ESTATÍSTICA II – Economia/Finanças

Duração: 2h – Época Normal – Maio 2021

Nome:____ Nº:_____________

Espaço reservado para classificações					
1a.(15)	2.(15)	3a.(10)	4.(15)	5a.(15)	5d.(10)
1bi.(15)		3b.(20)		5b.(10)	5e.(15)
1bii.(10)		3c.(20)		5c.(10)	5f.(20)

Atenção: todas as questões devem ser devidamente formalizadas e justificadas. Sempre que fizer um teste de hipóteses formule as hipóteses em teste e apresente a estatística de teste e sempre que fizer um intervalo de confiança apresente a variável fulcral.

1. Um operador de uma rede móvel, a NewMobile, assume que a duração de cada chamada dos seus clientes, em minutos, segue uma distribuição normal. Da observação de 15 chamadas escolhidas aleatoriamente, obteve-se um total de 102 minutos e um desvio padrão amostral corrigido de 1.7 minutos.
   a. **[15]** Um gestor de clientes afirma que a duração média de uma chamada não ultrapassa os 5 minutos. Teste a afirmação ao nível de significância de 10%.

$X \sim N(\mu, \sigma^2) \quad n = 15 \quad \sum_{i=}^{15} x_i = 102; \quad s' = 1.7$

$H_0: \mu \leq 5 \ \ contra \ H_1: \mu > 5 \qquad \alpha = 0.10$

Estatística de teste: $T = \frac{\bar{X}-\mu}{S'/\sqrt{n}} \sim t_{(n-1)}$ ou seja $T = \frac{\bar{X}-5}{S'/\sqrt{15}} \sim t_{(14)}$

$W = \{t: t > 1.345\}$ e $t_{obs} = \frac{6.8-5}{1.7/\sqrt{15}} = 4.1 \in W$

ou

$Valor - p = P\left(t_{(14)} \geq 4.1\right) < 0.001$

Rejeita-se H_0 e, portanto, a afirmação do gestor de clientes é de rejeitar.

   b. Admita que o desvio padrão da população é 1.5.
      i. **[15]** Construa um intervalo de confiança a 95% para a duração média de uma chamada.

Variável Fulcral - $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

$IC_{\mu}^{95\%} = \left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}; \ \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$ Com $z_{\alpha/2}: P(Z > z_{\alpha/2}) = \frac{\alpha}{2}$ donde $z_{\alpha/2} = 1.96$

$$IC_{\mu}^{95\%} = \left(6.8 - 1.96 * {1.5}/{\sqrt{15}}; \ 6.8 + 1.96 * {1.5}/{\sqrt{15}}\right) = (6.0409, ; 7.559)$$

      ii. **[10]** Quantas chamadas devem ser acrescentadas à amostra para reduzir a amplitude do intervalo obtido na alínea anterior para metade, mantendo o nível de confiança?

Amplitude do obtido: $2\, z_{\alpha/2}\frac{\sigma}{\sqrt{15}} = 1.5182$

Procura-se então o menor n inteiro tal que $2\, z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \frac{1.5182}{2} \Leftrightarrow \sqrt{n} \geq \frac{4\, z_{\frac{\alpha}{2}}\, \sigma}{1.5182} = 7.7460$ e portanto $n \geq 60.00075$ isto é $n = 61$.

Pelo que para reduzir a amplitude do intervalo obtido na alínea anterior para metade, mantendo o nível de confiança se teria de acrescentar à amostra 46 observações.

2. **[15]** A NewMobile está a equacionar lançar um novo pacote de preços, especialmente vocacionado para assinantes que fazem chamadas de longa duração. O departamento comercial deste operador alega não valer a pena lançar esse pacote, pois considera que a proporção de aderentes a uma eventual campanha de lançamento desse pacote não seria superior a 15%. Para avaliar a suposição do departamento comercial, foi efetuado um inquérito a 200 assinantes, escolhidos ao acaso, tendo 34 deles manifestado interesse em usufruir do dito pacote. Teste a suposição do departamento comercial. Considere uma dimensão de 5%.

