Estatística II - Licenciaturas em Economia e Finanças

ER - 1 Julho 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.

TÓPICOS DE RESOLUÇÃO

1. Pretende-se estudar a "performance" de activos financeiros segundo o risco. Para isso foram recolhidas amostras casuais de activos nas categorias de risco baixo e de risco elevado e observado o seu retorno percentual no período de um ano. Os dados estão resumidos abaixo:

	Média	Desvio padrão corrigido	Dimensão da amostra
Risco baixo	3.52	1.02	153
Risco elevado	7.01	2.93	231

a) Construa um intervalo de confiança a 95% para o retorno médio dos activos de risco elevado.

Resolução:

Variável fulcral:

$$\frac{\overline{X}_E - \mu_E}{\frac{S'_E}{\sqrt{n_E}}} \overset{a}{\sim} N(0,1)$$

Forma do intervalo de confiança para μ_E:

$$\left[\bar{x}_E - z_{0.25}\frac{s'_E}{\sqrt{n_E}};\ \bar{x}_E + z_{0.25}\frac{s'_E}{\sqrt{n_E}}\right]$$

$\bar{x}_E = 7.01$, $s_E = 2.93$, $n_E = 231$

Intervalo de confiança:

$$\left[7.01 - 1.96 \times \frac{2.93}{\sqrt{231}};\ 7.01 + 1.96 \times \frac{2.93}{\sqrt{231}}\right] = [6.63; 7.39].$$

b) Existe a convicção entre os analistas financeiros que os activos de risco baixo têm um retorno médio de 3%. Através da realização de um teste adequado, avalie se existe evidência estatística que contrarie esta convicção.

Resolução:

$H_0 : \mu_B = 3 \quad H_1 : \mu_B \neq 3$

Estatística teste:

$$Z = \frac{\overline{X}_B - \mu_{B0}}{\frac{S'_B}{\sqrt{n_B}}} \overset{a}{\sim} N(0,1)$$

$\bar{x}_B = 3.52$, $s'_B = 1.02$, $n_B = 153$

$$z_{obs} = \frac{3.52 - 3}{\frac{1.02}{\sqrt{153}}} = 6.31$$

Como $W_Z = \{z : z < -1.96 \vee z > 1.96\}$, $z_{obs} \in W_Z$ e rejeita-se H_0, isto é, rejeita-se que os activos de risco baixo tenham um retorno médio de 3%.

c) Os mesmos analistas supõem que o retorno médio dos activos de risco elevado excede o dos activos de risco baixo em pelo menos 4 pontos percentuais. Confirme ou infirme a suposição realizando o teste de hipóteses conveniente (admita que as amostras são independentes).

Resolução:

$H_0 : \mu_M - \mu_B \geq 4 \quad H_1 : \mu_M - \mu_B < 4$

Estatística teste:

$$Z = \frac{(\overline{X}_M - \overline{X}_B) - (\mu_M - \mu_B)_0}{\sqrt{\frac{S_M'^2}{n_M} + \frac{S_B'^2}{n_B}}} \overset{a}{\sim} N(0,1)$$

$\bar{x}_B = 3.52$, $\bar{x}_M = 7.01$, $s'_B = 1.02$, $s'_M = 2.93$, $n_B = 153$, $n_M = 231$

$$z_{obs} = \frac{7.01 - 3.52 - 4}{\sqrt{\frac{2.93^2}{231} + \frac{1.02^2}{153}}} = -2.43$$

Como $W_Z = \{z : z < -1.645\}$, $z_{obs} \in W_Z$ e rejeita-se H_0, isto é, que o retorno médio dos activos de risco elevado exceda o dos activos de risco baixo em pelo menos 4 pontos percentuais.

2. Num estudo complementar, procura-se avaliar a associação entre o perfil de risco dos investidores e o seu nível de rendimento (anual, em milhares de euros). Para isso, foi recolhida aleatoriamente uma amostra de investidores com os resultados abaixo:

Perfil \ Rendimento	Até 30	Mais de 30
Conservador	20	20
Moderado	20	40
Agressivo	40	60

Avalie, através da realização do teste apropriado, se o perfil do investidor e o nível de rendimento estão associados.

