ESTATÍSTICA II - 2º Ano/2º Semestre Gestão do Desporto

Exame época normal 03. 06. 2020 - 1hora.

(14 valores)

Nome:____ **Nº**___________

Espaço reservado para classificações

1 a) (10)	2 a) (15)	3 a) (15)	4 a) (10)	4 c) (10)
1 b) (20)	2 b) (15)	3 b) (15)	*4 b) (*15)	4 d) (15)
				T:

NOTA: em todos os testes utilize um nível de significância de 5%.

1. Admite-se que o tempo de funcionamento, em milhares de horas, de determinada componente eletrónica tem uma distribuição $G\left(3,\ \frac{1}{\theta}\right)$. Selecionou-se uma amostra aleatória de dimensão 5: $(15, 8, 10, 5,\ 17)$.

a) Sabendo que $E(X) = 3\theta$, determine o estimador e estimativa de θ pelo método dos momentos.

$$E(X) = \bar{X} \Leftrightarrow 3\theta = \bar{X} \Leftrightarrow \tilde{\theta}_{(X_1, X_2, \cdots, X_n)} = \frac{\bar{X}}{3}$$

$$\tilde{\theta}_{(x_1, x_2, \cdots, x_5)} = \frac{\bar{x}}{3} = \frac{11}{3}$$

b) Para estimar o parâmetro θ definiram-se os seguintes 2 estimadores:

$$T_1 = \frac{\bar{X}}{3} \text{ e } T_2 = \frac{n\bar{X}+1}{3n}$$

Considerando uma amostra aleatória $(X_1, X_2, \cdots, X_n)$ estude o enviesamento e consistência destes estimadores.

$E(T_1) = E\left(\frac{\bar{X}}{3}\right) = \frac{E(\bar{X})}{3} = \frac{3\theta}{3} = \theta$ então T_1 é um estimador não enviesado para θ

$E(T_2) = E\left(\frac{n\bar{X}+1}{3n}\right) = E\left(\frac{n\bar{X}}{3n}\right) + E\left(\frac{1}{3n}\right) = \frac{E(\bar{X})}{3} + \frac{1}{3n} = \theta + \frac{1}{3n}$ então T_2 é um estimador enviesado para θ

$$\lim_{n\to\infty} E(T_1) = \lim_{n\to\infty} \theta = \theta \qquad \lim_{n\to\infty} E(T_2) = \lim_{n\to\infty}\left(\theta + \frac{1}{3n}\right) = \theta + \lim_{n\to\infty}\frac{1}{3n} = \theta$$

$$Var(T_1) = Var\left(\frac{\bar{X}}{3}\right) = \frac{Var(\bar{X})}{9} = \frac{\sigma^2/n}{9} = \frac{3\theta^2}{9n} = \frac{\theta^2}{3n} \Rightarrow \lim_{n\to\infty}\frac{\theta^2}{3n} = 0$$

$$Var(T_2) = Var\left(\frac{n\bar{X}+1}{3n}\right) = Var\left(\frac{n\bar{X}}{3n}\right) = \frac{Var(\bar{X})}{9} = \frac{\theta^2}{3n} \Rightarrow \lim_{n\to\infty}\frac{\theta^2}{3n} = 0$$

Então ambos os estimadores são consistentes

2. Uma amostra de 10 gerentes de controlo de qualidade com mais de 15 anos de experiência conduziu a um salário médio anual de 54000€ com um desvio padrão corrigido de 15000€.

   a) Determine o intervalo de confiança a 95% para o salário médio anual dos gerentes de controlo de qualidade com mais de 15 anos de experiência. Interprete o resultado obtido.

Confidence Interval Estimate for the Mean

Data	
Sample Standard Deviation	**15000**
Sample Mean	**54000**
Sample Size	**10**
Confidence Level	**95%**

Intermediate Calculations	
Standard Error of the Mean	4743.41649
Degrees of Freedom	9
t Value	2.2622
Interval Half Width	10730.3536

Confidence Interval	
Interval Lower Limit	**43269.65**
Interval Upper Limit	**64730.35**

Se se selecionarem um grande número de amostras de dimensão 10 desta população e se calcularem os IC para essas amostras, o salário médio anual dos gerentes de controlo de qualidade com mais de 15 anos de experiência estará entre **43269.65 e 64730.35 em 95% dos casos**

   b) Qual a dimensão mínima da amostra necessária para reduzir para metade a margem de erro?