$H_0: \theta \leq 0.15 \quad contra\ H_1: \theta > 0.15 \qquad \alpha = 0.05$

Estatística de teste: $Z = \frac{\bar{X}-\theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} \dot{\sim} N(0,1)$ ou seja $Z = \frac{\bar{X}-0.15}{0.02525} \dot{\sim} N(0,1)$

$W = \{z: z > 1.645\}$ já que, sob H_0, $P(Z > 1.645) = 0.05$. Como $z_{obs} = \frac{\left(\frac{34}{200}\right)-0.15}{0.02525} = 0.7921 \notin W$, não se rejeita H_0, isto é, não se rejeita a afirmação do departamento comercial.
Alternativamente,

$$Valor - p = P(Z \geq z_{obs}) = P(Z \geq 0.7921) \approx 0.2148 > 0.05$$

3. A duração de uma chamada de longa duração, em minutos, é uma variável aleatória *X* que toma valores superiores a 10 minutos e tem função densidade de probabilidade dada por $f(x|\theta) = \theta\, e^{-\theta\,(x-10)}$, $x > 10$, com $\theta > 0$ e valor esperado $E(X) = \frac{1+10\,\theta}{\theta}$. Da observação das chamadas de longa duração de 50 clientes escolhidos aleatoriamente obteve-se um total de 978.87 minutos.

   a. **[10]** Obtenha a estimativa dos momentos para θ.

$E(X) = \overline{X} \Leftrightarrow \frac{1+10\,\theta}{\theta} = \overline{X} \Leftrightarrow \frac{1}{\theta} + 10 = \overline{X} \Leftrightarrow \theta = \frac{1}{\overline{X}-10}$

Logo, o estimador dos momentos será $\tilde{\theta} = \frac{1}{\overline{X}-10}$.

A estimativa vem $\tilde{\theta} = \tilde{\Theta}_{(x_1,x_2,\cdots,x_n)} = \frac{1}{\bar{x}-10} = \frac{1}{\frac{978.87}{50}-10} = \frac{1}{9.5774} = 0.1044$

   b. **[20]** Mostre que o estimador de máxima verosimilhança para θ é igual ao estimador dos momentos e obtenha, justificando, a estimativa de máxima verosimilhança para a duração esperada de uma chamada de longa duração.

$L(\theta) = \prod_{i=1}^{n} \theta\, e^{-\theta\,(x-10)} = \theta^n\, e^{-\theta \sum_{i=1}^{n}(x_i-10)}$ logo $l(\theta) = \ln\,(L(\theta) = n \ln\theta - \theta \sum_{i=1}^{n}(x_i - 10)$

$l'(\theta) = \frac{n}{\theta} - \sum_{i=1}^{n}(x_i - 10) = 0$

Ou alternativamente

$l(\theta) = \sum_{i=1}^{n} \ln\left(\theta\, e^{-\theta\,(x_i-10)}\right) = \sum_{i=1}^{n}(\ln\theta - \theta\,(x_i - 10)\,)$

$l'(\theta) = \sum_{i=1}^{n}\left(\frac{1}{\theta} - (x_i - 10)\,\right) = \frac{n}{\theta} - \sum_{i=1}^{n}(x_i - 10)$

Em qualquer dos casos

$l'(\theta) = 0 \Leftrightarrow \frac{n}{\theta} = \sum_{i=1}^{n}(x_i - 10) \Leftrightarrow \theta = \frac{n}{\sum_{i=1}^{n}(x_i-10)} = \frac{1}{\bar{x}-10} = 0.1044$

E como $l''(\theta) = -\frac{n}{\theta^2} < 0$, $\hat{\theta} = \frac{1}{\bar{x}-10}$ é a estimativa de MV para θ

O estimador será então $\hat{\theta} = \frac{1}{\bar{X}-10}$ que é igual ao estimador dos momentos.

