LISBOA SCHOOL OF ECONOMICS & MANAGEMENT

Estatística II

Licenciatura em Gestão

1º Semestre – 2018/2019
Época de Recurso
04/02/2019

Nome: ______________________________ Nº______________

Espaço reservado a classificações						
1.a)	1.b)	1.c)	1.d)	2.	3.a)	3.b)
4.a)	4.b)	4.c)	4.d)	4.e)	4.f)	4.g)
				TOTAL:		

Se necessitar de mais espaço, pode utilizar a última página do enunciado, indicando claramente a respectiva questão.
É expressamente proibido destacar as folhas do enunciado!

Boa Sorte!

1. Uma *start-up* de refrigerantes em início de actividade pretende abrir uma loja de retalho em Portugal. Devido a restrições de capital, apenas pode abrir uma loja, estando indecisa entre Lisboa e Porto. Para tal, decidiu basear a sua decisão na proporção de consumidores de refrigerantes nas duas cidades, reunindo uma amostra casual de 150 indivíduos em Lisboa e 100 indivíduos no Porto, e na qual observou 90 consumidores em Lisboa e 40 no Porto.

   a) Teste ao nível de 5% se a proporção de consumidores de refrigerantes em Lisboa é igual a 0.5 contra a alternativa de ser superior a 0.5. O que pode concluir? *[2.0]*

$$H_0: \theta_1 = 0.5 \quad \text{vs} \quad H_1: \theta_1 > 0.5$$

$$\textbf{ET}: \; Z = \frac{\bar{X}_1 - \theta_{10}}{\sqrt{\frac{\theta_{10}(1-\theta_{10})}{n}}} \overset{a}{\sim} N(0,1)$$

$$W_{0.05} = \{z_{obs} : z_{obs} > 1.645\}$$

$$z_{obs} = \frac{0.6 - 0.5}{\sqrt{\frac{0.5(1-0.5)}{150}}} \approx 2.4495$$

Porque $z_{obs} \in W_{0.05}$, rejeita-se H_0 ao nível de 5%. A evidência estatística é favorável à hipótese de a proporção de consumidores de refrigerantes em Lisboa ser superior a 0.5.

   b) Do teste anterior sabe-se que…
   [Resposta certa: 1 / Resposta errada: -0.25]

… o valor-p é igual a 0.95.	
… a potência não depende da proporção de consumidores sob H_1.	
… a probabilidade de rejeitar H_0 quando H_0 é verdadeira é igual a 0.025.	
… a probabilidade de não rejeitar H_0 quando H_0 é verdadeira é igual a 0.95.	**X**

c) Com base num intervalo de confiança a 95% verifique em qual das duas cidades deve a *start-up* abrir a sua loja. *[2.0]*

IC a 95% para $\theta_1 - \theta_2$ (Grandes Amostras – Populações de Bernoulli)

$$\textbf{VF}: \ Z = \frac{\bar{X}_1 - \bar{X}_2 - (\theta_1 - \theta_2)}{\sqrt{\frac{\bar{X}_1(1-\bar{X}_1)}{m} + \frac{\bar{X}_2(1-\bar{X}_2)}{n}}} \overset{a}{\sim} N(0,1)$$

$$\textbf{IC}: \left((\bar{x}_1 - \bar{x}_2) \pm z_{0.025} \times \sqrt{\frac{\bar{x}_1(1-\bar{x}_1)}{m} + \frac{\bar{x}_2(1-\bar{x}_2)}{n}} \right) =$$

$$= \left((0.6 - 0.4) \pm 1.96 \times \sqrt{\frac{0.6(1-0.6)}{150} + \frac{0.4(1-0.4)}{100}} \right) = (0.076, 0.324)$$

Com uma confiança de 95% pode afirmar-se que a diferença das proporções de consumidores de refrigerantes entre Lisboa e Porto se situa entre 0.076 e 0.324. Porque o intervalo de confiança obtido engloba valores estritamente positivos, pode concluir-se que a proporção em causa é superior em Lisboa. Assim, de acordo com este resultado, a *start-up* deve abrir a sua loja na cidade de Lisboa.

d) Se se quisesse aumentar o grau de confiança do intervalo anterior, o que aconteceria à respectiva amplitude?
[Resposta certa: 1 / Resposta errada: -0.25]

Diminuiria.	
Aumentaria.	**X**
Manter-se-ia inalterada.	

