ESTATÍSTICA II - 2º Ano/Gestão do Desporto 2ª Prova Intercalar-1ª parte 14. 06. 18

1hora. (10 valores)

Nome:____ **Nº**_________

Espaço reservado para classificações

1 a. (10)	2 a. (5).	2 d. *(10)*	3 (15)
1 b. (10)	2 b.(5)	2 e. (10)	
1 c. (10)	2 c.*(10)*	2 f. (15)	T:

Atenção: 1. As folhas EXCEL no écran do computador tem os dados para a regressão e os dados para resolução das questões 1 e 3.

2. Devem apresentar na folha de exame a formalização e Justificação dos cálculos efectuados no EXCEL.

3. Devem fazer os cálculos no ficheiro EXCEL em folhas separadas para cada questão.

1. A quantidade de lixo, em toneladas, processada diariamente numa central de reciclagem pode ser considerada uma variável aleatória com distribuição normal com variância igual a 9. A direção da central comprometeu-se a reciclar uma quantidade média de pelo menos 10 toneladas e está sujeita a uma multa pesada se não cumprir.

   a. Formule um ensaio de hipóteses quanto ao cumprimento do compromisso da direção? Justifique a sua opção

   $$H_0: \mu \leq 10 \text{ contra } H_0: \mu > 10$$

   Porque o erro tipo 1 é o erro que conseguimos controlar na formulação das hipóteses deve optar-se pela situação mais gravosa no caso de se rejeitar H_0 e H_0 verdadeira. Neste caso tal significaria aceitar que se estava a cumprir o compromisso quando de facto tal não acontecia, estando assim a central sujeita a uma multa pesada.

   b. Interprete os erros de 1ª e 2ª espécie.

   Erro tipo 1 – considerar que se estão a reciclar mais de 10 toneladas quando na realidade se reciclam menos de 10.

   Erro tipo 2 - considerar que se estão a reciclar menos de 10 toneladas quando na realidade se reciclam 10 ou mais

   c. Determine a região crítica de dimensão 5% e enuncie a decisão que tomaria face ao resultado obtido.

   Ensaio – EXCEL

$$W = \{\bar{x}: \bar{x} > 10.8224\} \text{ como } \bar{x} = 10.89226 \in W \Rightarrow Rejeita - se\ H_0 \text{ ou}$$

$$valor - p = P(Z > 1.7845) = 0.0372 < 0.05 \Rightarrow Rejeita - se\ H_0$$

$$Para\ o\ ensaio\ H_0: \mu \geq 10 \text{ contra } H_0: \mu < 10$$

$$W = \{\bar{x}: \bar{x} < 10.8224\} \text{ como } \bar{x} = 10.89226 \notin W \Rightarrow Não\ se\ Rejeita\ H_0 \text{ ou}$$

$$valor - p = P(Z < 1.7845) = 0.9628 > 0.05 \Rightarrow Não\ se\ Rejeita\ H_0$$

2. Considere os dados na folha EXCEL para estimar o modelo:

$$Ln(Rend_i) = \beta_0 + \beta_1 Ln(Escol_i) + \beta_2 Género_i + \beta_3 EstCiv_i + \beta_4 Id_i + u_i$$

Onde: $Rend$ - rendimento anual em euros

$Escol$ - número de anos de escolaridade

$Género$ - <u>1</u> se Homem e <u>0</u> se Mulher

$EstCiv$ - estado civil <u>1</u> se casado <u>0</u> se outro

Id - idade

a. Estime o modelo

Regression Statistics	
Multiple R	0.4604
R Square	0.2119
Adjusted R Square	0.1513
Standard Error	0.8489
Observations	57

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	10.0780	2.5195	3.4962	0.0133
Residual	52	37.4732	0.7206		
Total	56	47.5512			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	5.6537	1.4315	3.9494	0.0002
Ln(Escol)	1.1907	0.4572	2.6046	0.0120
Género	0.5873	0.2265	2.5934	0.0123
Estado civil	0.2719	0.2273	1.1964	0.2370
Idade	0.0161	0.0118	1.3609	0.1794

b. Interprete o significado dos coeficientes estimados das variáveis $Escol$ e $Género$.

$Escol$ - Por cada acréscimo de 1% do nível de escolaridade, em média, tudo o resto constante, o rendimento cresce 1.19%.

