Instituto Superior de Economia e Gestão
UNIVERSIDADE TÉCNICA DE LISBOA

Estatística II – Licenciaturas em Economia e Finanças
EN – 7 Junho 2013
Duração: 2 horas

Instruções

1. Verifique que o seu telemóvel está **desligado**. Ter o telemóvel ligado durante a prova **é motivo para anulação** da mesma
2. Durante o decorrer da prova não serão prestados esclarecimentos. Se tiver alguma dúvida apresente-a por escrito.
3. **Formalize e fundamente as suas respostas**. Nomeadamente, refira, sempre que necessário, a estatística de teste e a sua distribuição. Se nada for dito em contrário utilize uma dimensão de 5% nos testes estatísticos que efectuar.

1. Seja X uma população com função densidade $f(x\,|\,\theta)=\dfrac{5x^4}{\theta}\exp\left(-\dfrac{x^5}{\theta}\right)$, $x>0$, $\theta>0$.

   a) Com base numa amostra casual de dimensão n, mostre que o estimador de máxima verosimilhança para θ é dado por $\hat{\theta}=\dfrac{\sum_{i=1}^{n} X_i^5}{n}$.

   b) Sabendo que X^5 tem distribuição exponencial de parâmetro $\lambda = 1/\theta$ estude a consistência de $\hat{\theta}$ como estimador de θ.

2. Até ao momento uma cadeia de "fast food" abastece-se de batatas junto dos fornecedores locais. Numa tentativa para reduzir custos está a pensar centralizar o fornecimento de batatas. No entanto tal opção só será vantajosa se o número médio de doses de batatas fritas encomendadas diariamente em cada loja for superior a 500. Para testar se tal alternativa é razoável dado que a mudança no sistema de abastecimento envolve custos de alguma monta o departamento de estudos da cadeia de "fast food" decide efectuar um teste de hipóteses com base numa amostra de dimensão 200. Com base na amostra sabe-se que $\bar{x}=515$, $s'=75.2$.

   a) Formalize o teste adequado para $\alpha = 0.01$ e conclua.

   b) Construa um intervalo de confiança a 95% para μ.

3. Para aferir dos sentimentos independentistas em certa região um determinado partido político efectua sondagens regulares onde questiona cada eleitor sobre ser ou não favorável à independência. Sabendo que os resultados nas 2 últimas sondagens foram aqueles que constam do quadro anexo será que pode concluir que a proporção de pessoas favoráveis à independência na população cresceu?

	Nº de Inquiridos	Nº de respostas pró independência
Penúltimo inquérito	1000	450
Último inquérito	700	350

4. O gerente de uma loja de conveniência quer testar se o tempo de espera entre chegadas consecutivas (em minutos) de clientes segue uma distribuição exponencial. Para tal ele recolheu uma amostra casual de 50 tempos tendo obtido $\bar{x}=5.0$ o que originou a estimativa de máxima verosimilhança $\hat{\lambda}=0.2$. No quadro que se segue apresentam-se os valores observados bem como algumas das frequências esperadas. Complete o quadro e efectue o teste. Conclua comentando a fiabilidade da conclusão obtida.

Intervalo	Até 1	De 1 a 7	De 7 a 20	Mais de 20
Nº observações	4	34	9	3
Freq esperada	?	28.6	11.4	?

5. Dada a quebra verificada na venda de veículos automóveis de passageiros a associação do sector decidiu estimar um modelo explicativo do preço dos veículos. Depois de várias tentativas chegou-se ao seguinte modelo:

$$\log(preco) = \beta_0 + \beta_1 \log(cons) + \beta_2\, co2 + \beta_3 \log(pot) + \beta_4\, vel + \beta_5\, acel + u$$

Onde $\log$ representa o logaritmo neperiano, $preco$ o preço (euros) do veículo em novo, $cons$ o consumo médio (litros) aos 100 km, $co2$ a emissão de dióxido de carbono, pot a potência do veículo (cavalos), vel a velocidade máxima (km/h) e $acel$ o tempo de aceleração (segundos) dos 0 aos 100 km/h.

