LISBOA SCHOOL OF ECONOMICS & MANAGEMENT

Instituto Superior de Economia e Gestão

Universidade de Lisboa

Licenciaturas em Economia e em Finanças

Estatística II - Época Normal - 04/06/2015 – Duração 2 horas

Nome:____ Número:____________

Notas: Certifique-se de que o seu telemóvel está desligado. Se não estiver, é motivo suficiente para anulação da prova. As perguntas de escolha múltipla valem 1 valor; respostas erradas são penalizadas em 0.25. Caso nada seja dito em contrário utilize um nível de significância de 5% nos testes de hipóteses que efetuar. Fundamente e formalize devidamente todas as respostas. Pode usar a última página para continuar qualquer questão.

Espaço reservado para classificações

1. Sejam T_1 e T_2 dois estimadores centrados para θ , sendo T_1 estimador mais eficiente. Nestas condições, pode concluir-se que:

	$Var(T_1) \geq Var(T_2)$;
	$E(T_1) < E(T_2)$;
x	$EQM(T_1) \leq EQM(T_2)$;
	$Env(T_1) > Env(T_2)$.

2.[**1.5**] Considere uma variável aleatória, X , com distribuição $G\left(1, \frac{1}{\theta}\right)$. Com base numa amostra casual de dimensão n, definiu-se o seguinte estimador para o parâmetro θ ,

$$T = \frac{\bar{X}}{3}$$

Estude a consistência deste estimador.

T é estimador consistente para θ se

$\lim\limits_{n->+\infty} E(T) = \theta$ e $\lim\limits_{n->+\infty} Var(T) = 0$

$$E(T) = E\left(\frac{\bar{X}}{3}\right) = \frac{1}{3} \times \frac{1}{n} \times n \times E(X) = \frac{1}{3} \times 3\theta \Rightarrow \lim_{n->+\infty} E(T) = \lim_{n->+\infty} \theta = \theta$$

$$Var(T) = Var\left(\frac{\bar{X}}{3}\right) = \frac{1}{9} \times \frac{1}{n^2} \times n \times Var(X) = \frac{1}{3} \times \frac{1}{n} \times 3\theta^2 = \frac{\theta^2}{n} \Rightarrow \lim_{n->+\infty} Var(T) = \lim_{n->+\infty} \frac{\theta^2}{n} = 0$$

Logo o estimador T é consistente para θ.

3. Uma determinada marca de carros publicita que os seus carros percorrem em média mais de 25 km por cada litro de combustível. Para analisar a validade da afirmação foi recolhida uma amostra aleatória de 40 carros da referida marca tendo-se observado uma distância média percorrida de 26.81 km por litro de combustível. Admita que a distância percorrida em km por cada litro de combustível é uma v. a. que segue uma distribuição normal com um desvio padrão de 4.2 km por litro de combustível e média desconhecida.

a) [**1.5**] Formalize e efetue um teste de hipóteses que permita concluir sobre a validade da afirmação.

$H_0 : \mu = 25$ vs $H_1 : \mu > 25$

Estatística de teste: $Z = \dfrac{\overline{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$

$W_{5\%} = \{z : z > z_{0.05}\} = \{z : z > 1.645\}$

$$z_{obs} = \frac{26.81 - 25}{4.2 / \sqrt{40}} = 2.726$$

Como $z_{obs} \in W_{5\%}$, rejeita-se H_0 ao nível de significância de 5%, isto é, existe evidência estatística de que a afirmação da marca de carros é verdadeira.

b) [**1.5**] Calcule a probabilidade de cometer um erro se o verdadeiro número médio de quilómetros percorridos por cada litro de combustível for 30.

Probabilidade de cometer um erro de 2ª espécie P(Não rejeitar H0 |H0 falsa) quando $\mu = 30$

$$W_{5\%} = \{z : z > z_{0.05}\} = \{z : z > 1.645\} = \left\{ \bar{x} : \frac{\bar{x} - 25}{4.2 / \sqrt{40}} > 1.645 \right\} = \{\bar{x} : \bar{x} > 26.092\} =$$

$$P(\bar{x} < 26.092 \mid \mu = 30) = P\left(Z < \frac{26.092 - 30}{4.2 / \sqrt{40}} \mid \mu = 30 \right) = \Phi(-5.884) = 0$$

4. Para reduzir a amplitude de um intervalo de confiança para a média de uma população normal de variância conhecida pode-se,

	diminuir a dimensão da amostra mantendo-se fixo o grau de confiança;
x	diminuir o grau de confiança mantendo-se fixa a dimensão da amostra;
	aumentar o grau de confiança e reduzir a dimensão da amostra;
	Todas as hipóteses anteriores.

5. Nos testes de hipóteses simples paramétricas é particularmente importante,

x	o Lema de Neyman-Pearson;
	o Teorema de Breusch-Pagan;

☐ o método da variável fulcral;

☐ Todas as hipóteses anteriores.

6. O valor-p,

☒ Mede a evidência favorável à hipótese nula;

☐ é igual à probabilidade do erro de 1ª espécie;

☐ é igual à probabilidade do erro de 2ª espécie;

☐ é igual à probabilidade de tomar uma decisão correta.

