ESTATÍSTICA II – Economia/Finanças

Duração: 2h – Época de Recurso – Junho 2021

Nome:____ Nº:_____________

Espaço reservado para classificações

1.(15)	4a.(15)	5a.(10)	5e.(10)
2.(20)	4b.(15)	5b.(10)	5f.(15)
3.(20)	4c.(15)	5c.(10)	5g.(20)
	4d.(10)	5d.(15)	

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 concurso televisivo tem por objetivo conseguir um share de audiência de pelo menos 30%. Para verificar se o objetivo é cumprido, o canal de TV contratou uma empresa de audiometria que recolheu, à hora do concurso, uma amostra aleatória. Dos 300 inquiridos que estavam a ver televisão, 70 estavam a ver o concurso. Pensa que o objetivo estava a ser cumprido. Responda com base num teste adequado.

2. A associação de consumidores de determinado país decidiu testar se o nível esperado de colesterol dos consumidores da dieta A era inferior ou igual ao nível esperado de colesterol dos consumidores da dieta B. Para tal recolheu uma amostra de 9 indivíduos que consomem regularmente a dieta A, tendo observado um nível médio de colesterol de 197.6 e um desvio padrão corrigido de 56.3, e outra, independente da primeira, de 8 indivíduos que consomem regularmente a dieta B, tendo observado um nível médio de colesterol de 202.3 e um desvio padrão corrigido de 63.5. Assumindo que os níveis de colesterol em cada grupo têm distribuição normal com variâncias iguais, qual a conclusão a tirar ao nível de 5%.

3. Numa fábrica de doçarias produzem-se pacotes de gomas com 5 sabores diferentes: maçã, morango, cereja, limão e laranja. O produtor afirma que o número de gomas de cada sabor que vão para a secção de enchimento é o mesmo. Observada uma amostra aleatória de 1000 gomas, obteve-se:

Sabor	maçã	morango	cereja	limão	laranja
Nº de gomas	180	250	120	225	225

   Teste a afirmação do produtor ao nível de significância de 1%.

4. Seja X uma população com função densidade dada por $f(x|\mu) = \frac{128}{3}\ \mu^{-4}\ x^3\ e^{-4x/\mu}$, $x > 0$, da qual se recolheu uma amostra de dimensão n com o propósito de estimar μ. Sabe-se ainda que e $E(X) = \mu$ e que $var(X) = \mu^2/4$. A amostra recolhida, com 256 observações, originou $\bar{x} = 48$ e $s' = 35$.

   a. Mostre que o estimador de máxima verosimilhança para μ é a média da amostra.
   b. Mostre que o estimador de máxima verosimilhança é o mais eficiente.
      Recorda-se que $I_X(\mu) = -E\left(\frac{d^2}{d\mu^2}\ln f(X|\mu)\right)$.
   c. Obtenha uma estimativa de máxima verosimilhança para $var(X)$.
   d. Utilizando a melhor aproximação possível, obtenha um intervalo de confiança a 90% para μ.

5. Com o intuito de estudar os determinantes da taxa de criminalidade em cidades de média dimensão de determinado país, um criminologista definiu o modelo:

$$LTxC = \beta_0 + \beta_1\, LPD + \beta_2\, LPC + \beta_3\, LD + \beta_4\, PPJM + \beta_5\, TxD + u$$

Que se assume verificar as hipóteses habituais.

As variáveis têm o seguinte significado:

LTxC – Logaritmo da taxa de criminalidade (Nº de crimes participados por habitante) na cidade *i*.

LPD – Logaritmo da probabilidade de ser detido na cidade *i*.

LPC – Logaritmo da probabilidade de ser condenado, uma vez detido, na cidade *i*.

LD – logaritmo da densidade populacional (nº habitantes/km2) na cidade *i*.

PPJM – Proporção da população que corresponde a Jovens do sexo masculino na cidade *i*.

TxD – Taxa de desemprego (em % da população ativa) cidade *i*.

