ESTATÍSTICA II - 2º Ano/2º Semestre GDES

Exame Época Recurso 02. 07. 2020 – 1:30 hora.

(14 valores)

Nome:____ **Nº**__________

Espaço reservado para classificações					
1 a) (10)	2 a) (15)	3 a) (15)	4 a) (10)	4 c) (10)	5. (15)
1 b) (15)	2 b) (15)	3 b) (15)	*4 b) (*10)	4 d) (10)	
					T:

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

1. Seja uma população com distribuição Uniforme no intervalo $(0, \theta)$. Selecionou-se uma amostra aleatória de dimensão 10 a partir da qual se obteve uma média de 1.65. Sabe-se que $E(X) = \frac{\theta}{2}$ e $Var(X) = \frac{\theta^2}{12}$.

   a) Determine o estimador e estimativa de θ pelo método dos momentos.

   $$E(X) = \overline{X} \Leftrightarrow \frac{\theta}{2} = \overline{X} \Leftrightarrow \tilde{\theta}_{(X_1,X_2,\cdots,X_n)} = 2\overline{X}, \qquad \tilde{\theta}_{(x_1,x_2,\cdots,x_5)} = 2\overline{x} = 3.3$$

   b) Considerando uma amostra aleatória $(X_1, X_2, \cdots, X_n)$, estude o enviesamento e consistência do estimador $T_1 = 2\bar{X}$ para θ.

   $E(T_1) = E(2\bar{X}) = 2E(\bar{X}) = 2 * \frac{\theta}{2} = \theta \Rightarrow T_1$ é um estimador não enviesado para θ.

   $$Var(T_1) = Var(2\bar{X}) = 4Var(\bar{X}) = 4 * \frac{\theta^2/12}{n} = \frac{\theta^2}{3n}$$

   $\lim_{n\to\infty} E(T_1) = \lim_{n\to\infty} \theta = \theta, \quad \lim_{n\to\infty} Var(T_1) = \lim_{n\to\infty} \frac{\theta^2}{3n} = 0 \Rightarrow T_1$ é um estimador consistente para θ.

2. De uma população com distribuição normal de parâmetros desconhecidos foi selecionada uma amostra casual de dimensão 15 para a qual se obteve $\sum_{i=1}^{15} x_i = 30$ e $s'^2 = 0.16$.

a) Determine o intervalo de confiança a 95% para a variância da população. Comente a afirmação de que a variância da população pertence ao intervalo de confiança calculado com uma probabilidade de 95%?

IC variância

Data	
Sample Size	**15**
Sample Standard Deviation	**0.4**
Confidence Level	**95%**

Intermediate Calculations	
Degrees of Freedom	14
Sum of Squares	2.24
Single Tail Area	0.025
Lower Chi-Square Value	5.6287
Upper Chi-Square Value	26.1189

Results	
Interval Lower Limit for Variance	**0.0858**
Interval Upper Limit for Variance	**0.3980**

Não, porque a probabilidade da variância da população pertencer ao intervalo confiança é 0 ou 1. Se se selecionarem várias amostras da população e para cada uma se calcular o $IC_{\sigma^2}^{95\%}$, em 95% desses intervalos a variância da população estará contida no intervalo.

b) A partir da mesma amostra e recorrendo à metodologia habitual, obteve-se o intervalo de confiança para a média da população (1.86, 2.14). Qual o grau de confiança associado a este intervalo?

$$\bar{x} - t_{\alpha/2} \frac{s'}{\sqrt{n}} = 1.86 \Leftrightarrow 2 - t_{\alpha/2} \frac{0.4}{\sqrt{15}} = 1.86 \Leftrightarrow t_{\alpha/2} = -1.3555$$

$$\Rightarrow \alpha/2 = P\left(t_{(14)} < -1.3555\right) = 0.098355 \Rightarrow \alpha = 0.19671 \Rightarrow 1 - \alpha \approx 80\%$$

Nota: existem outras resoluções possíveis e igualmente certas.

3. O tempo de execução de determinada tarefa, em minutos, é uma variável aleatória com distribuição Normal. Numa amostra casual de 20 realizações dessa tarefa obteve-se um tempo médio de execução de 10.8 minutos e uma variância corrigida igual a 7.1.

a) Com base num teste de hipóteses adequado, com um nível de significância de 5%, poder-se-á concluir que o tempo médio de execução da tarefa não excede os 10 minutos?

