

Nome: \_\_\_\_\_ Número: \_\_\_\_\_

Cotação:

Pergunta	1	2	3	4a	4b	5a	5b	6a	6b	7a	7b	8a	8b	9a	9b
Cotação Total	1.5	1.5	2	1	1.5	1	1.5	1	1.5	1	1.5	1	1.5	1	1.5
Cotação Atribuída															

Atenção:

- Justifique detalhadamente as respostas.
- As respostas de escolha múltipla bem como as questões de verdadeiro ou falso não são penalizadas quando as respostas estão erradas.

1. Classifique as seguintes afirmações como verdadeiras (V) ou falsas (F). Cada resposta correta vale 0.25.

- |  | V                                   | F                                   |
|--|-------------------------------------|-------------------------------------|
| (a) Quando num teste de hipóteses com $\alpha = 0.01$ se obtém um valor-p de 0.00025, rejeita-se $H_0$ .   | <input checked="" type="checkbox"/> | <input type="checkbox"/>            |
| (b) No modelo de regressão linear $E(y_t   x_2) = \beta_1 + \beta_2 x_{t2}$ , a elasticidade de $y$ em relação a $x_2$ é constante.  | <input type="checkbox"/>            | <input checked="" type="checkbox"/> |
| (c) A eficiência, tal como foi vista no curso, é uma propriedade que só se aplica a estimadores centrados.   | <input checked="" type="checkbox"/> | <input type="checkbox"/>            |
| (d) Se num processo repetitivo se forem construindo intervalos de confiança para $\theta$ com base em amostras casuais simples independentes umas das outras e todas da mesma dimensão, é legítimo concluir que a proporção de intervalos que vão incluir $\theta$ se vai aproximando do nível de confiança escolhido. | <input checked="" type="checkbox"/> | <input type="checkbox"/>            |
| (e) O modelo $z_t = \exp(\alpha_1 + \alpha_2 x_{t2} + \alpha_3 x_{t3} + \alpha_4 x_{t3}^2) \times \nu_t^{\alpha_5}$ , com $t = 1, 2, \dots, n$ , é linearizável.   | <input checked="" type="checkbox"/> | <input type="checkbox"/>            |
| (f) A potência de um teste é igual ao seu valor-p.   | <input type="checkbox"/>            | <input checked="" type="checkbox"/> |

2. Escolha a opção correta em cada uma das seguintes questões de escolha múltipla. Cada resposta correta vale 0.5 valores.

a) A desigualdade de Fréchet-Cramér-Rao (FCR) é um resultado teórico importante no estudo:

Do modelo de regressão linear

Do enviesamento de um estimador

Dos testes de hipóteses

Da eficiência de um estimador

b) Considere o MRL  $y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + \beta_4 x_{t4} + u_t$  e assuma que verifica as hipóteses habituais. Para testar

$$H_0 : \beta_2 + 3\beta_3 = 0 \quad \text{contra} \quad H_1 : \beta_2 + 3\beta_3 \neq 0,$$

utilizando o procedimento apresentado no curso, qual das regressões auxiliares se deve utilizar?

$y_t = \beta_1 + \beta_3(x_{t3} - 3x_{t2}) + \beta_4 x_{t4} + u_t$

$y_t = \beta_1 + \beta_3(x_{t3} + 3x_{t2}) + \beta_4 x_{t4} + u_t$

$y_t = \beta_1 + \beta_3(x_{t3} - 3x_{t2} + x_{t4}) + u_t$

$y_t = \beta_1 + \beta_3(x_{t3} - 3x_{t2} - x_{t4}) + u_t$

c) A amplitude de um intervalo de confiança para a média de um universo normal de variância conhecida, construído com base no processo habitual, pode ser reduzida por:

Uma diminuição da dimensão da amostra

Uma diminuição do nível de confiança

As opções acima estão ambas corretas

Nenhuma das alternativas é correta

3. Seja  $X$  um universo para o qual existe média  $\mu$  e variância  $\sigma^2$ . Com base numa amostra aleatória de dimensão  $n$ , definiu-se  $T = \frac{1}{n-1} \sum_{i=1}^n X_i$  como estimador para  $\mu$ . Mostre que o erro quadrático médio (EQM) de  $T$  é dado por  $\frac{\mu^2 + n\sigma^2}{(n-1)^2}$ . Mostre também que o estimador  $T$  é consistente em média quadrática.