2. Verdadeiro ou Falso? *[Cada resposta certa: 0.25 / Cada resposta errada: -0.25]*

	V	F
A propriedade da invariância é exclusiva aos estimadores da máxima verosimilhança.	**X**	
Sejam T_1 e T_2 dois estimadores para um parâmetro desconhecido θ, onde T_1 é enviesado e T_2 é centrado. Então, T_1 é um melhor estimador que T_2 se: $Var(T_1) \leq Var(T_2), \forall\theta \in \Theta$		**X**
O teste de hipóteses mais potente é aquele que para uma determinada P(rejeitar $H_0 \| H_0$ verdadeira) se minimiza P(não rejeitar $H_0 \| H_0$ falsa).	**X**	
Se (t_1, t_2) é um intervalo de confiança a $(1-\alpha)\cdot 100\%$ para um parâmetro desconhecido θ (onde $0 < \alpha < 1$) então: $P[\theta \in (t_1, t_2)] = 1 - \alpha$.		**X**

3. Seja X uma variável aleatória discreta, onde:

$$E(X^k) = \begin{cases} 0, \text{se } k \text{ é ímpar} \\ \theta, \text{se } k \text{ é par, onde } 0 < \theta < 1 \end{cases}$$

Admita ainda que se retirou uma amostra casual de dimensão n desta população.

a) Mostre que os estimadores para θ obtidos pelo método dos momentos apenas existem sob determinada condição e são da forma: $\left(\sum_{i=1}^{n} X_i^k\right)/n$. *[1.0]*

Com a informação dada sabe-se que os momentos de ordem k ímpar são iguais a zero, pelo que não é possível aplicar o método dos momentos neste caso. Mas quando consideramos os momentos de ordem k par, sabe-se que $E(X^k) = \theta$, o que nos permite então obter directamente os estimadores pelo método dos momentos para θ:

$$E(X^k) = \frac{\sum_{i=1}^{n} X_i^k}{n} \Leftrightarrow \tilde{\theta} = \frac{\sum_{i=1}^{n} X_i^k}{n}$$

b) Sabe-se também que a distribuição de X^k é tal que:

$$X^k \sim \begin{cases} X, \text{se } k \text{ é ímpar} \\ B(1,\theta), \text{se } k \text{ é par} \end{cases}$$

Tendo em conta esta informação, avalie a consistência dos estimadores do método dos momentos da alínea anterior. *[2.0]*

Considerando k par, sabe-se que: $X^k \sim B(1,\theta)$, pelo que $E(X^k) = \theta$ e $Var(X^k) = \theta(1-\theta)$.

Assim,

$$\lim_{n\to\infty} E(\tilde{\theta}) = \lim_{n\to\infty} E\left(\frac{\sum_{i=1}^{n} X_i^k}{n}\right) = \lim_{n\to\infty}\left[\frac{1}{n}\sum_{i=1}^{n} E(X_i^k)\right] = \lim_{n\to\infty}\left[\frac{1}{n}\sum_{i=1}^{n} E(X^k)\right] = \lim_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n}\theta\right) = \lim_{n\to\infty}\left(\frac{1}{n}\cdot n\theta\right)$$

$$= \lim_{n\to\infty}\theta = \theta$$

$$\lim_{n\to\infty} Var(\tilde{\theta}) = \lim_{n\to\infty} Var\left(\frac{\sum_{i=1}^{n} X_i^k}{n}\right) = \lim_{n\to\infty}\left[\left(\frac{1}{n}\right)^2\sum_{i=1}^{n} Var(X_i^k)\right] = \lim_{n\to\infty}\left[\frac{1}{n^2}\sum_{i=1}^{n} Var(X^k)\right]$$

$$= \lim_{n\to\infty}\left[\frac{1}{n^2}\sum_{i=1}^{n}\theta(1-\theta)\right] = \lim_{n\to\infty}\left[\frac{1}{n^2}\cdot n\theta(1-\theta)\right] = \lim_{n\to\infty}\left[\frac{\theta(1-\theta)}{n}\right] = 0$$

Porque $\lim_{n\to\infty} E(\tilde{\theta}) = \theta$ e $\lim_{n\to\infty} Var(\tilde{\theta}) = 0$, conclui-se que os estimadores do método dos momentos para θ são consistentes.

4. Um analista do mercado imobiliário pretende estudar o preço de arrendamento dos escritórios de certa região litoral com base no seguinte modelo de regressão linear

$$\ln(PRECO) = \beta_1 + \beta_2 \ln(AREA) + \beta_3 SALAS + \beta_4 WC + \beta_5 IDADE + u$$

Onde:

- $\ln(PRECO)$ – logaritmo do preço de arrendamento do escritório (em que o preço está em euros);
- $\ln(AREA)$ – logaritmo da área do escritório (em que a área está em m^2);
- $SALAS$ – número de salas do escritório;
- WC – número de casas de banho do escritório;
- $IDADE$ – idade do escritório, em anos.

