95-6.01)-0}{\sqrt{\frac{0.018}{73}+\frac{0.02}{61}}}=-2.503$$

4. Decisão

Como -2.503<-1.96, rejeita-se a Hipótese Nula, ou seja, o comprimento médio das peças produzidas pela duas máquinas é significativamente diferente.

2. Perante a suspeita que o hábito de fumar da mãe pode influenciar o peso do recém-nascido foram recolhidos os dados referentes a 2053 mães e respectivos bebés. Os resultados encontram-se na tabela seguinte:

	Peso do bebé						ni.
Mãe Fumadora?	Menor que P_{10}	eij	Entre P_{10} e P_{90}	eij	Maior que P_{90}	eij	
Sim	117	78,06	529	542,88	19	44,05	**665**
Não	124	162,94	1147	1133,12	117	91,95	**1388**
n.j	**241**		**1676**		**136**		**2053**

Que pode concluir sobre estes dados para α = 0.010?

Resposta: Teste de independência do qui quadrado

H0: O peso do bebé é independente do facto da mãe ser fumadora:

H1:~H0

R.R:Q>c

Região crítica: $c=\chi^2_{2,0.010}=9,21034$

qij	Menor que P_{10}	Entre P_{10} e P_{90}	Maior que P_{90}
Sim	19,42	0,36	14,25
Não	9,30	0,17	6,83

Q=50,32

Decisão: Como Q>c, rejeita-se a h0, pelo existe evidência estatística para afirmar que o peso dos bebés está relacionado com o facto da mãe ser fumadora.

3. Uma organização de consumidores quis comparar o preço (em unidades monetárias, u.m.) de um brinquedo particular em 3 tipos de lojas: hipermercados, bazares, e lojas de brinquedos. Seleccionaram-se aleatoriamente 5 hipermercados, 5 bazares e 5 lojas de brinquedos e os preços foram registados para cada uma delas. Os resultados da ANOVA encontram-se na tabela seguinte.

ANOVA

Fonte	Soma dos quadrados	gl	Média dos quadrados	F	Sig.
Entre os grupos	19.734	2	9.867	4.23	0.041
Resíduos	27.999	12	2.333		
Total	47.733	14			

a) Qual o tipo de planeamento utilizado?

Resposta: O Planeamento Completamente Aleatório com amostras equilibradas (3 grupos com igual dimensão, k=3 e $n_j = 5$).

b) Indique quais os pressupostos para a resolução do problema.

Resposta: Os pressupostos são:
- Normalidade da variável dependente
- homogeneidade das variâncias entre os grupos

Equivale a dizer que $e_{ij} \sim IN(0,\sigma^2)$ **(resíduos independentes normalmente distribuídos com média zero e variância constante)**

c) Formule as hipóteses associadas ao teste e complete a tabela ANOVA.

Resposta:
H_0: Não existem diferenças significativas no valor médio do preço dos brinquedos praticado nas 3 lojas.
H_1: Existem….
R.R: F >c

d) Quais as conclusões que pode retirar para α = 0.05.
Resposta:
Decisão: Como valor p (Sig.) < α rejeita-se H_0 para um nível de significância de 5%, pelo que existem diferenças estatisticamente significativas nos valores médios dos preços praticados pelas 3 lojas.

e) Considerando que os valores dos preços médios nas amostras retiradas é de 10.4 u.m. nos hipermercados, 15.0 u.m. nos bazares e 17.20 u.m. nas lojas de brinquedos, verifique se existem diferenças significativas entre o preço médio dos brinquedos nos hipermercados e nas lojas dos brinquedos (α = 0,05).

Resposta:
Recorrendo ao formulário, e atendendo que se pretende manter a significância global, o teste a posteriori pode ser do tipo
$H_0 : \mu_1 - \mu_2 = 0$ isto é, não existem diferenças significativas no valor médio dos preços dos brinquedos praticados pelos hipermercados (1) e pelas lojas de brinquedos (2)
$H_1 : \mu_1 - \mu_2 \neq 0$, isto é, existem diferenças significativas no valor médio dos preços dos brinquedos praticados pelos hipermercados (1) e pelas lojas de brinquedos (2)
$R.R: |T| < c$

com $c = t_{\alpha/2; N-k} = t_{0,025;12} \overset{tab.6}{=} 2,179$

$$T=\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{MQR\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \qquad T=\frac{(10.4-17.2)-0}{\sqrt{2.333*\left(\frac{1}{5}+\frac{1}{5}\right)}}=\frac{-6.8}{0.966}=--7.039$$

Decisão: Como 7.039>2.179, rejeita-se a H_0 para um nível de significância de 5%, pelo que os preços médios dos brinquedos praticados pelos hipermercados são significativamente inferiores aos praticados pelas lojas de brinquedos

4. Os seguintes dados referem-se aos teores de histaminas (μg/g de matéria seca de saliva) em amostras de saliva de 9 indivíduos com problemas de alergias e de 12 indivíduos sem problemas de alergia. A tabela resume os resultados obtidos. Verifique se as distribuições das duas amostras dos teores de histaminas são idênticas (α=0.01).

