Probabilidade e Estatística

LEAer, LEBiom, LEMec

2º Semestre – 2023/2024
18/06/2024
18:00–20:00

Duração: **120** minutos

Exame Época Normal – C

- As questões de escolha múltipla têm uma e apenas uma resposta correta. Caso assinale uma resposta errada ou mais de uma resposta, serão descontados 0.5 valores.
- As respostas às questões de escolha múltiplas têm de ser assinaladas na folha de instruções previamente distribuída.
- Para as restantes questões, terá de justificar convenientemente as suas respostas.
- Só pode sair da sala uma hora após o início do exame, devendo nesse caso entregar a sua prova ou desistir da mesma.

Pergunta 1 2 valores

Admita que a probabilidade de uma mulher selecionada ao acaso ter cancro da mama é 0.05. A mamografia é um teste de diagnóstico do cancro da mama com sensibilidade e especificidade iguais a 0.85 e 0.90, respetivamente. Relembramos que: a sensibilidade é a probabilidade de o resultado do teste ser positivo, quando aplicado a uma paciente com cancro da mama; a especificidade é a probabilidade de o resultado do teste ser negativo, quando aplicado a uma paciente sem cancro da mama.

Calcule a probabilidade de uma paciente ter cancro da mama, sabendo que o resultado da sua mamografia foi positivo.

- **Quadro de acontecimentos e probabilidades**

Acontecimento	**Probabilidade**	
P = resultado da mamografia é positivo		
$\bar{P}$ = resultado da mamografia é negativo		
C = a paciente ter cancro da mama	$P(C) = 0.05$	
	$P(P \mid C) = 0.85$	(sensibilidade)
	$P(\bar{P} \mid \bar{C}) = 0.90$	(especificidade)

- **Probabilidade pedida**

  Pelo Teorema de Bayes,

$$\begin{aligned}
P(C \mid P) &= \frac{P(P \mid C) \times P(C)}{P(P)} \\
&= \frac{P(P \mid C) \times P(C)}{P(P \mid C) \times P(C) + P(P \mid \bar{C}) \times P(\bar{C})} \\
&= \frac{P(P \mid C) \times P(C)}{P(P \mid C) \times P(C) + [1 - P(\bar{P} \mid \bar{C})] \times P(\bar{C})} \\
&= \frac{0.85 \times 0.05}{0.85 \times 0.05 + (1-0.90) \times (1-0.05)} = \frac{0.0425}{0.1375} \\
&\simeq 0.309091.
\end{aligned}$$

Pergunta 2 2 valores

Um estudante arranjou um trabalho de verão em que tem de contactar telefonicamente ex-alunos da sua universidade. Diariamente, o estudante tem como objetivo contactar com sucesso 10 ex-alunos. O trabalho feito em anos anteriores indica que a probabilidade de um contacto ser bem sucedido é 0.3. Seja X a variável aleatória que representa o número de telefonemas a fazer num dia até contactar com sucesso 10 ex-alunos. X possui função de probabilidade dada por $P(X=x)=\binom{x-1}{10-1}\,0.3^{10}\,(1-0.3)^{x-10}$, para $x=10,11,\ldots$

Calcule a probabilidade de o estudante ter de ligar a mais de 14 ex-alunos num dia para atingir o seu objetivo.

- **V.a. e f.p.**

  $X=$ número de telefonemas a fazer num dia até estabelecer com sucesso 10 contactos

  $P(X=x)=\binom{x-1}{10-1}\,0.3^{10}\,(1-0.3)^{x-10}, \quad x=10,11,\ldots$

- **Probabilidade pedida**

$$P(X>14) \;=\; 1-P(X\leq 14)=1-\sum_{x=10}^{14}\binom{x-1}{10-1}0.3^{10}\,(1-0.3)^{x-10} \overset{calc.}{\simeq} 0.998334.$$

  [Alternativamente, o evento $\{X>4\}$ é equivalente ao acontecimento $\{Y<10\}$ em que Y representa o número de contactos bem sucedidos nos primeiros 14 telefonemas e $Y\sim \text{binomial}(14,0.3)$. Assim, $P(X>14)=P(Y<10)=P(Y\leq 9)=F_{binomial(14,0.3)}(9)\overset{tabelas,calc.}{\simeq} 0.9983.$]

Pergunta 3	2 valores

Considere uma variável aleatória X com distribuição exponencial de valor esperado $1/3$.

