Instituto Superior de Economia e Gestão - UL

Licenciaturas em Economia e Finanças

Estatística II, ER - 25 Junho de 2015

Duração da prova: 2h

Tópicos de Resolução

1. **[2]** Considere a variável aleatória X caracterizada pela seguinte função de densidade:

$$f_X(x;\theta) = \frac{e^{-\frac{x}{\theta}}}{\theta},\ \text{para } x \geq 0\,(\theta > 0).$$

Dada uma amostra de dimensão n proveniente da população de X, mostre que o estimador da máxima verosimilhança para θ é $\bar{X}$ e determine o estimador da máxima verosimilhança para a seguinte função do parâmetro θ:

$$\tau(\theta) = \theta^{\frac{5}{2}}.$$

$L(\theta) = \prod_{i=1}^{n} f_X(x_i;\theta) = exp(-\frac{1}{\theta}\sum x_i)/\theta^n, \quad l(\theta) = logL(\theta) = -\frac{1}{\theta}\sum x_i - nln(\theta)$

$\frac{\delta}{\delta\theta} l(\theta) = 0 \Leftrightarrow \frac{1}{\theta^2}\sum x_i = \frac{n}{\theta} \Leftrightarrow \hat{\theta} = \bar{X}$

Pela propriedade da invariância do EMV: $\tau\hat{(\theta)} = \tau\left(\hat{\theta}\right) = \hat{\theta}^{5/2} = \bar{X}^{5/2}$

2. **[1]** Seja $\tau\hat{(\theta)}$ o estimador da máxima verosimilhança para $\tau(\theta)$ que obteve na pergunta anterior. Admita que o dito estimador é **centrado**. Nestas circunstâncias indique a resposta **FALSA**:

[ ] Se existir o estimador mais eficiente para $\tau(\theta)$ então esse estimador é $\tau\hat{(\theta)}$.

[ ] Se $\tau\hat{(\theta)}$ for o estimador mais eficiente para $\tau(\theta)$ então este tem variância estritamente igual ao limite inferior de Cramer-Rao.

[ ] Pode ou não existir o estimador mais eficiente para $\tau(\theta)$.

[X] $\tau\hat{(\theta)}$ poderá não ser consistente.

3. **[1.5]** Dada a amostra aleatória (X_1, X_2, X_3, X_4) da população X, com $Var(X) = \sigma^2$ e $E[X] = \mu$, e os seguintes estimadores para μ,

$$\begin{aligned} T_1 &= \bar{X} \\ T_2 &= X_1 + X_2 - X_3 \\ T_3 &= 2X_4 - X_1 \end{aligned}$$

comente a seguinte afirmação: T_1 *é mais eficiente do que* T_3 *mas* T_3 *é mais eficiente do que* T_2. Apresente os cálculos que fundamentam a sua resposta.

$Var(T_1) = Var(\bar{X}) = \frac{1}{n^2}\sum Var(X_i) \overset{iid}{=} \frac{1}{n}Var(X) = \frac{\sigma^2}{n}$
$Var(T_2) = Var(X_1 + X_2 - X_3) \overset{i}{=} Var(X_1) + Var(X_2) + Var(X_3) \overset{id}{=} 3\sigma^2$
$Var(T_3) = Var\left(2X_4 - X_1\right) \overset{i}{=} 4Var(X_4) + Var(X_1) \overset{id}{=} 5\sigma^2$
$\therefore$ $Var(T_1) < Var(T_2) < Var(T_3)$, a afirmação é falsa pois T_1 *é mais eficiente do que* T_3 *e* T_3 *é menos eficiente do que* T_2.

4. **[1.5]** Seja X uma v.a. com distribuição desconhecida, observou-se a amostra aleatória $(X_1, \cdots, X_{225})$ $(n = 225)$ e registaram-se as quantidades $\bar{x} = 10$ e $s^2 = 1$. Parece-lhe pertinente afirmar-se com 95% de confiança que 9 é um valor admissível para μ?

