Estatística II – Licenciatura em Gestão
EE - 2 Setembro 2014

Parte teórica

Nome: ____ **Nº**________

1. Perguntas de Verdadeiro/Falso (1.5 valores) - Para cada afirmação, assinale se esta é Verdadeira (V) ou Falsa (F). Uma resposta certa vale 0.3 e uma resposta errada penaliza em idêntico valor.

	V	F
Num teste de dimensão $\alpha = 0.01$ em que o valor-*p*=0.02, rejeita-se H_0.		X
Num teste do χ^2 à bondade do ajustamento a região de rejeição pode ser bilateral		X
Se $\hat{\theta}$ é um estimador não enviesado de θ, então pode-se concluir que $\lim_{n\to\infty} E(\hat{\theta}) = \theta$	X	
No MRL $y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + u_t$ pode-se garantir que a soma dos resíduos MQ é nula, isto é, que $\sum_{t=1}^{n} \hat{u}_t = 0$	X	
No MRL $y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + u_t$, $t = 1,2,\ldots,50$, $R^2 = 0.9$, pode acontecer que não se rejeite individualmente $H_0 : \beta_2 = 0$	X	

2. Perguntas de escolha múltipla (2.25 valores) - Para cada pergunta escolha (assinalando com X) **a alternativa correcta**. Uma resposta certa vale 0.75 valores e uma resposta errada penaliza em 0.25 valores.

a. Suponha que a região crítica ($\alpha = 0.05$) de um determinado teste de hipóteses é dada por $W = \{(x_1, x_2, \cdots, x_n) : q_{obs} > 21.026\}$, tendo a estatística de teste, quando H_0 é verdadeiro, uma distribuição do Qui-Quadrado com 12 graus de liberdade. Observada uma amostra casual verificou-se que $q_{obs} = 20.65$. Então

|___| H_0 deve ser rejeitada

|___| A região de não rejeição é dada por $\bar{W} = \{(x_1, x_2, \cdots, x_n) : q_{obs} > 1/21.026\}$

| X | O valor-p é dado pela área por baixo de uma função densidade do $\chi^2_{(12)}$ e à direita de 20.65

|___| Nenhuma das alternativas anteriores é verdadeira.

b. O lema de Neyman-Pearson é um resultado particularmente interessante para a:

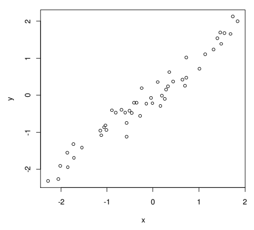
|___| construção de intervalos de confiança

| X | definição de testes de hipóteses

|___| estimação pela máxima verosimilhança

|___| estimação pelo método dos momentos

c. A figura representa o diagrama de dispersão dos valores y_t e x_t. Qual poderá ser o valor do coeficiente de determinação que se obtém ao estimar a regressão $y_t = \beta_1 + \beta_2 x_t + u_t$?



| X | 0.95

|___| -0.95

|___| 0.05

|___| -0.05

3. Perguntas de desenvolvimento (2.25 valores) – A alínea a) vale 1 valor e a alínea b) 1.25.

a. Seja uma população X de média μ e variância σ^2. Com o intuito de estimar μ recolheu-se uma amostra casual de dimensão $n=4$, tendo-se definido o estimador $T=\frac{2X_1+X_2+X_3-X_4}{4}$. Deduza o valor esperado e a variância do estimador explicitando os passos que for dando.

$$E(T)=E\left(\frac{2X_1+X_2+X_3-X_4}{4}\right)=\frac{1}{4}\left(2E(X_1)+E(X_2)+E(X_3)-E(X_4)\right)$$
$$=\frac{1}{4}(2\mu+\mu+\mu-\mu) \qquad \text{idêntica distribuição e } E(X)=\mu$$
$$=\frac{3}{4}\mu$$

$$\text{var}(T)=\text{var}\left(\frac{2X_1+X_2+X_3-X_4}{4}\right)=\frac{1}{16}\left(4\,\text{var}(X_1)+\text{var}(X_2)+\text{var}(X_3)+\text{var}(X_4)\right) \qquad \text{independentes}$$
$$=\frac{7}{16}\sigma^2 \qquad \text{idêntica distribuição e } \text{var}(X)=\sigma^2$$

b. Seja o modelo $y_t=\beta_1+\beta_2\,x_{t2}+\beta_3\,x_{t3}+u_t$ que foi estimado pelos MQ. Defina a regressão auxiliar que se utiliza para prever $E(y\,|\,x_2=3,x_3=1)$ explicando como é obtida.