**Nota:** não se esqueça de definir o conceito de EQM e de justificar os passos que for dando.

---

### RESPOSTA 3

O EQM de um estimador  $T$  de um parâmetro  $\mu$  é definido por:

$$\text{EQM}(T) = \mathbb{E} [(T - \mu)^2] = (E[T - \mu])^2 + \text{Var}(T).$$

No nosso caso, como  $T = \frac{1}{n-1} \sum_{i=1}^n X_i$  temos que

$$\mathbb{E}[T] = \frac{n}{n-1} E[X] = \frac{n}{n-1} \mu \quad \text{e} \quad \mathbb{E}[T] - \mu = \frac{\mu}{n-1}$$

e

$$\text{Var}(T) = n \left( \frac{\sigma}{n-1} \right)^2$$

Consequentemente,

$$\text{EQM}(T) = \mathbb{E} [(T - \mu)^2] = \left( \frac{\mu}{n-1} \right)^2 + n \left( \frac{\sigma}{n-1} \right)^2 = \frac{\mu^2 + n\sigma^2}{(n-1)^2}$$

O estimador é consistente porque

$$\lim_{n \rightarrow +\infty} \frac{\mu^2 + n\sigma^2}{(n-1)^2} = 0.$$

4. Seja  $X \sim \text{Gama}(4, \lambda)$ , com função densidade de probabilidade:

$$f(x|\lambda) = \frac{\lambda^4}{6} x^3 e^{-\lambda x}, \quad x > 0.$$

Suponha-se uma amostra aleatória de dimensão  $n$ .

a) Qual das seguintes variáveis é uma variável fulcral?

A estatística  $T_1 = \sum_{i=1}^n X_i$ , com distribuição Gama( $4n, \lambda$ )

A estatística  $T_2 = \frac{\bar{X}}{\lambda}$ , com distribuição Gama( $4n, \lambda^2$ )

A estatística  $T_3 = 2n\lambda\bar{X}$ , com distribuição  $\chi^2(8n)$

A estatística  $T_1 = \bar{X}$ , com distribuição Gama( $4n, n\lambda$ )

b) Calcule a estimativa de máxima verosimilhança de  $\lambda$ , sabendo que a média amostral é  $\bar{x} = 100$ .

### RESPOSTA 4.b)

A função de verosimilhança é dada por

$$L(\lambda) = \prod_{i=1}^n f(x_i|\lambda) = \prod_{i=1}^n \left( \frac{\lambda^4}{6} x_i^3 e^{-\lambda x_i} \right).$$

Consequentemente, a função de log-verosimilhança é dada por

$$\begin{aligned} \ell(\lambda) = \ln L(\lambda) &= \sum_{i=1}^n \ln \left( \frac{\lambda^4}{6} x_i^3 e^{-\lambda x_i} \right) = \sum_{i=1}^n (-\ln 6 + 4 \ln \lambda + 3 \ln x_i - \lambda x_i) \\ &= -n \ln 6 + 4n \ln \lambda + \sum_{i=1}^n 3 \ln x_i - \lambda \sum_{i=1}^n x_i \end{aligned}$$

A derivada da função log-verosimilhança em ordem a  $\lambda$  é dada por

$$\ell'(\lambda) = \frac{4n}{\lambda} - \sum_{i=1}^n x_i.$$

Os seus pontos críticos são portanto obtidos:

$$\ell'(\lambda) = 0 \Rightarrow \frac{4n}{\lambda} = \sum_{i=1}^n x_i \Rightarrow \lambda = \frac{4n}{\sum x_i} = \frac{4}{\bar{x}}$$

Como a média amostral verifica  $\bar{x} = 100$ , temos:

$$\hat{\lambda} = \frac{4}{100} = 0.04$$

Para sabermos se  $\hat{\lambda}$  representa realmente um máximo podemos verificar a concavidade da função:

$$\ell''(\lambda) = -\frac{4n}{\lambda^2} < 0 \Rightarrow \text{Máximo global}$$

Conclusão: A estimativa de máxima verosimilhança para  $\lambda$  é  $\hat{\lambda} = 0.04$ .

