ESTATÍSTICA II – Economia/Finanças

Duração: 2h – Época Especial – Setembro 2021

Nome:____ Nº:____________

Espaço reservado para classificações				
1.(15)	3a.(15)	4.(20)	5a.(15)	5e.(15)
2a.(15)	3b.(15)		5b.(10)	5f.(10)
2b.(15)	3c.(10)		5c.(10)	5g.(10)
	3d.(10)		5d.(15)	

Atenção: todas as questões devem ser devidamente formalizadas e justificadas. Sempre que fizer um teste de hipóteses formule as hipóteses em teste e apresente a estatística de teste e sempre que fizer um intervalo de confiança apresente a variável fulcral.

1. Numa amostra aleatória com 625 observações obtiveram-se 200 respostas defendendo a realização das eleições autárquicas a 3 de outubro em vez da data fixada pelo governo, 26 de setembro. Construa um intervalo de confiança a 90% para a proporção de eleitores na população favoráveis à realização das eleições a 3 de outubro.

2. Considere a variável aleatória X com função densidade $f(x|\theta) = \frac{2}{\theta^2}(\theta - x)$, $0 < x < \theta$. Sabe-se ainda que $E(X) = \frac{\theta}{3}$ e $var(X) = \frac{\theta^2}{18}$.
   a. Obtenha uma estimativa para θ utilizando o método dos momentos sabendo que se observou a seguinte amostra aleatória (3.2 0.5; 1.2; 12.7; 2.4). Comente o resultado obtido.
   b. Analise a consistência do estimador dos momentos.

3. Uma empresa que explora dois campos de Padel situados em duas localidades diferentes recolheu de forma aleatória e independente as seguintes amostras relativas à idade dos seus clientes (uma amostra em cada localidade) que se supõe seguirem uma distribuição normal.

   Campo 1: 24 27 26 21 24

   Campo 2: 27 28 22 31

   Sabe-se ainda que, para o campo 1, a média amostral é 24.4 e a variância corrigida da amostra é 5.3.
   a. Será sensato admitir que a variância da idade dos jogadores é a mesma nos dois campos? Utilize um teste estatístico de dimensão 0.05.
   b. Assumindo igualdade de variâncias, será sensato admitir que a idade média dos jogadores é a mesma em ambos os campos? Utilize um teste estatístico de dimensão 0.05.
   c. Calcule o valor-p para o teste da alínea anterior e interprete o seu significado.
   d. Construiu-se, pelo processo habitual, um intervalo de confiança para a média das idades dos utilizadores do campo 1 tendo-se obtido (22.82168,25.97832). Diga, justificando, qual a confiança atribuída a esse intervalo.

4. Um analista financeiro defende que os retornos mensais de um determinado ativo seguem uma distribuição normal de média nula. Da observação de 10 anos (120 meses), que se supõem constituir uma amostra aleatória, observaram-se os valores seguintes:

Retornos (%)	$(-\infty;\ -0.3]$	$(-0.3;\ -0.15]$	$(-0.15;\ 0]$	(0; 0.15]	(0.15; 0.3]	(0.3; ∞]
Nº meses	14	18	38	26	16	8

Sabe-se ainda que $s^2 = 0.059$. Recordando que, para a distribuição normal, o estimador de máxima verosimilhança para σ^2 é dado por S^2, teste a afirmação do analista financeiro ($\alpha = 0.05$).

5. A Associação do Comércio Automóvel de determinado país construiu o seguinte modelo par explicar o preço dos automóveis de serviço

$$preço = \beta_0 + \beta_1\, cil + \beta_2\, rpm + \beta_3\, ccid + \beta_4\, cest + \beta_5\, comb + u$$

Onde

$preço$ – Preço em euro de o automóvel se serviço
cil – cilindrada do veículo (em cm3)
rpm – máximo para o número de rotações por minutos do veículo
$ccid$ – consumo do veículo (litros por 100 km) em circuito urbano
$cest$ - consumo do veículo (litros por 100 km) em estrada
$comb$ – variável que assume o valor 1 se o veículo é movido a gasolina e 0 se for movido a diesel.