Recolhendo uma amostra de 130 escritórios, o analista estimou o modelo anterior, cujos resultados se encontram na **Equação 1**, em anexo.

a) Interprete as estimativas dos coeficientes associados com os regressores $\mathbf{ln}(\boldsymbol{AREA})$ e $\boldsymbol{IDADE}$ na **Equação 1**. *[1.5]*

$b_2 = 1.0541 \rightarrow$ *Ceteris paribus*, estima-se em média, que se a área do escritório aumentar em 1%, o preço de arrendamento do escritório aumenta aproximadamente 1.0541%. Estimativa MQ da elasticidade do preço de arrendamento em relação à área.

$b_5 = -0.0088 \rightarrow$ *Ceteris paribus*, estima-se em média, que se o escritório for 1 ano mais velho, o preço de arrendamento do escritório diminui aproximadamente 0.88%. Estimativa MQ da semi-elasticidade do preço de arrendamento em relação à idade.

b) Teste ao nível de 5% a significância individual dos regressores $\boldsymbol{SALAS}$ e $\boldsymbol{WC}$ na **Equação 1**. *[1.5]*

Significância individual de *SALAS:*

$$H_0: \beta_3 = 0 \quad \text{vs} \quad H_1: \beta_3 \neq 0$$

$$\textbf{ET}: \ t_3 = \frac{b_3}{s_{b_3}} \sim t(125)$$

Do output obtém-se $p_{3,obs} = 0.0304$. Porque $p_{3,obs} < \alpha = 0.05$, rejeita-se H_0 ao nível de 5%. Logo, o regressor *SALAS* é estatisticamente significativo.

Significância individual de *WC:*

$$H_0: \beta_4 = 0 \quad \text{vs} \quad H_1: \beta_4 \neq 0$$

$$\textbf{ET}: \ t_4 = \frac{b_4}{s_{b_4}} \sim t(125)$$

Do output obtém-se $p_{4,obs} = 0.1068$. Porque $p_{4,obs} > \alpha = 0.05$, não se rejeita H_0 ao nível de 5%. Logo, o regressor *WC* não é estatisticamente significativo.

c) Teste a significância global da **Equação 1** ao nível de 1%. O que pode concluir? *[1.5]*

$$H_0: \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0 \quad \text{vs} \quad H_1: \exists \beta_j \neq 0 \ (j = 2, 3, 4, 5)$$

$$\textbf{ET}: \ F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} \sim F(4, 125)$$

Do output, obtém-se: $p_{obs} = 0$. Porque $p_{obs} < \alpha = 0.01$, rejeita-se H_0 ao nível de 1%. Logo, conclui-se que a Equação 1 é globalmente significativa.

d) De maneira a testar a significância conjunta dos regressores ***SALAS*** e ***WC***, estimou-se o modelo restrito, tendo-se obtido $R^2 = 0.6017$. Indique o modelo restrito em questão e efectue o respectivo teste com um nível de significância de 1%. O que pode concluir? *[1.5]*

$$H_0: \beta_3 = \beta_4 = 0 \quad \text{vs} \quad H_1: \exists \beta_j \neq 0 \ (j = 3, 4)$$

$$\textbf{ET}: \ F = \frac{(R^2 - R_0^2)/m}{(1-R^2)/(n-k)} \sim F(2, 125)$$

Onde R_0^2 é o coeficiente de determinação do modelo restrito:

$$\ln(PRECO) = \beta_1 + \beta_2 \ln(AREA) + \beta_3 IDADE + u$$

Assim,

$$f_{obs} = \frac{(0.6321 - 0.6017)/2}{(1 - 0.6321)/125} \approx 5.164$$

$$W_{0.01} = \{f_{obs}: f_{obs} > 4.61\}$$

Porque $f_{obs} \in W_{0.01}$, rejeita-se H_0 ao nível de 1%. Logo, conclui-se que os regressores $SALAS$ e WC são conjuntamente significativos.

e) O analista pretende descobrir se, em média, o efeito *ceteris paribus* sobre o preço de arrendamento de uma sala adicional é idêntico ao efeito de uma casa de banho adicional. Sabendo que $\widehat{Cov}(b_3, b_4) = -0.000105$ efectue um teste a 1% que lhe permita responder à questão do analista. O que pode concluir? *[1.5]*

$$H_0: \beta_3 = \beta_4 \;\; \text{vs} \;\; H_1: \beta_3 \neq \beta_4 \;\Leftrightarrow\; H_0: \beta_3 - \beta_4 = 0 \;\; \text{vs} \;\; H_1: \beta_3 - \beta_4 \neq 0$$

Defina-se um novo parâmetro: $\delta = \beta_3 - \beta_4$. Então, tem-se de forma equivalente:

$$H_0: \delta = 0 \;\; \text{vs} \;\; H_1: \delta \neq 0$$

$$\textbf{ET}: \; t_{\hat{\delta}} = \frac{\hat{\delta}}{s_{\hat{\delta}}} \sim t(125)$$

$$W_{0.01} = \left\{ t_{\hat{\delta},obs}: \; t_{\hat{\delta},obs} < -2.576 \vee t_{\hat{\delta},obs} > 2.576 \right\}$$