Margem de erro= $\text{Interval Half Width} = t\alpha/_2 \frac{s'}{\sqrt{n}}$=10730.3536

n =?: Margem de erro= $\frac{10730.3536}{2} = 5365.1768 = 2.2622 * \frac{15000}{\sqrt{n}}$

$$5365.1768 \geq 2.2622 * \frac{15000}{\sqrt{n}} \Leftrightarrow \sqrt{n} \geq \frac{2.2622 * 15000}{5365.1768} = 6.3246 \Leftrightarrow n \geq 6.3246^2$$

$$\Leftrightarrow n \geq 41$$

3. A duração de lâmpadas de 100 W fabricadas por determinada companhia segue uma distribuição normal. A companhia garante que as lâmpadas de 100 W por ela fabricadas duram, em média, no mínimo 800 horas. Foi selecionada uma amostra casual de 50 lâmpadas de um lote tendo-se obtido uma duração média de 770 horas e um desvio padrão corrigido de 120 horas.

   a) Se se admitir uma devolução indevida com uma probabilidade de 1%, o lote devia ser devolvido?

      $H_0: \mu \geq 800$ contra $H_0: \mu < 800$ Estatística teste - $\bar{X}$ e como variância da população é desconhecida a distribuição por amostragem da Estatística teste é $\frac{\bar{X}-\mu}{S'/\sqrt{n}} \sim t_{(n-1)}$

      admitir uma devolução indevida=erro tipo 1⇒ $P(\text{admitir uma devolução indevida}) = \alpha$

      A- Resolução com determinação região de rejeição

      $P(\text{admitir uma devolução indevida}) = P(\bar{X} < k|H_0) = 0.01$

      Então $\frac{k-800}{120/\sqrt{50}} = invt(0.01, 49) = -2.40489 \Leftrightarrow k = 759.1877 \Rightarrow W_{\bar{X}} = \{\bar{x}: \ \bar{x} < 759.1877\}$

      $\bar{x} = 770 \in \overline{W_{\bar{x}}} \Rightarrow$ não se rejeita H_0 pelo que o lote não deve ser devolvido

      B – Resolução pelo valor-p

      $$valor - p = P\left(t_{(n-1)} < t_{observ.}\right) = P\left(t_{(n-1)} < -1.7678\right) = 0.0417 > 0.01$$

      ⇒ não se rejeita H_0 pelo que o lote não deve ser devolvido

   b) Determine a potência do ensaio se a verdadeira duração média das lâmpadas for de 790 e 750 horas. Comente os resultados obtidos.

      $$\beta(790) = P(Rej. H_0|H_0\, falsa) = P(\bar{X} < 759.1877|\mu = 790) = 0.037774$$

      $$\beta(750) = P(Rej. H_0|H_0\, falsa) = P(\bar{X} < 759.1877|\mu = 750) = 0.704654$$

4. um estudo sobre o desempenho dos alunos universitários foi recolhida no final do ano lectivo, uma amostra aleatória de 141 alunos do terceiro ano, tendo-se estimado o seguinte modelo::

$$média_i = \beta_0 + \beta_1 médiasec_i + \beta_2 educpai_i + \beta_3 educmãe_i + \beta_4 faltas_i + u_i$$

Onde: $média$ – média obtida no final do ano lectivo
$médiasec_i$ – média obtida no Ensino Secundário
$educpai\backslash mãe$ –número de anos de escolaridade de pai e mãe
$faltas$ – nº aulas a que o aluno não assistiu no último semestre

Tendo-se obtido o resultado apresentado no anexo 1.

a) Interprete as estimativas para os parâmetros β_1, β_4.

