Estatística II

Licenciatura em Gestão

1º semestre – 2015/2016

ER - 03/02/2016

Nome____ **N^o**__________

Espaço reservado a classificações

A utilização do telemóvel, em qualquer circunstância, é motivo suficiente para a anulação da prova.

Perguntas de escolha múltipla: apenas uma opção é correta
cada resposta certa vale **1** valor
cada resposta errada vale **-0.25** valores

Pergunta Verdadeiro/Falso: cada resposta certa vale **0.25** valores
cada resposta errada vale **-0.25** valores

Se necessitar de espaço utilize a última página do enunciado, indicando com clareza a respetiva questão.

É expressamente proíbido destacar as folhas do enunciado.

Parte I

1. Para cada afirmação assinale se esta é verdadeira **(V)** ou falsa **(F)**. (1.0)

	V	**F**
Dado um parâmetro desconhecido θ, um estimador diz-se centrado para θ se, em média, o seu valor for igual ao valor de θ.	X	
Uma estimativa é uma variável aleatória.		X
O nível de significância de um teste estatístico, α, é a probabilidade de cometer o erro de 2^a espécie		X
Num teste estatístico em que o valor-$p = 0$, rejeita-se H_0, qualquer que seja a dimensão α.	X	

2. O número de chamadas telefónicas por minuto no *call center* do atendimento ao cliente de uma determinada empresa segue uma distribuição de Poisson com valor esperado λ. Tendo por base a contabilização do número de chamadas em 60 intervalos de um minuto, selecionados ao acaso, foram contabilizadas 69 chamadas.

(a) Sabendo que $\widehat{\lambda} = \bar{X}$ é estimador de máxima verosimilhança de λ, qual a estimativa de máxima verosimilhança para a probabilidade de não haver chamadas num minuto. Justifique. (1.5)

X=número de chamadas telefónicas por minuto

$X \sim Poisson(\lambda)$, amostra: $n = 60$ e $\sum_{i=1}^{60} x_i = 69$, $\bar{x} = \frac{69}{60} = 1.15$

$\hat{\lambda} = \bar{X}$ é estimador de máxima verosimilhança de λ, logo $\bar{x}$ é estimativa de máxima verosimilhança de λ:

$$\widehat{P(X=0)} = \widehat{\frac{e^{\lambda}\lambda^0}{0!}} \underbrace{=}_{\text{prop. invariância}} e^{-\bar{x}} = e^{-1.15} = 0.3166$$

A estimativa de máxima verosimilhança de $P(X = 0)$ é 0.3166.

(b) Determine o intervalo de confiança aproximado a 95% para o número médio de chamadas por minuto. (1.5)

Variável fulcral: $Z = \dfrac{\bar{X}-\lambda}{\sqrt{\frac{\bar{X}}{n}}} \overset{a}{\sim} N(0,1)$

Logo, o intervalo de confiança aproximado a 95%para o número médio de chamadas é:

$$\bar{x} \pm z_{\alpha/2}\sqrt{\frac{\bar{x}}{n}} = 1.15 \pm 1.96\sqrt{\frac{1.15}{60}} = (1.15 \pm 0.2713) = (0.8787, 1.4213)$$

(c) Os funcionários do *call center* afirmam que o número médio de chamadas por minuto é 1.5. Com base na alínea anterior, o que pode concluir sobre essa afirmação? (se não resolveu a alínea anterior, considere que o intervalo de confiança a 95% para o número médio de chamadas por minuto é $(0.9, 1.4)$) (1.0)

- ☐ A hipótese é suportada pelos dados ao nível de significância de 10%
- ☒ A hipótese não é suportada pelos dados ao nível de significância de 5%
- ☐ A hipótese é suportada pelos dados ao nível de significância de 5%
- ☐ Não se pode concluir sobre a afirmação, com base na informação dada

(d) No sentido de otimizar os recursos da empresa, um gestor afirma que o número médio de chamadas por minuto não é superior a 1. É aceitável a suposição do gestor? Decida com base no valor-p. (1.5)

$H_0: \lambda \leqslant 1$ vs $H_1: \lambda > 1$

Grandes amostras ($n > 30$), VF: $Z = \dfrac{\bar{X}-\lambda}{\sqrt{\frac{\lambda}{n}}} \overset{a}{\sim} N(0,1)$

Estatística teste: $Z|H_0 = \dfrac{\bar{X}-1}{\sqrt{\frac{1}{n}}} \overset{a}{\sim} N(0,1)$

Valor observado da estatística teste: $z_{obs} = \dfrac{\bar{x}-1}{\sqrt{\frac{1}{60}}} = \dfrac{1.15-1}{0.1291} = \dfrac{0.15}{0.1291} = 1.162$

valor-p= $P(Z \geqslant 1.162|H_0) = 1 - \Phi(1.162) = 0.123$

Como valor-p= 12.3% é superior aos níveis usuais de significância (nomeadamente 12.3% > 5%), não se rejeita H_0. Há evidência estatística que suporta a suposição do gestor.