Estimado o modelo pelos mínimos quadrados, ele obteve o seguinte output

$$\widehat{preço} = \underset{(3122)}{-16900} + \underset{(61.5)}{7.153}\ cil + \underset{(0.6026)}{1.766}\ rpm + \underset{(349.2)}{1381}\, ccid - \underset{(479.7)}{307.7}\ cest - \underset{(90.69)}{5583}\, comb$$

$n = 199$ $\qquad$ $R^2 = 0.8402$

a. Teste a significância estatística do parâmetro associado com consumo em estrada do veículo e, recorrendo ao teste LM (multiplicadores de Lagrange), teste a significância global do modelo.
b. Teste a eliminação simultânea das variáveis relativas ao consumo de combustível.
c. Será aceitável considerar que um aumento do 500 cm3 na cilindrada tem um impacto esperado de pelo menos 3500 euros, *ceteris paribus*. Responda por meio de um teste com $\alpha = 0.05$.
d. Para construir um intervalo de previsão a 95% para o preço esperado de um veículo a gasolina com uma cilindrada de 1400 cm3, 5500 rpm, consumo de cidade de 11 litros aos 100 km e consumo de estrada de 7 litros aos100 km recorreu-se à regressão auxiliar cujo output consta do anexo 2. Defina as variáveis explicativas desta regressão e apresente o intervalo.
e. Será que o impacto no preço esperado de um veículo de serviço de um aumento do consumo em cidade em 1 litro (aos 100 km) é compensado por um aumento de 3 litros no consumo em estrada? Formalize o teste, apresente a regressão auxiliar que utilizaria e explique como procederia.
f. Explique quais as alterações que se verificariam no output apresentado no enunciado caso se dividissem por 1000 todos os valores da variável cilindrada.
g. Um economista da Associação decidiu fazer um teste RESET ao modelo. Qual a ideia do economista? Apresente a regressão auxiliar que utilizaria e admitindo que ao estimar a regressão auxiliar se obteve $R^2 = 0.9293$, realize o teste.

Anexo 1 – Variável dependente: preço

Regression Statistics	
Multiple R	0.89140569
R Square	0.7946041
Adjusted R Square	0.79144416
Standard Error	3643.70936
Observations	199

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	1.0016E+10	3338564992	251.462007	9.4834E-67
Residual	195	2588940495	13276617.9		
Total	198	1.2605E+10			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	-22572.343	3250.40421	-6.9444725	5.517E-11
cil	10.6728293	0.3923225	27.2042246	2.677E-68
rpm	3.30882839	0.63354166	5.22274789	4.503E-07
comb	-3668.2187	978.602726	-3.7484248	0.0002344

Anexo 2 – Variável dependente: preço

Regression Statistics	
Multiple R	0.916629963
R Square	0.840210489
Adjusted R Square	0.836070864
Standard Error	3230.432158
Observations	199

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	5	1059054693	211810937	202.96779	7.79463E-75
Residual	193	2014088542	10435691.9		
Total	198	12604635471			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	10279.95358	875.5531943	11.7410954	2.27561E-24
cil0	7.153187205	0.615010788	11.63099468	4.86039E-24
rpm0	1.766318704	0.60256949	2.931311215	0.003783028
cons_cid0	1381.090288	349.1525398	3.955549884	0.000107181
cons_est0	-307.6731622	479.6958467	-0.64139217	0.52202887
Comb0	-5582.822909	906.8787173	-6.15608549	4.23311E-09