$H_0: \mu \leq 10$ contra $H_0: \mu > 10$ Estatística teste - $\bar{X}$ e como variância da população é desconhecida a distribuição por amostragem da Estatística teste é $\frac{\bar{X}-\mu}{S'/\sqrt{n}} \sim t_{(n-1)}$

A- Resolução com determinação região de rejeição

$$k = ?: P(\bar{X} > k|H_0) = 0.05 \Leftrightarrow P\left(\frac{\bar{X}-\mu_0}{S'/\sqrt{n}} > \frac{k-10}{\sqrt{\frac{7.1}{20}}}\right) = 0.05$$

Então $\frac{k-10}{\sqrt{\frac{7.1}{20}}} = invt(0.95, 19) = 1.7291 \Leftrightarrow k = 11.03 \Rightarrow W_{\bar{X}} = \{\bar{x}: \ \bar{x} > 11.03\}$

$\bar{x} = 10.8 \in \overline{W_{\bar{x}}} \Rightarrow$ não se rejeita H_0 pelo que se pode concluir que o tempo médio de execução da tarefa não excede os 10 minutos

B – Resolução pelo valor-p

$$t_{observ.} = \frac{10.8-10}{\sqrt{\frac{7.1}{20}}} = 1.343$$

$$valor - p = P\left(t_{(n-1)} > t_{observ.}\right) = P\left(t_{(n-1)} > 1.343\right) = 0.0976 > 0.05$$

$\Rightarrow$ não se rejeita H_0 pelo que se pode concluir que o tempo médio de execução da tarefa não excede os 10 minutos

b) Admita que a variância da população é conhecida e igual a 8 e que pretende testar $H_0: \mu = 10$ contra $H_1: \mu > 10$. Qual deverá ser a dimensão mínima da amostra, para que garantindo uma dimensão do ensaio de 5%, a potência do ensaio seja igual a 0.995 quando $\mu = 11$?

$$n = ?: \beta(11) = 0.995 \Leftrightarrow P(\bar{X} > 11.03|\mu = 11) = 0.995$$
$$\Leftrightarrow P(\bar{X} \leq 11.03|\mu = 11) = 0.005$$
$$\Leftrightarrow P\left(\frac{\bar{X}-11}{\sqrt{\frac{\sigma^2}{n}}} \leq \frac{11.03-11}{\sqrt{\frac{8}{n}}}\right) = 0.005$$

$$\Rightarrow \frac{11.03-11}{\sqrt{\frac{8}{n}}} = norm.s.inv(0.005) \Leftrightarrow \frac{11.03-11}{\sqrt{\frac{8}{n}}} = -2.57583 \Leftrightarrow n \geq 58977$$

Para explicar o salário anual dos directores gerais de empresas, um analista financeiro desenvolveu o seguinte modelo econométrico:

$$lsal_i = \beta_0 + \beta_1 lucros_i + \beta_2 lvendas_i + \beta_3 exper_i + \beta_4 \text{exper}_i^2 + u_i$$

Onde: sal – salário anual do director (milhares €)
$lsal$ – logaritmo do salário anual do director
$lucros$ – lucros (milhões €)
$vendas$ –volume de vendas da empresa (milhões €)
$lvendas$ – logaritmo do volume de vendas
$exper$ –número de anos como director da empresa

Tendo-se obtido o resultado apresentado no anexo 1.

a) Interprete as estimativas para os parâmetros β_1 e β_2.
β_1- um crescimento dos lucros de 1 milhão de euros induz, em média, tudo o resto constante, um acréscimo de aproximadamente 0.018% no salário anual dos directores de empresa
β_2 – um crescimento de 1% no volume de vendas da empresa induz, em média, tudo o resto constante, um acréscimo de aproximadamente 0.196% no salário anual dos directores de empresa

b) Teste a significância estatística das variáveis lucros e volume de vendas.
Nota: embora estando os valor-p calculados no anexo teriam de ter explicitado as hipóteses e formalizado o ensaio.
$H_0: \beta_1 = 0$ contra $H_0: \beta_1 \neq 0$ $\qquad t_{observ.} = \frac{0.00018}{0.00012} = 1.5$

$$valor - p = P\left(t_{(172)} \leq -1.5\right) + P(Z \geq 1.5) = 2 * P\left(t_{(172)} \geq 1.5\right) = 2 * 0.0677 = 0.1354 > 0.05$$

