ESTATÍSTICA II - 2º Ano/2º sem. Gestão do Desporto Época Recurso 01. 07. 2019

2horas. (20 valores)

Nome:____

Espaço reservado para classificações

1.(15)	2 a) (10)	3 a) (10)	5 (10)	*7.* (15)	*8a) (10)*	*8.d)* (10)
	2 b) (15)	3 b) (15)	6 a) (15)		*8 b)* (10)	*8.e)* (15)
	2 c) (15)	4. (10)	6 b) (15)		*8 c)* (10)	T:

Atenção: 1. A folhas EXCEL no écran do computador tem os dados para a resolução de todas as questões do exame.

2. Intervalos de confiança, ensaios de hipóteses e regressão tem de ser feitos usando o EXCEL.

3. As questões não podem ser respondidas usando o método de tentativa e erro.

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

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

1. O que pode dizer acerca da variância da amostra S^2, como estimador da variância da população σ^2.

A variância da amostra é um estimador enviesado para σ^2 ■

A variância da amostra subestima, em média, a variância da população. ■

A variância da amostra sobrestima, em média, a variância da população. □

A variância da amostra não é um estimador consistente para a variância da população. ■

2. Seja X uma população com função densidade

$$f_X(x,p) = p * (1-p)^{x-1} \quad 0 < p < 1, \ \ x = \ 1, 2, \cdots$$

com $E(X) = \frac{1}{p}$ e $Var(X) = \frac{1-p}{p^2}$

(a) Determine o estimador pelo método dos momentos para p.

$$E(X) = \overline{X} \Leftrightarrow \frac{1}{p} = \overline{X} \Leftrightarrow \tilde{p} = \frac{1}{\overline{X}}$$

(b) Estude o enviesamento e consistência do estimador para a média da população.

$\overline{X}$ é o estimador para a média da população

$E(\overline{X}) = \mu$ pelo que $\overline{X}$ é estimador não enviesado para a média da população

$$\lim_{n\to+\infty} E(\overline{X}) = \lim_{n\to+\infty} \mu = \mu$$

$$\lim_{n\to+\infty} Var(\overline{X}) = \lim_{n\to+\infty} \frac{\sigma^2}{n} = 0$$

então, $\overline{X}$ é estimador consistente para a média da população

(c) Seja $T = \frac{X_1+X_n}{2}$ um outro estimador para a média da população. Estude a eficiência relativa deste estimador em relação ao usado na alínea b)

$$E(T) = E\left(\frac{X_1+X_n}{2}\right) = \frac{1}{2}E(X_1+X_n)$$

$$= \overbrace{\frac{1}{2}[E(X_1)+E(X_n)] = \frac{1}{2}(\mu+\mu)}^{Porque\ X_1, X_n\ \text{sáo elementos de uma amostra casual logo i.d.a } X} = \mu$$

pelo que T é estimador não enviesado para a média da população.

$$Var(T) = Var\left(\frac{X_1+X_n}{2}\right) = \overbrace{\frac{1}{4}Var(X_1+X_n) = \frac{1}{4}[Var(X_1)+Var(X_n)]}^{Porque\ X_1, X_n\ \text{são elementos de uma amostra casual logo ind.}},$$

$$= \frac{1}{4}(\sigma^2+\sigma^2) = \frac{\sigma^2}{2}$$

Então $Var(\overline{X}) < Var(T)$ pelo que, $\overline{X}$ é estimador relativamente mais eficiente que T para a média da população

3. O modo de comunicarmos alterou-se dramaticamente com o uso dos telemóveis e computadores. Um sociólogo pretende determinar a proporção da população portuguesa que usa o email como meio privilegiado de comunicação. Para tal recolheu uma amostra aleatória de dimensão 350 [Ver ficheiro EXCEL – Questão 2].

(a) Construa um intervalo de confiança a 90% para a proporção da população portuguesa que usa o email como meio privilegiado de comunicação. Interprete a estimativa calculada.

$$IC_{\theta}^{90\%} = (0.1809, 0.2534)$$

Em 90% dos intervalos de confiança calculados para esta população o verdadeiro valor da proporção da população portuguesa que usa o email como meio privilegiado de comunicação pertence a este intervalo.

(b) Se pretendesse aumentar a precisão da estimativa por intervalos calculada para um valor de 0.05 mantendo constante o nível de confiança, qual a dimensão da amostra a considerar?

$$n = ?: \Delta IC_{\theta}^{90\%} = 0.05 \Rightarrow z_{0.1/2} * \sqrt{\frac{\bar{x}(1-\bar{x})}{n}} = 0.025$$

$$\Leftrightarrow 1.645 * \sqrt{\frac{0.217(1-0.217)}{n}} = 0.025 \Leftrightarrow n = 736$$

4. Sabendo que o número de alunos que não passam a Estatística 2 segue uma distribuição binomial de parâmetro θ, um investigador pretende testar a hipótese de que a percentagem de insucessos é superior a 20%, isto é que $P(\bar{X} > 0.2)$. Esta é uma hipótese estatística? Se sim, explique a hipótese se não explique porquê.