Regressão auxiliar: $y_t=\theta+\beta_2\,(x_{t2}-3)+\beta_3\,(x_{t3}-1)+u_t$

O nosso objectivo é estimar $\theta=E(y\,|\,x_2=3,x_3=1)=\beta_1+\beta_2\times 3+\beta_3\times 1$. Daqui tira-se $\beta_1=\theta-3\beta_2-\beta_3$ que se substitui na equação original

$$y_t=\beta_1+\beta_2\,x_{t2}+\beta_3\,x_{t3}+u_t$$
$$=\theta-3\beta_2-\beta_3+\beta_2\,x_{t2}+\beta_3\,x_{t3}+u_t$$
$$=\theta+\beta_2\,(x_{t2}-3)+\beta_3\,(x_{t3}-1)+u_t$$

Ao estimar a regressão auxiliar estar-se-á a estimar θ.

Estatística II – Licenciatura em Gestão
EE - 2 Setembro 2014

Parte Prática

Nome: ____ **Nº**________

Espaço reservado para classificações				
				T:
1. (15)	3. (15)	4a. (10)	4d. (15)	**P:** _____
2a. (15)		4b. (15)	4e. (15)	
2b. (15)		4c. (10)	4f. (15)	

Em todos os testes de hipóteses que fizer, formule as hipóteses em teste, indique a estatística de teste e a sua distribuição. Se nada for dito em contrário utilize $\alpha = 0.05$. Para os intervalos de confiança proceda de forma semelhante referindo a variável fulcral.
Se necessitar de espaço dispõe de uma página em branco no fim do enunciado

1) Seja uma população com distribuição dada por $f(x \mid \theta) = \theta x^{\theta-1}$, $0 < x < 1$, $\theta > 0$.
Obtenha o estimador para θ através do método da máxima verosimilhança.

$$L(\theta) = \prod_{i=1}^{n} f(x_i \mid \theta) = \prod_{i=1}^{n} \theta x_i^{\theta-1}$$

$$\ell(\theta) = \ln L(\theta) = \sum_{i=1}^{n} \left(\ln \theta + (\theta - 1) \ln x_i \right)$$

$$\ell'(\theta) = \sum_{i=1}^{n} \left(\frac{1}{\theta} + \ln x_i \right) \text{ logo } \ell'(\theta) = 0 \Leftrightarrow \sum_{i=1}^{n} \left(\frac{1}{\theta} + \ln x_i \right) = 0 \Leftrightarrow \frac{n}{\theta} = -\sum_{i=1}^{n} \ln x_i \Leftrightarrow \theta = -\frac{n}{\sum_{i=1}^{n} \ln x_i}$$

Como $\ell''(\theta) = \sum_{i=1}^{n} \left(-\frac{1}{\theta^2} \right) = -\frac{n}{\theta^2} < 0$, o estimador de MV será $\hat{\theta} = -\frac{n}{\sum_{i=1}^{n} \ln X_i}$

2) No início do semestre o Professor de Estatística 2 afirmou que o estudo ao longo do semestre tinha como consequências a melhoria dos resultados obtidos em exame. Um dos alunos decidiu testar esta teoria, tendo observado a média e a variância corrigida das notas finais de alguns alunos, uns dos quais estudaram ao longo do semestre (grupo 1) e outros concentraram o estudo nas vésperas do exame (grupo 2). Os valores observados, no que se refere à nota final na disciplina foram:

Grupo 1 – Estudo ao longo do semestre	$n_1 = 11$	$\bar{x}_1 = 13.5$	$s_1'^2 = 1.0$
Grupo 2 – Estudo concentrado	$n_2 = 21$	$\bar{x}_2 = 12.8$	$s_2'^2 = 5.0$

Assuma que em cada grupo as notas têm distribuição normal.

a. Considere apenas o grupo dos alunos que estudaram regularmente ao longo do semestre e construa um intervalo de confiança a 95% para a nota esperada.