Estimado o modelo pelos mínimos quadrados, obteve-se

$$\widehat{LTxC} = \underset{(0.372)}{-5.212} - \underset{(0.108)}{0.415}\, LPD - \underset{(0.074)}{0.347}\, LPC + \underset{(0.076)}{0.214}\, LD + \underset{(1.703)}{2.048}\, PPJM + \underset{(0.025)}{0.065}\, TxD$$

$$n = 90 \qquad R^2 = 0.6306 \qquad \hat{\sigma} = 0.3433$$

a. Teste a significância global do modelo e a significância individual da variável *ppjm*.
b. Construa um intervalo de confiança a 90% para o parâmetro associado com a taxa de desemprego. Será que este intervalo lhe permite concluir sobre a significância estatística do parâmetro?
c. Teste a eliminação simultânea das variáveis *PPJM* e *LD* do modelo.
d. Mostre que, no teste de nulidade conjunta de q coeficientes, as estatísticas $F_1 = \frac{SSR_r - SSR_{ur}}{SSR_{ur}} \frac{n-k-1}{q}$ e $F_2 = \frac{R^2_{ur} - R^2_r}{1 - R^2_{ur}} \frac{n-k-1}{q}$ são iguais.
e. Assumindo que a variável residual tem distribuição normal, qual a **estimativa pontual** para a taxa de criminalidade numa cidade em que a probabilidade de ser detido é 0.3, a probabilidade de ser condenado, uma vez detido, é 0.4, a densidade populacional é 250, a proporção da população que corresponde a jovens do género masculino é 0.1 e a taxa de desemprego é 9%?
f. O criminologista defende que uma diminuição de 5 pontos percentuais na taxa de desemprego tem o mesmo efeito na taxa de criminalidade do que um aumento de 1% na probabilidade de detenção. Realize o teste ao nível de significância de 5%, sabendo que $cov(\hat{\beta}_1, \hat{\beta}_5) = -0.000226$.
g. Considere agora o modelo
   $LTxC = \beta_0 + \beta_1\, LPD + \beta_2\, TxD + u$ com $var(u|LPD, TxD) = \sigma^2\, TxD$
   i. Quais as consequências se estimarmos o modelo pelos mínimos quadrados?
   ii. Qual a solução que aponta neste caso concreto para corrigir eventuais problemas.

ANEXO

SUMMARY OUTPUT

Variável Dependente: $LTxC$

Regression Statistics	
Multiple R	0.762654501
R Square	0.581641888
Adjusted R Square	0.567048
Standard Error	0.361068689
Observations	90

ANOVA

	Df	*SS*	*MS*	*F*
Regression	3	15.58782748	5.195942493	39.85517099
Residual	86	11.21187147	0.130370598	
Total	89	26.79969894		

	Coefficients	*Standard Error*	*t Stat*
Intercept	-5.81748965	0.223479589	-26.0314138
LPD	-0.58471632	0.099910413	-5.85240622
LPC	-0.45366149	0.070004227	-6.48048712
TxD	0.11047214	0.018661279	5.919859367

RESOLUÇÃO

1.
$H_0: \theta \geq 0.3$ contra $H_1: \theta < 0.3$
Estatística de teste: $Z = \frac{\bar{X}-0.3}{\sqrt{0.3\times 0.7/300}} \dot{\sim}\, n(0;1)$

$Z_{obs} = \frac{(\frac{70}{300})-0.3}{\sqrt{0.3\times 0.7/300}} = -2.51 \qquad valor - p = P(Z \leq -2.51) = 0.006$

Rejeita-se claramente a hipótese nula, isto é, rejeita-se que o objetivo esteja a ser concluído.

2.
μ_1 valor esperado do nível de colesterol para quem toma habitualmente a dieta A
μ_2 valor esperado do nível de colesterol para quem toma habitualmente a dieta B

$H_0: \mu_1 \leq \mu_2$ contra $H_1: \mu_1 > \mu_2$ isto é $H_0: \mu_1 - \mu_2 \leq 0$ contra $H_1: \mu_1 - \mu_2 > 0$

Estatística de teste: $T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{1}{9}+\frac{1}{8}}\sqrt{\frac{8\,S'^2+7\,S'^2}{9+8-2}}} \sim t_{(15)}$

Região crítica a 5%: $W_{0.05} = \{t: t > 1.753\}$

$T_{obs} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{1}{9}+\frac{1}{8}}\sqrt{\frac{8\,s_1'^2+7\,s_2'^2}{9+8-2}}} = \frac{197.6-202.3}{\sqrt{\frac{1}{9}+\frac{1}{8}}\sqrt{\frac{8\times 56.3^2+7\times 63.5^2}{9+8-2}}} = -0.1519$

