ESTATÍSTICA II (Eco+Fin) - 2º Ano/1º sem.

Época Normal – 18.01.2020

1.5horas (14 valores)

Nome:____

Espaço reservado para classificações

1 a) (15)	2 a) (10)	3 a) (20)
1 b) (10)	2 b) (5)	3 b) (15)
	2 c) (10)	3 c) i) (10)
	2 d) (10)	3 c) ii) (20)
	2 e) (15)	
T:		

1. A distância percorrida diariamente por determinado autocarro é uma variável aleatória X com média igual a $\theta + 2$, variância igual a θ^2 e função densidade dada por:

$$f(x|\theta) = \frac{1}{\theta} e^{-\frac{x_i - 2}{\theta}} \quad x > 2, \theta > 0$$

(a) Mostre que o estimador de máxima verosimilhança para θ é $\bar{X} - 2$.

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\theta} e^{-\frac{x_i-2}{\theta}} = \left(\frac{1}{\theta}\right)^n e^{-\frac{\sum_{i=1}^{n} x_i - 2n}{\theta}} \Rightarrow l(\theta) = -nln(\theta) - \frac{1}{\theta}\left(\sum_{i=1}^{n} x_i - 2n\right)$$

$$\frac{dl(\theta)}{d\theta} = -\frac{n}{\theta} + \frac{1}{\theta^2}\left(\sum_{i=1}^{n} x_i - 2n\right) = 0 \Leftrightarrow -n\theta + \sum_{i=1}^{n} x_i - 2n = 0 \Leftrightarrow \theta = \bar{X} - 2$$

$$\frac{l''(\theta)}{d\theta^2} < 0$$

(b) Estude a consistência do estimador encontrado.

$$\lim_{n\to\infty} E(\bar{X} - 2) = \lim_{n\to\infty} E(\bar{X}) - 2 = \lim_{n\to\infty} \theta + 2 - 2 = \theta$$

$$\lim_{n\to\infty} Var(\bar{X} - 2) = \lim_{n\to\infty} Var(\bar{X}) = \lim_{n\to\infty} \frac{\theta^2}{n} = 0$$

Pelo que o estimador é consistente.

2. A companhia aérea *TravelAir* analisou as horas de chegada de 51 dos seus voos, tendo observado que a média dos atrasos, em minutos, foi de 67, com um desvio padrão corrigido de 59.

(a) Construa um intervalo de confiança a 95% para a variância do atraso nos voos da companhia *TravelAir*.

$X -$ atraso, em minutos, nos vôos da $TravelAir \sim N(\mu, \sigma^2\ );\ \bar{x} = 67;\ s' = 59$

$$Variável\ fulcral: T = \frac{(n-1)S'^2}{\sigma^2} \sim \chi^2_{(n-1)};\ IA^{95\%}_{\sigma^2} = \left(\frac{(n-1)\boldsymbol{S'^2}}{q_2}, \frac{(n-1)\boldsymbol{S'^2}}{q_1}\right)$$

$q_1: P\left(\chi^2_{(n-1)} < q_1\right) = 0.025 \Leftrightarrow q_1 = chiinv(0.025, 50) = 32.3574;$

$q_2: P\left(\chi^2_{(n-1)} > q_2\right) = 0.025 \Leftrightarrow q_2 = chiinv(0.975, 50) = 71.4202$

$$IC^{95\%}_{\sigma^2} = \left(\frac{50 * 59^{\boldsymbol{2}}}{71.4202}, \frac{50 * 59^{\boldsymbol{2}}}{32.3574}\right) = (2436.986,\ \ 5378.992)$$

(b) A probabilidade de a variância do atraso dos voos da companhia *TravelAir* estar contida no intervalo encontrado na alínea anterior é:

	95%		Qualquer valor entre 0 e 1
	5%	X	0 ou 1

(c) Os responsáveis da companhia afirmam que o atraso médio dos seus voos não excede uma hora. Concorda com a afirmação? Responda com base num teste ao nível de significância de 5%.