Resolução:

$H_0 : \forall i,j \; p_{ij} = p_{i0} \cdot p_{0j} \quad H_1 : \exists i,j : p_{ij} \neq p_{i0} \cdot p_{0j} \quad (i = 1,2,3; j = 1,2)$

Estatística teste:

$$Q = \sum_{i=1}^{3} \sum_{j=1}^{2} \frac{(N_{ij} - n \cdot \hat{p}_{i0} \cdot \hat{p}_{0j})^2}{n \cdot \hat{p}_{i0} \cdot \hat{p}_{0j}} \overset{a}{\sim} \chi^2((3-1)(2-1))$$

Perfil \ Rendimento	Até 30	Mais de 30	Total
Conservador	20 (16)	20 (24)	40
Moderado	20 (24)	40 (36)	60
Agressivo	40 (40)	60 (60)	100
Total	80	120	200

$$Q_{obs} = \frac{(20-16)^2}{16} + \cdots + \frac{(60-60)^2}{60} = 2.78$$

Como $W_Q = \{q : q > 5.991\}$, $Q_{obs} \notin W_Q$ e não se rejeita H_0, isto é, não se rejeita a independência dos atributos “Perfil” e “Rendimento”.

3. Considere a função massa de probabilidade f_X

$$f_X(x|p) = \begin{cases} p(1-p)^{x-1}, & x = 1, 2, \ldots \\ 0, & \text{outros valores de } x \end{cases},$$

onde $0 < p < 1$ é uma constante.

Deduza o estimador de máxima verosimilhança do parâmetro p.

Resolução:

$L(p) = \prod_{i=1}^{n} p(1-p)^{x_i - 1}$

$l(p) = \sum_{i=1}^{n} (\ln p + (x_i - 1)\ln(1-p)) = n \ln p + \ln(1-p)\left(\sum_{i=1}^{n} x_i - n\right)$

$l'(p) = \dfrac{n}{p} + \dfrac{n - \sum_{i=1}^{n} x_i}{1-p}$

$$l'(p) = 0 \Leftrightarrow \frac{n}{p} + \frac{n - \sum_{i=1}^{n} x_i}{1-p} = 0 \Leftrightarrow \frac{\sum_{i=1}^{n} x_i - n}{1-p} = \frac{n}{p} \quad \Leftrightarrow \left(\sum_{i=1}^{n} x_i - n\right)p = n(1-p)$$
$$\Leftrightarrow p = \frac{n}{\sum_{i=1}^{n} x_i}$$

$l''(p) = -\dfrac{n}{p^2} + \dfrac{n - \sum_{i=1}^{n} x_i}{(1-p)^2} < 0 \;\forall p$

Em conclusão, o estimador de máxima verosimilhança de p é

$$\hat{p} = \frac{n}{\sum_{i=1}^{n} X_i} = \frac{1}{\overline{X}}.$$

4. Foi extraída uma amostra de 200 famílias de Chicago para investigar qual a distância que as famílias americanas tendem a percorrer quando vão de férias. Foi considerado o modelo

$$\text{MILES} = \beta_0 + \beta_1 \text{INCOME} + \beta_2 \text{AGE} + \beta_3 \text{KIDS} + u,$$

onde

MILES: milhas percorridas em férias por ano

INCOME: rendimento anual do agregado familiar em US\$1000

AGE: idade média dos adultos no agregado familiar

KIDS: número de crianças no agregado familiar

O resultado da estimação OLS do modelo encontra-se sob **Equação 1** no anexo.

a) Teste a significância estatística global do modelo.

Resolução:

$H_0 : \beta_1 = \beta_2 = \beta_3 = 0 \qquad H_1 : \exists j : \beta_j \neq 0 \;\; (j = 1, 2, 3)$

Estatística teste:

$$F = \frac{R^2_{ur}}{1 - R^2_{ur}} \left(\frac{n-k-1}{k}\right) \overset{a}{\sim} F\left(k, n-k-1\right)$$

$R^2_{ur} = 0.340605$, $n = 200$, $k = 3$

$$F_{obs} = \frac{0.340605}{1 - 0.340605}\left(\frac{196}{3}\right) = 33.74740$$

e o correspondente valor-p é 0.0000, pelo que se rejeita a hipótese nula, isto é, rejeita-se a ausência de significância estatística global do modelo.

b) Interprete $\hat{\beta}_2$ e teste a significância estatística do correspondente regressor AGE.