Como $T_{obs} \notin W$ não se rejeita H_0, isto é, não se rejeita que o nível esperado de colesterol dos consumidores da dieta A seja inferior ou igual ao de os consumidores da dieta B

3.
Seja p_j a probabilidade de uma goma de sabor j ir para a secção de enchimento ($j = 1,2,..,5$) com $j = 1$ para maçã, $j = 2$ para morango,… até $j = 5$ para laranja.
$H_0: p_j = 0.2$ para $j = 1,2,...,5$ contra $H_1: \exists\, j: p_j \neq 0.2$

Estatística de teste: $Q = \frac{\sum_{j=1}^{5}\left(N_j - E_j\right)^2}{E_j} \dot{\sim} \chi^2_{(4)}$

Região crítica a 1%: $W_{0.01} = \{q: q > 13.277\}$

$$Q_{obs} = \frac{(180-200)^2}{200} + \frac{(250-200)^2}{200} + \frac{(120-200)^2}{200} + \frac{(225-200)^2}{200} + \frac{(225-200)^2}{200}$$
$$= 2 + 12.5 + 32 + 3.125 + 3.125 = 52.75$$

Como $Q_{obs} \in W$ rejeita-se H_0, isto é, a afirmação do produtor não é aceitável.

4.
a)
$$\ell(\mu) = \sum_{j=1}^{n} \ln\left(\frac{128}{3}\,\mu^{-4}\,x_i^3\,e^{-4x_i/\mu}\right) = \sum_{j=1}^{n}\left(\ln\left(\frac{128}{3}\right) - 4\ln\mu + 3\ln x_i - \frac{4\,x_i}{\mu}\right)$$

$\ell'(\mu) = \sum_{j=1}^{n}\left(-\frac{4}{\mu} + \frac{4\,x_i}{\mu^2}\right) = -\frac{4n}{\mu} + \frac{4n\,\bar{x}}{\mu^2}$

$\ell'(\mu) = 0 \Leftrightarrow \frac{4n}{\mu} = \frac{4n\,\bar{x}}{\mu^2} \Leftrightarrow \mu = \bar{x}$ logo o estimador de MV será $\hat{\mu} = \bar{X}$

Já que $\ell''(\mu) = \frac{4n}{\mu^2} - \frac{8n\,\bar{x}}{\mu^3}$ e portanto $\ell''(\bar{x}) = \frac{4n}{\bar{x}^2} - \frac{8n\,\bar{x}}{\bar{x}^3} = -\frac{4n}{\bar{x}^2} < 0$ **ou**

Já que $\ell'(\mu) > 0$ para $\mu < \bar{x}$ e $\ell'(\mu) < 0$ para $\mu > \bar{x}$

b)
Sendo o EMV, $\hat{\mu}$, centrado, ($E(\hat{\mu}) = E(\overline{X}) = E(X) = \mu$) para ser o mais eficiente a sua variância terá de ser igual ao valor mínimo dado pela desigualdade de FCR, isto é, a $1/(n\, I_X(\mu))$.

$var(\hat{\mu}) = var(\overline{X}) = \frac{var(X)}{n} = \frac{\mu^2}{4\,n}$

$I_X(\mu) = -E\left(\frac{d^2}{d\,\mu^2}\ln f(X|\mu)\right) = -E\left(\frac{d}{d\,\mu}\left(-\frac{4}{\mu} + \frac{4\,X}{\mu^2}\right)\right) = -E\left(\frac{4}{\mu^2} - \frac{8\,X}{\mu^3}\right) = -\frac{4}{\mu^2} + \frac{8\,E(X)}{\mu^3} = \frac{4}{\mu^2}$

Logo $\frac{1}{n\,I(\mu)} = \frac{\mu^2}{4\,n}$

Assim $\hat{\mu}$ é o estimador mais eficiente para μ.

c)
Como $var(X) = \frac{\mu^2}{4}$, a estimativa de máxima verosimilhança virá (propriedade da invariância dos estimadores MV) $\widehat{var}(X) = \frac{\hat{\mu}^2}{4} = \frac{48^2}{4}$=576

d)

Melhor solução:

$$Z = \frac{\hat{\mu} - \mu}{\sqrt{var(\hat{\mu})}} = \frac{\hat{\mu} - \mu}{\mu/\sqrt{4n}} \mathrel{\dot{\sim}} n(0;1)$$