Calcule a variância de $Y=e^X$.

- **V.a. e distribuição**

  $X\sim \text{exponencial}(\lambda)$

  $\lambda: E(X)=1/\lambda=1/3 \Leftrightarrow \lambda=3$

- **Momento de ordem k ($k=1,2$) de $Y=e^X$**

$$\begin{aligned} E(Y^k) &= \int_0^{+\infty}(e^x)^k\times 3e^{-3x}\,dx \\ &= \int_0^{+\infty}3e^{-(3-k)x}\,dx \\ &= \frac{3}{3-k}\int_0^{+\infty}(3-k)\times e^{-(3-k)x}\,dx \\ &= \frac{3}{3-k}\int_0^{+\infty}f_{exponencial(3-k)}(x)\,dx \\ &= \frac{3}{3-k} \end{aligned}$$

- **Variância de Y**

$$V(Y) \;=\; E(Y^2)-E^2(Y)=\frac{3}{3-2}-\left(\frac{3}{3-1}\right)^2=3-\frac{9}{4}=\frac{3}{4}.$$

Pergunta 4	2 valores

Suponha que o par aleatório (X,Y) possui função de probabilidade conjunta dada pela tabela à direita.

Calcule o coeficiente de correlação entre X e Y e interprete o seu valor.

	Y		
X	0	1	2
0	3/8	3/16	1/4
1	1/8	1/16	0

- **Par aleatório**

  (X,Y)

- **Correlação pedida**

$$corr(X,Y) = \frac{cov(X,Y)}{\sqrt{V(X)\,V(Y)}} = \frac{E(XY)-E(X)\times E(Y)}{\sqrt{V(X)\,V(Y)}}.$$

Logo são necessários alguns cálculos auxiliares que envolverão a f.p. conjunta de (X,Y) e as f.p. marginais de X e Y dadas por $P(X=x)=\sum_y P(X=x,Y=y)$ e $P(Y=y)=\sum_x P(X=x,Y=y)$.

	Y			
X	0	1	2	$P(X=x)$
0	3/8	3/16	1/4	13/16
1	1/8	1/16	0	3/16
$P(Y=y)$	8/16	4/16	4/16	1

- **Valor esperado e variância de X**

$$\begin{aligned} E(X) &= \sum_{x=0}^{1} x\,P(X=x) = 0\times 13/16 + 1\times 3/16 = 3/16; \\ V(X) &= E(X^2)-E^2(X) \\ &= \sum_{x=0}^{2} x^2\,P(X=x) - E^2(X) = \left(0^2\times 13/16 + 1^2\times 3/16\right) - (3/16)^2 = 3/16 - 9/256 = 39/256 \end{aligned}$$

- **Valor esperado e variância de Y**

$$\begin{aligned} E(Y) &= \sum_{y=0}^{2} y\,P(Y=y) = 0\times 8/16 + 1\times 4/16 + 2\times 4/16 = 3/4; \\ V(Y) &= E(Y^2)-E^2(Y) \\ &= \sum_{y=0}^{2} y^2\,P(Y=y) - E^2(Y) = \left(0^2\times 8/16 + 1^2\times 4/16 + 2^2\times 4/16\right) - (3/4)^2 = 20/16 - 9/16 = 11/16 \end{aligned}$$

- **Valor esperado de XY**

$$E(XY) = \sum_{x=0}^{1}\sum_{y=0}^{2} x\,y\,P(X=x,\,Y=y) = 1\times 1\times 1/16 = 1/16.$$

- **Covariância entre X e Y**

$$cov(X,Y) = E(XY)-E(X)\times E(Y) = \frac{1}{16} - \frac{3}{16}\times\frac{3}{4} = -\frac{5}{64}$$

- **Correlação pedida** (cont.)

$$corr(X,Y) = \frac{cov(X,Y)}{\sqrt{V(X)\,V(Y)}} = \frac{-5/64}{\sqrt{(39/256)\times(11/16)}} \simeq -0.241402.$$

- **Comentários**
  - [É sabido que: caso X e Y sejam v.a. independentes, então $corr(X,Y)=0$.]
    Uma vez que $corr(X,Y)\neq 0$, concluímos que X e Y são v.a. dependentes.
  - Dado que $corr(X,Y)<0$, podemos adiantar que X e Y tenderão a variar em sentidos opostos relativamente aos respectivos valores esperados.
  - Como $|corr(X,Y)|\simeq 0.24$ dista bastante de 1, podemos afirmar que as v.a. estão fracamente correlacionadas. [Importa notar que o coeficiente de correlação quantifica a associação linear entre as v.a. X e Y, logo não captura uma eventual relação não linear entre estas duas v.a.]