Resolução:

Interpretação de $\hat{\beta}_2 = 15.74092$: a variação de 1 ano na idade média dos adultos na família origina, em média e ceteris paribus, uma variação no mesmo sentido de 15.74092 milhas percorridas em férias por ano.

Significância do regressor AGE:

$H_0 : \beta_2 = 0 \quad H_1 : \beta_2 \neq 0$

Estatística teste:

$$t = \frac{\hat{\beta}_j - \beta_{j0}}{\mathrm{se}\left(\hat{\beta}_j\right)} \overset{a}{\sim} t(n-k-1)$$

$\hat{\beta}_2 = 15.74092$, $\mathrm{se}\left(\hat{\beta}_2\right) = 3.757370$, $n = 200$, $k = 3$

$$t_{obs} = \frac{15.74092 - 0}{3.757370} = 4.189346$$

e o correspondente valor-p é 0.0000, pelo que se rejeita a hipótese nula, isto é, rejeita-se a ausência de significância estatística do regressor AGE.

c) Teste a hipótese $\beta_1 \geq \beta_2$ contra a hipótese $\beta_1 < \beta_2$.

Resolução:

$H_0 : \beta_1 - \beta_2 \geq 0 \qquad H_1 : \beta_1 - \beta_2 < 0$

Estatística teste:

$$t = \frac{(\hat{\beta}_1 - \hat{\beta}_2) - (\beta_1 - \beta_2)_0}{\mathrm{se}\left(\hat{\beta}_1 - \hat{\beta}_2\right)} \overset{a}{\sim} t\left(n-k-1\right)$$

$\hat{\beta}_1 - \hat{\beta}_2 = -1.539590$, $n = 200$, $k = 3$

$$\begin{aligned}\mathrm{se}\left(\hat{\beta}_1 - \hat{\beta}_2\right) &= \sqrt{\widehat{\mathrm{Var}}\left(\hat{\beta}_1\right) + \widehat{\mathrm{Var}}\left(\hat{\beta}_2\right) - 2\widehat{\mathrm{Cov}}\left(\hat{\beta}_1, \hat{\beta}_2\right)} \\ &= \sqrt{3.240923 + 14.11783 - 2 \cdot (-1.543798)} \\ &= 4.521764\end{aligned}$$

$$t_{obs} = \frac{-1.539590 - 0}{4.521764} = -0.340484$$

Como $W_t = \{t : t < -1.645\}$, $t_{obs} \notin W_Q$ e não se rejeita H_0, isto é, não se rejeita $\beta_1 \geq \beta_2$.

d) Com vista à detecção de heterocedasticidade, considerou-se a equação

$$\begin{aligned}\hat{u}_i^2 =& \delta_0 + \delta_1 \mathrm{INCOME}_i + \delta_2 \mathrm{AGE}_i + \delta_3 \mathrm{KIDS}_i + \delta_4 \mathrm{INCOME}_i^2 + \delta_5 \mathrm{AGE}_i^2 + \delta_6 \mathrm{KIDS}_i^2 \\ &+ \delta_7 \mathrm{INCOME{*}AGE}_i + \delta_8 \mathrm{INCOME{*}KIDS}_i + \delta_9 \mathrm{AGE{*}KIDS}_i + e_i\end{aligned}$$

cuja estimação OLS se encontra sob **Equação 2** no anexo.

d1) Indique as hipóteses nula e alternativa do teste relevante, assim como a estatística de teste, e conclua.

Resolução:

Teste de White.