$-z_{\alpha/2} < \frac{\hat{\mu}-\mu}{\mu/\sqrt{4n}} < z_{\alpha/2} \Leftrightarrow -\frac{z_{\alpha/2}}{\sqrt{4n}} < \frac{\hat{\mu}}{\mu} - 1 < \frac{z_{\alpha/2}}{\sqrt{4n}}$ logo $\frac{\hat{\mu}}{1+\frac{z_{\alpha/2}}{\sqrt{4n}}} < \mu < \frac{\hat{\mu}}{1-\frac{z_{\alpha/2}}{\sqrt{4n}}}$ e portanto o IC vem (45.65; 50.60)

Segunda melhor solução:

$$Z = \frac{\hat{\mu} - \mu}{\sqrt{\widehat{var}(\hat{\mu})}} = \frac{\hat{\mu} - \mu}{\hat{\mu}\,/\sqrt{4n}} \mathrel{\dot{\sim}} n(0;1)$$

$-z_{\alpha/2} < \frac{\hat{\mu}-\mu}{\hat{\mu}/\sqrt{4n}} < z_{\alpha/2} \Leftrightarrow -\frac{z_{\alpha/2}}{\sqrt{4n}} < \frac{\hat{\mu}-\mu}{\hat{\mu}} < \frac{z_{\alpha/2}}{\sqrt{4n}} \Leftrightarrow -\frac{z_{\alpha/2}}{\sqrt{4n}}\hat{\mu} < \hat{\mu} - \mu < \frac{z_{\alpha/2}}{\sqrt{4n}}\hat{\mu}$ logo
$\hat{\mu} - \frac{z_{\alpha/2}}{\sqrt{4n}}\hat{\mu} < \mu < \hat{\mu} + \frac{z_{\alpha/2}}{\sqrt{4n}}\hat{\mu}$ e portanto o IC vem (45.53; 50.47)

Terceira solução possível (pior)

$$Z = \frac{\overline{X} - \mu}{S'/\sqrt{n}} \mathrel{\dot{\sim}} n(0;1)$$

Neste caso o IC vem dado por $\overline{x} \pm z_{\alpha/2}\frac{s'}{\sqrt{n}}$, isto é, (44.40; 51.60)

Nota: Tirando partido de a distribuição de X ser uma gama de parâmetros $\alpha = 4$ e $\lambda = 4/\mu$ poder-se-ia obter um IC com base na distribuição exata de $\hat{\mu} = \overline{X}$ já que $\frac{8\,n\,\overline{X}}{\mu} \sim \chi^2_{(8n)}$

O IC viria (45.63; 50.57)

5)
a)
Teste global: $H_0\colon \beta_j = 0$ para $j = 1,2,\ldots,5$ contra $H_1\colon \exists\, j\colon \beta_j \neq 0$

Estatística de teste: $F = \frac{R^2}{1-R^2}\,\frac{90-5-1}{5} \sim F(5;84)$

Região crítica a 5%: $W_{0.05} = \{f\colon f > 2.323\}$ (na tabela com 5 e 60 gl vem 2.37 ou 2.29 com 5; 120 gl)

$F_{obs} = \frac{0.6306}{1-0.6306} \times \frac{84}{5} = 28.68$ Rejeita-se claramente H_0 logo o modelo é globalmente significante

Teste individual: $H_0\colon \beta_4 = 0$ contra $H_1\colon \beta_4 \neq 0$

Estatística de teste: $T = \frac{\widehat{\beta}_4}{se(\widehat{\beta}_4)} \sim t(84)$

Região crítica a 5%: $W_{0.05} = \{t: t < -1.989 \; ou \; t > 1.989\}$ (na tabela com 80 gl vem 1.99)

$t_{obs} = \frac{2.048}{1.703} = 1.2026$ Não se rejeita H_0 logo a variável $PPJM$ **não é** significante

b)

IC a 90% para β_5

Variável fulcral $T = \frac{\hat{\beta}_5 - \beta_5}{se(\hat{\beta}_5)} \sim t(84)$

IC dado por $\hat{\beta}_5 \pm t_{0.05} \, se(\hat{\beta}_5)$ ou seja (com $t_{0.05} = 1.663$) vem (0.023; 0.107)

