Probabilidade e Estatística

LEAmb, LEAN, LEEC, LEGM, LEMat; LEE, LERC

2º Semestre – 2023/2024
18/06/2024
10:30–12:30

Duração: **120** minutos

Exame Época Normal – A

- 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

Considere que há três estradas (1,2,3) entre as cidades A e B, uma estrada (4) entre as cidades B e C e que a ligação por estrada entre as cidades A e C é feita unicamente através da cidade B. Cada uma das quatro estradas pode estar não transitável num dado período. Sejam: P_i o evento a indicar estrada i $(i = 1,2,3)$ não transitável entre as cidades A e B; e P_4 o evento a indicar estrada 4 não transitável entre as cidades B e C. Considere que os quatro eventos são mutuamente independentes com probabilidade comum e igual a 0.1.

Qual é a probabilidade de não ser possível transitar entre as cidades A e C nesse período?

- **Quadro de acontecimentos e probabilidades**

Acontecimento	Probabilidade
P_i = estrada i não transitável num dado período $(i = 1,2,3,4)$	$P(P_i) = 0.1$
D = não é possível transitar entre as cidades A e C nesse período	$P(D) = ?$

- **Probabilidade pedida**

Sabendo que $P_1, \ldots, P_4$ são eventos mutuamente independentes com probabilidade comum e igual a 0.1, temos

$$\begin{aligned}
P(D) &= P[(P_1 \cap P_2 \cap P_3) \cup P_4] \\
&= P(P_1 \cap P_2 \cap P_3) + P(P_4) - P(P_1 \cap P_2 \cap P_3 \cap P_4) \\
&= P(P_1) \times P(P_2) \times P(P_3) + P(P_4) - P(P_1) \times P(P_2) \times P(P_3) \times P(P_4) \\
&= 0.1^3 + 0.1 - 0.1^4 \\
&= 0.1009.
\end{aligned}$$

Pergunta 2 2 valores

Suponha que n máquinas operam de modo independente e que a probabilidade de ao final de um dia uma máquina ainda esteja em operação é p. Considere que o número X de máquinas em operação ao final desse dia é tal que $E(X) = 7.8$ e $E(X^2) = 63.96$. Identifique a distribuição de X e obtenha os valores n e p.

- **V.a.**

X = número de máquinas em operação ao final do dia

- **Distribuição**

$X \sim \text{binomial}(n, p)$, pois estamos a contabilizar o número de sucessos (i.e., máquinas operando) em n provas de Bernoulli i.i.d. com probabilidade de sucesso comum p.

- **Obtenção dos parâmetros n e p**

  Da consulta do formulário, temos

$$(n,p) \quad : \quad \begin{cases} E(X) = 7.8 \\ E(X^2) = 63.96 \end{cases}$$

$$\begin{cases} n\,p = 7.8 \\ E(X^2) = V(X) + E^2(X) = n\,p\,(1-p) + (n\,p)^2 = 63.96 \end{cases}$$

$$\begin{cases} n = \frac{7.8}{p} \\ 7.8 \times (1-p) + 7.8^2 = 63.96 \end{cases}$$

$$\begin{cases} p = \frac{7.8+7.8^2-63.96}{7.8} = 0.6 \\ n = \frac{7.8}{0.6} = 13. \end{cases}$$

Pergunta 3	2 valores

Seja X uma variável aleatória contínua que representa o desvio da medida dos parafusos produzidos por uma máquina em relação à norma especificada pelo mercado, com função de densidade de probabilidade

$$f_X(x) \quad = \quad \begin{cases} \frac{3x^2}{2}, & -1 < x < 1 \\ 0, & \text{caso contrário.} \end{cases}$$

Obtenha o terceiro quartil de X .