Observada uma amostra aleatória e estimado o modelo obteve-se

Dependent Variable: LOG(PRECO)
Method: Least Squares
Date: 05/29/13 Time: 17:57
Sample: 1 292
Included observations: 292

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.453246	0.572970	7.772216	0.0000
LOG(CONS)	-1.101470	0.136055	-8.095758	0.0000
CO2	0.007405	0.000895	8.271487	0.0000
LOG(POT)	1.359302	0.096689	14.05849	0.0000
VEL	-0.000979	0.001007	-0.972230	0.3318
ACEL	0.026052	0.014464	1.801151	0.0727

R-squared	0.909896	Mean dependent var	10.60976
Adjusted R-squared	0.908321	S.D. dependent var	0.646263
S.E. of regression	0.195680	Akaike info criterion	-0.404343
Sum squared resid	10.95108	Schwarz criterion	-0.328793
Log likelihood	65.03408	Hannan-Quinn criter.	-0.374081
F-statistic	577.6209	Durbin-Watson stat	1.388108
Prob(F-statistic)	0.000000		

Todas as perguntas se referem sempre ao modelo base embora possa utilizar, sempre que necessário os outputs que constam em anexo. Tenha presente que nem todos os outputs serão necessários para responder às questões postas.

a) Interprete as estimativas obtidas para os coeficientes β_1 e β_5. Teste a significância estatística do regressor $acel$.
b) Será legítimo concluir que um aumento da potência do veículo origina um aumento no seu preço. Responda com base num teste adequado (não se esqueça de definir a região de rejeição).
c) Teste a nulidade conjunta dos parâmetros β_4 e β_5
d) Um técnico defende que, tudo o resto igual, um aumento de 2% na potência do veículo se traduz num aumento de 3% no preço. Qual a sua opinião (fundamentada num teste adequado)?
e) Escreva a regressão auxiliar adequada para testar $\theta = \beta_1 + \beta_3 = 0$. Explique sucintamente como procederia ao teste com base na regressão auxiliar.
f) Obtenha um intervalo de previsão a 90% para o $\log(preco)$ para um veículo com um consumo médio de 8L/100km, um emissão de $co2$ de 200, uma potência de 200 cavalos, uma velocidade máxima de 180km/h e um tempo de aceleração de 9 segundos dos 0 aos 100 km/h.

Alínea	**1a**	**1b**	**2a**	**2b**	**3**	**4**	**5a**	**5b**	**5c**	**5d**	**5e**	**5f**
Cotação	20	15	15	15	20	20	15	15	15	15	20	15

ANEXO

Dependent Variable: LOG(PRECO)
Method: Least Squares
Date: 05/17/13 Time: 17:41
Sample: 1 292
Included observations: 292

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.682387	0.198784	28.58570	0.0000
LOG(CONS)	-1.174558	0.134294	-8.746194	0.0000
CO2	0.008511	0.000803	10.60397	0.0000
LOG(POT)	1.118050	0.043741	25.56080	0.0000

R-squared	0.907439	Mean dependent var	10.60976
Adjusted R-squared	0.906475	S.D. dependent var	0.646263
S.E. of regression	0.197639	Akaike info criterion	-0.391141
Sum squared resid	11.24967	Schwarz criterion	-0.340774
Log likelihood	61.10652	Hannan-Quinn criter.	-0.370966
F-statistic	941.1545	Durbin-Watson stat	1.396008
Prob(F-statistic)	0.000000		

Dependent Variable: LOG(PRECO)
Method: Least Squares
Date: 05/23/13 Time: 12:21
Sample: 1 292
Included observations: 292

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	10.90420	0.032767	332.7818	0.0000
LOG(CONS)-LOG(8)	-1.101470	0.136055	-8.095758	0.0000
CO2-200	0.007405	0.000895	8.271487	0.0000
LOG(POT)-LOG(200)	1.359302	0.096689	14.05849	0.0000
VEL-180	-0.000979	0.001007	-0.972230	0.3318
ACEL-9	0.026052	0.014464	1.801151	0.0727

R-squared	0.909896	Mean dependent var	10.60976
Adjusted R-squared	0.908321	S.D. dependent var	0.646263
S.E. of regression	0.195680	Akaike info criterion	-0.404343
Sum squared resid	10.95108	Schwarz criterion	-0.328793
Log likelihood	65.03408	Hannan-Quinn criter.	-0.374081
F-statistic	577.6209	Durbin-Watson stat	1.388108
Prob(F-statistic)	0.000000		

Tópicos de solução

1a

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

$\ell(\theta)=\ln L(\theta)=\sum_{i=1}^{n}\ln f(x_i \mid \theta)=\sum_{i=1}^{n}\ln\left(\frac{5x_i^4}{\theta}\exp\left(-\frac{x_i^5}{\theta}\right)\right)=\sum_{i=1}^{n}\left(\ln\left(5x_i^4\right)-\ln(\theta)-\frac{x_i^5}{\theta}\right)$

$\ell'(\theta)=\sum_{i=1}^{n}\left(-\frac{1}{\theta}+\frac{x_i^5}{\theta^2}\right)=-\frac{n}{\theta}+\frac{\sum_{i=1}^{n}x_i^5}{\theta^2}$

$\ell'(\theta)=0 \Leftrightarrow \frac{n}{\theta}=\frac{\sum_{i=1}^{n}x_i^5}{\theta^2} \Leftrightarrow \theta=\frac{\sum_{i=1}^{n}x_i^5}{n}$