7. [**2.0**] Uma empresa de estudos de mercado foi contratada para analisar se a escolha do modo de transporte de uma pessoa na deslocação para o local de trabalho é independente do seu género. Para o efeito foram inquiridos 100 indivíduos, aleatoriamente escolhidos, e aplicado um teste de hipóteses. A tabela seguinte resume os resultados obtidos no inquérito.

	Deslocação por Transporte público	Deslocação por meios próprios	Total
Mulher	20	15	35
Homem	20	45	65
Total	40	60	100

O que pode concluir através do valor-p associado à estatística de teste?

Teste de independência: $H_0: p_{ij} = p_{i\bullet} \times p_{\bullet j}, i=1,2, j=1,2$ vs $H_1: \exists(i,j) p_{ij} \neq p_{i\bullet} \times p_{\bullet j}, i=1,2, j=1,2$

Estatística de teste: Q= $\sum_{i=1}^{2}\sum_{j=1}^{2}\frac{(N_{ij}-fe_{ij})^2}{fe_{ij}} \overset{a}{\sim} \chi^2(1)$

Frequência esperadas:

	Deslocação por Transporte público	Deslocação por meios próprios
Mulher	14	21
Homem	26	39

$$q_{obs} = \frac{(20-14)^2}{14} + \ldots + \frac{(45-39)^2}{39} = 6.59$$

$$valor-p = P(Q > 6.59 \mid H0verd) = 0.01$$

Como valor-p<0.05 rejeita-se a hipótese nula de independência entre o género e o modo de deslocação.

8. Para explicar os custos totais de empresas produtoras de eletricidade considerou-se o seguinte modelo:

$$\log(CT_i) = \beta_0 + \beta_1 \log(prod_i) + \beta_2 \log(ptrab_i) + \beta_3 \log(penerg_i) + \beta_4 \log(pcapital_i) + u_i \quad i = 1,2,\ldots,n$$

onde $\log(CT_i)$ é o logaritmo dos custos totais de produção na empresa i, $\log(prod_i)$ é o logaritmo da quantidade de energia elétrica produzida pela empresa i, $\log(ptrab_i)$ é o logaritmo do preço do factor trabalho na empresa i, $\log(penerg_i)$ é o logaritmo do preço do factor energia na empresa i e $\log(pcapital_i)$ é o logaritmo do preço do factor capital na empresa i. No **anexo** são disponibilizadas algumas estimações relacionadas com este modelo. Suponha que se verificam as hipóteses clássicas do Modelo de Regressão Linear Múltipla (MRLM) a não ser que haja indicação do contrário.

a) **1.5**] Interprete a estimativa obtida para o coeficiente β_1 na equação 1 e o valor do coeficiente de determinação.

Se a quantidade de energia elétrica produzida pela empresa aumentar 1% ,mantendo tudo o resto constante, estima-se que os custos totais de produção aumentem aproximadamente 0.72%.

Coeficiente de determinação - R^2: a regressão da equação 1 explica cerca de 92.6% da variação total do logaritmo dos custos totais de produção.

b) O intervalo de confiança a 95% para o coeficiente da variável log($ptrab$) é igual a:

☐ [0.69,0.75] ☐ [-0.04,0.92] ☐ [-0.62,1.49] ☒ [-0.13,1.01]

c) [**1.0**] Teste a significância global do modelo da equação 1 supondo um nível de significância de 1%.

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0 \qquad H_1: H_0 \text{ falsa}$$

$$F_{obs} = \frac{0.925955}{1-0.925955} \times \frac{145-4-1}{4} = 437.6855$$

$$W_{0.01} = \{F : F > F_{0.01}(4,140) = 3.456\}$$

Rejeita-se H_0 existindo evidência de que as variáveis explicativas são conjuntamente significativas.

d) [**1.5**] A equação 2 apresentada no anexo resulta da estimação do modelo incorporando as restrições, $\beta_2 = 1, \beta_3 = 0, \beta_4 = 0$

Deduza o modelo restrito da equação 2 e retire as conclusões possíveis.

Incorporando as restrições $\beta_2 = 1, \beta_3 = 0, \beta_4 = 0$ no modelo inicial obtém-se:

$$\log(CT) = \beta_0 + \beta_1 \log(prod) + \log(ptrab) + v \Leftrightarrow$$

$$\log(CT) - \log(ptrab) = \beta_0 + \beta_1 \log(prod) + v \Leftrightarrow$$

$$\log\left(\frac{CT}{ptrab}\right) = \beta_0 + \beta_1 \log(prod) + v$$

Conclusão: teste F para escolher entre o modelo com restrições sob H_0 (equação 2) e o modelo inicial sem restrições (equação 1).

$$H_0: \beta_2 = 1, \beta_3 = 0, \beta_4 = 0 \qquad H_1: H_0 \text{ falsa}$$

$$F_{obs} = \frac{24.36327 - 21.55201}{21.55201} \times \frac{145-4-1}{3} = 6.087234$$

$$W_{0.05} = \{F : F > F_{0.05}(3,140) = 2.67\}$$

Rejeita-se H_0, as restrições não são validadas pela evidência empírica. Rejeita-se assim o modelo da equação 2, sendo mais adequado o modelo da equação 1 (sem restrições).

e)[**1.5**] Verifique se se justifica a utilização do estimador de White no contexto da equação 1.