$\boldsymbol{H_0}: \mu \leq 60$ contra $\boldsymbol{H_1}: \mu > 60$; Estatística teste: $\frac{\bar{X}-\mu_0}{S'/\sqrt{n}} \sim t(n-1);\ \alpha = 0.05$

Região crítica -$W_{\bar{X}} = \{\bar{x}: \bar{x} > k\ \}\ com\ k: P(\bar{X} > k|\boldsymbol{H_0}) = 0.05$

$$\Rightarrow k = invt(0.95, 50) * \frac{59}{\sqrt{51}} + 60 = 1.676 * \frac{59}{\sqrt{51}} + 60 = 73.8465$$

como $\bar{x} = 67 \in W_{\bar{X}}$,

ou

$$valor - p = P(\bar{X} > 67|\mu \leq 60) = P\left(\frac{\bar{X} - \mu_0}{S'/\sqrt{n}} > \frac{67 - 60}{59/\sqrt{51}}\right) = P\left(t_{(51-1)} > 0.8473\right)$$

$$= 0.2004 > 0.05$$

Então, não se rejeita $\boldsymbol{H_0}$ pelo que se pode concluir que a afirmação dos responsáveis da companhia é verdadeira.

(d) Calcule a potência do ensaio se a verdadeira média da população for igual a 61 e 70. O que conclui com base nos resultados sobre a qualidade do ensaio realizado na alínea anterior?

$$\beta(\mathbf{61}) = P(rejeitar\boldsymbol{H_0}|\mu = 61) = P(\bar{X} > 73.8465|\mu = 61)$$

$$= P\left(\frac{\bar{X}-\mu_0}{S'/\sqrt{n}} > \frac{73.8465\ -61}{59/\sqrt{51}}\right) = P\big(t_{(51-1)} > 1.555\big) = 0.0631$$

$$\beta(\mathbf{65}) = P(rejeitar\boldsymbol{H_0}|\mu = 70) = P(\bar{X} > 73.8465|\mu = 70)$$

$$= P\left(\frac{\bar{X}-\mu_0}{S'/\sqrt{n}} > \frac{73.8465\ -70}{59/\sqrt{51}}\right) = P\big(t_{(51-1)} > 0.4656\big) = 0.3218$$

como a probabilidade de rejeitar $\boldsymbol{H_0}$ quando $\boldsymbol{H_0}$ é falsa aumenta quando a média da população se afasta de $\boldsymbol{H_0}$, no sentido da alternativa, o teste é um bom teste.

(e) Com o objetivo de testar se os atrasos dos voos seguem uma distribuição normal com média 60 e desvio padrão 59, o responsável pela Satisfação do Cliente reanalisou os dados dos 51 voos recolhidos, tendo-os classificado da seguinte forma:

Atraso	< 60	[60, 120)	[120, 180)	≥ 180
Número de voos	20	18	8	5
$P(A_j)$	0.5	0.3454	0.1336	0.02098
Freq. esperadas	25.5	17.616	6.814	1.07

Concorda com a hipótese de que os atrasos seguem uma distribuição normal com os parâmetros indicados?

$\boldsymbol{H_0}: \boldsymbol{X} \sim \boldsymbol{N}(\mathbf{60}, \mathbf{59^2})$

Como a frequência esperada da última célula é inferior a 5:

Atraso	< 60	[60, 120)	≥ 120
Número de voos	20	18	8+5=13
$P(A_j)$	0.5	0.3454	0.1336
Freq. esperadas	25.5	17.616	6.814+1.07=7.884
q_j	1.1863	0.0084	3.3198

Estatística teste: $Q = \sum_{j=1}^{3} \frac{\left(N_j - n * p_{0_j}\right)^2}{n * p_{0_j}} \dot{\sim} \chi^2_{(3-1)}$

$$W = \{q: q > q_\alpha\}\ com\ q_\alpha: \mathrm{P}(Q > q_\alpha) = 0.05 \Rightarrow q_\alpha = 5.9915$$

como, $q_{obs} = 4.5144 < 5.9915 \Rightarrow$ não se rejeita $\boldsymbol{H_0}$

3. Por forma a estudar um indicador relativo à qualidade dos serviços de saúde nos países da OECD, como resposta à pandemia COVID-19, foi estimado o seguinte modelo (via Mínimos Quadrados Ordinários):

$$saude = \beta_0 + \beta_1 \log(desp) + \beta_2\ educ + \beta_3\ pibpc + \beta_4\ gini + u$$