- **V.a. e f.d.**

  X = desvio da medida dos parafusos em relação à norma especificada pelo mercado

$$f_X(x) = \begin{cases} \frac{3x^2}{2}, & -1 < x < 1 \\ 0, & \text{caso contrário} \end{cases}$$

- **Terceiro quartil de X**

$$\xi = F_X^{-1}(0.75) \in (-1,1) \quad : \quad F_X(\xi) = 0.75$$

$$\int_{-1}^{\xi} \frac{3x^2}{2}\,dx = 0.75$$

$$\left.\frac{x^3}{2}\right|_{-1}^{\xi} = 0.75$$

$$\frac{\xi^3+1}{2} = 0.75$$

$$\xi = 0.5^{1/3}$$

$$\xi \simeq 0.793701.$$

  [**Obs.**: Note que o valor do terceiro quartil é um número real no intervalo $(-1, 1)$.]

Pergunta 4	2 valores

Considere uma caixa com 2 peças perfeitas e 4 peças defeituosas. Suponha que são extraídas duas peças ao acaso e sem reposição. Sejam: X a variável aleatória que representa um indicador de ocorrência de peça defeituosa na primeira extração; e Y a variável aleatória que representa o número de peças defeituosas nas duas extrações, respetivamente. Este par aleatório possui função de probabilidade conjunta dada pela tabela à direita.

	Y		
X	0	1	2
0	2/30	8/30	0
1	0	8/30	12/30

Obtenha a variância de $(X - Y)$.

- **Par aleatório**

  (X,Y)

  X = indicador de ocorrência de peça defeituosa na primeira extração (de duas sem reposição)

  Y = número de peças defeituosas nas duas extrações sem reposição

- **V.a. de interesse**

  $X-Y$

- **Variância pedida**

  Uma vez que $V(X-Y)=V(X)+V(Y)-2\,cov(X,Y)$, 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			$P(X=x)$
X	0	1	2	
0	2/30	8/30	0	10/30
1	0	8/30	12/30	20/30
$P(Y=y)$	2/30	16/30	12/30	1

- **Valores esperados e variâncias de X e de Y**

$$
\begin{aligned}
E(X) &= \sum_{x=0}^{1} x\times P(X=x) = 1\times 20/30 = 2/3\\
E(X^2) &= \sum_{x=0}^{1} x^2\,P(X=x) = 1^2\times 20/30 = 2/3\\
V(X) &= E(X^2)-E^2(X) = 20/30-(2/3)^2 = 2/9\\
E(Y) &= \sum_{y=0}^{2} y\times P(Y=y) = 1\times 16/30+2\times 12/30 = 4/3\\
E(Y^2) &= \sum_{y=0}^{2} y^2\,P(Y=y) = 1^2\times 16/30+2^2\times 12/30 = 64/30 = 32/15\\
V(Y) &= E(Y^2)-E^2(Y) = 64/30-(4/3)^2 = 32/90 = 16/45
\end{aligned}
$$

- **Valor esperado de XY**

$$E(XY) = \sum_{x=0}^{1}\sum_{y=0}^{2} xy\times P(X=x,Y=y) = 1\times 1\times 8/30+1\times 2\times 12/30 = 32/30 = 16/15$$

- **Covariância entre X e Y**

$$cov(X,Y) = E(XY)-E(X)\times E(Y) = 32/30-(2/3)\times(4/3) = 16/90$$

- **Variância pedida** (cont.)

$$V(X-Y) = V(X)+V(Y)-2\,cov(X,Y) = 2/9+16/45-2\times 16/90 = (20+32-32)/90 = 2/9.$$

Pergunta 5	2 valores

Suponha que a probabilidade de um ensaio num laboratório ter reação *positiva* é igual a 0.4. Seja X_i o número de ensaios realizados independentemente até à ocorrência da primeira reação positiva no dia i, $i=1,\ldots,60$. Suponha que as variáveis X_i são independentes e identicamente distribuídas.

Indique a probabilidade aproximada de o total de ensaios nesses 60 dias não exceder 135.