Sim, o facto de o intervalo não incluir o valor 0 significa que no teste de significância individual se rejeitaria a hipótese nula.

c)

Teste de nulidade conjunta de β_3 e de β_4: $H_0: \beta_3 = \beta_4 = 0$ contra $H_1: \beta_3 \neq 0 \; ou \; \beta_4 \neq 0$

Estatística de teste: $F = \frac{R^2}{1-R^2} \frac{90-5-1}{5} \sim F(5; 84)$

Região crítica a 5%: $W_{0.05} = \{f: f > 3.1052\}$ (na tabela com 2 e 60 gl vem 3.15)

$F_{obs} = \frac{0.6306 - 0.5816}{1 - 0.6306} \times \frac{84}{2} = 5.571$

Rejeita-se H_0, logo não de deve eliminar simultaneamente LD e $PPJM$.

d)

Como a variável dependente é a mesma no modelo com restrições e no modelo sem restrições, $SST_r = SST_{ur}$. Recordando que $R^2 = 1 - \frac{SSR}{SST}$ tem-se

$$\frac{R_{ur}^2 - R_r^2}{1 - R_{ur}^2} = \frac{\left(1 - \frac{SSR_{ur}}{SST_{ur}}\right) - \left(1 - \frac{SSR_r}{SST_r}\right)}{1 - \left(1 - \frac{SSR_{ur}}{SST_{ur}}\right)} = \frac{\frac{SSR_r - SSR_{ur}}{SST_{ur}}}{\frac{SSR_{ur}}{SST_{ur}}} = \frac{SSR_r - SSR_{ur}}{SSR_{ur}}$$

E portanto as estatísticas serão iguais.

e)

$$\begin{aligned}\widehat{LTxC}_0 &= -5.212 - 0.415 \ln 0.3 - 0.347 \ln 0.4 + 0.214 \ln 250 + 2.048 \times 0.1 + 0.065 \times 9 \\ &= -2.423\end{aligned}$$

Assim

$$\widehat{TxC}_0 = \exp\left(\widehat{LTxC}_0 + \frac{\hat{\sigma}^2}{2}\right) = \exp\left(-2.423 + \frac{0.3433^2}{2}\right) = 0.094$$

f)

A teoria do criminologista traduz-se em $-5 \, \beta_5 = \beta_1$

Teste individual: $H_0: \beta_1 + 5 \, \beta_5 = 0$ contra $H_1: \beta_1 + 5 \, \beta_5 \neq 0$

Estatística de teste: $T = \frac{\hat{\beta}_1 + 5 \, \hat{\beta}_5}{se(\hat{\beta}_1 + 5 \, \hat{\beta}_5)} \sim t(84)$

Região crítica a 5%: $W_{0.05} = \{t: t < -1.989 \; ou \; t > 1.989\}$ (na tabela com 80 gl vem 1.99)

$$\begin{aligned}\widehat{var}(\hat{\beta}_1 + 5\hat{\beta}_5) &= \widehat{var}(\hat{\beta}_1) + 25 \, \widehat{var}(\hat{\beta}_5) + 10 \, \widehat{cov}(\hat{\beta}_1, \hat{\beta}_5) \\ &= 0.108^2 + 25 \times 0.025^2 + 10 \times (-0.000226) = 0.02503\end{aligned}$$

$t_{obs} = \frac{-0.415 + 5 \times 0.065}{\sqrt{0.02503}} = \frac{-0.09}{0.1582} = -0.569$

Não se rejeita H_0 logo não se rejeita a teoria do criminologista.

g)

Parte 1 -> Os estimadores dos coeficientes da regressão continuam a ser centrados e consistentes mas deixam de ser eficientes. Os erros-padrão passam a estar incorretamente estimados e a inferência estatística deixa de ser válida.

Parte 2 -> dividir todas as variáveis (incluindo o termo independente) por raiz de TxD.

$$\frac{LTxC}{\sqrt{TxD}} = \frac{\beta_0}{\sqrt{TxD}} + \beta_1 \frac{LPD}{\sqrt{TxD}} + \beta_2 \frac{TxD}{\sqrt{TxD}} + \frac{u}{\sqrt{TxD}}$$

O modelo transformado sendo homocedástico, pode-se estimar os parâmetros recorrendo aos mínimos quadrados.