TÓPICOS DE RESOLUÇÃO

1. Seja θ a proporção de eleitores na população favoráveis à realização das eleições a 3 de outubro
   Sabe-se que $Z = \frac{\overline{X}-\theta}{\sqrt{\overline{X}(1-\overline{X})/625}} \dot{\sim} n(0;1)$
   logo o IC a 90% virá dado por $\overline{x} \pm z_{0.05}\sqrt{\frac{\overline{x}\,(1-\overline{x})}{625}}$ isto é $0.32 \pm 1.645\sqrt{\frac{0.32\times 0.68}{625}}$ ou seja
   (0.2893; 0.3507)
2. $f(x|\theta) = \frac{2}{\theta^2}(\theta - x)$, $0 < x < \theta$. Sabe-se ainda que $E(X) = \frac{\theta}{3}$ e $var(X) = \frac{\theta^2}{18}$
   a. Equação a resolver para obter o estimador pelo método dos momentos: $E(X|\theta) = \overline{X}$
      Assim $\frac{\theta}{3} = \overline{X}$ logo $\tilde{\theta} = 3\,\overline{X}$.
      A estimativa é então $\tilde{\theta} = 3\,\overline{x} = 12$ já que $\overline{x} = 4$.
      A estimativa não é admissível uma vez que não respeita a condição $0 < x < \theta$ já que $x_4 > \tilde{\theta}$.
   b. $\lim_{n\to\infty} E(\tilde{\theta}) = \lim_{n\to\infty} E(3\overline{X}) = 3\lim_{n\to\infty} E(\overline{X}) = 3\lim_{n\to\infty} E(X) = \theta$
      $\lim_{n\to\infty} var(\tilde{\theta}) = \lim_{n\to\infty} var(3\overline{X}) = 9\lim_{n\to\infty} var(\overline{X}) = 9\lim_{n\to\infty}\frac{var(X)}{n} = 9\lim_{n\to\infty}\frac{\theta^2}{18\,n} = 0$
      Assim, o estimador dos momentos é consistente
3.
   a. $H_0\colon \sigma_1^2 = \sigma_2^2$ contra $H_1\colon \sigma_1^2 \neq \sigma_2^2$ ou seja $H_0\colon \frac{\sigma_2^2}{\sigma_1^2} = 1$ contra $H_1\colon \frac{\sigma_2^2}{\sigma_1^2} \neq 1$
      Estatística de teste: $F = \frac{S_1^{'2}}{S_2^{'2}} \sim F(4;3)$
      Região de rejeição: $W = \{f\colon f < 0.1002 \; ou \; f > 15.10\}$
      Já que $f_{0.025} = 15.10$ e $f_{0.975} = \frac{1}{9.98} = 0.1002$ tendo em conta que $\frac{1}{F} \sim F(3;4)$
      Como $F_{obs} = \frac{5.3}{14} = 0.3786$ não se rejeita a hipótese nula e, portanto, será "sensato" assumir a igualdade de variâncias.
   b. $H_0\colon \mu_1 = \mu_2$ contra $H_1\colon \mu_1 \neq \mu_2$
      Estatística de teste: $T = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{1}{5}+\frac{1}{4}}\sqrt{\frac{4\,S_1'^2 + 3\,S_2'^2}{7}}} \sim t(7)$
      Região de rejeição: $W = \{t\colon t < -2.365 \; ou \; t > 2.365\}$
      Como $T_{obs} = \frac{24.4-27}{\sqrt{\frac{1}{5}+\frac{1}{4}}\sqrt{\frac{4\times 5.3 + 3\times 14}{7}}} = -1.2899$, não se rejeita a hipótese nula e, portanto, será "razoável" assumir a igualdade de médias.

   c. Assuma-se H0 verdadeiro e calcule-se
      $$valor - p = P(T \le -|T_{obs}|) + P(T \ge |T_{obs}|) = 2 \times P(T > 1.2899) > 0.2$$
      Assumindo H0 verdadeiro, o valor-p traduz a probabilidade de se observar uma amostra que conduza a um valor da estatística de teste tão ou mais desfavorável (para H0) do que aquele que se observou.
   d. A variável fulcral utilizada foi $T = \frac{\overline{X}-\mu}{S'/\sqrt{5}} \sim t(4)$ que origina o IC para μ dado por $\overline{x} \pm t_{\alpha/2}\frac{s'}{\sqrt{5}}$ isto é, $24.4 \pm t_{\alpha/2} \times 1.029563$.
      Da equação $24.4 + t_{\alpha/2} \times 1.029563 = 25.97832$ vem $t_{\alpha/2} = 1.533$ e portanto o grau de confiança será dado por $1 - 2 \times P(T > 1.533) = 1 - 2 \times 0.1 = 0.8$