Aplicando a propriedade da invariância dos estimadores da máxima verosimilhança, tem-se:

$$\widehat{\boldsymbol{E}(\boldsymbol{X})} = \frac{1+10\,\hat{\theta}}{\hat{\theta}} = \frac{1+10 \times 0.1044}{0.1044} = 19.58$$

c. **[20]** Determine o valor da constante a por forma a que o estimador $T = a + \bar{X}$ seja centrado para $\frac{1}{\theta}$. Será que o quadrado do estimador encontrado, T^2, é estimador centrado para $\frac{1}{\theta^2}$? E assintoticamente centrado? Justifique.

$E(T) = E(a + \bar{X}) = a + E(\bar{X}) = a + E(X) = a + \frac{1}{\theta} + 10$ logo para que $E(T) = \frac{1}{\theta}$ virá $a = -10$.
Assim $T = -10 + \bar{X}$.

$E(T^2) = var(T) + (E(T))^2 = var(-10 + \bar{X}) + \left(\frac{1}{\theta}\right)^2 = var(\bar{X}) + \frac{1}{\theta^2}$

O estimador só seria centrado para $(1/\theta^2)$ se $var(\bar{X}) = 0$ o que não e possível já que X é uma v.a. não degenerada.

Em termos assintóticos, ter-se-ia de verificar $\lim_{n\to\infty} E(T^2) = 1/\theta^2$.

Assumindo que $var(X)$ existe (o que se pode verificar), então $var(T) = var(\bar{X}) = \frac{var(X)}{n}$ e portanto

$\lim_{n\to\infty} E(T^2) = \lim_{n\to\infty}\left(var(\bar{X}) + \frac{1}{\theta^2}\right) = \lim_{n\to\infty} var(\bar{X}) + \frac{1}{\theta^2} = \frac{1}{\theta^2} + \lim_{n\to\infty}\frac{var(X)}{n} = \frac{1}{\theta^2}$ já que $\lim_{n\to\infty}\frac{var(X)}{n} = 0$.

Assim T^2 será assintoticamente centrado para $\frac{1}{\theta^2}$.

4. **[15]** Com o objetivo de estudar a relação entre a prática de desporto e a frequência de consumo de suplementos alimentares, foram selecionados de forma aleatória e posteriormente inquiridos 100 indivíduos, tendo-se obtido as seguintes contagens:

	Consumo de Suplementos Alimentares	
	Diariamente	Ocasionalmente
Desportistas	36	23
Não desportistas	18	23

Com base num teste estatístico de dimensão 0.05, pode-se afirmar que o consumo de suplementos alimentares está associado à prática de desporto?

Seja i o indicador das linhas ($i = 1$ desportistas e 2 não desportistas) e j o das colunas ($j = 1$ diariamente e 2 ocasionalmente) e procede-se a um reste de independência do qui-quadrado.

$$H_0: p_{ij} = p_{i.} \times p_{.j} \;\; \forall(i,j)\; i,j = 1,2 \; contra \; H_1: p_{ij} \neq p_{i.} \times p_{.j} \;\; algum\; (i,j)\; i,j = 1,2$$

Estatística de teste:

$Q = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{\left(Obs_{ij} - Esp_{ij}\right)^2}{Esp_{ij}} \sim \chi_1^2$ já que $(r-1)(s-1) = 1$

$W = \{q : q > 3.841\}$ para $\alpha = 0.05$

Matriz dos $\widehat{p_{ij}} = \widehat{p_{i.}} \times \widehat{p_{.j}}$ (nº esperado de casos assumindo H_0 verdadeira)

	Consumo de Suplementos Alimentares	
	Diariamente	Ocasionalmente
Desportistas	31.86	27.14
Não desportistas	22.14	18.86

$$q_{obs} = \frac{(36-31.86)^2}{31.86} + \frac{(23-27.14)^2}{27.14} + \frac{(18-22.14)^2}{22.14} + \frac{(23-18.86)^2}{18.86}$$
$$= 0.5380 + 0.6315 + 0.7741 + 0.9088 = 2.8524$$

Como $q_{obs} \notin W$, não se rejeita que haja independência entre o consumo de suplementos alimentares e a natureza do consumidor (desportista ou não desportista)