(e) O mesmo gestor afirma que a variância do número de chamadas por minuto não é superior a 1. É aceitável a suposição do gestor, ao nível de significância $\alpha = 0.05$? (1.0)

Como $X \sim Poisson(\lambda)$, então $V(X) = E(X) = \lambda$.
Logo testar $H_0 : \sigma^2 \leqslant 1$ vs $H_1 : \sigma^2 > 1$ é equivalente a testar $H_0 : \lambda \leqslant 1$ vs $H_1 : \lambda > 1$.
Logo, pela alínea anterior, não se rejeita H_0. A evidência estatística aponta para aceitar a suposição do gestor.

3. Num teste de hipóteses em que a região crítica é $W = \{t : t > 15.5\}$, para uma determinada estatística teste T, então (1.0)

- ☐ O valor$-p$ do teste é $p = 2P(T \geqslant 15.5|H_0)$
- ☐ A dimensão do teste é $\alpha = 2P(T \geqslant 15.5|H_0)$
- ☒ A dimensão do teste é $\alpha = P(T \geqslant 15.5|H_0)$
- ☐ O valor$-p$ do teste é $p = P(T \geqslant 15.5|H_0)$

4. A administração de uma fábrica de baterias decidiu estudar o tempo de autonomia das mesmas. Assumiu-se para tal que o tempo de autonomia, em horas, de cada bateria é uma variável aleatória X com função densidade de probabilidade dada por: (1.5)

$$f_X(x) = \frac{x^2}{2\beta^3} e^{-\frac{x}{\beta}}, \quad x \geqslant 0$$

onde $\beta > 0$ é desconhecido. Determine o estimador de máxima verosimilhança para β (considere apenas a condição de 1º ordem).

X= tempo de autonomia de uma bateria. Seja $X_1, \ldots, X_n$ amostra casual.
Função de verosimilhança:

$$L(\beta) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \frac{x_i^2}{2\beta^3} e^{-\frac{x_i}{\beta}} = \frac{1}{2^n \beta^{3n}} \prod_{i=1}^{n} x_i^2 e^{-\frac{1}{\beta}\sum_{i=1}^{n} x_i}$$

Função log-verosimilhança:

$$l(\beta) = \ln L(\beta) = -n\ln(2) - 3n\ln(\beta) + \ln\left(\prod_{i=1}^{n} x_i^2\right) - \frac{1}{\beta}\sum_{i=1}^{n} x_i$$

Estimativa de máxima verosimilhança para β é $\hat{\beta}$ tal que $\max l(\beta) = l(\hat{\beta})$.
Condição de 1^a ordem:

$$\frac{dl(\beta)}{d\beta} = 0 \Leftrightarrow \frac{-3n}{\beta} + \frac{\sum_{i=1}^{n} x_i}{\beta^2} = 0 \Leftrightarrow \frac{-3n\beta + \sum_{i=1}^{n} x_i}{\beta^2} = 0 \Leftrightarrow -3n\beta + \sum_{i=1}^{n} x_i \Leftrightarrow \beta = \frac{\sum_{i=1}^{n} x_i}{3n} = \frac{\bar{x}}{3}$$

Logo, o estimador de máxima verosimilhança é

$$\hat{\beta} = \frac{\hat{X}}{3}$$

Parte II

1. Para cada afirmação assinale se esta é verdadeira **(V)** ou falsa **(F)**. (1.0)

	V	**F**
No modelo de regressão linear $y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$, $t = 1, 2, \ldots, n$, em que o coeficiente de determinação, R^2, é elevado, pode concluir-se que x_2 tem significância estatística		X
O modelo $y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{2t}^2 + u_t$, $t = 1, 2, \ldots, n$, é linear nos parâmetros.	X	
Nos testes não paramétricos estudados, a região de rejeição situa-se sempre na aba direita da distribuição.	X	
Uma das hipóteses do MRL é a exogeneidade dos regressores.	X	

2. Para um determinado cruzamento rodoviário, registaram-se o número mensal de acidentes lá ocorridos ao longo dos últimos 10 anos (120 meses). Os resultados encontram-se na tabela de frequências abaixo. (2.0)

número de acidentes	0	1	2	3 ou mais
frequência observada	81	27	10	2

Será que os dados suportam a hipótese de o número mensal de acidentes nesse cruzamento seguir uma distribuição de Poisson de parâmetro 0.5, ao nível de significância de 5%?