$VF : Z = \frac{\bar{X}-\mu}{S/\sqrt{n}} \overset{a}{\sim} N\left(0, 1\right)$
$\mu \in [\bar{x} \pm 1.96 \times s/\sqrt{n}] = [10 \pm 1.96/15] = [9.87; 10.13]$
$\therefore$ Como $9 \notin IC$ não é um valor admissível a 95% de confiança

5. **[1]** De uma população normal com variância conhecida e igual a 25 extraiu-se uma amostra aleatória de dimensão 100. Através do método da Variável Fulcral obteve-se o seguinte intervalo de confiança para μ: $[2,1775;\ 3,8225]$. Qual o nível de confiança que se deve atribuir ao dito intervalo?

- [ ] 99%.
- [ ] 95%.
- [X] 90%.
- [ ] 5%.

6. **[1]** Seja X uma v.a. proveniente de uma população Normal onde se verifica que $n = Var(X)$. Admita ainda que com base numa amostra *iid* de dimensão n se pretende ensaiar o seguinte teste:

$$\begin{cases} H_0 : & \mu = a \\ H_1 : & \mu = b \end{cases}$$

Tenha ainda em ponderação a seguinte região de rejeição:

$$W_\alpha = \{(x_1, \cdots, x_n) : \bar{x} < c\}$$

para $b < c < a$. Nestas circunstâncias a potência do teste é dada por:

- [X] $\Phi\left(c - b\right)$.
- [ ] $\Phi\left(c - a\right)$.
- [ ] 1-$\Phi\left(a - c\right)$.
- [ ] 1-$\Phi\left(c - b\right)$.

7. **[2]** Foi recolhida uma amostra aleatória de 200 portugueses e observou-se que dos 150 que defendem a saída de Portugal da Zona Euro 80 têm formação superior. Constatou-se ainda que os indivíduos sem formação superior e que defendem que Portugal deve permanecer na Zona Euro são 40. Podemos afirmar que o facto dos indivíduos terem ou não formação superior é independente da sua opinião em relação à saída de Portugal da Zona Euro?

Frequências Observadas	Saída do Euro	Permanência no Euro	Totais
Com Formação Superior	80	10	**90**
Sem Formação Superior	70	40	**110**
Totais	**150**	**50**	**200**

Frequências Esperadas	Saída do Euro	Permanência no Euro	Totais
Com Formação Superior	67.5	22.5	**90**
Sem Formação Superior	82.5	27.5	**110**
Totais	**150**	**50**	**200**

$\begin{cases} H_0: & p_{i\circ}p_{\circ j} = p_{ij} \\ H_1: & H_0 \text{ falsa} \end{cases}$ sob H_0, $Q = \sum_i \sum_j \frac{(N_{ij} - n\hat{p}_{i\circ}\hat{p}_{\circ j})^2}{n\hat{p}_{i\circ}\hat{p}_{\circ j}} \overset{a}{\sim} \chi^2_{[(r-1)(s-1)]}$

$Q_{obs} = 16.835$ $W_{5\%} = \{Q_{obs} > 3.84\} \Rightarrow$ Rejeito H_0

$\therefore$ A evidência empírica sugere que a formação dos indivíduos não é independente da sua opinião em relação à saída de Portugal da Zona Euro.

8. **[1]** Indique a resposta **FALSA**. O Modelo de Regressão Linear Múltipla:

[ ] Exibe um valor para o R^2 superior ou igual ao do Modelo de Regressão Linear Simples, para a mesma variável dependente.

[ ] Apresenta uma soma do quadrado dos resíduos inferior ou igual àquela que é obtida pelo Modelo de Regressão Linear Simples, para a mesma variável dependente.

[ ] Permite efetuar análise *ceteris paribus* ao contrário do Modelo de Regressão Linear Simples.

[X] Não permite que o estimador OLS seja obtido através do Método dos Momentos ao contrário do Modelo de Regressão Linear Simples.

