Estatística II – Lic. Economia e Finanças

ER – 2 de Fevereiro de 2017 – duração: 2 horas

TESTE I – Tópicos de resolução

Nome: ____ **Nº**_________

Espaço reservado para classificações							
1a)	1b)	1c)	1d)	1e)	2a)	2b)	**Total**: _____

Perguntas de escolha múltipla: cada resposta certa vale 10 pontos; cada resposta errada vale -2.5 pontos; assinale a resposta escolhida com uma cruz no quadrado adequado.
As cotações das restantes perguntas são indicadas no enunciado.

1. Admita que a quantidade de açúcar em gramas (g) contida numa embalagem de 100 gramas da nova marca de tostas de trigo tem distribuição normal de média μ e variância igual a 0.64. Para testar a garantia do fornecedor de que a quantidade média de açúcar por embalagem (de 100 g) não ultrapassa 3 g, foi recolhida uma amostra aleatória de 16 embalagens, tendo-se obtido uma média de 3.2 g.

a) **[15]** Teste, ao nível de 5%, a garantia do fornecedor.

Seja X a variável aleatória que representa a quantidade de açúcar, em gramas, contida numa embalagem de 100 gramas de tostas de trigo. Tem-se: $X \sim N(\mu, 0.64)$, onde μ é a quantidade média de açúcar por embalagem de 100 g. Pretende-se testar:

$H_0: \mu \leq 3$ contra $H_1: \mu > 3$

Sob H_0,

$$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

Como $z_{obs} = \frac{\bar{x}-3}{\frac{\sqrt{0.64}}{\sqrt{16}}} = \frac{3.2-3}{\frac{0.8}{4}} = 1$, e $W_{5\%} = \{z: z > 1.645\}$, $z_{obs} \notin W_{5\%}$ e portanto não se rejeita H_0 a 5%: é de admitir, ao nível de 5%, a garantia do fornecedor.

b) **[20]** Calcule a probabilidade de rejeitar a hipótese nula do teste da alínea anterior se a quantidade média de açúcar por embalagem for igual a 3.1 g. Interprete o resultado obtido.

Pretende-se calcular o valor da função potência quando $\mu = 3.1$, ou seja $\beta(3.1)$.

$P(\text{rejeitar } H_0 | \mu = 3.1) = P(\bar{X} > \mu_0 + 1.645 \times \frac{\sigma}{\sqrt{n}} | \mu = 3.1) = P(\bar{X} > 3 + 1.645 \times \frac{0.8}{4} | \mu = 3.1) =$

$P(\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} > \frac{3.329-\mu}{\frac{\sigma}{\sqrt{n}}} | \mu = 3.1) = P(Z > \frac{3{,}329-3.1}{0.2}) = P(Z > 1.145)$, com $Z \sim N(0,1)$.

Obtém-se: $P(Z > 1.145) = 1 - \Phi(1.145) \approx 1 - \Phi(1.15) = 1 - 0.8749 = 0.1251$.
Assim, $\beta(3.1) = 0.1251$.

A probabilidade de não cometer o erro de 2ª espécie, quando $\mu = 3.1$ é igual a 0.1251.

c) A estimativa de máxima verosimilhança para a proporção de embalagens (de 100 g) de tostas de trigo com mais de 3 g de açúcar é igual a:

☐ 0.1587; ☐ 0.4013; ☒ 0.5987; ☐ 0.8413.

d) **[15]** Foi recolhida uma segunda amostra, independente da primeira, de 25 embalagens (de 100 g) de tostas de trigo. Sendo $\bar{X}_1$ e $\bar{X}_2$ as médias da primeira e segunda amostras respectivamente, considere o seguinte estimador para a quantidade média de açúcar por embalagem, $T = a\bar{X}_1 + b\bar{X}_2$, onde a e b são reais. Determine os valores de a e b para os quais T é um estimador centrado para μ.

T é um estimador centrado para μ se $E(T) = \mu$.
Ora, como $T = a\bar{X}_1 + b\bar{X}_2$ e X_{1i} e X_{2i} são iid $N(\mu, 0.64)$, tem-se: $E(\bar{X}_1) = E(\bar{X}_2) = E(X) = \mu$.