5. Um hospital público está a monitorizar o número médio de quilómetros que os seus veículos de transporte de doentes percorrem diariamente. De acordo com as previsões orçamentais em vigor, este valor não deverá ultrapassar 275 km por dia. Para averiguar se as previsões orçamentais ainda se mantêm adequadas, foi recolhida uma amostra aleatória de 500 dias de atividade dos veículos hospitalares, tendo-se obtido uma média amostral de 278.5 km e uma variância amostral corrigida de 6430.5.

a) Qual das seguintes opções é **falsa**?

i)  $(\frac{1}{500} \sum_{i=1}^{500} x_i)^2 = 77562.25$

ii)  $\frac{1}{500} \sum_{i=1}^{500} x_i^2 = 83979.83$

iii)  $s^2 = 6417.64$

iv)  $\sum_{i=1}^{500} x_i = 137500$

b) Recorra a um teste de hipóteses de dimensão  $\alpha = 0.05$  para avaliar se é de proceder à revisão dos montantes orçamentados para esta atividade.

**RESPOSTA 5.b)** Hipóteses:

$$H_0 : \mu \leq 275 \quad (\text{n\~{a}o \acute{e} necess\~{a}rio rever o or\~{c}amento)}$$

contra

$$H_1 : \mu > 275 \quad (\text{\'{e} necess\~{a}rio rever o or\~{c}amento)}$$

Dado termos uma amostra aleatória de grande dimensão, a estatística de teste é dada por

$$Z = \frac{\bar{X} - 275}{\sqrt{\frac{S'^2}{n}}} \underset{a}{\sim} N(0, 1)$$

O  $z_{obs}$  é dado por

$$z_{obs} = \frac{\bar{x} - 275}{\sqrt{\frac{s'^2}{n}}} = \frac{278.5 - 275}{\sqrt{\frac{6430.5}{500}}} \approx 0.976$$

A região de rejeição é dada por  $W : \{z : z > 1.645\}$  onde 1.645 verifica  $P(Z > 1.645) = 0.05$ . Como  $0.976 \notin W$  não rejeitamos  $H_0$  e portanto, não há evidência estatística para revermos o orçamento com base nesta amostra.

6. Uma consultora de investimentos pretende analisar os lucros anuais obtidos por startups em dois segmentos distintos: empresas que operam através de modelos de subscrição e empresas que utilizam modelos de receita baseada em publicidade. Para isso, foram selecionadas duas amostras casuais, cada uma com 16 empresas de cada segmento. Para cada grupo, calcularam-se a média amostral e a variância corrigida dos lucros líquidos anuais, tendo-se obtido os seguintes resultados:

$$\text{Modelo de Subscrição: } \bar{x}_1 = 0.215 \quad s_1'^2 = 0.024$$

$$\text{Modelo Publicitário: } \bar{x}_2 = 0.185 \quad s_2'^2 = 0.018$$

Admita-se que os lucros anuais, expressos em milhões de euros, em ambos os segmentos seguem uma distribuição normal com variâncias iguais.

a) Considere o teste de hipóteses  $H_0 : \sigma_1^2 = 0.025$  contra  $H_1 : \sigma_1^2 > 0.025$  ao nível de significância de 5%, onde  $\sigma_1^2$  é a variância da variável aleatória que representa os lucros anuais de uma startup que utiliza o modelo de subscrição. Qual das seguintes opções mostra os quantis relevantes para construir o teste de hipóteses?

i) 6.262 e 27.488

ii) 7.261

iii) 24.996

iv) 7.261 e 24.996

b) Construa um intervalo de confiança a 95% para a diferença entre a média dos lucros líquidos dos dois modelos de negócio. O que pode concluir a consultora?

