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]*

   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.	

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]*

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.	
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.		
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: $$Var(T_1) \leq Var(T_2), \forall\theta \in \Theta$$		
O teste de hipóteses mais potente é aquele que para uma determinada $P(\text{rejeitar } H_0 \mid H_0 \text{ verdadeira})$ se minimiza $P(\text{não rejeitar } H_0 \mid H_0 \text{ falsa})$.		
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: $P[\theta \in (t_1, t_2)] = 1 - \alpha$.		

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]*

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]*

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

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

d) De maneira a testar a significância conjunta dos regressores $\boldsymbol{SALAS}$ e $\boldsymbol{WC}$, 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]*

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]*

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 m^2, 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]*

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	

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