Em vez de se testar a hipótese H_0

$H_0 : \mathrm{Var}\left(u_i|X\right) = \sigma^2 \qquad H_1 : \mathrm{Var}\left(u_i|X\right) = \sigma_i^2$ (com σ_i^2 não constante) vai-se testar a hipótese derivada H_0'

$H_0' : \delta_1 = \delta_2 = \cdots = \delta_9 = 0 \qquad H_1 : \exists \delta_j \neq 0,\ j = 1, \ldots, 9$

Estatística teste:

$$F = \frac{R^2_{\hat{u}^2}}{(1 - R^2_{\hat{u}^2})} \left(\frac{n-p-1}{p} \right) \overset{a}{\sim} F\,(p, n-p-1)$$

$R^2_{\hat{u}^2} = 0.199964$, $n = 200$, $p = 9$

$$F_{obs} = \frac{0.199964}{1 - 0.199964} \left(\frac{190}{9} \right) = 5.276598$$

e o correspondente valor-p é 0.000002, pelo que se rejeita a hipótese nula, ou seja, existe evidência de heterocedasticidade.

d2) Admita que há evidência de heterocedasticidade. Quais as consequências na estimação OLS dos coeficientes do modelo e erros padrão?

Resolução:

Os estimadores OLS dos coeficientes β permanecem centrados e consistentes. Todavia deixam de ser BLUE: $\widehat{\mathrm{Var}(\hat{\beta}|X)}_{OLS}$ é um estimador inconsistente de $\mathrm{Var}(\hat{\beta}|X)$.

Na presença de heterocedasticidade a estimação OLS produz subavaliação dos erros padrão de $\hat{\beta}$, inviabilizando a inferência estatística.

e) Sob **Equação 3** no anexo encontram-se os resultados da estimação OLS da equação

$$\text{MILES}_i = \alpha_0 + \alpha_1 \text{INCOME}_i + \alpha_2 \text{AGE}_i + \alpha_3 \text{KIDS}_i + \delta_1 \text{FITTED}_i^2 + \delta_2 \text{FITTED}_i^3 + v_i,$$

onde $\text{FITTED}_i = \widehat{\text{MILES}}_i$.

Explique qual a finalidade desta regressão relativamente ao modelo em estudo e proceda ao teste relevante.

Resolução:

Finalidade: detecção de má especificação por omissão de termos não lineares no modelo.

Teste RESET.

$H_0 : \delta_1 = \delta_2 = 0 \qquad H_1 : \delta_1 \neq 0 \ \vee \delta_2 \neq 0$

Estatística teste:

$$F = \frac{SSR_r - SSR_{ur}}{SSR_{ur}} \left(\frac{n-k-1}{q} \right) \overset{a}{\sim} F\,(q, n-k-1)$$

$SSR_r = 40098973$, $SSR_{ur} = 39879974$, $n = 200$, $k = 5$, $q = 2$

$$F_{obs} = \frac{40098973 - 39879974}{39879974} \left(\frac{194}{2} \right) = 0.532671$$

e o correspondente valor-p é 0.587891, pelo que não se rejeita a hipótese nula, ou seja, não há evidência de má especificação do modelo (por omissão de não linearidades).

5. Considere o modelo de regressão linear simples

$$y_i = \alpha + \beta x_i + u_i \qquad (i = 1, 2, \ldots, n)$$

e o estimador de β

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}.$$

Determine o enviesamento e a variância de $\hat{\beta}$.

Resolução:

$$Env\big(\hat{\beta}|X\big) = \mathrm{E}\left(\hat{\beta}|X\right) - \beta = \mathrm{E}\left(\frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}\middle|X\right) - \beta = \frac{\sum_{i=1}^{n} x_i \,\mathrm{E}(y_i|X)}{\sum_{i=1}^{n} x_i^2} - \beta$$

$$= \frac{\sum_{i=1}^{n} x_i \,\mathrm{E}(\alpha + \beta x_i + u_i|X)}{\sum_{i=1}^{n} x_i^2} - \beta = \frac{\sum_{i=1}^{n} x_i(\alpha + \beta x_i)}{\sum_{i=1}^{n} x_i^2} - \beta = \alpha \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i^2}.$$

Assim, $\hat{\beta}$ é não enviesado apenas se $\alpha = 0$ ou $\bar{x} = 0$. O sinal do enviesamento depende de α e de $\sum_{i=1}^{n} x_i$.

