INSTITUTO SUPERIOR DE ECONOMIA E GESTÃO

ESTATÍSTICA II – LICENCIATURA EM GESTÃO

Época de Recurso – 1 de Fevereiro de 2013

Parte Prática

Nome: ____ **Nº**________

Espaço reservado para classificações				
1. (20)	3. (15)	5a. (15)	5c. (20)	**T:**
2. (15)	4. (20)	5b. (15)	5d. (20)	**P:** _____

Em todos os testes de hipóteses que fizer, formule as hipóteses em teste, indique a estatística de teste e a sua distribuição. Para os intervalos de confiança proceda de forma semelhante para a variável fulcral. Se necessitar de espaço dispõe de uma página em branco no fim do enunciado

1. De uma população com distribuição de Poisson foi observada uma amostra de dimensão 5 que apresentou os seguintes valores: 3, 5, 2, 9, 1. Atendendo a que função probabilidade da distribuição de Poisson é $f(x \mid \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, x = 0, 1, 2, \dots$, obtenha uma estimativa para o parâmetro λ através do método da máxima verosimilhança. Apenas considere a condição de primeira ordem.

$$L(\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} = \frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} e^{-n\lambda}$$

$$l(\lambda) = \log L(\lambda) = \sum_{i=1}^{n} x_i \ \log \lambda - \sum_{i=1}^{n} \log x_i! - n\lambda$$

$$\frac{dl(\lambda)}{d\lambda} = \frac{\sum_{i=1}^{n} x_i}{\lambda} - n$$

$$\frac{dl(\lambda)}{d\lambda} = 0 \iff \hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}$$

Com base na amostra obtem-se a seguinte estimativa para λ: (3 + 5 + 2 + 9 + 1) / 5 = 4

2. Um gestor pretende estudar as diferenças entre o salário dos trabalhadores não qualificados no sector do calçado, X_1, e no sector têxtil, X_2. Para esse fim, inquiriu 120 trabalhadores não qualificados do sector do calçado e 150 do sector têxtil.
Os salários médios dos trabalhadores inquiridos foram:

- Sector do calçado: $\bar{x}_1$ = 965 euros
- Sector têxtil: $\bar{x}_2$ = 890 euros

As variâncias corrigidas dos salários dos trabalhadores inquiridos foram:

- Sector do calçado: s'^2_1 = 8650
- Sector têxtil: s'^2_2 = 8400

Admitindo que os salários pagos nos dois sectores seguem distribuições normais *com variâncias iguais*, construa um intervalo de confiança a 95% para a diferença entre o salário médio dos dois sectores. Com base no resultado, pode o gestor afirmar que existem diferenças estatisticamente significativas entre o salário médio dos dois sectores?

Variável fulcral para diferença das médias de populações normais com variância desconhecida mas iguais.

$$\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{m} + \frac{1}{n}}\sqrt{\frac{(m-1)S'^2_1 + (n-1)S'^2_2}{m+n-2}}} \sim t(m+n-2)$$

s^* = (1/120 + 1/150)$^{1/2}$. { [(120 – 1). 8650 + (150 – 1) . 8400] / (120 + 150 – 2)}$^{1/2}$ = 11.2989
graus de liberdade = 120 + 150 – 2 = 268 $t_{0.025}$ = 1.969
Intervalo de confiança: (965 – 890 – 1.969 . 11.2989, 965 – 890 + 1.969 . 11.2989) = (52.754, 97.246)
O intervalo de confiança não contém o valor zero. Logo, pode dizer-se que há diferenças significativas entre o salário médio dos dois sectores, com um nível de confiança de 95%.

3. A Alice e o Bruno conversavam sobre o número médio de "amigos" dos utilizadores da rede social Facebook. A Alice garantia que tinha lido no *site* do Facebook que o número médio de "amigos" dos utilizadores do Facebook é 245. "Não pode ser", dizia o Bruno, "o número médio de "amigos" é *superior* a 245". Para tirar teimas, seleccionaram aleatoriamente 30 utilizadores do Facebook e registaram o número de amigos de cada um.

- O número médio de "amigos" da amostra foi: $\bar{x}$ = 260
- A variância corrigida do número de "amigos" da amostra foi: s'^2 = 3100

Os resultados da amostra dão razão ao *site* do Facebook? Construa um teste de hipótese adequado para responder a esta questão. Considere uma dimensão de teste de 0.05 e que o número de "amigos" dos utilizadores do Facebook segue uma distribuição normal.

Pretende testar-se: H_0 : μ = 245 contra H_1 : μ > 245.