Alérgicos	67.6	39.6	1651.0	100.0	65.9	1112.0	31.0	102.4	64.7			
Não alérgicos	34.3	27.3	35.4	48.1	5.2	29.1	4.7	41.7	48.0	6.6	18.9	32.4

Resposta: Teste às distribuições de Smirnov para duas amostras independentes
H0: As distribuições das duas amostras dos teores de histaminas são idênticas
H1: As distribuições das duas amostras dos teores de histaminas não são idênticas

X_1	X_2	S_1	S_2	$\lvert S_1-S_2\rvert$
	4,7	0	1/12	0,0833
	5,2	0	1/6	0,1667
	6,6	0	1/4	0,2500
	18,9	0	1/3	0,3333
	27,3	0	5/12	0,4167
	29,1	0	1/2	0,5000
31		1/9	1/2	0,3889
	32,4	1/9	7/12	0,4722
	34,3	1/9	2/3	0,5556
	35,4	1/9	3/4	0,6389
39,6		2/9	3/4	0,5278
	41,7	2/9	5/6	0,6111
	48	2/9	11/12	0,6944
	48,1	2/9	1	0,7778
64,7		1/3	1	0,6667
65,9		4/9	1	0,5556
67,6		5/9	1	0,4444
100		2/3	1	0,3333
102,4		7/9	1	0,2222
1112		8/9	1	0,1111
1651		1	1	0,0000

T=0.7778
c=2/3 (tabela 14, n1=9, n2=12, 1-α=0.99, teste bilateral)

Decisão: Como 0.7778>2/3; rejeita-se H0 para um nível de significância de 1%, pelo que as distribuições das duas amostras dos teores de histaminas são diferentes.

5. Os seguintes resultados referem-se à análise dos dados referentes às intensidades de precipitação (mm/hora) registados num posto udométrico localizado numa bacia hidrográfica, e ao caudal (m^3/s) escoado numa secção de um curso de água a jusante desta bacia.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Precipitação	10	15	42	27.90	9.351
Caudal	10	2.20	3.40	2.7200	0.38528
Valid N (listwise)	10				

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0.960[a]	0.921	0.911	0.11472

a. Predictors: (Constant), Precipitação

b. Dependent Variable: Caudal

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1.231	1	1.231	93.521	0.000[a]
	Residual	0.105	8	0.013		
	Total	1.336	9			

a. Predictors: (Constant), Precipitação

b. Dependent Variable: Caudal

Coefficients[a]

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B	
	B	Std. Error	Beta			Lower Bound	Upper Bound
1 (Constant)	1.617	0.120		13.503	0.000	1.341	1.893
Precipitação	0.040	0.004	0.960	9.671	0.000	0.030	0.049

a. Dependent Variable: Caudal

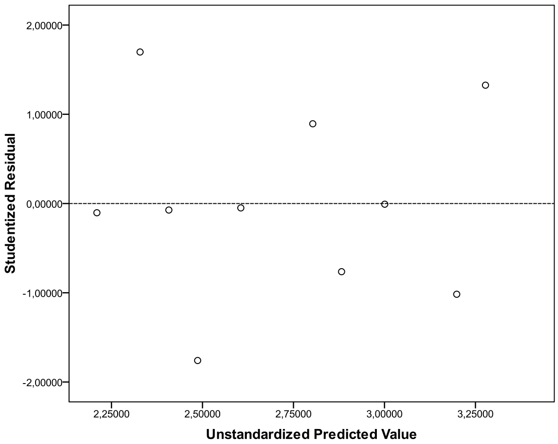
Descriptives			Statistic
Studentized Residual	Mean		0.0145207
	95% Confidence Interval for Mean	Lower Bound	-0.7457937
		Upper Bound	0.7748350
	Std. Deviation		1.06284617
	Minimum		-1.75902
	Maximum		1.69791

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Studentized Residual	0.208	10	0.200*	0.959	10	0.770

a. Lilliefors Significance Correction

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



a) Escreva a equação estimada para a recta.

Resposta:

$$\hat{Y}_i = 1.617 + 0.04 * X_i$$

b) Teste o valor do coeficiente β_0 =0 (use α=0.05).

Resposta:

$$H_0 : \beta_0 = 0$$
$$H_1 : \beta_0 \neq 0$$
$$R.R : |T| > c$$

Na tabela de coeficientes do modelo o teste a este parâmetro dá um valor para T=13.503, com valor p < 0.000. Como valor p <0.05 rejeita-se a hipótese deste coeficiente ser nulo para o modelo linear considerado.

c) De acordo com os resultados avalie a qualidade do modelo justificando.

Resposta:

- Em termos de r^2 (0,960) verifica-se que o modelo apresenta uma boa qualidade para o ajuste pois cerca de 96% da variável resposta pode ser explicada pela variação em X;
- Segundo a tabela ANOVA o modelo é estatisticamente significativo (valor p < 0,05, rej. A H_0, pelo que o modelo de regressão linear considerado é válido);
- Normalidade dos resíduos: os resíduos segundo o teste de ajuste à Normalidade de Kolmogorov (tipo Lilliefors) conferem a normalidade dos resíduos (valor p=0.200>0.05, logo não rejeitamos a hipótese da normalidade);
- A média dos erros apresenta valor médio zero com significância estatística de 5%, que pode ser conferida pelos valores do intervalo de confiança a 95% para os resíduos [-0.7457937;0.7748350]

- Homogeneidade da variância: o modelo linear apresenta um padrão para os resíduos no gráfico que pode ser considerado aleatório pelo que o pressuposto da homogeneidade é validado neste caso.

d) Qual o valor previsto para o caudal sabendo que o registo revelou uma precipitação de 12 mm/hora? E se esse registo for de 20 mm/hora?

Resposta:

$\hat{Y}_0 = 1.617 + 0.04 * 20 = 2.417$

Para o valor 12 não é possível prever pelo modelo de regressão dado que este valor encontra-se fora dos limites de X para a estimação do modelo.