O *output* do modelo estimado:

Dependent Variable: SAUDE
Method: Least Squares
Included observations: 217

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	39.08905	3.133142	12.47599	0.0000
LOG(DESP)	6.692624	0.746897	8.960568	0.0000
EDUC	1.160908	0.222859	5.209165	0.0000
PIBPC	-0.000242	0.000115	-2.099302	0.0370
GINI	-18.29416	5.024727	-3.640826	0.0003

R-squared	0.840799	Mean dependent var	75.00928
Adjusted R-squared	0.837795	S.D. dependent var	12.46736
S.E. of regression	5.021185	Akaike info criterion	6.087981
Sum squared resid	5345.008	Schwarz criterion	6.165859
Log likelihood	-655.5459	Hannan-Quinn criter.	6.119440
F-statistic	279.9124	Durbin-Watson stat	1.502682
Prob(F-statistic)	0.000000		

Em que:

- $saude$, é o indicador de saúde do país *i*;
- $desp$, corresponde às despesas de saúde do país *i* (em milhares de euros);
- $educ$, corresponde à mediana do nível educacional do país *i* (em anos);
- $pibpc$, corresponde ao Produto Interno Bruto (PIB), *per capita*, do país *i* (em milhões de euros, por habitante);
- $gini$, corresponde ao índice de Gini do país *i*.

(a) Interprete as estimativas para β_1 e para β_4 e teste a sua significância individual

$\beta_1 -$ acréscimo de 1% das despesas de saúde do país gera, *ceteris paribus* um acréscimo médio da qualidade da saúde de aproximadamente 0.0669 unidades

$$H_0: \beta_1 = 0\ \ contra\ H_0: \beta_1 \neq 0$$

$T = \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\hat{\beta}_1}} \sim t\left(\overbrace{217 - 4 - 1}^{212}\right)$ como $\underbrace{valor - p = 0.0000}_{dado\ no\ quadro\ 3c1} \Rightarrow rejeita - se\ H_0$

$\beta_4 -$ acréscimo do índice de Gini em 1 unidade gera, *ceteris paribus* um decréscimo médio da qualidade da saúde de aproximadamente 18.29 unidades

$$H_0: \beta_4 = 0 \;\; contra \; H_0: \beta_4 \neq 0$$

$$T = \frac{\hat{\beta}_4 - \beta_4}{\hat{\sigma}_{\hat{\beta}_4}} \sim t\left(\overbrace{217 - 4 - 1}^{212}\right) \quad \text{como } v\underbrace{alor - p = 0.0003}_{dado\ no\ quadro\ 3c1} \Rightarrow rejeita - se\ H_0$$

(b) Considere a matriz de covariâncias dos coeficientes do modelo estimado:

	C	LOG(DESP)	EDUC	PIBPC	GINI
C	9.8166	-1.4591	0.0053	0.0002	-8.2232
LOG(DESP)	-1.4591	0.5579	-0.0998	-0.0001	-1.0466
EDUC	0.0053	-0.0998	0.0497	≈0.0000	0.4496
PIBPC	0.0002	-0.0001	≈0.0000	≈0.0000	0.0002
GINI	-8.2232	-1.0466	0.4496	0.0002	25.2479

Teste a hipótese que uma variação de 5 anos da mediana do nível educacional tem um efeito simétrico ao de uma variação unitária no índice de Gini. Responda com base num teste ao nível de significância de 1%.

$H_0: 5\beta_2 = -\beta_4 \Leftrightarrow H_0: 5\beta_2 + \beta_4 = 0 \Leftrightarrow \delta = 5\beta_2 + \beta_4$ contra $H_0: 5\beta_2 + \beta_4 \neq 0$

Estatística teste: $T = \frac{\hat{\delta} - \delta}{\hat{\sigma}_{\hat{\delta}}} \sim t\left(\underbrace{n - k - 1}_{217-4-1}\right)$;

$$\hat{\sigma}_{\delta}^2 = var\left(5\hat{\beta}_2 + \hat{\beta}_4\right) = 5^2 var\left(\hat{\beta}_2\right) + var\left(\hat{\beta}_4\right) + 2 * 5 * covar\left(\hat{\beta}_2, \hat{\beta}_4\right)$$

$$t_{obs} = \frac{5 * 1.160908 - 18.29416}{sqrt(25 * 0.0497 + 25.2479 + 2 * 5 * 0.4496)} = \frac{-12.4896}{5.5655} = -2.2437$$

$valor - p = P(|T| > |-2.2437|) = 0.0259 > 0.01$, então não se rejeita H_0.