### RESPOSTA 6.b)

**Distribuições normais** — Intervalo de confiança para  $\mu_1 - \mu_2$ , quando as variâncias são desconhecidas mas iguais.

A variável fulcral para a construção de um intervalo de confiança para  $\mu_1 - \mu_2$  é dada por

$$\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{1}{m} + \frac{1}{n}\right) \left(\frac{(m-1)S_1'^2 + (n-1)S_2'^2}{m+n-2}\right)}} \sim t(m+n-2)$$

O intervalo de confiança aleatório é então dado por

$$\left( (\bar{X}_1 - \bar{X}_2) - t_{\alpha/2} \cdot S^*, (\bar{X}_1 - \bar{X}_2) + t_{\alpha/2} \cdot S^* \right) \quad \text{com} \quad S^* = \sqrt{\left(\frac{1}{m} + \frac{1}{n}\right) \left(\frac{(m-1)S_1'^2 + (n-1)S_2'^2}{m+n-2}\right)}$$

No nosso problema temos que  $X_1$  (resp.,  $X_2$ ) representa os lucros anuais em milhões de euros para uma empresa que utiliza modelos de subscrição (resp., modelo publicitário). Assim,

$$m = n = 16, \quad \bar{x}_1 - \bar{x}_2 = 0.215 - 0.185 = 0.03 \quad e \quad s^* = \sqrt{\left(\frac{1}{16} + \frac{1}{16}\right) \left(\frac{15 \cdot 0.024 + 15 \cdot 0.018}{30}\right)} = 0.05123$$

Dado que o quantil da  $t$  é igual a

$$t_{0.025}(30) = 2.042$$

obtemos

$$(0.03 - 2.042 \cdot 0.05123, 0.03 + 2.042 \cdot 0.05123) = (0.03 - 0.1046; 0.03 + 0.1046) = (-0.0746; 0.1346)$$

Conclusão: Como o intervalo contém o valor zero, não se pode concluir que exista diferença significativa entre os resultados médios dos dois grupos, com um nível de confiança de 95%.

**VSFF**

7. Com o objetivo de apoiar o planeamento de infraestruturas rodoviárias, a Direção Nacional de Vias de Comunicação de um país europeu decidiu estudar o volume médio mensal de tráfego automóvel como função da densidade populacional da região, do número de faixas de rodagem e da classificação funcional da estrada. Consideraram-se quatro níveis de classificação: Local, Regional, Itinerário Complementar (IC) e Itinerário Principal (IP). Foi então definido o seguinte modelo base, assumindo que se verificam as hipóteses do Modelo de Regressão Linear:

$$NV_t = \beta_1 + \beta_2 POP_t + \beta_3 FAIXAS_t + \beta_4 REG_t + \beta_5 IC_t + \beta_6 IP_t + u_t$$

em que:  $NV_t$  representa o número de veículos por mês numa determinada via ( $t$ ), em milhares;  $POP_t$  representa a população residente da região onde a estrada está inserida (em milhares de habitantes);  $FAIXAS_t$  representa o número de faixas de rodagem da estrada;  $REG_t$  representa a variável dummy que assume o valor 1 se a estrada é Regional, 0 caso contrário;  $IC_t$  representa a variável dummy que assume o valor 1 se a estrada é Itinerário Complementar, 0 caso contrário e  $IP_t$  representa a variável dummy que assume o valor 1 se a estrada é Itinerário Principal, 0 caso contrário. Foi recolhida uma amostra de 126 estradas distribuídas por diferentes regiões do país, tendo-se ajustado o modelo descrito acima aos dados observados. Obteve-se  $R^2 = 0.75130880$  e  $s = 10.16973$ .