Cálculo de $t_{\hat{\delta},obs}$:

- $\hat{\delta} = b_3 - b_4 = 0.0366 - 0.0357 = 0.0009$
- $s_{\hat{\delta}} = \sqrt{\widehat{Var}(\hat{\delta})} = \sqrt{\widehat{Var}(b_3 - b_4)} = \sqrt{\widehat{Var}(b_3) + \widehat{Var}(b_4) - 2\widehat{Cov}(b_3, b_4)} =$

  $\sqrt{s_{b_3}^2 + s_{b_4}^2 - 2\widehat{Cov}(b_3, b_4)} = \sqrt{0.0167^2 + 0.0220^2 - 2 \times (-0.000105)} \approx 0.03119$

Logo,

$$t_{\hat{\delta},obs} = \frac{0.0009}{0.03119} \approx 0.0289$$

Porque $t_{\hat{\delta},obs} \notin W_{0.01}$, não se rejeita $H_0: \beta_3 = \beta_4$ ao nível de 1%. Evidência estatística favorável a que o efeito de uma sala adicional sobre o preço de arrendamento seja idêntico ao efeito de uma casa de banho adicional.

f) O analista deseja também obter um intervalo de previsão a 90% para o preço médio de arrendamento de apartamentos com 30 anos de idade, 250 m², 3 salas e 2 casas de banho. Para tal, estimou a **Equação 2**, em anexo. Construa o intervalo de previsão respectivo e interprete. *[1.5]*

<u>Previsão em média</u>

1) IP a 90% para: $\theta = E(\ln(PRECO)\,|AREA = 250, SALAS = 3, WC = 2, IDADE = 30)$

$$\textbf{VF}:\ t_{\hat{\theta}} = \frac{\hat{\theta} - \theta}{s_{\hat{\theta}}} \sim t(125)$$

IP: $\left(\hat{\theta} \pm t_{0.05} \times s_{\hat{\theta}}\right) = (7.6587 \pm 1.645 \times 0.0146) = (7.6347, 7.6827)$

Onde $\hat{\theta}$ e $s_{\hat{\theta}}$ obtêm-se da Equação 2, linha "Intercept" e colunas "Coefficients" e "Standard Error", respectivamente.

2) IP a 90% para $E(PRECO|AREA = 250, SALAS = 3, WC = 2, IDADE = 30)$

$$(e^{7.6347}, e^{7.6827}) = (2068.75, 2170.47)$$

Logo, com uma confiança de 90% prevê-se que o preço de arrendamento médio para escritórios com estas características se situa entre 2068.75 euros e 2170.47 euros.

g) Suponha agora que o analista incluía na regressão uma variável binária para "escritórios com vista para o mar", e outra variável binária para "escritórios sem vista para o mar". Nesta situação, qual das hipóteses básicas do modelo de regressão linear está o analista a violar?
[Resposta certa: 1 / Resposta errada: -0.25]

Exogeneidade	
Homocedasticidade condicionada	
Ausência de autocorrelação	
Ausência de multicolinearidade exacta	**X**

ANEXO

Equação 1: $\ln(PRECO) = \beta_1 + \beta_2 \ln(AREA) + \beta_3 SALAS + \beta_4 WC + \beta_5 IDADE + u$

Regression Statistics	
Multiple R	0.7950
R Square	0.6321
Adjusted R Square	0.6203
Standard Error	0.1343
Observations	130

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	3.8738	0.9685	53.6889	0.0000
Residual	125	2.2548	0.0180		
Total	129	6.1286			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	1.9197	0.8294	2.3144	0.0223
ln(AREA)	1.0541	0.1587	6.6406	0.0000
SALAS	0.0366	0.0167	2.1896	0.0304
WC	0.0357	0.0220	1.6246	0.1068
IDADE	-0.0088	0.0010	-8.6293	0.0000

Equação 2: $\ln(PRECO) = \theta + \beta_2[\ln(AREA) - \ln(250)] + \beta_3(SALAS - 3) + \beta_4(WC - 2) + \beta_5(IDADE - 30) + u$

Regression Statistics	
Multiple R	0.7950
R Square	0.6321
Adjusted R Square	0.6203
Standard Error	0.1343
Observations	130

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	3.8738	0.9685	53.6889	0.0000
Residual	125	2.2548	0.0180		
Total	129	6.1286			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	7.6587	0.0146	523.8415	0.0000
ln(AREA)-ln(250)	1.0541	0.1587	6.6406	0.0000
SALAS-3	0.0366	0.0167	2.1896	0.0304
WC-2	0.0357	0.0220	1.6246	0.1068
IDADE-30	-0.0088	0.0010	-8.6293	0.0000