Então, $E(T) = E(a\bar{X}_1 + b\bar{X}_2) = a\,E(\bar{X}_1) + b\,E(\bar{X}_2) = a\,\mu + b\mu = (a+b)\mu$.

T é um estimador centrado para μ se e só se $(a+b)\,\mu = \mu \Leftrightarrow a + b = 1 \Leftrightarrow a = 1 - b, \forall b \in R$.

e) Com base na segunda amostra obteve-se um intervalo de confiança para μ de amplitude igual a 0.5264. Então,

o grau de confiança associado a esse intervalo é igual a 0.90.	X
o grau de confiança associado a esse intervalo é igual a 0.95.	
o grau de confiança associado a esse intervalo é igual a 0.99.	
o grau de confiança associado a esse intervalo não pode ser calculado porque a média da segunda amostra, $\bar{x}_2$, não é conhecida.	

2. Considere uma amostra casual simples de dimensão n retirada de uma população X de média $E(X) = \frac{1}{\theta+2}$ e função de densidade, $f(x;\theta) = (\theta + 2)\exp(-(\theta + 2)\,x),\;\; x > 0,\;\; \theta > -2.$

a) **[20]** Determine o estimador de máxima verosimilhança para o parâmetro θ.

Função de verosimilhança $L(\theta)$:

$$L(\theta) = \prod_{i=1}^{n} f(x_i\,|\theta) = \prod_{i=1}^{n} (\theta + 2)\exp(-(\theta + 2)x_i) = (\theta + 2)^n \exp\left(-(\theta + 2)\sum x_i\right)$$

$$l(\theta) = \ln L(\theta) = n\,ln(\theta + 2) - (\theta + 2)\sum x_i$$

Estando satisfeitas as usuais condições de regularidade, podemos obter o máximo da função $L(\theta)$ através do cálculo das duas primeiras derivadas:

$$\frac{dl(\theta)}{d\theta} = 0 \Leftrightarrow \frac{n}{\theta+2} - \sum x_i = 0 \;\Leftrightarrow\; \frac{n}{\theta+2} = \sum x_i \;\Leftrightarrow\; (\theta + 2)\sum x_i = n \;\Leftrightarrow\; \hat{\theta} = \frac{n-2\sum x_i}{\sum x_i} = \frac{1}{\bar{X}} - 2$$

$$\frac{d^2 l(\theta)}{d\theta^2} = -\frac{n}{(\theta+2)^2} < 0$$

b) Admita que se obteve $\bar{x} = 5$, então a estimativa para θ, obtida pelo método dos momentos …

… não é admissível porque o valor obtido é negativo.	
… é igual a -1.8.	X
… não pode ser calculada porque a dimensão da amostra não é conhecida.	
… é igual a 5.	

Estatística II – Lic. Economia e Finanças

2 de Fevereiro de 2017

TESTE II

Nome: ____ **Nº**__________

Espaço reservado para classificações							
3.	4a)	4b)	4c)	4d)	4e)	4f)	**Total**: _____

3. O gerente de vendas de determinado supermercado recolheu dados sobre as vendas de vários tipos de tostas para realizar um teste não paramétrico. Os resultados obtidos, com base numa amostra casual de 1500 vendas, são:

Tostas de trigo	Tostas de trigo integral	Tostas de trigo sarraceno	**Total**
600	495	405	1500

O gerente calculou as frequências esperadas de cada classe e obteve: 750 na primeira classe e 375 em cada uma das restantes classes. Então, o seu estudo teve como objectivo:

realizar um teste de independência entre os vários tipos de tostas.	
testar $H_0: p_1 = 0.4 \wedge p_2 = 0.33 \wedge p_3 = 0.27$ contra $H_1: H_0$ falsa.	
testar $H_0: p_1 = p_2 = p_3 = 0.33(3)$ contra $H_1: \exists\, p_j \neq 0.33(3),\ j = 1,2,3.$	
testar $H_0: p_1 = 0.5 \wedge p_2 = 0.25 \wedge p_3 = 0.25$ contra $H_1: H_0$ falsa.	X