X=número mensal de acidentes

$H_0 : X \sim Poisson(0.5)$ vs $H_1 : X \nsim Poisson(0.5)$

A frequência esperada na última classe ($X \geqslant 3$) é:

$$np_j = 120P(X \geqslant 3|H_0) = 120\left[1 - \left(e^{-0.5} + 0.5e^{-0.5} + \frac{0.5^2 e^{-0.5}}{2}\right)\right] = 1.7265 < 5$$

logo é necessário agrupar as duas últimas classes.

$$p_1 = P(X = 0|H_0) = e^{-0.5} = 0.6065, \quad p_2 = P(X = 1|H_0) = 0.5e^{-0.5} = 0.3033, \quad p_3 = P(X \geqslant 2|H_0) = 1 - (p_1 + p_2) = 0.0902$$

número de acidentes classes	frequência observada n_j	p_j	frequência esperada np_j	q_j
0	81	0.6065	72.78	0.9284
1	27	0.3033	36.396	2.4257
$\geqslant 2$	12	0.0902	10.824	0.1278
totais	120	1	120	$q_{obs} = 3.4819$

Estatística teste (sob H_0), $k = 3$: $Q = \sum_{i=1}^{3} \frac{(N_j - np_j)^2}{p_j} \overset{a}{\sim} \chi^2_{(3-1)}$ e região crítica: $W_{5\%} = \{q : q > 5.991\}$.

Como $q_{obs} = 3.4819 < 5.991$ ($q_{obs} \notin W$), não se rejeita H_0 a 5%. Há evidência estatística de que o número mensal de acidentes tem distribuição de Poisson de média 0.5.

3. Para estudar os determinantes do crédito concedido ($cred$) para a aquisição de habitação, um investigador analisou os processos de clientes de uma instituição financeira, tendo especificado o seguinte modelo:

$$lcred_t = \beta_1 + \beta_2 lrend_t + \beta_3 lpreco_t + \beta_4 peso_t + \beta_5 senior_t + u_t$$

onde $rend$ é o rendimento anual do cliente , $preco$ é o preço de compra da habitação, $peso$ é a percentagem das outras obrigações do cliente no seu rendimento (medido de 0 a 100), e $senior$ é uma variável artificial que assume o valor 1 se o cliente tem mais de 50 anos. Os resultados obtidos encontram-se em anexo, onde o prefixo l representa o logaritmo natural.

(a) Interprete as estimativas de β_2, β_4 e β_5 da **Equação 1** (1.5)

$b_2 = 0.2308$: estimativa da elasticidade do crédito relativamente ao rendimento. Estima-se que, se o rendimento aumentar 1%, o crédito concedido aumenta, em média, aproximadamente 0.2308%, todas as outras variáveis se mantendo constantes.

$b_4 = 0.0029 \approx 0.3\%$: Estima-se que, se o peso das outras obrigações no rendimento aumentar 1 ponto percentual, o crédito concedido aumenta, em média 0.3%, para valores constantes das restantes variáveis.

$b_5 = -0.0286 = -2.86\%$ Estima-se que, para valores constantes das restantes variáveis, o crédito concedido é inferior, em média, em cerca de 2.86% para clientes com idades superior a 50 anos, face a clientes com idade inferior a 50 anos.

(b) Teste a significância conjunta de β_4 e de β_5 a 5%. (1.5)

$H_0 : \beta_4 = \beta_5 = 0$ vs $H_1 : \beta_4 \neq 0$ ou $\beta_5 \neq 0$

Estatística teste (sob validade de H_0): $F = \dfrac{(VR_0 - VR_1)/2}{VR_1/(n-k)} \sim F_{(2,232)}$ e região crítica: $W = \{f : f > 3.0\}$

$$f_{obs} = \frac{(28.3325 - 27.9519)/2}{27.9519/232} = \frac{0.1903}{0.1204823} = 1.579485$$

Como $f_{obs} = 1.579485 < 3.0$ ($f_{obs} \notin W$), não se rejeita H_0 a 5%. Os dados estatísticos suportam a hipótese de nulidade conjunta de β_4 e β_5.