Variável fulcral $\frac{\bar{X} - \mu}{S' / \sqrt{n}} \sim t_{(n-1)}$ com $n = 11$

O IC a 95% para μ_1 vem $\bar{x} \pm t_{0.025} \frac{s'}{\sqrt{11}}$ isto é $13.5 \pm 2.228 \times 1 / \sqrt{11}$ já que $\alpha = 0.05$ ou seja

(12.83; 14.17)

b. Assumindo que a variância populacional é a mesma nos 2 grupos, teste ($\alpha = 0.05$) a teoria do professor em relação às médias das notas.

$H_0: \mu_1 \le \mu_2$ contra $H_1: \mu_1 > \mu_2$ ou $H_0: \mu_1 = \mu_2$ contra $H_1: \mu_1 > \mu_2$

Estatística de teste $T = \dfrac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}} \sqrt{\dfrac{(n_1 - 1) S_1'^2 + (n_2 - 1) S_2'^2}{n_1 + n_2 - 2}}} \sim t_{(n_1 + n_2 - 2)}$

Como $T_{obs} = \dfrac{13.5 - 12.8}{\sqrt{\dfrac{1}{11} + \dfrac{1}{21}} \sqrt{\dfrac{10 \times 1.0 + 20 \times 5.0}{30}}} = 0.9822$ tem-se $valor - p = 0.1669$ logo não se rejeita H_0

ou seja a teoria do professor não está correcta.

Alternativamente $W = \{t : t > 1.697\}$ e tira-se a mesma conclusão já que $T_{obs} \notin W$

3) Observada uma amostra aleatória de dimensão $n = 100$ referente à variável aleatória *X* – número de idas ao centro de saúde/ano em determinado país – quer-se testar se esta tem distribuição de Poisson de parâmetro 1.5. Os valores observados foram

Nº de visitas	0	1	2	3	4
Nº de observações	20	35	25	12	8

Efectue o teste para $\alpha = 0.05$.

Teste do qui-quadrado à bondade do ajustamento

$H_0: X \sim Po(1.5)$ contra $H_1: H_0 falsa$ que vai ser testada por $H_0': p_j = p_{0j}\ (j = 1, 2, \cdots, k)$

Estatística de teste $Q = \sum_{j=1}^{k} \dfrac{\left(N_j - n \times p_{0j}\right)^2}{n \times p_{0j}} \sim \chi^2_{(k-1)}$

Com

$p_{0j} = \Pr(X = j)$ para $j = 1,2,3,4$ assumindo H_0

$p_{05} = \Pr(X \ge 4)$ assumindo H_0

Nº visitas	0	1	2	3	4+
j	1	2	3	4	5
Nº observado	20	35	25	12	8
Nº esperado	22.31	33.47	25.10	12.55	6.57

$q_{obs} = \dfrac{(20 - 22.31)^2}{22.31} + \dfrac{(35 - 33.47)^2}{33.47} + \cdots + \dfrac{(8 - 6.57)^2}{6.57} = 0.6449$

Logo $valor - p = 0.9579$, logo não se rejeita que a distribuição seja uma Poisson de parâmetro 1.5.

4) Com o objectivo de estudar o sucesso dos alunos na cadeira de Estatística 2 de determinada escola (dados inventados), construiu-se o seguinte modelo:

$NotaEst2_t = \beta_1 + \beta_2\, idade_t + \beta_3\, rapaz_t + \beta_4\, NotaEst1_t + \beta_5\, FaltasSem + \beta_6\, HorasEst + u_t$

Onde:

- $NotaEst2$ – Nota na cadeira de Estatística 2;
- $idade$ – Idade do aluno na data do exame;
- $rapaz$ – Variável binária que toma o valor 1 quando o aluno é de sexo masculino;
- $NotaEst1$ – Nota na cadeira de estatística 1;

- $FaltasSem$ – Número médio de aulas a que o aluno faltou por semana;
- $HorasEst$ – Número de horas de estudo ao longo do semestre (extra aulas).