Pergunta 5 — 2 valores

Suponha que os montantes diários (em milhares de euros) de encomendas de material informático numa loja especializada são variáveis aleatórias independentes e identicamente distribuídas a X com valor esperado $\mu = 8$ e desvio padrão $\sigma = 2$.

Indique o número mínimo de dias para que a probabilidade aproximada de o montante total de encomendas exceder 1000 milhares de euros seja de pelo menos 90%.

A: 129 **B**: 61 **C**: 122 **D**: 66

- **V.a.; valor esperado e variância comuns**

  X_i = montante encomendas no dia i, $\quad i = 1, \ldots, n$

  $E(X_i) = E(X) = \mu = 8$

  $V(X_i) = V(X) = \sigma^2 = 2^2$

- **V.a. de interesse**

  $S_n = \sum_{i=1}^n X_i$

  $E(S_n) = \cdots = n\mu, \qquad V(S_n) = \cdots = n\sigma^2$

- **Distribuição aproximada de S_n**

  De acordo com o TLC,

$$Z \;=\; \frac{S_n - E(S_n)}{\sqrt{V(S_n)}} = \frac{S_n - n\mu}{\sqrt{n\sigma^2}} \overset{a}{\sim} \text{normal}(0,1).$$

- **Número mínimo pedido** (valor aproximado)

$$n \;:\; P(Z > 1000) = 1 - P\left(Z \leq \frac{1000 - n\mu}{\sqrt{n\sigma^2}}\right) \overset{TLC}{\simeq} 1 - \Phi\left(\frac{1000 - n\mu}{\sqrt{n\sigma^2}}\right)$$

$$1 - \Phi\left(\frac{1000 - 8n}{\sqrt{4n}}\right) \geq 0.90 \quad \Leftrightarrow \quad \frac{1000 - 8n}{\sqrt{4n}} \geq \Phi^{-1}(1 - 0.90) = -\Phi^{-1}(0.90) \overset{tabelas/calc.}{=} -1.2816$$

$$1000 - 8n \geq -1.2816 \times 2\sqrt{n} \quad \Leftrightarrow \quad 8n - 1.2816 \times 2 \times \sqrt{n} - 1000 \leq 0$$

  A raiz positiva de $8n - 1.2816 \times 2 \times \sqrt{n} - 1000 = 0$ é $\sqrt{n} = \frac{1.2816 \times 2 + \sqrt{(1.2816 \times 2)^2 - 4 \times 8 \times (-1000)}}{2 \times 8} \simeq 11.3417$.
  Logo, o número mínimo de encomendas pedido é $n^\star = \lceil 11.3417^2 \rceil \simeq \lceil 128.634 \rceil = 129$.

Pergunta 6 | 2 valores

Considere uma variável aleatória X com função de densidade de probabilidade:

$$f_X(x) = \begin{cases} \frac{1}{\sqrt{2\pi}x}\, e^{-\frac{(\ln x - \mu)^2}{2}}, & x > 0 \\ 0, & \text{outros valores de } x, \end{cases}$$

onde μ é uma constante real desconhecida.

Deduza o estimador de máxima verosimilhança de μ, com base numa amostra aleatória de dimensão n proveniente de X.

- **V.a. de interesse e f.d.p.**

$$X, \quad f_X(x) = \begin{cases} \frac{1}{\sqrt{2\pi}x}\, e^{-\frac{(\ln x - \mu)^2}{2}}, & x > 0 \\ 0, & \text{c.c.} \end{cases}$$

- **Parâmetro desconhecido e espaço paramétrico**

  $\mu, \quad \Theta = \mathbb{R}$

- **A.a.**

  $(X_1, X_2, \ldots, X_n)$, amostra aleatória de dimensão n proveniente de X.

- **Obtenção da estimativa de MV de μ**

  Será representada por $\hat{\mu}$ e $L(\hat{\mu} \mid \underline{x}) = \max_{\mu \in \Theta} L(\mu \mid \underline{x})$, onde $L(\mu \mid \underline{x})$ representa a f. de verosimilhança.