Teste de White simplificado (equação 3).

$H_0 : Var(u\,|\,\mathbf{x}) = \sigma^2 \qquad H_1 : Var(u\,|\,\mathbf{x}) = \sigma^2 h(\mathbf{x})$

$F_{obs} = 41.09177$ com $p_{obs} = 0 < 0.05 \Rightarrow$ rejeita-se H_0

Há evidência de heterocedasticidade logo o estimador OLS da matriz de variâncias-covariâncias da equação 1 não é válido (é inconsistente, e por conseguinte os erros padrão da equação 1 são inválidos). Como tal, para que se possa fazer inferência válida, é necessário usar o estimador robusto de White da matriz de variâncias-covariâncias que é consistente em presença de heterocedasticidade.

9. Os estimadores dos mínimos quadrados dos coeficientes de regressão deixam de ser BLUE quando,

- [ ] não se verifica a hipótese de normalidade;
- [ ] existe multicolinearidade não perfeita;
- [x] existe heterocedasticidade;
- [ ] todas as opções anteriores.

10.[**2.0**] Considere o MRLM $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u$. Supondo que se pretende testar,

$$H_0 : \beta_1 = \beta_2 = 0 \qquad H_1 : \beta_1 \neq 0 \;\; ou \;\; \beta_2 \neq 0 \;.$$

Escreva a estatística de teste com base no R^2 do modelo com restrições e no R^2 do modelo sem restrições e <u>demonstre</u> que esta expressão é equivalente à baseada na soma dos quadrados dos resíduos dos dois modelos.

$$F = \frac{R_{UR}^{\;2} - R_R^{\;2}}{1 - R_{UR}^{\;2}} \times \frac{n-k-1}{2} \sim F(2, n-k-1) \quad (1)$$

$$R_{UR}^{\;2} = 1 - \frac{SSR_{UR}}{SST} \;\; \text{e} \;\; R_R^{\;2} = 1 - \frac{SSR_R}{SST}$$

Substituindo na expressão de F em (1) obtém-se

$$F = \frac{1 - \frac{SSR_{UR}}{SST} - \left(1 - \frac{SSR_R}{SST}\right)}{1 - \left(1 - \frac{SSR_{UR}}{SST}\right)} \times \frac{n-k-1}{2} = \frac{\frac{SSR_R - SSR_{UR}}{SST}}{\frac{SSR_{UR}}{SST}} \times \frac{n-k-1}{2} = \frac{SSR_R - SSR_{UR}}{SSR_{UR}} \times \frac{n-k-1}{2} \quad q.e.d.$$

Anexo

Equação 1

Dependent Variable: LOG(CT)
Included observations: 145

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.526503	1.774367	-1.987471	0.0488
LOG(PROD)	0.720394	0.017466	41.24448	0.0000
LOG(PTRAB)	0.436341	0.291048	1.499209	0.1361
LOG(PENERG)	0.426517	0.100369	4.249483	0.0000
LOG(PCAPITAL)	-0.219888	0.339429	-0.647819	0.5182

R-squared	0.925955	Mean dependent var	1.724663
Adjusted R-squared	0.923840	S.D. dependent var	1.421723
S.E. of regression	0.392356	Akaike info criterion	1.000578
Sum squared resid	21.55201	Schwarz criterion	1.103224
Log likelihood	-67.54189	Hannan-Quinn criter.	1.042286

Equação 2

Dependent Variable: LOG(CT/PTRAB)
Method: Least Squares
Included observations: 145

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.579138	0.122787	-29.14920	0.0000
LOG(PROD)	0.706479	0.017983	39.28696	0.0000

R-squared	0.915207	Mean dependent var	1.052996
Adjusted R-squared	0.914614	S.D. dependent var	1.412561
S.E. of regression	0.412762	Akaike info criterion	1.081806
Sum squared resid	24.36327	Schwarz criterion	1.122864
Log likelihood	-76.43094	Hannan-Quinn criter.	1.098489
F-statistic	1543.465	Durbin-Watson stat	1.190295
Prob(F-statistic)	0.000000		

Equação 3

Dependent Variable: RES^2
Included observations: 145

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.302844	0.042800	7.075847	0.0000
FITTED	-0.265318	0.029551	-8.978359	0.0000
FITTED^2	0.062769	0.010279	6.106682	0.0000

R-squared	0.366590	Mean dependent var	0.148635
Adjusted R-squared	0.357669	S.D. dependent var	0.386672
S.E. of regression	0.309900	Akaike info criterion	0.515340
Sum squared resid	13.63741	Schwarz criterion	0.576927
Log likelihood	-34.36213	Hannan-Quinn criter.	0.540365
F-statistic	41.09177	Durbin-Watson stat	1.713219
Prob(F-statistic)	0.000000		

onde RES e FITTED são os resíduos e os valores ajustados da equação 1, respetivamente.