4. $H_0\colon X \sim n(0;\,\sigma^2)$ contra $H_1\colon H_0\,falsa$
   O estimador para σ^2 é dado por S^2 e portanto a sua estimativa será $\hat{\sigma}^2 = 0.059$
   Dados os intervalos propostos, calcula-se

$\hat{P}(X < -0.3) = \Phi\left(\frac{-0.3}{\sqrt{0.059}}\right) = \Phi(-1.2351) \approx \Phi(-1.24) = 0.1075$
$\hat{P}(X < -0.15) = \Phi\left(\frac{-0.15}{\sqrt{0.059}}\right) = \Phi(-0.61754) \approx \Phi(-0.62) = 0.2676$
$\hat{P}(X < 0) = \Phi(0) = 0.5$
$\hat{P}(X < 0.15) = 0.7324$ Simetria
$\hat{P}(X < 0.3) = 0.8925$ Simetria

E testa-se
$H'_0: p_1 = p_6 = 0.1075; p_2 = p_5 = 0.1601;\ p_3 = \ p_4 = 0.2324;$ contra $H'_1: H_0\ falsa$
Estatística de Teste: $Q = \sum_{j=1}^{6} \frac{(Obs_j - Esp_j)^2}{Esp_j} \dot{\sim} \chi^2_{(4)}$
$W_{0.05} = \{q: q > 9.4877\}$

Intervalo	Obs	Prob	Esp (120*Prob)	χ^2
$(-\infty;\ -0.3]$	14	0.1075-0 =0.1075	12.9	0.0938
$(-0.3;\ -0.15]$	18	0.2676-0.1075=0.1601	19.2	0.0750
$(-0.15;\ 0]$	38	0.5 -0.2676=0.2324	27.9	3.6563
$(0; 0.15]$	26	0.2324	27.9	0.1294
$(0.15; 0.3]$	16	0.1601	19.2	0.5333
$(0.3; \infty]$	8	0.1075	12.9	1.8612

Não se rejeita H0 já que $Q_{obs} = 6.3490$

5. $\widehat{preço} = -16905 + 7.153\ cil + 1.766\ rpm + 1381\ ccid - 307.7\ cest - 5583\ comb$
(3122) (0.6150) (0.6026) (349.2) (479.7) (906.8)

$n = 199$ $R^2 = 0.8402$

a)
Teste individual: $H_0: \beta_4 = 0$ contra $H_1: \beta_4 \neq 0$
Estatística de teste: $T = \frac{\hat{\beta}_4}{se(\hat{\beta}_4)} \sim t(193)$
Região crítica a 5%: $W_{0.05} = \{t: t < -1.96\ ou\ t > 1.96\}$
$t_{obs} = \frac{-307.7}{479.7} = -0.6414$ Não se rejeita H_0 logo a variável $ccest$ **não é** significante
Teste global pelos ML: $H_0: \beta_j = 0$ para $j = 1,2, \ldots, 5$ contra $H_1: \exists\ j: \beta_j \neq 0$
Estatística de teste: $Q = 199\ R^2 \dot{\sim} \chi^2_{(5)}$
Região crítica a 5%: $W_{0.05} = \{q: q > 11.070\}$
$Q_{obs} = 199 \times 0.8402 = 167.2$ Rejeita-se claramente H_0 logo o modelo é globalmente significante
b)
Teste de nulidade conjunta de β_3 e de β_4: $H_0: \beta_3 = \beta_4 = 0$ contra $H_1: \beta_3 \neq 0\ ou\ \beta_4 \neq 0$
Estatística de teste: $F = \frac{R^2_{ur} - R^2_r}{1 - R^2_{ur}} \sim F(2; 193)$
Região crítica a 5%: $W_{0.05} = \{f: f > 3.00\}$
$F_{obs} = \frac{0.8402 - 0.7946}{1 - 0.8402} \times \frac{193}{2} = 27.5369$
Rejeita-se H_0, logo não de deve eliminar simultaneamente as variáveis referentes ao consumo.
c)
$H_0: 500\ \beta_1 \geq 3500$, isto é $H_0:\ \beta_1 \geq 7$ contra $H_1: \beta_1 < 7$
Estatística de teste: $T = \frac{\hat{\beta}_1 - 7}{se(\hat{\beta}_1)} \sim t(193)$
Região crítica a 5%: $W_{0.05} = \{t: t < -1.96\ \}$
$t_{obs} = \frac{7.153 - 7}{61.5} = 0.0025$ Não se rejeita H_0 logo a hipótese feita parece aceitável.