Passo 1 — Função de verosimilhança

$$
\begin{aligned}
L(\mu \mid x_1, x_2, \ldots, x_n) &= f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) \overset{X_i\, indep}{=} \prod_{i=1}^{n} f_{X_i}(x_i) \overset{X_i \sim X}{=} \prod_{i=1}^{n} f_X(x_i) \\
&= \prod_{i=1}^{n} \left[\frac{1}{\sqrt{2\pi}x} e^{-\frac{(\ln x - \mu)^2}{2}} \right] = (2\pi)^{-n/2} \left(\prod_{i=1}^{n} \frac{1}{x_i} \right) \exp\left[-\frac{1}{2} \sum_{i=1}^{n} (\ln x_i - \mu)^2 \right]
\end{aligned}
$$

Passo 2 — Função de log-verosimilhança

$$
\ln L(\mu|\underline{x}) = -\frac{n}{2}\ln(2\pi) - \sum_{i=1}^{n} \ln(x_i) - \frac{1}{2}\sum_{i=1}^{n} \left(\ln x_i - \mu\right)^2
$$

Passo 3 — Maximização

$$
\hat{\mu} : \begin{cases} \left.\frac{d \ln L(\mu|\underline{x})}{d\mu}\right|_{\mu=\hat{\mu}} = 0 & \text{(ponto de estacionaridade)} \\ \left.\frac{d^2 \ln L(\mu|\underline{x})}{d\mu^2}\right|_{\mu=\hat{\mu}} < 0 & \text{(ponto de máximo)} \end{cases} \Leftrightarrow \begin{cases} \sum_{i=1}^{n} \ln x_i - n\hat{\mu} = 0 \\ -n < 0 \quad \text{(prop. verdadeira)} \end{cases}
$$

Passo 4 — Estimativa e estimador de MV de μ

$$
\begin{aligned}
\hat{\mu} &= \frac{\sum_{i=1}^{n} \ln x_i}{n} \\
EMV(\mu) &= \frac{\sum_{i=1}^{n} \ln X_i}{n}.
\end{aligned}
$$

Pergunta 7	2 valores

De modo a estimar o valor esperado da massa X (em kg) dos carros estacionados em certa zona recolheu-se uma amostra casual de dimensão $n = 12$, tendo-se obtido uma média $\bar{x} = 1200$ kg e um desvio padrão igual a $s = 320$ kg. Suponha que X é uma v.a. com distribuição normal com valor esperado e desvio padrão, μ e σ, desconhecidos.

Obtenha um intervalo de confiança a 95% para μ.

- **V.a. de interesse**

  X = massa do carro estacionado em certa zona

- **Situação**

  $X \sim \text{normal}(\mu, \sigma^2)$, $\quad \mu = E(X)$ DESCONHECIDO, $\quad \sigma^2 = V(X)$ desconhecida

- **Obtenção de IC aproximado para μ**

Passo 1 - Seleção da v.a. fulcral

$$
Z = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{(n-1)}
$$

Passo 2 — Obtenção dos quantis de probabilidade

Como $(1-\alpha) \times 100\% = 95\% \Leftrightarrow \alpha = 0.05$, lidaremos com os quantis seguintes:

$$
\begin{cases} a_\alpha = F^{-1}_{t_{(11)}}(\alpha/2) = -F^{-1}_{t_{(12-1)}}(1-\alpha/2) = -F^{-1}_{t_{(11)}}(0.975) \overset{tabela/calc.}{=} -2.201 \\ b_\alpha = F^{-1}_{t_{(12-1)}}(1-\alpha/2) = \Phi^{-1}(0.975) = 2.201. \end{cases}
$$

Passo 3 — Inversão da desigualdade $a_\alpha \le Z \le b_\alpha$

$$
P(a_\alpha \le Z \le b_\alpha) = 1 - \alpha
$$

$$
P\left(a_\alpha \le \frac{\bar{X} - \mu}{S/\sqrt{n}} \le b_\alpha\right) = 1 - \alpha
$$

$$P\left(\bar{X} - b_\alpha \times \frac{S}{\sqrt{n}} \le \mu \le \bar{X} - a_\alpha \times \frac{S}{\sqrt{n}}\right) = 1-\alpha$$