$Género$ $-$ um indivíduo do sexo masculino tem, em média, tudo o resto constante, um rendimento 58% superior ao de um indivíduo do sexo feminino com idênticas características.

c. Analise a qualidade global do modelo

$H_0: \beta_1 = \cdots = \beta_4 = 0$ contra $H_1: \exists \beta_j \neq 0 \quad j = 1,2,3,4$

Estatística teste - $\frac{VE/(k)}{VR/(n-k-1)} \sim F_{(k,n-k-1)}$

$\underbrace{P\left(F_{(4,52)} > 3.4962\right) = 0.0133}_{2.5} \Rightarrow$ rejeita-se H_0

d. Teste a hipótese $H_0: \beta_2 \geq 0.5$ contra $H_1: \beta_2 < 0.5$.

Estatística teste -$T = \frac{b_j - \beta_j}{s_{\beta_j}} \sim t_{(n-k-1)}$

$$P(T < t_{obs}) = P\left(T < \frac{0.5873 - 0.5}{0.2265}\right) = P(T < 0.3855) = 0.6493 \Rightarrow$$

não se rejeita H_0

e. Formalize a hipótese para testar se as variáveis $EstCiv$.e Id contribuem para explicar o rendimento anual e realize o ensaio.

$H_0: \beta_3 = \beta_4 = 0$ contra $H_1: \exists \beta_j \neq 0 \quad j = 3,4$

Estatística teste - $F = \frac{(VR_0 - VR_1)/\left(\overbrace{k-p}^{m}\right)}{VR_1/(n-k-1)} \sim F_{(m, n-k-1)}$

$$F_{obs} = \frac{(39.7860 - 37.4732)/2}{37.4732/52} = 1.605 \Rightarrow valor - p = P\left(F_{(2,52)} > 1.605\right) = 0.2107$$

$valor - p > 0.05 \Rightarrow não\ se\ rejeita\ H_0$

f. Calcule o intervalo de confiança a 95% para a previsão em média do rendimento auferido por um individuo do sexo feminino, solteira, com 35 anos de idade e 15 anos de escolaridade.

$$\hat{E}(ln(Rend_0)) = 5.6537 + 1.1907 * \ln(15) + 0.0161 * 35 = 9.442$$

$$t\alpha/_2: P\left(T_{(52)} > t\alpha/_2\right) = 0.05 \Leftrightarrow t\alpha/_2 = 2.006;\ s_0 = \frac{s}{\sqrt{n}} = \frac{0.8489}{\sqrt{57}} = 0.1124$$

$$(9.442 - 2.006 * 0.1124, 9.442 + 2.006 * 0.1124) = (9.2165, 9.6675)$$

$$I.C._{E(Rend)}^{0.95} = \left(e^{9.2165}, e^{9.6675}\right) = (10047.538, 15817.012\ )$$

3. No seu livro “Outliers” Malcolm Gladwell argumenta que a maioria dos jogadores de “baseball” americanos nasceram nos meses entre Agosto e Dezembro, porque esta é a data limite de inscrição nas ligas de “baseball” não escolares. Os registos de uma amostra de 4512 jogadores americanos da liga de “baseball” constam da tabela seguinte:

Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
387	329	366	344	336	313	313	503	421	434	398	371

Usando um nível de significância de 5% existe evidência suficiente para rejeitar a hipótese de que os jogadores americanos da liga de “baseball” nasceram nos diferentes meses do ano com igual probabilidade?

$H_0: p_j^0 = \frac{1}{2} \quad j = 1, \cdots, 12$ contra $H_1: p_j^0 \neq \frac{1}{2} \quad algum\ j = 1, \cdots, 12$

$$Estatística\ teste - Q = \frac{\left(n_j - n * p_j^0\right)^2}{n * p_j^0} \sim \chi^2_{(m-1)}$$

	Jan	Fev	Mar	Abr	Mai	Jun
Freq.Obs.	387	329	366	341	336	313
Freq.Esp.	376	376	376	376	376	376
	0.321809	5.875	0.265957	3.257979	4.255319149	10.55585

	Jul	Ago	Set	Out	Nov	Dez	Total
Freq.Obs.	313	503	421	434	398	371	4512
Freq.Esp.	376	376	376	376	376	376	
	10.55585	42.89628	5.385638	8.946809	1.287234043	0.066489	93.67021

$Q_{obs} = 93.67021 \Rightarrow P\left(\chi^2_{(12-1)} > 93.67021\right) \approx 0 \Rightarrow$rejeita-se H_0