9. Um Economista do Desenvolvimento pretende explicar a esperança média de vida num conjunto de países, para o efeito estimou a seguinte equação:

$$Esp_i = \beta_0 + \beta_1 LPIBpc_i + \beta_2 Gini_i + \beta_3 Mort_i + \beta_4 Mortinf_i + u_i$$

onde as variáveis têm o seguinte significado:

- Esp_i: Esperança média de vida no país i;
- $LPIBpc_i$: logaritmo do PIB *per capita* do país i;
- $Gini$: Índice de Gini, em %, do país i;
- $Mortinf$: Taxa de mortalidade infantil do país i em %;
- $Mort$: Taxa de mortalidade total do país i em %;

Tendo em conta os resultados da Estimação 1 e da Estimação 2 disponíveis no Anexo, responda as seguintes questões assumindo que se verificam as hipóteses clássicas (a não ser que tenha explícita evidência do contrário).

(a) **[2]** Interprete as estimativas dos coeficientes das variáveis $LPIBpc_i$ e $Mortinf$. Teste individualmente a sua significância estatística.

Estima-se que em média quando o $PIBpc$ aumenta 1% a esperança média de vida aumenta 0.01 anos mantendo tudo o resto constante.
$H_0 : \beta_1 = 0$ contra $H_1 : \beta_1 \neq 0$ sob H_0, $t = \frac{\hat{\beta}_1}{SE(\hat{\beta}_1)} \sim t(n-k-1)$, $t_{obs} = 3.79$,
$W_{5\%} = \left\{t : |t_{obs}| > t_{\frac{\alpha}{2}}\right\}$, $p-value = 0.0004 \Rightarrow$Rejeito a hipótese nula, i.e, o PIB_{pc} é estatisticamente significativo.

Estima-se que em média quando a mortalidade infantil aumenta 1pp a esperança média de vida diminui em 6.62 anos mantendo tudo o resto constante.
$H_0 : \beta_4 = 0$ contra $H_1 : \beta_4 \neq 0$ sob H_0, $t = \frac{\hat{\beta}_4}{SE(\hat{\beta}_4)} \sim t(n-k-1)$, $t_{obs} = -7.04$,
$W_{5\%} = \left\{t : |t_{obs}| > t_{\frac{\alpha}{2}}\right\}$ $p-value = 0.0000 \Rightarrow$Rejeito a hipótese nula, i.e., a $Mortinf$ é estatisticamente significativa.

(b) **[0.5]** Mostre como se obteve o valor da *F-statistic* na Estimação 1.

$F = \frac{R^2/k}{(1-R^2)/(n-k-1)} \sim F(k, n-k-1)$
$F-statistic = F_{obs} = \frac{0.888440/4}{(1-0.888440)/(63-4-1)} = 115.4749$

(c) **[2]** Com o objetivo de se testar se um acréscimo de 1 ponto percentual na mortalidade infantil diminui mais a esperança média de vida do que um acréscimo de 2 pontos percentuais na mortalidade total, o economista obteve a Estimação 2. Deduza esta equação e retire as devidas conclusões.

$\begin{cases} H_0: & \beta_4 = 2\beta_3 \\ H_1: & \beta_4 < 2\beta_3 \end{cases} \Leftrightarrow \begin{cases} H_0: & \theta = 0 \\ H_1: & \theta < 0 \end{cases}$ onde $\theta \equiv \beta_4 - 2\beta_3$
fazendo $\beta_4 = \theta + 2\beta_3$ vem

$$\begin{aligned}
Esp_i &= \beta_0 + \beta_1 LPIBpc_i + \beta_2 Gini_i + \beta_3 Mort_i + (\theta + 2\beta_3)\, Mortinf_i + u_i \Leftrightarrow \\
Esp_i &= \beta_0 + \beta_1 LPIBpc_i + \beta_2 Gini_i + \beta_3 Mort_i + \theta Mortinf_i + 2\beta_3 Mortinf_i + u_i \Leftrightarrow \\
Esp_i &= \beta_0 + \beta_1 LPIBpc_i + \beta_2 Gini_i + \theta Mortinf_i + \beta_3 (Mort_i + 2Mortinf_i) + u_i \Leftrightarrow
\end{aligned}$$

$t_{obs} = -0.133534$, $p-value = 0.8942/2 \Rightarrow$ Não rejeito H_0

$\therefore$ Um acréscimo de 1 ponto percentual na mortalidade infantil não diminui mais a esperança média de vida do que um acréscimo de 2 pontos percentuais na mortalidade total