Recolhida uma amostra de 141 alunos, estimou-se o modelo através dos mínimos quadrados (**Anexo 1**). As perguntas referem-se sempre ao modelo apresentado podendo utilizar os restantes anexos, sempre que necessário. Em todos os outputs apresentados a variável dependente é sempre $NotaEst2$.

a. Interprete as estimativas obtidas para os coeficientes de regressão referentes às variáveis $NotaEst1$ e $rapaz$

$b_2 = 0.5437$ Tudo o resto igual, um acréscimo de 1 valor na nota final de Estatística 1 origina um acréscimo de 0.5437 na nota esperada em Estatística 2

$b_4 = 0.2602$ Tudo o resto igual, a nota esperada em Estatística 2 de um rapaz situa-se 0.2602 pontos acima da de uma rapariga.

b. No seu conjunto, as variáveis presentes no modelo são estatisticamente relevantes para explicar a nota final a estatística 2? Efectue um teste adequado. Teste também a relevância individual das variáveis $HorasEst$ e $rapaz$ na explicação da nota obtida em Estatística 2.

$H_0 : \beta_2 = \beta_3 = \cdots = \beta_6 = 0$ contra $H_1 : \exists \beta_j \neq 0 \quad j = 2,3,\cdots,6$

Teste F. $F = \dfrac{R^2/(k-1)}{(1-R^2)/(n-k)} \sim F(k-1, n-k)$

$F_{obs} = 21.7644$ $valor-p \approx 0$ rejeita-se H_0 logo as variáveis presentes são, no seu conjunto, estatisticamente relevantes.

$H_0 : \beta_j = 0$ contra $H_1 : \beta_j \neq 0$ Est. Teste $T = \dfrac{b_j - \beta_j}{s_{b_j}} \sim t_{(135)}$

Para $HorasEst$ vem $t_{obs} = 8.6765$, $valor-p \approx 0$ rejeita-se H_0 logo a variável é estatisticamente relevante.

Para $rapaz$ vem $t_{obs} = 1.1547$, $valor-p = 0.25021$, não se rejeita H_0 logo a variável não é estatisticamente relevante.

c. Formalize o teste de hipóteses relacionado com a estimação presente no **Anexo 2** e retire a sua conclusão. Independentemente da conclusão que tirar, as restantes alíneas referem-se ao modelo original dado no Anexo 1.

$H_0 : \beta_2 = \beta_3 = 0$ contra $H_1 : \beta_2 \neq 0 \text{ ou } \beta_3 \neq 0$

Teste F. $F = \dfrac{(VR_0 - VR_1)/2}{VR_1/(n-k)} \sim F(2, n-k)$

$F_{obs} = \dfrac{(228.6450 - 222.3281)/2}{1.6469} = 1.9178$ $valor-p = 0.1509$

Não se rejeita H_0 logo as variáveis $idade$ e $rapaz$ não são, no seu conjunto, estatisticamente relevantes.

d. Teste $H_0 : \beta_4 = 0.5$ contra $H_1 : \beta_4 > 0.5$ para $\alpha = 0.05$. Interprete o significado do teste.

$H_0 : \beta_4 = 0.5$ contra $H_1 : \beta_4 > 0.5$ Est. Teste $T = \dfrac{b_4 - 0.5}{s_{b_j}} \sim t_{(135)}$

$T_{obs} = \dfrac{0.5437 - 0.5}{0.09} = 0.4856$ $valor-p = 0.3140$

logo não se rejeita H_0, isto é, não se rejeita que o impacto de 1 valor adicional em Estatística 1 seja inferior ou igual a meio valor na nota esperada de Estatística 2, tudo o resto igual

e. A Ana, aluna de 22 anos que acabou de obter 11 a Estatística 1 e que vai frequentar estatística 2 no próximo semestre, pede-lhe que construa um intervalo de previsão a 90% para a nota que ela irá obter se não faltar a nenhuma aula e estudar 30 horas ao longo do semestre (extra-aulas).