(c) Pretende-se testar $H_0 : \beta_3 = 1$ contra $H_1 : \beta_3 > 1$. Com pase no *output* da **Equação 1**, pode concluir-se, a 5%: (1.0)

- ☐ que se rejeita H_0 porque o valor de b_3 é diferente de 1.
- ☐ que se rejeita H_0 porque o valor da estatística de teste é 13.4009 e pertence à região de rejeição.
- ☒ que não se rejeita H_0 porque o valor da estatística de teste é −5.94 e não pertence à região de rejeição.
- ☐ que a informação dada não permite calcular o valor da estatística de teste e tomar uma decisão.

(d) Para testar, na **Equação 1**, se a elasticidade do crédito relativamente ao preço é o triplo da elasticidade do crédito relativamente ao rendimento, foi estimado outro modelo. Qual? (1.0)

- ☐ $lcred_t = \alpha_1 + \alpha_2 lrend_t + \alpha_3 lpreco_t + erro_t$
- ☒ $lcred_t = \delta_1 + \delta_2(lrend_t + 3 \times lpreco_t) + \delta_3 peso_t + \delta_4 senior_t + erro_t$
- ☐ $lcred_t = \gamma_1 + \gamma_2(3 \times lpreco_t) + \gamma_3 peso_t + \gamma_4 senior_t + erro_t$
- ☐ $lcred_t = \theta_1 + \theta_2 peso_t + \theta_3 senior_t + erro_t$

(e) Utilizando o modelo da **Equação 2**, pretende-se testar se há evidência de regressões idênticas para dois tipos de clientes: clientes com outras aplicações financeiras além dos depósitos e clientes com apenas depósitos nessa instituição financeira. Dos 237 clientes da amostra, 160 clientes só têm depósitos. Indique como procederia para realizar esse teste. Apresente as regressões necessárias, as hipóteses H_0 e H_1 e a estatística de teste utilizada. Sabendo que o valor da estatística de teste é igual a 3.62, o que pode concluir, ao nível de significância de 5%? (2.0)

Trata-se de um teste de Chow aplicado à equação $lcred_t = \beta_1 + \beta_2 lrend_t + \beta_3 lpreco_t + u_t$. É necessário estimar 3 regressões:

i) a **Equação 2** com as 237 observações

ii) o modelo só para clientes com outras aplicações: $n = 77$ e $lcred_t = \alpha_1 + \alpha_2 lrend_t + \alpha_3 lpreco_t + erro_t$

iii) o modelo só para clientes com depósitos: $n = 160$ e $lcred_t = \theta_1 + \theta_2 lrend_t + \theta_3 lpreco_t + erro_t$

Hipóteses do teste estatístico: $H_0 : \alpha_j = \theta_j,\ j = 1,2,3$ vs $H_1 : \exists \alpha_i \neq \theta_j,\ j = 1,2,3$

Estatística teste (sob validade de H_0): $F = \dfrac{[VR - (VR_1)]/3}{(VR_1)/(237-6)} \sim F_{(3,231)}$ com:

VR_1**:** soma das variações residuais dos modelis ii) e iii)

VR**:** variação residual do modelo i)

Região crítica: $W_{5\%} = \{f : f > 2.6\}$.

Valor observado da estatística: $f_{obs} = 3.62 > 2.6 \Longrightarrow f_{obs} \in W$.
Logo, rejeita-se H_0 a 5%. Há evidência estatística contra a hipótese de regressões idênticas.

Anexo

Equação 1: $lcred_t = \beta_1 + \beta_2 lrend_t + \beta_3 lpreco_t + \beta_4 peso_t + \beta_5 senior_t + u_t$

Regression Statistics	
Multiple R	0.8103
R Square	0.6566
Adjusted R Square	0.6507
Standard Error	0.3471
Observations	237

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	53.4502	13.3625	110.9087	0.0000
Residual	232	27.9519	0.1205		
Total	236	81.4021			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.8235	0.5207	1.5816	0.1151
lrend	0.2308	0.0474	4.8722	0.0000
lpreco	0.6929	0.0517	13.4009	0.0000
peso	0.0029	0.0017	1.7052	0.0895
senior	-0.0286	0.0456	-0.6269	0.5314

Equação 2: $lcred_t = \beta_1 + \beta_2 lrend_t + \beta_3 lpreco_t + u_t$

Regression Statistics	
Multiple R	0.8074
R Square	0.6519
Adjusted R Square	0.6490
Standard Error	0.3480
Observations	237

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	53.0696	26.5348	219.1524	0.0000
Residual	234	28.3325	0.1211		
Total	236	81.4021			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.8698	0.5201	1.6724	0.0958
lrend	0.2360	0.0474	4.9839	0.0000
lpreco	0.6835	0.0515	13.2782	0.0000