5. Com o intuito de modelar as despesas declaradas ao seguro de saúde por apólice considerou-se o seguinte modelo econométrico:
$ldesp_i = \beta_0 + \beta_1\, idade_i + \beta_2\, imc_i + \beta_3\, dep_i + \beta_4\, fumador_i + \beta_5\, dist_i + u_i,\ i = 1,2,\dots,n$
Onde:
   - $ldesp_i$: logaritmo das despesas participadas ao seguro pela apólice *i*.
   - $idade_i$: idade, em anos, do tomador da apólice *i*
   - imc_i: índice de massa corporal do tomador da apólice *i* (*kg* /*m*2)
   - dep_i: número de dependentes menores de idade cobertos pela apólice *i*
   - $fumador_i$: variável que toma o valor 1 se o segurado *i* for fumador e 0 no caso contrário.
   - $dist_i$: distância da residência do segurado i ao hospital mais próximo.

Os resultados da estimação pelo método dos mínimos quadrados encontram-se em anexo. Sempre que necessário pode utilizar os resultados em anexo para responder às diferentes questões.

a. **[15]** Interprete a estimativa obtida para β_1 e teste a significância estatística do coeficiente. Será que o coeficiente tem relevância prática?

Um acréscimo de 1 ano na idade do segurado induz, em média, ceteris paribus, um aumento de aproximadamente 3.56 % nas despesas participadas ao seguro.

$H_0: \beta_1 = 0 \quad contra\ H_1: \beta_1 \neq 0$

$T = \frac{\widehat{\beta_1} - \beta_1}{s_{\widehat{\beta_1}}} = \frac{\widehat{\beta_1}}{s_{\widehat{\beta_1}}} \sim t_{(294)}$ $\qquad W_{0.05} = \{t: t < -1.96\ ou\ t > 1.96\}$

$t_{obs} = 20.64$ logo rejeita-se claramente a hipótese nula. A variável idade tem não só relevância estatística como relevância prática já que induz variação substancial no valor das despesas participadas.

b. **[10]** Teste se um aumento do índice de massa corporal origina, em média um crescimento das despesas participadas ao seguro, *ceteris paribus*.

$H_0: \beta_2 \leq 0 \quad contra\ H_1: \beta_2 > 0$ (ou $H_0: \beta_2 = 0$) $\qquad T = \frac{\widehat{\beta_2} - \beta_2}{s_{\widehat{\beta_2}}} = \frac{\widehat{\beta_2}}{s_{\widehat{\beta_2}}} \sim t_{(294)}$

$W_{0.05} = \{t: t > 1.645\}$ Como $t_{obs} = 2.844$ rejeita-se H_0 e pode concluir-se que um aumento do índice de massa corporal origina, em média um crescimento das despesas participadas ao seguro, *ceteris paribus.*

c.[10] Teste a nulidade conjunta dos coeficientes β_2 e β_5. Qual a conclusão a tirar?

$H_0: \beta_2 = \beta_5 = 0 \quad contra\ H_1: \exists \beta_j \neq 0 \quad j = 2, 5$

Estatística de teste: $F = \frac{(SSR_{\,r} - SSR_{\,nr})/2}{SSR_{\,nr}/294} \sim F_{(2,\ 294)}$ ou $F = \frac{(R^2 - R_*^2)/2}{(1-R^2)/294} \sim F_{(2,\ 294)}$

$W_{0.05} = \{f: f > 3\}$

$f_{obs=} \frac{(SSR_{\,r} - SSR_{\,nr})/q}{SSR_{\,nr}/(n-k-1)} = \frac{(55.436652 - 53.604126)/2}{53.604126/(300-5\quad)} = 5.025$ ou $f_{obs=} \frac{(0.805219 - 0.798560)/2}{(1-0.805219)/(294)} = 5.025$

Como $f_{obs} \in W$ (ou $valor - p < 0.05$ para quem tenha optado por esta via), rejeita-se H_0 e conclui-se que as variáveis *imc* e *dist* são conjuntamente significantes ao nível 5%.

d.[10] Uma analista da companhia afirma que o aumento do agregado em um dependente leva, *ceteris paribus*, a um aumento das despesas esperadas em pelo menos 10%. Teste a afirmação ao nível de 1%.