(c) Considere os seguintes *outputs*:

Output 3c1

Dependent Variable: SAUDE
Method: Least Squares
Included observations: 217

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-70.44850	61.84258	-1.139158	0.2559
LOG(DESP)	-30.27991	27.05471	-1.119210	0.2643
EDUC	-5.042263	4.708939	-1.070785	0.2855
PIBPC	0.001574	0.000989	1.591809	0.1129
GINI	71.31145	73.51598	0.970013	0.3332
FITTED^2	0.095406	0.055860	1.707960	0.0891
FITTED^3	-0.000537	0.000254	-2.111983	0.0359

R-squared	0.869741	Mean dependent var	75.00928
Adjusted R-squared	0.866020	S.D. dependent var	12.46736
S.E. of regression	4.563465	Akaike info criterion	5.905768
Sum squared resid	4373.295	Schwarz criterion	6.014797
Log likelihood	-633.7758	Hannan-Quinn criter.	5.949811

F-statistic	233.6963	Durbin-Watson stat	1.576708
Prob(F-statistic)	0.000000		

Output 3c2

Dependent Variable: RES^2
Method: Least Squares
Included observations: 217

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-16.80448	22.63510	-0.742408	0.4587
LOG(DESP)	-2.187680	5.395891	-0.405434	0.6856
EDUC	-1.189399	1.610023	-0.738747	0.4609
PIBPC	0.000724	0.000834	0.868075	0.3863
GINI	150.6778	36.30068	4.150827	0.0000

R-squared	0.133526	Mean dependent var	24.63137
Adjusted R-squared	0.117177	S.D. dependent var	38.60751
S.E. of regression	36.27509	Akaike info criterion	10.04291
Sum squared resid	278967.0	Schwarz criterion	10.12079
Log likelihood	-1084.656	Hannan-Quinn criter.	10.07437
F-statistic	8.167445	Durbin-Watson stat	1.830773
Prob(F-statistic)	0.000004		

i. Identifique os testes em cada um dos *outputs* acima.

Teste RESET e Teste Breush-Pagan para detecção de heterocesdasticidade

ii. Faça os testes respetivos, a um nível de significância de 5%. Com base nas conclusões dos testes, discuta a validade dos resultados obtidos até aqui.

Teste RESET $H_0: \delta_1 = \delta_2 = 0$

$$F = \frac{(SSR_{\,r} - SSR_{\,nr})/\left(\overbrace{k-p}^{m}\right)}{SSR_{\,nr}/(n-k-1)} \sim F_{\left(\underbrace{m}_{2},\ \underbrace{n-k-1}_{217-6-1}\right)} \text{ ou } F = \frac{\left(R^2_{nr} - R^2_{r}\right)/\left(\overbrace{k-p}^{m}\right)}{\left(1-R^2_{nr}\right)/(n-k-1)} \sim F_{\,(m,\, n-k-1)}$$

$$f_{obs} = \frac{(5345.008-4373.295)/2}{4373.295/(217-6-1)} = 23.3302 \text{ ou } f_{obs} = \frac{(0.8697-0.8408)/2}{(1-0.8697)5/(217-6-1)} = 23.2886$$

$valor - p = P\left(F_{(2,210)} > 23.3302\right) \cong 0$ ou $valor - p = P\left(F_{(2,210)} > 23.2886\right) \cong 0$ então rejeita-se H_0 pelo que o modelo está mal especificado

Teste Breush-Pagan para detecção de heterocesdasticidade

$$H_0: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0 \quad contra \ \exists \alpha_j \neq 0 \; j = 1, 2, 3, 4$$

$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)} = \frac{SSE/k}{SSR/(n-k-1)} \sim F_{\,(k,\, n-k-1)}$$

$valor - p = 0.000004 \Rightarrow$rejeita-se homocedasticidade