Estatística de teste:

$$\frac{\bar{X} - \mu_0}{S'/\sqrt{n}} \sim t(n-1)$$

t_{obs} = (260 – 245) / (3100 / 30$^{1/2}$) = 1.476
graus de liberdade = 30 – 1 = 29

$t_{0.05}$ = 1.699 : região de rejeição: $W_T = \{ t : t > t_{0.05} \} = \{ t : t > 1.699\}$
t_{obs} não se encontra na região de rejeição. Logo, não se rejeita a hipótese nula. Os dados da amostra suportam a afirmação da Alice.
Alternativamente, valor-p = $P(T > t_{obs} \mid H_0)$ = 0.0754 > 0.05 : não se rejeita a hipótese nula.

4. O número de aplicações não gratuitas instaladas num telemóvel da marca *iPhone* é bem modelada por uma distribuição de Poisson. Numa amostra aleatória de 200 utilizadores do telemóvel *iPhone* contou-se as seguintes frequências de aplicações não gratuitas:

Número de aplicações não gratuitas:	0	1	2	3	≥4
Número de utilizadores:	48	68	46	26	12

Com uma dimensão de 0.05, teste a hipótese de o número de aplicações não gratuitas num telemóvel *iPhone* seguir uma distribuição de Poisson com média de 1.5.

Seja X o número de aplicações não gratuitas num telemóvel *iPhone*.
A hipótese nula é: H_0: $X \sim Po(1.5)$
A hipótese aparentada é: $H'_0 : p_j = p_{0j}$, j = 1,…,5 contra $H'_1 : p_j \neq p_{0j}$, para algum j.
Estatística de teste:

$$Q = \sum_{j=1}^{m} \frac{\left(N_j - np_{0j}\right)^2}{np_{0j}} \underset{\sim}{a} \chi^2(m-1)$$

x	p(x)	$n\,p_{0j}$	N_j	$(N_j - n\,p_{0j})^2 / n\,p_{0j}$
0	0.223	44.63	48	0.2551
1	0.335	66.94	68	0.0168
2	0.251	50.20	46	0.3521
3	0.126	25.10	26	0.0321
≥4	0.066	13.12	12	0.0961
q_{obs} =	0.7522			

graus de liberdade = m – 1 = 4; α = 0.05; $Q_{0.05}$ = 9.488 Região de rejeição: $W_Q = \{ q : q > 9.488\}$
valor-p = $P(\chi^{(4)} > q_{obs} \mid H_0)$ = 0.945
Não se rejeita a hipótese nula de o número de aplicações não gratuitas num telemóvel *iPhone* seguir uma distribuição de Poisson com média de 1.5.

5. Uma equipa de investigadores decidiu viajar para uma região sub-desenvolvida com o intuito de estudar o potencial empresarial daquela zona e, em particular, os factores determinantes do salário médio praticado pelas empresas. Após alguma reflexão, decidem formular o seguinte modelo:

$$\ln(sal_t) = \beta_1 + \beta_2 antig_t + \beta_3 \ln(prod_t) + \beta_4 est_t + \beta_5 exp_t + u_t.$$

Onde,

- $\ln(sal)$ – logaritmo do salário médio por trabalhador (em unidades monetárias);
- $antig$ – antiguidade (idade) da empresa, em anos;
- $\ln(prod)$ – logaritmo da produtividade média por trabalhador (em unidades monetárias);
- est – variável artificial (=1 se o dono da empresa é estrangeiro / =0 caso contrário);
- exp – variável artificial (=1 se a empresa exporta / =0 caso contrário).

Analisada uma amostra constituída por 168 empresas, o modelo acima é estimado recorrendo ao EXCEL. Os resultados encontram-se no **Modelo 1**, em anexo.

a) Interprete a estimativa do coeficiente do regressor "$antig$" e analise a sua significância estatística com uma dimensão de teste de 0.05. Não se esqueça de apresentar o teste de hipóteses e a estatística de teste em questão.

$b_2 = 0.0126$ → se a idade da empresa aumentar um ano, estima-se que o salário médio aumente cerca de 1.26%, ceteris paribus. Estimativa MQ da semi-elasticidade do salário médio em relação à idade da empresa. (Variação exacta: $100\left(e^{b_2} - 1\right) = 100(e^{0.0126} - 1) \approx 1.268\%$)

$$H_0: \beta_2 = 0 \quad vs \quad H_1: \beta_2 \neq 0,$$

ET: $t_2 = \frac{b_2 - \beta_2^0}{s_{b_2}} \sim t(n-k) = t(163)$

Pelo output,

$t_{2,obs} = 2.3215 \Rightarrow p_{obs} = 2P\left(t_2 \geq |t_{2,obs}| \middle| H_0\right) = 0.0215 < \alpha = 0.05$

Logo, rejeita-se H_0 ao nível de 5%, o que nos leva a concluir que o regressor "$antig$" é estatisticamente significativo.

b) Apresente um intervalo de confiança a 90% para a elasticidade do salário médio em relação à produtividade média.

Quer-se um IC para β_3.