$H_0: \beta_3 \geq 0.1 \quad contra\ H_1: \beta_3 < 0.1$

Estatística de teste: $T = \frac{\widehat{\beta_3} - 0.1}{s_{\widehat{\beta_3}}} \sim t_{(294)}$

$W = \{t: t < -2.326\} \qquad t_{obs} = \frac{0.089340 - 0.1}{0.020740} = -0.5140$

Não se rejeita H_0 ao nível de 1%, pelo que não se nega que o aumento do agregado em um dependente leve a um aumento das despesas esperadas em pelo menos 10%, *ceteris paribus*.

Alternativamente, $valor - p = P(T \leq -0.5140)$ que vem entre 0.4 e 0.25 levando à mesma conclusão

e.[15] Pretende-se testar $H_0: 6\,\beta_1 + 3\,\beta_3 \leq 0.3$ contra $H_1: 6\,\beta_1 + 3\,\beta_3 > 0.3$. Realize o teste usando um nível de significância de 5%.

$H_0: \delta \leq 0.3\ contra\ H_1: \delta > 0.3$ com $\delta = 6\,\beta_1 + 3\,\beta_3$

Sol 1: $\delta = 6\,\beta_1 + 3\,\beta_3 \Leftrightarrow \beta_3 = \frac{\delta}{3} - 2\beta_1$

$$ldesp_i = \beta_0 + \beta_1\, idade_i + \beta_2\, imc_i + \left(\frac{\delta}{3} - 2\beta_1\right)\frac{dep_i}{3} + \beta_4\, fumador_i + \beta_5\, dist_i + u_i$$

$$ldesp_i = \beta_0 + \beta_1(idade_i - 2dep_i) + \beta_2\, imc_i + \delta\frac{dep_i}{3} + \beta_4\, fumador_i + \beta_5\, dist_i + u_i$$

Sol 2: $\delta = 6\,\beta_1 + 3\,\beta_3 \Leftrightarrow \beta_1 = \frac{\delta}{6} - \frac{\beta_3}{2}$

$$ldesp_i = \beta_0 + \left(\frac{\delta}{6} - \frac{\beta_3}{2}\right) idade_i + \beta_2\, imc_i + \beta_3 dep_i + \beta_4\, fumador_i + \beta_5\, dist_i + u_i$$

$$ldesp_i = \beta_0 + \delta\,\frac{idade_i}{6} + \beta_2\, imc_i + \beta_3\left(dep_i - \frac{idade_i}{2}\right) + \beta_4\, fumador_i + \beta_5\, dist_i + u_i$$

Em qualquer dos casos:

Estatística de teste: $T = \frac{\hat{\delta} - 0.3}{s_{\hat{\delta}}} \sim t_{(294)} \quad W_{0.05} = \{t: t > 1.645\} \qquad t_{obs} = \frac{0.481393 - 0.3}{0.062346} = 2.9095$

Rejeita-se H_0.

Alternativamente, $valor - p = P(T > 2.9095) < 0.05$ pelo que se rejeita H_0

f.[20] Construa uma regressão auxiliar que lhe permita testar a existência de heterocedasticidade neste modelo. Sabendo que o R2 obtido na estimação da regressão auxiliar que definiu foi 0.094, qual a conclusão a tirar (refira eventuais consequências na qualidade dos estimadores propostos).

Várias soluções possíveis de acordo com o teste escolhido.

White simplificado:

$$\hat{u}_i^{\,2} = \gamma_0 + \gamma_1\, \widehat{ldesp_i} + \gamma_2\, \widehat{ldesp_i}^{\,2} + e_i$$

Estatística teste – $LM = nR^2_{\hat{u}^2} \sim \chi^2_{(2)}$

$LM_{obs} = 300 * 0.094 = 28.2$

$valor - p = P\left(\chi^2_{(2)} > 28.2\right) < 0.001$ então rejeita-se existência de homocedasticidade

Rejeitando-se a homocedasticidade, o estimador OLS dos βs conyinua a ser centrado e consistente, mas deixa de ser eficiente e, mais grave, os erros-padrão estão mal calculados. Assim, a inferência realizada sobre o modelo não é válida.