4. Num estudo sobre o desempenho dos alunos universitários foi recolhida, no final do ano lectivo, uma amostra aleatória de 141 alunos do terceiro ano, tendo-se estimado o seguinte modelo:

$$media = \beta_0 + \beta_1 esmedia + \beta_2 educpai + \beta_3 educmae + \beta_4 faltas + u,$$

onde $media$ é a classificação média obtida até ao final do ano lectivo, $esmedia$ é a média obtida no ensino secundário, $educpai$ e $educmae$ são o número de anos de escolaridade do pai e da mãe, respectivamente e $faltas$ é o número de aulas a que o aluno não assistiu no último semestre. Os resultados obtidos na estimação de várias regressões encontram-se no anexo.

a) **[15]** Na **equação 1**, e supondo satisfeitas as hipóteses do modelo de regressão linear, teste a significância individual do coeficiente β_4 e interprete a sua estimativa. Teste também a significância global da regressão.

- $\hat{\beta}_4 = -0.024689$: estima-se que, por cada falta adicional, a classificação média do aluno diminui, em média, aproximadamente, 0.025 pontos, para valores constantes das restantes variáveis.
- Teste de significância individual:

$H_0: \beta_4 = 0$ contra $H_1: \beta_4 \neq 0$. Sob H_0, $T = \frac{\hat{\beta}_4}{se(\hat{\beta}_4)} \sim t(136)$.

Como $t_{obs} = -2.848802$, e $W_{5\%} = \{t: |\, t\,| > 1.98\}$ na tabela da $T(120)$, $t_{obs} \in W_{5\%}$ e portanto rejeita-se H_0 a 5%: β_4 é estatisticamente significativo ao nível de 5%.

- Teste de significância global:

  $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ contra $H_1: \exists\, \beta_j \neq 0\,, j = 1, 2, 3, 4.$

Sob H_0, $F = \frac{R^2/4}{(1-R^2)/136} \sim F(4, 136)$.

Como $f_{obs} = 10.01$, e $W_{5\%} = \{f: f > 2.45\}$, na tabela da F(4, 120), $f_{obs} \in W_{5\%}$ e portanto rejeita-se H_0 a 5%: conclui-se que a regressão é globalmente significativa ao nível de 5%.

b) **[15]** Com base num teste adequado, comente a seguinte afirmação: "O número de anos de escolaridade dos pais de um estudante é um factor estatisticamente relevante para explicar a classificação média obtida no final do ano lectivo".

O número de anos de escolaridade dos pais é um factor relevante para explicar a classificação média obtida se o número de anos de escolaridade da mãe ou o número de anos de escolaridade do pai é um factor relevante. Assim, o teste a realizar é:

$H_0: \beta_2 = \beta_3 = 0$ contra $H_1: \exists\, \beta_j \neq 0\,, j = 2, 3$.

Sob H_0, $F = \frac{(SSR_r - SSR_{ur})/2}{SSR_{ur}/136} \sim F(2, 136)$.

Como $f_{obs} = \frac{(241.3395 - 239.8745)/2}{239.8745/136} = 0.415$, e $W_{5\%} = \{f: f > 3.07\}$, na F(2, 120), $f_{obs} \notin W_{5\%}$ e portanto não se rejeita H_0 a 5%: o número de anos de escolaridade dos pais não parece ser um factor relevante no desempenho escolar do aluno.

c) Admita que se obteve um intervalo de confiança para β_5, coeficiente do novo regressor x_5 do modelo da **equação 1**. Então, para qualquer grau de confiança, pode-se afirmar que …

… o valor zero pertence sempre ao intervalo obtido.	
… o coeficiente β_5 pertence sempre ao intervalo obtido.	
… a estimativa do coeficiente β_5 pertence sempre ao intervalo obtido.	X
… a média amostral de x_5, $\bar{x}_5$ pertence sempre ao intervalo obtido.	

d) **[20]** Utilizando os resíduos do modelo da **equação 2** realizou-se um determinado teste. Os resultados obtidos com o programa EViews são:

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	1.628385	Prob. F(2,138)	0.2000

Qual é o objectivo do teste efectuado? Escreva a regressão auxiliar utilizada. Formalize e efectue o respectivo teste.