Variável	Coefficiente	Erro-padrão	Estatística t	Valor-p
Intercepto	-2.2010	3.8746	-0.5681	0.5710
POP	0.0460	0.0172	2.6698	0.0086
FAIXAS	12.5107	1.4998	8.3417	< 0.0001
REG	1.9217	2.4300	0.7908	0.4306
IC	5.4977	2.8903	1.9021	0.0596
IP	38.3286	2.4267	15.7945	< 0.0001

- a) Considere o teste  $H_0 : \beta_5 = 0$  contra  $H_1 : \beta_5 \neq 0$  ao nível de significância de 5%. Seja  $T$  a estatística de teste quando  $H_0$  é verdadeira. Qual das seguintes opções é verdadeira?

- i) Rejeita-se  $H_0$   ii)  $P(T > 1.9021) = 0.0596$   
 iii)  $P(|T| > 1.9021) = 0.05$   iv)  $P(T > 1.9021) = 0.0298$

- b) Teste a significância global da regressão ao nível de significância de 5%.

**RESPOSTA 7.b)** Pretendemos testar as seguintes hipóteses:

$$H_0 : \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0 \quad \text{vs} \quad H_1 : \exists j : \beta_j \neq 0, \quad j = 2, 3, \dots, 6$$

Estatística de teste:

$$F = \frac{R^2/5}{(1 - R^2)/120} \sim F(5; 120)$$

Estatística observada:

$$f_{obs} = \frac{0.75130880/5}{(1 - 0.75130880)/120} = 72.50522$$

Região de rejeição:

$$W = \{f : f > 2.29\}$$

onde 2.29 é tal que  $P(F > 2.29) = 0.05$ . Devemos rejeitar a hipótese nula pois  $f_{obs} \in W$ , o que significa que o modelo é globalmente significativo.

**VSFF**

8. Considere o enunciado e todos os dados apresentados na Pergunta 7. Responda às seguintes questões.

a) Com base no modelo de regressão linear múltiplo ajustado para explicar o volume médio mensal de tráfego ( $NV_t$ ), pretende-se construir um intervalo de confiança a 95% para o número médio de veículos por mês numa estrada com as seguintes características:

- População residente: 20000 residentes;
- Número de faixas: 2;
- Classificação da estrada: Itinerário Complementar.

Para construir o intervalo de confiança a 95%, procedem-se a vários cálculos intermédios. Alguns deles encontram-se nas opções abaixo. Assuma que  $s_{NV} = s\sqrt{c(X^T X)^{-1}c^T} = 2.4609794$ . Qual das seguintes opções é **falsa**?

i)  $\hat{NV} = 29.2381$

ii)  $t_{0.025} = 1.98$

iii)  $IC = (9.102115, 49.37423)$

iv)  $s_d = 10.46326$

b) Para testar a hipótese conjunta:  $H_0 : \beta_4 = 0 \wedge \beta_6 = 6\beta_5$  vs  $H_1 : \beta_4 \neq 0 \vee \beta_6 \neq 6\beta_5$  ao nível de significância de 1%, foi estimado uma regressão auxiliar para a qual se obteve  $VR = 13024.5687328$ . Descreva a regressão auxiliar que foi estimada e realize o teste de hipóteses. Interprete o resultado deste teste de hipóteses.

### RESPOSTA 8.b)

A regressão a ser estimada é dada por

$$NV_t = \beta_1 + \beta_2 POP_t + \beta_3 FAIXAS_t + \beta_5 (IC_t + 6IP_t) + u_t$$

Pare realizar o teste de hipóteses  $H_0 : \beta_4 = 0 \wedge \beta_6 = 6\beta_5$  vs  $H_1 : \beta_4 \neq 0 \vee \beta_6 \neq 6\beta_5$ , precisamos da estatística de teste:

$$F = \frac{\frac{VR_0 - VR_1}{2}}{\frac{VR_1}{120}} \sim F(2; 120)$$

Dado que  $VR_1 = 10.16973^2 \times 120 = 12410.81$ , o valor observado da estatística é dado por:

$$f_{obs} = \frac{(13024.5687328 - 12410.81)/2}{12410.81/120} = 2.967278$$

A região de rejeição é dada por  $W = \{f : f > 4.79\}$ . Como  $f_{obs} \notin W$  não podemos rejeitar a hipótese nula ao nível de significância de 1%, ou seja não podemos rejeitar a hipótese de que o impacto de uma estrada regional é igual ao de uma estrada local e que o efeito marginal de uma estrada do tipo IP é seis vezes superior ao de uma estrada tipo IC no termo constante.