A: 0.8413 **B**: 0.5232 **C**: 0.1587 **D**: 0.4768

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

  X_i = número de ensaios até à ocorrência da primeira reação positiva no dia i, $\quad i = 1, \ldots, n$

  $n = 60 \geq 30$

  $X_i \overset{i.i.d.}{\sim} X \sim \text{geométrica}(p = 0.4), \quad i = 1, \ldots, n$

  $E(X_i) = E(X) = \mu \overset{form.}{=} \frac{1}{p} = \frac{1}{0.4} = 2.5$

  $V(X_i) = V(X) = \sigma^2 \overset{form.}{=} \frac{1-p}{p^2} = \frac{1-0.4}{0.4^2} = 3.75$

- **V.a. de interesse**

  $S_n = \sum_{i=1}^n X_i$ = número total de ensaios em n dias

- **Valor esperado e variância de S_n**

  $E(S_n) = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) \overset{X_i \sim X}{=} n E(X) = n\mu$

  $V(S_n) = V\left(\sum_{i=1}^n X_i\right) \overset{X_i\, indep.}{=} \sum_{i=1}^n V(X_i) \overset{X_i \sim X}{=} n V(X) = n\sigma^2$

- **Distribuição aproximada de S_n**

  Segundo o teorema do limite central (TLC),

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

- **Probabilidade pedida (valor aproximado)**

  $$\begin{aligned} P(S_n \leq 135) &= P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq \frac{135 - n\mu}{\sqrt{n\sigma^2}}\right) \overset{TLC}{\simeq} \Phi\left(\frac{135 - n\mu}{\sqrt{n\sigma^2}}\right) = \Phi\left(\frac{135 - 150}{\sqrt{225}}\right) \\ &= \Phi(-1) = 1 - \Phi(1) \overset{tabelas/calc.}{=} 1 - 0.8413 = 0.1587. \end{aligned}$$

Pergunta 6 — 2 valores

Seja X uma variável aleatória com função de densidade de probabilidade

$$f_X(x) = \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}}, \quad x \in \mathbb{R},$$

onde λ é um parâmetro real positivo desconhecido. A partir de uma amostra de dimensão 40 desta população obteve-se $\sum_{i=1}^{40} |x_i| = 47.5408$.

Determine a correspondente estimativa de máxima verosimilhança da variância de X, $\quad V(X) = 2\lambda^2$.

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

  X

  $f_X(x) = \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}}, \quad x \in \mathbb{R}$

- **Parâmetro desconhecido**

  $\lambda \quad (\lambda > 0)$

- **Amostra**

  $\underline{x} = (x_1, \ldots, x_n)$, amostra de dimensão $n = 40$ da população X e tal que $\sum_{i=1}^{40} |x_i| = 47.5408$.

- **Obtenção da estimativa de MV de λ**

  Passo 1 — Função de verosimilhança

$$L(\lambda \mid \underline{x}) \quad = \quad f_{\underline{X}}(\underline{x}) \stackrel{X_i\, indep}{=} \prod_{i=1}^{n} f_{X_i}(x_i) \stackrel{X_i \sim X}{=} \prod_{i=1}^{n} f_X(x_i) = \prod_{i=1}^{n} \left(\frac{1}{2\lambda} e^{-\frac{|x_i|}{\lambda}} \right) = (2\lambda)^{-n} e^{-\frac{\sum_{i=1}^{n} |x_i|}{\lambda}}$$

  Passo 2 — Função de log-verosimilhança

$$\ln L(\lambda \mid \underline{x}) \quad = \quad -n \ln(2\lambda) - \frac{\sum_{i=1}^{n} |x_i|}{\lambda}$$

  Passo 3 — Maximização

  A estimativa de MV de λ é representada por $\hat{\lambda}$ e

$$\hat{\lambda} \quad : \quad \begin{cases} \left.\frac{d \ln L(\lambda|\underline{x})}{d\lambda}\right|_{\lambda=\hat{\lambda}} = 0 & \text{(ponto de estacionaridade)} \\ \left.\frac{d^2 \ln L(\lambda|\underline{x})}{d\lambda^2}\right|_{\lambda=\hat{\lambda}} < 0 & \text{(ponto de máximo)} \end{cases} \quad \Leftrightarrow \quad \begin{cases} -\frac{n}{\hat{\lambda}} + \frac{\sum_{i=1}^{n} |x_i|}{\hat{\lambda}^2} = 0 \\ \frac{n}{\hat{\lambda}^2} - 2\frac{\sum_{i=1}^{n} |x_i|}{\hat{\lambda}^3} < 0 \end{cases}$$

$$\begin{cases} \hat{\lambda} = \frac{\sum_{i=1}^{n} |x_i|}{n} \\ -\frac{n^3}{(\sum_{i=1}^{n} |x_i|)^2} < 0 & \text{(prop. verdadeira)} \end{cases}$$