Pretende-se testar a presença de heterocedasticidade nos erros do modelo da equação 2:
$H_0: var(u_i|esmedia_i, faltas_i) = \sigma^2$ contra $H_1: var(u_i|esmedia_i, faltas_i) = \sigma_i^2$.

Com base na regressão auxiliar do teste de BP:

$\hat{u}^2 = \alpha_0 + \alpha_1 esmedia + \alpha_2 faltas + v$, onde $\hat{u}$ são os resíduos da equação 2, tem-se:

$H'_0: \alpha_1 = \alpha_2 = 0$ contra $H'_1: \alpha_1 \neq 0 \vee \alpha_2 \neq 0$

Sob H_0, $F = \frac{R^2/2}{(1-R^2)/138} \sim F(2, 138)$. Pelo *output* fornecido, tem-se: $f_{obs} = 1.628$, com valor-p = 0.20. Aos níveis de significância habituais, a hipótese H_0 não é rejeitada, não havendo evidência de heterocedasticidade nos erros do modelo da equação 2.

e) Utilizando o *output* do EViews da alínea anterior, pode-se concluir …

… que o coeficiente de determinação, R^2, da regressão auxiliar é igual a 0.011663	
… que o coeficiente de determinação, R^2, da regressão auxiliar é igual a 0.023056.	X
… que o coeficiente de determinação, R^2, da regressão auxiliar é igual a 0.222733.	
... que a informação dada não permite obter esse R^2.	

f) **[20]** Indique como procederia para testar se a forma funcional da **equação 2** está mal especificada. Deve escrever a(as) equação(ões) que necessita empregar para realizar esse teste, especificar as hipóteses H_0 e H_1 e indicar a distribuição da estatística de teste sob H_0. Admitindo que o valor-p associado a essa estatística é igual a 0.49, que conclusão pode retirar? Apresente também a expressão do valor-p acima referenciado.

Teste RESET de especificação da forma funcional do modelo da equação 2 permite testar:
H_0: especificação correcta contra H_1: presença de erros de especificação

Por exemplo, com base na seguinte regressão auxiliar,

$$media = \beta_0 + \beta_1 esmedia + \beta_2 faltas + \delta\, \widehat{media}^2 + \varepsilon$$

Tem-se: H'_0: $\delta = 0$ contra H'_1: $\delta \neq 0$.

Sob H_0, $T = \frac{\hat{\delta}}{se(\hat{\delta})} \sim t(137)$. Valor-p $= 2 \times P(T > |t_{obs}| | H_0) = 0.49$. Como valor-p é superior aos níveis de significância habituais, não se rejeita H_0: não há evidência de má especificação da forma funcional do modelo da equação 2.

Equação 1

Dependent Variable: MEDIA
Included observations: 141

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.789119	1.349656	4.289329	0.0000
ESMEDIA	0.459202	0.088072	5.213945	0.0000
EDUCPAI	0.030556	0.042054	0.726585	0.4687
EDUCMAE	0.004612	0.030764	0.149924	0.8810
FALTAS	-0.024869	0.008730	-2.848802	0.0051

R-squared	0.227451	S.E. of regression	1.328075
Adjusted R-squared	0.204729	S.D. dependent var	1.489241
Sum squared resid	239.8745	F-statistic	10.01017

Equação 2

Dependent Variable: MEDIA
Included observations: 141

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.316668	1.213468	5.205468	0.0000
ESMEDIA	0.458804	0.087691	5.232063	0.0000
FALTAS	-0.025812	0.008588	-3.005480	0.0032

R-squared	0.222733	S.E. of regression	1.322436
Adjusted R-squared	0.211468	S.D. dependent var	1.489241
Sum squared resid	241.3395	F-statistic	19.77258