β_1- acréscimo de 1 valor na média no Ensino Secundário induz, em média, tudo o resto constante um acréscimo de aproximadamente 0.459 valores na média no final do ano lectivo

β_4 – por cada aula a que o aluno não assistiu no último semestre a nota média no final do semestre decresce, em média, tudo o resto constante, aproximadamente 0.028 valores.

b) Poder-se-á afirmar que um acréscimo de 1 valor na média obtida no Ensino Secundário conduz a um acréscimo médio da média no final do ano lectivo superior a meio valor, tudo o resto constante?

$H_0: \beta_1 \geq 0.5$ contra $H_0: \beta_1 < 0.5$

$$valor - p = P\left(t_{(n-k-1)} < t_{observ.}\right) = P\left(t_{(136)} < -0.4632\right) = 0.32197$$

Não se rejeita H_0, pelo que é muito provável que um acréscimo de 1 valor na média obtida no Ensino Secundário conduza a um acréscimo médio da média no final do ano lectivo superior a meio valor

c) Teste a nulidade conjunta dos regressores.

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0 \quad contra \;\; H_1: \exists\, \beta_j \neq 0 \;\; (j = 1, \cdots, k)$$

$$F = \frac{R^2/(k)}{(1 - R^2)/(n - k - 1)} \sim \boldsymbol{\mathcal{F}}_{(\boldsymbol{k,n-k-1})}$$

$$\boldsymbol{f}_{\boldsymbol{observ.}} = \frac{\mathbf{0.227451/4}}{\mathbf{(1 - 0.227451)/136}} = \mathbf{10.01}$$

$Valor - p = P\left(\boldsymbol{\mathcal{F}}_{(\mathbf{4,136})} > \mathbf{10.01}\right) \approx 0$ pelo que se rejeita H_0 e o modelo é globalmente significativo

d) Com base num teste adequado, comente a afirmação: “O número de anos de escolaridade dos pais de um estudante é um factor estatísticamente relevante para explicar a classificação média obtida no final do ano lectivo”. Dados no anexo 2

$$H_0: \beta_2 = \beta_3 = 0 \quad contra \;\; H_1: \exists\, \beta_j \neq 0 \;\; (j = 2,3)$$

$$F = \frac{(VR_{\,0} - VR_{\,1})/\left(\overbrace{k - p}^{m}\right)}{VR_{\,1}/(n - k - 1)} \sim F_{\,(m,\, n-k-1)}$$

$$\boldsymbol{f}_{\boldsymbol{observ.}} = \frac{\mathbf{(241.3395 - 239.8745)/2}}{\mathbf{239.8745/136}} = \mathbf{0.4153}$$

$Valor - p = P\left(\boldsymbol{\mathcal{F}}_{(\mathbf{2,136})} > \mathbf{0.4153}\right) = 0.661$, então não se rejeita H_0 podendo concluir-se que o número de anos de escolaridade dos pais de um estudante não é um factor estatísticamente relevante para explicar a classificação média obtida no final do ano lectivo

Anexo 1

Dependent Variable: MEDIA
Method: Least Squares
Included observations: 141

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.789119	1.349656	4.289329	0.0000
ESMEDIA	0.459202	0.088072	5.213945	0.0000
EDUCPAI	0.030556	0.042054	0.726585	0.4687
EDUCMAE	0.004612	0.030764	0.149924	0.8810
FALTAS	-0.024869	0.008730	-2.848802	0.0051

R-squared	0.227451	S.E. of regression	1.328075
Adjusted R-squared	0.204729	S.D. dependent var	1.489241
Sum squared resid	239.8745	F-statistic	10.01017

Anexo 2

Dependent Variable: MEDIA
Method: Least Squares
Included observations: 141

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.316668	1.213468	5.205468	0.0000
ESMEDIA	0.458804	0.087691	5.232063	0.0000
FALTAS	-0.025812	0.008588	-3.005480	0.0032

R-squared	0.222733	S.E. of regression	1.322436
Adjusted R-squared	0.211468	S.D. dependent var	1.489241
Sum squared resid	241.3395	F-statistic	19.77258