VF: $t_3 = \frac{b_3 - \beta_3}{s_{b_3}} \sim t(n-k) = t(163)$

IC a 90% para β_3: $\left(b_3 \mp t_{\alpha/2} * s_{b_3}\right)$

- $b_3 = 0.7816$
- $1 - \alpha = 0.90 \Rightarrow \alpha = 0.10 \Rightarrow t_{\alpha/2} \approx 1.645$
- $s_{b_3} = 0.0468$

IC a 90% para β_3: $(0.7816 \mp 1.645 \times 0.0468) = (0.705{,}0.859)$

c) Um dos investigadores afirmou que se poderiam retirar ambas as variáveis artificiais do modelo e, como tal, decidiu estimar novamente o modelo sem as introduzir, tendo obtido uma variação residual $VR = 101.682$. Com esta informação, teste a nulidade conjunta dos coeficientes β_4 e β_5 e diga se concorda ou discorda com tal afirmação. Admita uma dimensão do teste de 0.05.

$$H_0: \beta_4 = \beta_5 = 0 \quad vs \quad H_1: \beta_4 \neq 0 \vee \beta_5 \neq 0$$

ET: $F = \frac{(VR_0 - VR_1)/m}{VR_1/(n-k)} \sim F(m, n-k) = F(2, 163)$

$$f_{obs} = \frac{(101.682 - 99.2949)/2}{99.2949/(168-5)} \approx 1.9593$$

$p_{obs} = P(F \geq f_{obs} | H_0) = P(F \geq 1.9593) \approx 0.1443$

Como $p_{obs} \approx 0.1443 > \alpha = 0.05$, não se rejeita H_0 ao nível de 5%. Portanto, não se deve colocar em causa a afirmação do investigador.

Nota: Valor crítico do teste (5%): $F_{0.05} = 3$

d) Com o objectivo de testar a presença de heterocedasticidade da variável residual do Modelo 1, estimou-se a seguinte regressão auxiliar de teste,

$$\hat{u}_t^2 = \alpha_1 + \alpha_2 \ln\widehat{(sal_t)} + \alpha_3 \ln\widehat{(sal_t)}^2 + v_t,$$

cujos resultados se encontram no **Modelo 2**, em anexo. Diga qual o nome do teste em questão e efectue o teste.

<u>Teste de White Simplificado</u>

$$H_0: \alpha_2 = \alpha_3 = 0 \quad vs \quad H_1: \alpha_2 \neq 0 \vee \alpha_3 \neq 0$$

ET: $W_s = nR_a^2 \overset{a}{\rightarrow} \chi^2(2)$

A partir da regressão auxiliar, obtém-se $n = 168$ e $R_a^2 = 0.000981$, pelo que:

$w_{s,obs} = 168 \times 0.000981 = 0.1648$

$p_{obs} = P\left(W_s \geq w_{s,obs} \middle| H_0\right) = P(W_s \geq 0.1648) \approx 0.921$

Como $p_{obs} \approx 0.921 > \alpha = 0.05$, não se rejeita H_0 ao nível de 5%. Logo, não se rejeita que a variável residual seja homocedástica.

Nota: Valor crítico do teste (5%): $\chi^2_{0.05} = 5.991$

Continuação da questão ______

Modelo 1

$$\ln(sal_t) = \beta_1 + \beta_2 antig_t + \beta_3 \ln(prod_t) + \beta_4 est_t + \beta_5 exp_t + u_t$$

Regression Statistics	
Multiple R	0.8248
R Square	0.6803
Adjusted R Square	0.6725
Standard Error	0.7805
Observations	168

ANOVA

	Df	*SS*	*MS*	*F*	*Significance F*
Regression	4	211.3102	52.8275	86.7204	0.0000
Residual	163	99.2949	0.6092		
Total	167	310.6051			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	-3.4591	0.4118	-8.4000	0.0000
antig	0.0126	0.0054	2.3215	0.0215
ln(prod)	0.7816	0.0468	16.6894	0.0000
est	0.2840	0.1699	1.6722	0.0964
exp	0.1188	0.1796	0.6614	0.5093

Modelo 2

$$\hat{u}_t^2 = \alpha_1 + \alpha_2 \ln\widehat{(sal_t)} + \alpha_3 \ln\widehat{(sal_t)}^2 + v_t$$

Regression Statistics	
Multiple R	0.031329
R Square	0.000981
Adjusted R Square	-0.011128
Standard Error	0.841849
Observations	168

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	0.114885	0.057442	0.081052	0.922182
Residual	165	116.936976	0.708709		
Total	167	117.051861			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.378929	0.570425	0.664293	0.507430
Predicted ln(sal)	0.117701	0.297495	0.395641	0.692881
[Predicted ln(sal)]^2	-0.014940	0.037109	-0.402604	0.687761

NOTAS: $\hat{u}_t$ – resíduos MQ do Modelo 1 ; $\ln\widehat{(sal_t)}$ – valores ajustados do Modelo 1.