Variável fulcral: $T = \frac{y_0 - \hat{y}_0}{s_d} \sim t_{(135)}$ com $s_d = \sqrt{s^2 + s_{\hat{\theta}}^2}$

O output necessário encontra-se no anexo 3 a partir do qual se obtém o seguinte intervalo de previsão: $12.1828 \pm 1.645 \times \sqrt{1.6469 + 0.2079^2}$, isto é (10.04; 14.32)

f. Admita que se decide averiguar a razoabilidade de se assumir que os erros são homocedásticos. Apresente uma regressão auxiliar adequada para o efeito bem como a hipótese em teste. Assumindo que da estimação da regressão auxiliar que apresenta se retira $R^2 = 0.11$, qual a conclusão que retira.

Duas soluções são aceitáveis: teste de White simplificado (preferível) ou teste de White
No 1º caso a regressão auxiliar é

$$\hat{u}_t^2 = \alpha_1 + \alpha_2 \, Nota\hat{E}st2_t + \alpha_3 \, Nota\hat{E}st2_t^2 + v_t$$

Em que $\hat{u}_t$ representa os resíduos da regressão original e $Nota\hat{E}st2_t$ os valores ajustados na mesma regressão.

$H_0: \alpha_2 = \alpha_3 = 0$ contra $H_1: \alpha_2 \neq 0$ ou $\alpha_3 \neq 0$ Est. Teste $W = n \times R^2 \overset{\circ}{\sim} \chi^2_{(2)}$

$W_{obs} = 141 \times 0.11 = 15.51$ e $valor - p = 0.0004$ logo rejeita-se H_0 já que existe evidência estatística de que o modelo sofre de heterocedasticidade.

Anexo 1

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.6681
R Square	0.44631
Adjusted R Square	0.42581
Standard Error	1.28331
Observations	141

ANOVA

	df	SS	MS	F	Significance F
Regression	5	179.2161	35.8432	21.7644	6.14599E-16
Residual	135	222.3281	1.6469		
Total	140	401.5443			

	Coefficients	Standard Error	t Stat	P-value
Intercept	1.8319	2.3349	0.7846	0.43410
idade	0.1369	0.0880	1.5549	0.12231
rapaz	0.2602	0.2254	1.1547	0.25021
NotaEst1	0.5437	0.0900	6.0400	0.00000
FaltasSem	-0.2720	0.1024	-2.6568	0.00881
HorasEst	0.0453	0.0052	8.6765	0.00000

Anexo 2

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.6562
R Square	0.4306
Adjusted R Square	0.4181
Standard Error	1.2919
Observations	141

ANOVA

	df	SS	MS	F	Significance F
Regression	3	172.8992	57.6331	34.5327	1.098E-16
Residual	137	228.6450	1.6689		
Total	140	401.5443			

	Coefficients	Standard Error	t Stat	P-value
Intercept	5.3923	1.0524	5.1237	0.0000
Nota_Est1	0.4950	0.0871	5.6837	0.0000
FaltasSem	-0.2646	0.1006	-2.6305	0.0095
HorasEst	0.0448	0.0052	8.5953	0.0000

Anexo 3 –

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.6681
R Square	0.4463
Adjusted R Square	0.4258
Standard Error	1.2833
Observations	141

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	5	179.2161	35.8432	21.7644	6.146E-16
Residual	135	222.3281	1.6469		
Total	140	401.5443			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	12.1828	0.2079	58.5876	0.0000
idade-22	0.1369	0.0880	1.5549	0.1223
rapaz	0.2602	0.2254	1.1547	0.2502
Nota_Est1-11	0.5437	0.0900	6.0400	0.0000
FaltasSem	-0.2720	0.1024	-2.6568	0.0088
HorasEst-30	0.0453	0.0052	8.6765	0.0000