Sim, porque $P(\bar{X} > 0.2)$ é uma função do parâmetro θ e como tal é uma hipótese estatística.

5. A potência de um dado ensaio é igual a 0.985 e a probabilidade de erro tipo 1 é 0.025. Qual das afirmações é verdadeira?

A probabilidade associada a uma decisão errada é 0.05

Se H_0 é verdadeira, a probabilidade de tomar uma decisão correta é 0.975. X

A probabilidade associada ao erro tipo 2 é 0.985

Se H_0 é falsa, a probabilidade de tomar uma decisão correta é 0.025.

6. Considere um controlador aéreo na torre de controlo do aeroporto Humberto Delgado em Lisboa. Se uma pequena mancha aparecer no écran, aproximando-se da rota de aterragem de um AIRBUS 300 o controlador tem de decidir entre:
- Está tudo sobre controlo pois trata-se apenas de uma pequena interferência no écran
- Está iminente uma colisão entre uma pequena avioneta e o AIRBUS 300

(a) Qual destas opções escolheria para hipótese nula num ensaio de hipóteses? Justifique

(b) Com base na sua escolha, em que consistiriam os erros de tipo 1 e 2?

7. Uma instituição de crédito pretende averiguar se a regularidade no cumprimento dos planos de amortização dos empréstimos para aquisição de habitação própria depende ou não do vínculo laboral do mutuário no momento do pedido do empréstimo. A análise de 200 empréstimos forneceu a seguinte informação:

Amortização do empréstimo	Vínculo laboral	
	Contrato a prazo	Contrato permanente
Pagamento em atraso	40	20
Pagamento feito	110	30

Utilizando um nível de significância de 2.5% dê a sua opinião e fundamente-a.

$H_0: p_{ij} = p_{i.} * p_{.j} \; i=1,2 \quad j=1,2$ contra $H_0: \exists p_{ij} \neq p_{i.} * p_{.j} \; i=1,2 \quad j=1,2$

Expected Frequencies			
	Column variable		
Row variable	C1	C2	Total
R1	45	15	60
R2	105	35	140
Total	150	50	200

(fo-fe)^2/fe	
0.555556	1.666667
0.238095	0.714286

Data	
Level of Significance	**0.025**
Number of Rows	2
Number of Columns	2
Degrees of Freedom	1

Results	
Critical Value	**5.023886**
Chi-Square Test Statistic	**3.174603**
***p*-Value**	**0.074791**
Do not reject the null hypothesis	

Expected frequency assumption is met.

8. Com o objectivo de conhecer os factores que determinam o rendimento de um indívíduo, use os dados do ficheiro EXCEL – Questão 5 para estimar a equação seguinte:

$$ln(rendimento_i) = \beta_0 + \beta_1 estado\ civil_i + \beta_2\ idade_i + \beta_3 \ln(anos\ escol._i) + \beta_4\ sexo_i + u_i$$

Com: $rendimento$ (€)
$idade\ \ (anos)$
$estado\ civil = 1 \quad se\ casado;$
$sexo = 1 \quad se\ homem$

a) Interprete os coeficientes das variáveis estado civil e anos de escolaridade.

$b_1 = 0.0684 -$ um indivíduo casado tem em média, tudo o resto constante, um rendimento 6.84% superior à de um solteiro

$b_3 = 1.2264 -$ um acréscimo de 1% no número de anos de escolaridade induz, em média, um acréscimo de 1.2264% no rendimento, tudo o resto constante

b) Teste a hipótese de o impacto da variável idade sobre o rendimento médio ser, tudo o resto constante, superior a 1%.

H_0: $\beta_2 \geq 0.01$ contra H_0: $\beta_1 < 0.01$ Estatística teste - $\frac{b_1 - \beta_1}{s_{\beta_1}} \sim t_{(152-4-1)}$

$t_{(obs)} = \frac{0.0168-0.01}{0.0071} = 0.9481$ Valor-p= $P(T_{147} < 0.9481) = 0.8277 > 0.05$

Pelo que não se rejeita H_0.

c) Estime a variância da variável residual.

$$\hat{\sigma}^2 = \frac{VR}{n-k-1} = \frac{132.0261}{147} = 0.8981$$

d) Teste a significância global do modelo.

H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ contra $H_1 \colon \exists\, \beta_j \neq 0 \quad j = 1,2,3,4\ \neq 0$

e) Calcule o intervalo de confiança para o rendimento médio de uma mulher casada, com 25 anos de idade e 10 anos de escolaridade.

$$IC^{95\%}_{rend.} = (996.61, 45413.6)$$