$H_0: \beta_2 = 0$ contra $H_0: \beta_2 \neq 0$ $\qquad t_{observ.} = \frac{0.19636}{0.03314} = 5.925$

$$valor - p = 2 * P\left(t_{(172)} \geq 5.925\right) \approx 2 * 0 = 0 < 0.05$$

Então pode concluir-se que a variável lucros não é estatísticamente significativa mas a variável volume de vendas é.

c) Teste a adequação global do modelo.

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

$$F = \frac{R^2/(k)}{(1-R^2)/(n-k-1)} \sim \boldsymbol{\mathcal{F}}_{(\boldsymbol{k},\boldsymbol{n-k-1})} \qquad \boldsymbol{f}_{\boldsymbol{observ.}} = \frac{\mathbf{0.3336/4}}{\mathbf{(1-0.3336)/172}} = \mathbf{21.526}$$

$Valor - p = P\left(\boldsymbol{\mathcal{F}}_{(\mathbf{4,172})} > \mathbf{21.526}\right) \approx 0$ pelo que se rejeita H_0 e o modelo é globalmente significativo

d) Alguém afirmou que o número de anos como director da empresa e o seu quadrado não são significativas na explicação do salário dos directores de empresa. Comente a afirmação com um teste adequado.

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

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

$$\boldsymbol{f_{observ.}} = \frac{\mathbf{(45.909 - 43.081)/2}}{\mathbf{43.081/172}} = \mathbf{5.6454}$$

$Valor - p = P\left(\boldsymbol{\mathcal{F}}_{\mathbf{(2,172)}} > \mathbf{5.6454}\right) = 0.004,$ pelo que se rejeita H_0 podendo concluir-se que as variáveis $exper$ e $exper^2$ também são estatísticamente significativas.

4. Numa amostra de 100 dias, o número de tentativas de intrusão nos sistemas informáticos de uma instituição do Ensino Superior teve a seguinte distribuição:

Número tentativas por dia	0	1	2	3 ou mais
Número de dias	40	35	20	5

Teste a hipótese, com uma dimensão de 5%, de o número de tentativas de intrusão nos sistemas informáticos seguir uma distribuição de Poisson.

$H_0: X \sim Po(\lambda)$

Número tentativas por dia	0	1	2	3 ou mais	
Número de dias	40	35	20	5	100
$\widehat{p_j}$	0.407	0.366	0.165	0.063	
$n * \widehat{p_j}$	40.657	36.591	16.466	6.286	
q_j	0.011	0.069	0.758	0.263	1.1012
				valor-p	0.576591

$$P(X \geq 3)$$

Média da amostra=$\hat{\lambda}$ =0.9

$\boldsymbol{Q} \dot{\sim} \boldsymbol{\chi^2_{(4-1-1)}}$

$\boldsymbol{q_{observ}} = \mathbf{1.1012}$ $\quad$ $\boldsymbol{valor - p = P\left(\chi^2_{(4-1-1)} > 1.1012\right) = 0.5766 > 0.05}$
então não se rejeita H_0

Anexo 1 – Variável dependente *lsal*

Regression Statistics	
R Square	0.3336
Adjusted R Square	[redacted]
Standard Error	0.5005
Observations	177

ANOVA

	df	*SS*
Regression	4	21.566
Residual	172	43.081
Total	176	64.646

	Coefficients	*Standard Error*
Intercept	4.90417	0.24397
lucros	0.00018	0.00012
lvendas	0.19636	0.03314
exper	0.04537	0.01422
exper2	-0.00123	0.00048

Anexo 2 – Variável dependente *lsal*

Regression Statistics	
R Square	0.290
Adjusted R Square	0.282
Standard Error	0.514
Observations	177

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	18.737	9.369	35.507	0.000
Residual	174	45.909	0.264		
Total	176	64.646			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	5.14509	0.23473	21.919	0.000
lucros	0.00018	0.00012	1.483	0.140
lvendas	0.19370	0.03400	5.697	0.000