Passo 4 — Concretização

A expressão geral do intervalo de confiança a $(1-\alpha)\times 100\%$ para μ é

$$IC_{(1-\alpha)\times 100\%}(\mu) = \left[\bar{x} - F^{-1}_{t_{(n-1)}}(1-\alpha/2) \times \frac{s}{\sqrt{n}},\ \bar{x} + F^{-1}_{t_{(n-1)}}(1-\alpha/2) \times \frac{s}{\sqrt{n}}\right].$$

Atendendo a que $n = 12$, $\alpha = 0.05$, $F^{-1}_{t_{(n-1)}}(1-\alpha/2) = 2.201$, $\bar{x} = 1200$ e $s = 320$, segue-se

$$IC_{95\%}(\mu) \quad = \quad \left[1200 - 2.201 \times \frac{320}{\sqrt{12}},\ 1200 + 2.201 \times \frac{320}{\sqrt{12}}\right] \simeq [996.68,\ 1403.32].$$

Pergunta 8 2 valores

Uma engenheira mecânica pretende inferir sobre o valor esperado da massa dos modelos SUV (*sport utility vehicle*) da frota de uma empresa de aluguer de carros. Recolhida uma amostra casual de 40 destes veículos obteve-se uma média e uma variância corrigida amostrais iguais a $\bar{x} = 2\,500$ kg e $s^2 = 540^2$ kg^2.

Teste a hipótese $H_0 : \mu = 2\,400$ contra $H_1 : \mu \neq 2\,400$. Obtenha o valor-p aproximado do teste. Deverá rejeitar-se H_0 ao nível de significância de 1%?

- **V.a. de interesse e situação**

  X = massa do SUV (em kg)

  X v.a. com distribuição arbitrária, $\mu = E(X)$ DESCONHECIDO, $\sigma^2 = V(X)$ desconhecida

- **Hipóteses**

  $H_0 : \mu = \mu_0 = 2\,400$

  $H_1 : \mu \neq \mu_0$

- **Estatística de teste**

$$T \quad = \quad \frac{\bar{X} - \mu}{S/\sqrt{n}} \overset{a}{\sim}_{H_0} \text{normal}(0,1)$$

- **Região de rejeição de** H_0 (para valores da estatística de teste)

  O teste é bilateral, logo a região de rejeição de H_0 é do tipo $W = (-\infty, -c) \cup (c, \infty)$.

- **Decisão** (com base no valor-p aproximado)

  Atendendo a que o valor observado da estatística de teste e do valor-p são iguais a

$$\begin{aligned} t &= \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{2\,500 - 2\,400}{540/\sqrt{40}} \simeq 1.17 \\ valor - p &= 2 \times P(Z > |t| \mid H_0) \simeq 2 \times [1 - \Phi(|1.17|)] = 2 \times (1 - 0.8790) = 0.242, \end{aligned}$$

  é suposto:

  - não rejeitar H_0 a qualquer n.s. $\alpha_0 \le valor - p = 24.2\%$, nomeadamente ao n.u.s. de 1% [ou a outros n.u.s. como 5% e 10%;
  - rejeitar H_0 a qualquer n.s. $\alpha_0 > valor - p = 24.2\%$.]

Pergunta 9 2 valores

A função de probabilidade $\frac{10-x}{55}$, para $x = 0, 1, \ldots, 9$, foi apontada por uma engenheira aeroespacial como adequada para descrever o número semanal de eventos de manutenção de aeronaves da frota comercial que ela supervisiona (hipótese H_0).

A engenheira organizou a seguinte tabela de frequências referente a 100 registos semanais selecionados casualmente para testar a adequação de tal distribuição.

Número semanal de eventos de manutenção	0	1	2	3	$4,5,\ldots,9$
Frequência absoluta observada	23	18	3	13	43
Frequência absoluta esperada sob H_0	18.18	16.36	14.55	12.73	38.18

Com base no teste de ajustamento do qui-quadrado, que decisão deverá tomar a engenheira aeroespacial?

A: Rejeitar H_0 a 1%, 5% e 10%.
B: Rejeitar H_0 a 5% e 10% e não rejeitar H_0 a 1%.
C: Rejeitar H_0 a 10% e não rejeitar H_0 a 1% e 5%.
D: Não rejeitar H_0 a 1%, 5% e 10%.