$$\mathrm{Var}\left(\hat{\beta}|X\right) = \mathrm{Var}\left(\frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}\middle|X\right) = \frac{\sum_{i=1}^{n} x_i^2 \,\mathrm{Var}(y_i|X)}{\left(\sum_{i=1}^{n} x_i^2\right)^2} = \frac{\sum_{i=1}^{n} x_i^2 \,\mathrm{Var}(\alpha + \beta x_i + u_i|X)}{\left(\sum_{i=1}^{n} x_i^2\right)^2}$$

$$= \frac{\sigma^2 \sum_{i=1}^{n} x_i^2}{\left(\sum_{i=1}^{n} x_i^2\right)^2} = \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}.$$

Observa-se que $\mathrm{Var}\left(\hat{\beta}|X\right) = \mathrm{Var}\left(\hat{\beta}_{OLS}|X\right)$.

Questão	**1a**	**1b**	**1c**	**2**	**3**	**4a**	**4b**	**4c**	**4d1**	**4d2**	**4e**	**5**
Cotação	15	15	20	20	20	10	15	15	20	10	20	20

ANEXO

EQUAÇÃO 1
Dependent Variable: MILES
Included observations: 200

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	−391.5480	169.7752	−2.306273	0.0221
INCOME	14.20133	1.800256	7.888506	0.0000
AGE	15.74092	3.757370	4.189346	0.0000
KIDS	−81.82642	27.12960	−3.016131	0.0029
R-squared	0.340605	Mean dependent var		1054.230
Adjusted R-squared	0.330512	S.D. dependent var		552.7990
S.E. of regression	452.3125	Akaike info criterion		15.08642
Sum squared resid	40098973	Schwarz criterion		15.15239
Log likelihood	−1504.642	Hannan-Quinn criter.		15.11312
F-statistic	33.74740	Durbin-Watson stat		1.948060
Prob(F-statistic)	0.000000			

Coefficient Covariance Matrix

	C	INCOME	AGE	KIDS
C	28823.62	−142.6730	−446.6777	233.4264
INCOME	−142.6730	3.240923	−1.543798	0.843166
AGE	−446.6777	−1.543798	14.11783	−34.93170
KIDS	233.4264	0.843166	−34.93170	736.0150

EQUAÇÃO 2
Dependent Variable: RESIDUAL^2
Included observations: 200

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	−630678.4	523380.4	−1.205010	0.2297
INCOME	21644.67	7831.167	2.763914	0.0063
AGE	1529.975	25030.70	0.061124	0.9513
KIDS	81985.66	119500.9	0.686067	0.4935
INCOME^2	−44.32885	50.26254	−0.881946	0.3789
AGE^2	220.7181	329.2774	0.670310	0.5035
KIDS^2	14459.40	15165.47	0.953442	0.3416
INCOME*AGE	−315.0473	131.5589	−2.394724	0.0176
INCOME*KIDS	2101.332	923.2261	2.276076	0.0240
AGE*KIDS	−5269.675	2702.375	−1.950016	0.0526
R-squared	0.199964	Mean dependent var		200494.9
Adjusted R-squared	0.162068	S.D. dependent var		318053.0
S.E. of regression	291141.3	Akaike info criterion		28.04971
Sum squared resid	1.61E + 13	Schwarz criterion		28.21463
Log likelihood	−2794.971	Hannan-Quinn criter.		28.11645
F-statistic	5.276598	Durbin-Watson stat		2.052532
Prob(F-statistic)	0.000002			

EQUAÇÃO 3
Dependent Variable: MILES
Included observations: 200

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	−1360.553	1187.171	−1.146047	0.2532
INCOME	34.97383	24.03814	1.454931	0.1473
AGE	39.06872	26.89428	1.452678	0.1479
KIDS	−199.6796	138.5500	−1.441210	0.1511
FITTED^2	−0.001634	0.001720	−0.950078	0.3433
FITTED^3	5.49E − 07	5.50E − 07	0.998163	0.3194
R-squared	0.344206	Mean dependent var		1054.230
Adjusted R-squared	0.327304	S.D. dependent var		552.7990
S.E. of regression	453.3948	Akaike info criterion		15.10094
Sum squared resid	39879974	Schwarz criterion		15.19989
Log likelihood	−1504.094	Hannan-Quinn criter.		15.14099
F-statistic	20.36495	Durbin-Watson stat		1.952065
Prob(F-statistic)	0.000000			