(d) **[1]** Com o objetivo de averiguar a presença de heterocedasticidade nos erros do Modelo da Estimação 1, o Economista efetuou um teste conhecido e que foi estudado nas aulas de Estatística II. Para o efeito, estimou uma regressão do quadrado dos resíduos da Estimação 1 sobre um conjunto de regressores. Sabendo que a estatística teste tem distribuição do Qui-Quadrado com 14 graus de liberdade, podemos concluir que se trata de um teste de:

- [ ] Preusch-Bagan
- [ ] Breusch-Pagan
- [X] White
- [ ] Reset

(e) **[1.5]** Sabe-se que no teste anterior o Economista obteve o seguinte valor para a estatística teste observada: $LM_{obs} = 50$. Que implicações poderá ter este resultado nas propriedades do estimador OLS da Estimação 1 do anexo?

Teste de Heterocedasticidade de White

$$\begin{cases} H_0: & Var\left(u_i | X\right) = \sigma^2 \\ H_1: & Var\left(u_i | X\right) = \sigma_i^2 \end{cases}; \begin{cases} H_0': & \alpha_1 = \cdots = \alpha_{14} = 0 \\ H_1': & \exists \alpha_j \neq 0, j = 1, \cdots 14. \end{cases}$$

Regressão auxiliar:

$$\begin{aligned} \hat{u}^2 &= \alpha_0 + \alpha_1 LPIBpc_i + \alpha_2 Gini_i + \alpha_3 Mort_i + \alpha_4 Mortinf_i + \\ &+ \alpha_5 LPIBpc_i^2 + \alpha_6 Gini_i^2 + \alpha_7 Mort_i^2 + \alpha_8 Mortinf_i^2 + \\ &+ \alpha_9 LPIBpc_i \times Gini_i + \cdots \alpha_{14} Mort_i \times Mortinf_i + erro \end{aligned}$$

Sob H_0, $LM = nR^2_{\hat{u}^2} \overset{a}{\sim} \chi^2_{(14)}$
$W_{5\%} = \{LM_{obs} > 23.685\}$ $LM_{obs} \in W \Rightarrow$Rej H_o
$\therefore$ Existe evidência empírica de heterocedasticidade. O estimador para β_j, com $j = 0, 1, \ldots, 4$, não é BLUE, mas continua centrado e consistente (partindo do princípio de que se verificam as restantes hipóteses). Os erros-padrão OLS são inconsistentes, logo a inferência habitual deixa de ser válida.

(g) **[1]** Um outro Economista sugeriu a introdução de uma outra variável: a taxa de mortalidade dos adultos, isto é, a taxa de mortalidade total depois de lhe descontada a mortalidade infantil. Concorda? Comente as implicações desta afirmação sobre as propriedades do estimador dos mínimos quadrados.

Violar-se-ia a hipótese da ausência de multicolinearidade perfeita pois a taxa de mortalidade total é igual à taxa de mortalidade dos adultos mais a taxa de mortalidade infantil. Conclusão: o estimador OLS nestas circunstâncias não existe.

10. **[1]** Admita que o seu objetivo é o de estimar o seguinte modelo:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + u_i \tag{1}$$

Pelo meio estimou o modelo:

$$x_{i1} = \gamma_0 + \gamma_1 x_{i2} + \gamma_3 x_{i3} + v_i \tag{2}$$

tendo obtido um coeficiente de determinação (R^2) de 0.9. Nestas circunstâncias:

☐ Existe multicolinearidade fraca entre os regressores do modelo (**??**) logo o estimador OLS não é BLUE.

☐ Existe multicolinearidade perfeita entre os regressores do modelo (**??**) logo não é possível a estimação de todos os seus coeficientes.

☐ Existe multicolinearidade perfeita entre os regressores do modelo (**??**) logo o estimador OLS é BROWN.

[X] Existe um problema sério de multicolinearidade entre as variáveis do modelo (**??**) que pode implicar uma precisão pequena na estimação dos respetivos parâmetros.