- **V.a. de interesse; f.p. conjecturada**

  X = número semanal de eventos de manutenção de aeronaves

  $$P_0(X=x) = \begin{cases} \frac{10-x}{55}, & x = 0,1,\ldots,9 \\ 0, & \text{caso contrário.} \end{cases}$$

- **Hipóteses**

  $H_0 : P(X=x) \equiv P_0(X=x) \quad$ vs. $\quad H_0 : f_X(x) \not\equiv f_0(x)$

- **Estatística de teste**

  $$T = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \overset{a}{\sim}_{H_0} \chi^2_{(k-1)}$$

  onde: k = número de classes = 5; O_i = frequência absoluta observável da classe i; E_i = frequência absoluta esperada, sob H_0, da classe i.

- **Região de rejeição de H_0** (para valores de T)

  Ao efectuarmos um teste de ajustamento do qui-quadrado a região de rejeição de H_0 é um intervalo à direita $W = (c, +\infty)$.

- **Decisão (com base no valor-p)**

i	Classe	Freq. abs. obs. o_i	Freq. abs. esper. sob H_0 $E_i = n \times p_i^0$	Parcelas valor obs. estat. teste $\frac{(o_i-E_i)^2}{E_i}$
1	$\{0\}$	23	18.18	$\frac{(23-18.18)^2}{18.18} \simeq 1.278$
2	$\{1\}$	18	16.36	0.164
3	$\{2\}$	3	14.55	9.169
4	$\{3\}$	13	12.73	0.006
5	$\{4,\ldots,9\}$	43	38.18	0.608
		$\sum_{i=1}^{k} o_i = n = 100$	$\sum_{i=1}^{k} E_i = n = 100$	$t = \sum_{i=1}^{k} \frac{(o_i-E_i)^2}{E_i} \simeq 11.225$

  Atendendo a que a região de rejeição de H_0 é um intervalo à direita, temos

  $$valor-p \;=\; P(T > t = 11.225 \mid H_0) \simeq 1 - F_{\chi^2_{(5-1)}}(11.225) \overset{calc.}{\simeq} 0.024.$$

  Assim sendo, devemos decidir pela:

  - rejeição de H_0 a qualquer n.s. $\alpha_0 > 2.4\%$, designadamente aos n.u.s. de 5% e 10%;
  - não rejeição de H_0 a qualquer n.s. $\alpha_0 \le 2.4\%$, nomeadamente ao n.u.s. de 1%.

  Consequentemente, a resposta, das quatro apresentadas, que se coaduna o referido acima é:

  - **B**: Rejeita-se H_0 para 5% e 10% e não se rejeita H_0 para 1%.

[Alternativamente, a consulta das tabelas dos quantis de probabilidade da distribuição do qui-quadrado com $k-1=4$ graus de liberdade permitem concluir que

$$\begin{aligned} F^{-1}_{\chi^2_{(5-1)}}(0.975) = 11.14 &\quad < \quad 11.225 < 13.28 = F^{-1}_{\chi^2_{(5-1)}}(0.99) \\ 0.01 = 1-0.99 &\quad < \quad 1 - F_{\chi^2_{(5-1)}}(11.225) < 1-0.975 = 0.025. \end{aligned}$$

Logo, devemos:

- rejeitar H_0 a qualquer n.s. $\alpha_0 \geq 2.5\%$, designadamente aos n.u.s. de 5% e 10%;
- não rejeitar H_0 a qualquer n.s. $\alpha_0 \leq 1\%$, nomeadamente ao n.u.s. de 1%.

Deste modo, a resposta correcta é **B**: Rejeita-se H_0 para 5% e 10% e não se rejeita H_0 para 1%.]

Pergunta 10 2 valores

Um engenheiro afirma que o número de anos (x) de utilização de um veículo de determinada marca e o rácio (Y) entre o preço do veículo usado e o preço do veículo novo se relacionam de acordo com o modelo de regressão linear simples $Y = \beta_0 + \beta_1\, x + \epsilon$. Este profissional dispõe dos seguintes dados relativos a uma amostra casual de veículos da marca em causa:

$$\sum_{i=1}^{10} x_i = 55, \quad \sum_{i=1}^{10} x_i^2 = 385, \quad \sum_{i=1}^{10} y_i = 4.6, \quad \sum_{i=1}^{10} x_i\, y_i = 19.7, \quad \hat{\sigma}^2 \simeq 0.006735$$

Considere as hipóteses de trabalho convenientes e teste a hipótese $H_0 : E(Y \mid x_0 = 3) = 0.65$ contra $H_1 : E(Y \mid x_0 = 3) \neq 0.65$, ao nível de significância de 10%.