$\ell''(\theta)=\frac{n}{\theta^2}-2\frac{\sum_{i=1}^{n}x_i^5}{\theta^3}$ logo para $\theta=\frac{\sum_{i=1}^{n}x_i^5}{n}$ vem $\ell''(\theta)=-\frac{n}{\theta^2}<0$

Assim o estimador de MV é dado por $\hat{\theta}=\frac{\sum_{i=1}^{n}X_i^5}{n}$

1b

$E(\hat{\theta})=E\left(\frac{\sum_{i=1}^{n}X_i^5}{n}\right)=\frac{\sum_{i=1}^{n}E(X_i^5)}{n}=\frac{n\theta}{n}=\theta$

$\text{var}(\hat{\theta})=\text{var}\left(\frac{\sum_{i=1}^{n}X_i^5}{n}\right)=\frac{\sum_{i=1}^{n}\text{var}(X_i^5)}{n^2}$ independência dos X_i

$=\frac{n\theta^2}{n^2}=\frac{\theta^2}{n}$

Logo $\begin{cases}\lim_{n\to\infty}E(\hat{\theta})=\lim_{n\to\infty}\theta=\theta\\ \lim_{n\to\infty}\text{var}(\hat{\theta})=\lim_{n\to\infty}\frac{\theta^2}{n}=0\end{cases}$ encontrando-se verificadas as condições suficientes para que $\hat{\theta}$ seja estimador consistente de θ.

2a

$H_0:\mu\le 500$ contra $H_1:\mu>500$ (na dúvida não se faz o investimento)

Estatística de teste: $Z=\frac{\bar{X}-500}{S'/\sqrt{200}}\overset{a}{\sim}n(0,1)$

$z_{obs}=\frac{515-500}{75.2/\sqrt{200}}=2.821$ valor-p=0.002 **ou** $W_{Z;0.01}=\{z:z>2.326\}$

Rejeita-se H_0 logo avança-se com o investimento.

2b

Variável fulcral: $Z=\frac{\bar{X}-\mu}{S'/\sqrt{200}}\overset{a}{\sim}n(0,1)$

$z_{\alpha/2}=1.96$

O intervalo de confiança vem $\bar{x}\pm z_{\alpha/2}\,s'/\sqrt{n}$ isto é (504.58; 525.42)

3

θ_1: proporção de independentistas na população aquando da 1ª sondagem

θ_2: proporção de independentistas na população aquando da 2ª sondagem

$H_0:\theta_1=\theta_2$ (ou $H_0:\theta_1\ge\theta_2$) contra $H_1:\theta_1<\theta_2$

Estatística de teste: $Z = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\hat{\theta}(1-\hat{\theta})\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}} \overset{a}{\sim} n(0,1)$ com $\hat{\theta} = \dfrac{n_1 \bar{X}_1 + n_2 \bar{X}_2}{n_1 + n_2}$

$z_{obs} = \dfrac{0.45-0.5}{\sqrt{\left(\dfrac{1}{1000}+\dfrac{1}{700}\right)\times 0.4706 \times 0.5294}} = -2.033$ já que $\hat{\theta} = \dfrac{450+350}{1700} = 0.4706$

valor-p=0.021 **ou** $W_{Z;0.05} = \{z : z < -1.645\}$

Rejeita-se H_0, logo há evidência de que o sentimento independentista tenha crescida na população.

4

Teste do qui-quadrado à bondade do ajustamento

$H_0 : X \sim Ex(\lambda)$ contra $H_1 : H_0$ falsa

Intervalo	Até 1	De 1 a 7	De 7 a 20	Mais de 20
Nº observações	4	34	9	3
Freq esperada	**9.1**	28.6	11.4	**0.9**

Deve-se reagrupar a classe 4 com a classe 3.

Intervalo	Até 1	De 1 a 7	Mas de 7	**Soma**
Nº observações	4	34	12	50
Freq esperada	9.1	28.6	12.3	50
χ^2_{obs}	2.858	1.020	0.007	**3.885**

$Q = \sum_{j=1}^{m} \dfrac{(n_j - fe_j)^2}{fe_j} \overset{a}{\sim} \chi^2_{(m-1-1)}$ (1 parâmetro foi estimado por máxima verosimilhança)

$Q_{obs} = 3.885$ valor-p=0.0487 Rejeita-se H0

No entanto a conclusão está longe de ser "confortável" já que o valor-p está próximo do limiar de rejeição. Note-se também que número de classes utilizadas é muito pequeno, o que seria mais grave no caso de não rejeição.

5a.

$\hat{\beta}_1 = -1.101$ (elasticidade) Tudo o resto igual, quando o consumo do veículo aumenta 1% estima-se que o seu preço esperado diminua de 1.101% aproximadamente.