d)

$cil0 = cil - 1400; rpm0 = rpm - 5500; ccid0 = ccid - 11; cest0 = cest - 7; comb0 = comb - 1$

O IP para o preço esperado vem dado, **em termos da regressão auxiliar**, por $\hat{\beta}_0 \pm 1.96 \times se(\hat{\beta}_0)$, isto é , $10279.95 \pm 1.96 \times 875.5532$ ou seja (8563.87; 11996.04)

e)

A afirmação leva a testar $H_0: \theta = \beta_3 + 3\,\beta_4 = 0$ contra $H_1: \theta = \beta_3 + 3\,\beta_4 \neq 0$

Estatística de teste: $T = \frac{\hat{\beta}_3 + 3\,\hat{\beta}_4}{se(\hat{\beta}_3 + 3\,\hat{\beta}_4)} \sim t(193)$

Região crítica a 5%: $W_{0.05} = \{t: t < -1.96\ ou\ t > 1.96\}$

De H0 tira-se $\beta_3 = \theta - 3\,\beta_4$, o que leva à regressão auxilar

$$preço = \beta_0 + \beta_1\,cil + \beta_2\,rpm + (\theta - 3\,\beta_4)\,ccid + \beta_4\,cest + \beta_5\,comb + u$$

$$preço = \beta_0 + \beta_1\,cil + \beta_2\,rpm + \theta\,ccid + \beta_4\,(cest - 3\,ccid) + \beta_5\,comb + u$$

Estima-se esta regressão e testa-se $H_0: \theta = \beta_3 + 3\,\beta_4 = 0$ contra $H_1: \theta = \beta_3 + 3\,\beta_4 \neq 0$. Caso se rejeite H_0 conclui-se que não existe compensação e caso não se rejeite H_0 não se rejeita que a compensação possa existir.

f)

Neste caso apenas se alterariam a estimativa de β_1 e o erro-padrão associado a este estimador vindo ambos multiplicados por 1000.

g)

Testar se a forma funcional adotada é aceitável face aos valores observados. Se designarmos por $\hat{y}$ os valores ajustados que se obtêm com a regressão proposta, ir-se-ia estimar a regressão auxiliar

$$preço = \beta_0 + \beta_1\,cil + \beta_2\,rpm + \beta_3\,ccid + \beta_4\,cest + \beta_5\,comb + \beta_6\,\hat{y}^2 + \beta_7\,\hat{y}^3 + \varepsilon$$

e testar $H_0: \beta_6 = \beta_7 = 0$ contra $H_1: \beta_6 \neq 0\ ou\ \beta_7 \neq 0$

Estatística de teste: $F = \frac{R_{ur}^2 - R_r^2}{1 - R_{ur}^2} \sim F(2; 191)$

Região crítica a 5%: $W_{0.05} = \{f: f > 3.00\}$

$F_{obs} = \frac{0.9293 - 0.8402}{1 - 0.9293} \times \frac{191}{2} = 120.3543$

Rejeita-se claramente H0 concluindo que o modelo se encontra mal formulado!