- **[Modelo de RLS e hipóteses de trabalho**

  Y = rácio entre preço do veículo usado (PUV) e preço do veículo novo (PUN) $= PVU/PVN$ (v.a. resposta)

  x = anos de utilização do veículo (variável explicativa)

  $Y_i = \beta_0 + \beta_1\, x_i + \varepsilon_i, \quad i = 1, \ldots, n$

  $\epsilon_i \stackrel{i.i.d.}{\sim} \text{normal}(0, \sigma^2), \quad i = 1, \ldots, n$]

- **Estimativas de MQ de β_0 e β_1**

  Temos

  - $n = 10$
  - $\sum_{i=1}^n x_i = 55$

    $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i = \frac{55}{10} = 5.5$

    $\sum_{i=1}^n x_i^2 = 385$

    $\sum_{i=1}^n x_i^2 - n\,\bar{x}^2 = 385 - 10 \times 5.5^2 = 82.5$
  - $\sum_{i=1}^n y_i = 4.6$

    $\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i = \frac{4.6}{10} = 0.46$

    $\sum_{i=1}^n y_i^2 = 2.55$

    $\sum_{i=1}^n y_i^2 - n\,\bar{y}^2 = 2.55 - 10 \times 0.46^2 = 0.4340$
  - $\sum_{i=1}^n x_i\, y_i = 19.7$

    $\sum_{i=1}^n x_i\, y_i - n\,\bar{x}\,\bar{y} = 19.7 - 10 \times 5.5 \times 0.46 = -5.6.$

Logo, as estimativas de MQ dos parâmetros desconhecidos β_1 e β_0 são dadas por:

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y}}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2} = \frac{-5.6}{82.5} \simeq -0.067879; \\
\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \times \bar{x} \simeq 0.46 - (-0.067879) \times 5.5 \simeq 0.833335.
\end{aligned}$$

- **Hipóteses**

  $H_0 : \beta_0 + \beta_1\, x_0 = e_0$

  $H_1 : \beta_0 + \beta_1\, x_0 \neq e_0 \qquad (x_0 = 3, e_0 = 0.65)$

- **N.s.**

  $\alpha_0 = 0.01$

- **Estatística de teste**

$$T = \frac{\hat{\beta}_0 + \hat{\beta}_1\, x_0 - e_0}{\sqrt{\left[\frac{1}{n} + \frac{(\bar{x}-x_0)^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}\right]\hat{\sigma}^2}} \sim_{H_0} t_{(n-2)}$$

- **Região de rejeição de H_0 (para valores da estatística de teste)**

  O teste é bilateral, logo a região de rejeição de H_0 é do tipo $W = (-\infty, -c) \cup (c, +\infty)$, onde $c : P(\text{Rejeitar } H_0 \mid H_0) = \alpha_0$, ou seja,

$$c = F^{-1}_{t_{(n-2)}}(1 - \alpha_0/2) = F^{-1}_{t_{(8)}}(0.95) \overset{tabelas/calc.}{=} 1.860.$$

- **Decisão**

  Dado que $\hat{\sigma}^2 \simeq 0.006735$, o valor observado da estatística de teste a

$$\begin{aligned}
t &= \frac{(\hat{\beta}_0 + \hat{\beta}_1\, x_0) - e_0}{\sqrt{\left[\frac{1}{n} + \frac{(\bar{x}-x_0)^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}\right]\hat{\sigma}^2}} \simeq \frac{[0.833335 + (-0.067879) \times 3] - 0.65}{\sqrt{\left[\frac{1}{10} + \frac{(5.5-3)^2}{82.5}\right] \times 0.006735}} \simeq \frac{0.629698 - 0.65}{\sqrt{0.175758 \times 0.006735}} \\
&\simeq -0.590082.
\end{aligned}$$

  Como $t \simeq -0.590082 \not\in W = (-\infty, -1.860) \cup (1.860, +\infty)$, a hipótese H_0 não deve ser rejeitada ao n.s. $\alpha_0 = 10\%$ [nem a quaisquer n.s. inferiores a α_0].