$\hat{\beta}_5 = 0.02605$ (semi-elasticidade) Tudo o resto igual, quando o tempo de aceleração do veículo aumenta de 1 segundo estima-se que o seu preço esperado aumente de 2.61% aproximadamente.

Pode ser feito um teste bilateral **ou** um teste unilateral esquerdo

$H_0 : \beta_5 = 0$ contra $H_1 : \beta_5 \neq 0$ Estatística de teste $T = \hat{\beta} / se(\hat{\beta}) \sim t(286)$

valor-p=0.0727 logo não se rejeita H_0, isto é, o regressor $acel$ não tem significância estatística

$H_0 : \beta_5 = 0$ contra $H_1 : \beta_5 < 0$ Estatística de teste $T = \hat{\beta} / se(\hat{\beta}) \sim t(286)$

$t_{obs} = 1.801$ valor-p=0.96 logo não se rejeita H_0, isto é, o regressor $acel$ não tem significância estatística

5b

$H_0 : \beta_3 \leq 0$ contra $H_1 : \beta_3 > 0$ Estatística de teste $T = \hat{\beta}_3 / se(\hat{\beta}_3) \sim t(286)$

Região de rejeição $W_{T,0.05} = \{t : t > 1.650\}$ $T_{obs} = 14.05849$

logo rejeita-se muito claramente H_0, isto é, um aumento da potência encarece o preço do veículo

5c

$H_0 : \beta_4 = \beta_5 = 0$ contra $H_1 : \beta_4 \neq 0 \vee \beta_5 \neq 0$

Estatística de teste $F = \dfrac{SSR_r - SSR_{ur}}{SSR_{ur}} \times \dfrac{266}{2} \sim F(2,286)$ $F_{obs} = \dfrac{11.24967 - 10.95108}{10.95108} \times 143 = 3.8991$

Valor-p=0.021 ou $W_{F;0.05} = \{f : f > 3.03\}$ Rejeita-se H_0, isto é, rejeita-se a nulidade conjunta de β_4 e β_5.

5d

A teoria do técnico traduz-se em $2\times\beta_3 = 3$, isto é, $\beta_3 = 1.5$

$H_0 : \beta_3 = 1.5$ contra $H_1 : \beta_3 \neq 1.5$ Estatística de teste $T = (\hat{\beta}_3 - 1.5) / se(\hat{\beta}_3) \sim t(286)$

Região de rejeição $W_{T,0.05} = \{t : t < -1.968 \vee t > 1.968\}$ $T_{obs} = \dfrac{1.359302 - 1.5}{0.096689} = -1.45516$

logo não se rejeita H_0 , isto é, não se rejeita a afirmação feita pelo técnico.

5e

$\theta = \beta_1 + \beta_3 \Leftrightarrow \beta_1 = \theta - \beta_3$

Regressão auxiliar

$$\begin{aligned}\log(preco) &= \beta_0 + \beta_1 \log(cons) + \beta_2\, co2 + \beta_3 \log(pot) + \beta_4\, vel + \beta_5\, acel + u \\ &= \beta_0 + (\theta - \beta_3)\log(cons) + \beta_2\, co2 + \beta_3 \log(pot) + \beta_4\, vel + \beta_5\, acel + u \\ &= \beta_0 + \theta \log(cons) + \beta_2\, co2 + \beta_3 \left(\log(pot) - \log(cons)\right) + \beta_4\, vel + \beta_5\, acel + u\end{aligned}$$

Estimada a regressão utiliza-se um teste *t* para testar $H_0 : \theta = 0$ contra $H_1 : \theta \neq 0$ com base na estatística de teste $T = \hat{\theta} / se(\hat{\theta}) \sim t(286)$

5f

Recorrendo à regressão auxiliar

$$\log(preco) = \theta + \beta_1 \left(\log(cons) - \log(8)\right) + \beta_2 \left(co2 - 200\right) + \beta_3 \left(\log(pot) - \log(200)\right) + \beta_4 \left(vel - 180\right) + \beta_5 \left(acel - 9\right) + u$$

Cujo output é dado em anexo vem

$$\widehat{\log(preco \mid ...)} = 10.90420$$

Utilizando

$$t = \frac{y^0 - \hat{y}^0}{\sqrt{se(\hat{\theta})^2 + \hat{\sigma}^2}} \sim t(286)$$ em que $y^0 = \log(preco)$

Vem o intervalo

$\hat{y}^0 \pm t_{\alpha/2} \sqrt{se(\hat{\theta})^2 + \hat{\sigma}^2}$ isto é (10.577; 11.232)

Já que

$t_{\alpha/2} = 1.65$ $\sqrt{se(\hat{\theta})^2 + \hat{\sigma}^2} = \sqrt{0.032767^2 + 0.19568^2} = 0.198404$