9. Um economista está a estudar os fatores que influenciam o logaritmo das receitas anuais por trabalhador,  $\ln(\text{produt}_t)$  em pequenas empresas de comércio eletrónico. Para isso, estimou um modelo de regressão linear múltiplo com base numa amostra de 209 empresas, usando as seguintes variáveis explicativas:  $\ln(\text{vendas}_t)$ , que representa o logaritmo das vendas anuais totais,  $\text{rcp}_t$ , que representa a percentagem de receita proveniente de clientes premium,  $\text{ra}_t$ , que representa o investimento anual médio em publicidade digital (em milhares de euros) e  $\text{servpub}_t$  que é uma dummy que assume valor 1 se a empresa presta exclusivamente serviços públicos online (ex: portais governamentais), 0 caso contrário. Verificou-se também na estimação do modelo que  $VR = 44.0535$ .

Variável	Coefficiente	Erro-padrão	t estat.	Valor-p
Intercepto	4.6283	0.3126	14.8060	0.0000
$\ln(\text{vendas})$	0.2614	0.0343	7.6292	0.0000
$\text{rcp}$	0.0119	0.0042	2.8645	0.0046
$\text{ra}$	0.0003	0.0005	0.4804	0.6314
$\text{servpub}$	-0.3798	0.0904	-4.1989	0.0000

a) Qual das seguintes afirmações é verdadeira?

- Todas as variáveis são isoladamente estatisticamente significativas ao nível de significância de 10%.
- A variável  $\text{ra}$  é estatisticamente significativa ao nível de significância de 10%.
- A variável  $\text{servpub}$  é estatisticamente significativa ao nível de significância de 5%.
- A variável  $\text{rcp}$  não é estatisticamente significativa ao nível de significância de 1%.

b) Um analista quer avaliar se o modelo que explica o  $\ln(\text{produt}_t)$  é estruturalmente diferente consoante as empresas de comércio eletrónico operem apenas com serviços próprios ou através de plataformas de terceiros. Para isso, dividiu a amostra em dois grupos: (i) empresas com modelo próprio (número de observações: 171) e (ii) empresas que operam via plataformas de terceiros (número de observações: 38). Para ambos os grupos, ajustou-se o mesmo modelo de regressão acima descrito. As somas dos resíduos quadrados (VR) obtidas foram:

$$VR = 36.04 \quad (\text{modelo próprio}) \quad VR = 6.45 \quad (\text{plataformas})$$

Formalize e resolva o teste estatístico com dimensão de 5%. Comente.

### RESPOSTA 9.b)

Deve ser aplicado um teste de Chow, ou seja, um teste de alteração estrutural:

$$H_0 : \beta_{j1} = \beta_{j2} \quad \forall j = 1, \dots, 5 \quad \text{vs} \quad H_1 : \exists j : \beta_{j1} \neq \beta_{j2}$$

onde os índices 1 e 2 representam cada um dos grupos. A estatística de teste é dada por

$$F = \frac{(VR_0 - VR_1)/k}{VR_1/(n - 2k)} \sim F(k, n - 2k)$$

$$f_{obs} = \frac{(44.0535 - (36.04 + 6.45))/5}{42.49/(209 - 10)} = \frac{0.3127}{0.2135} \approx 1.4645$$

Comparando com o valor crítico  $F_{5,199,0.05} \approx 2.21$ , temos:

$$f_{obs} = 1.4645 < 2.21 \Rightarrow \text{N\~{a}o se rejeita } H_0$$

Conclus\~{a}o: N\~{a}o h\~{a} evid\~{e}ncia estat\~{i}stica de que o modelo explicativo do *lproduct* difere entre empresas que usam modelo pr\~{o}prio e as que operam via plataformas. N\~{a}o existe altera\~{c}\~{a}o de estrutura.

**FIM**