  Passo 4 — Estimativa de MV de λ

$$\hat{\lambda} \quad = \quad \frac{\sum_{i=1}^{n} |x_i|}{n} = \frac{47.5408}{40} = 1.18852$$

- **Outro parâmetro desconhecido**

  $h(\lambda) = V(X) = 2\,\lambda^2$

- **Estimativa de MV de $h(\lambda)$**

  Ao invocar a propriedade de invariância dos EMV, concluímos que a estimativa de MV de $h(\lambda)$ é

$$\widehat{h(\lambda)} \quad = \quad h(\hat{\lambda}) = 2\hat{\lambda}^2 = 2 \times 1.18852^2 \simeq 2.825160.$$

Pergunta 7	2 valores

O tempo (em segundos) de processamento das tarefas que chegam a um determinado *cluster* de computação é uma variável aleatória X que tem distribuição normal com valor esperado desconhecido μ e desvio padrão igual a 1.5 segundos.

Obtenha o intervalo aleatório de confiança a 95% para o valor esperado de X. Qual deverá ser a dimensão mínima da amostra a obter de modo a que este intervalo tenha uma amplitude inferior ou igual a 0.5 segundos?

- **V.a. de interesse**

  X = tempo de processamento de tarefas que chega a um determinado *cluster* de computação

- **Situação**

  $X \sim \text{normal}(\mu, \sigma^2)$

  μ DESCONHECIDO

  $\sigma = 1.5$ conhecido

- **Obtenção de IAC para μ**

  Passo 1 — Seleção da v.a. fulcral para μ

$$Z \quad = \quad \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim \text{normal}(0, 1)$$

pois é suposto obter intervalo aleatório de confiança (IAC) para o valor esperado de população normal, com variância conhecida.

Passo 2 — Obtenção dos quantis de probabilidade

Ao termos em consideração que $(1-\alpha)\times 100\% = 95\%$, faremos uso dos quantis

$$\begin{cases} a_\alpha = \Phi^{-1}(\alpha/2) = \Phi^{-1}(0.025) = 1-\Phi^{-1}(1-0.025) \overset{tabela/calc.}{=} -1.96 \\ b_\alpha = \Phi^{-1}(1-\alpha/2) = \Phi^{-1}(0.975) \overset{tabela/calc.}{=} 1.96. \end{cases}$$

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

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

$$P\left[a_\alpha \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq b_\alpha\right] = 1-\alpha$$

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

Passo 4 — IAC para μ

Atendendo aos extremos aleatórios obtidos no passo anterior, aos quantis acima e ao valor de σ^2, temos

$$IAC_{95\%}(\mu) = \left[\bar{X} - 1.96 \times \frac{\sigma}{\sqrt{n}},\ \bar{X} + 1.96 \times \frac{\sigma}{\sqrt{n}}\right].$$

- **Dimensão mínima pedida**

  Pretendemos a dimensão mínima da a.a. de modo a que amplitude do IAC para μ não seja superior a 1, i. e., o menor dos inteiros positivos

$$n \ : \ \left(\bar{X} + 1.96 \times \frac{\sigma}{\sqrt{n}}\right) - \left(\bar{X} - 1.96 \times \frac{\sigma}{\sqrt{n}}\right) \leq 0.5 \quad \Leftrightarrow \quad 2 \times 1.96 \times \frac{\sigma}{\sqrt{n}} \leq 0.5 \quad \Leftrightarrow \quad n \geq (4 \times 1.96 \times \sigma)^2$$
$$n \geq 138.298.$$

  Logo, a dimensão mínima pedida é $n_{min} = 139$.

Pergunta 8	2 valores

Um vendedor de um certo tipo de sementes alega que 90% das sementes, vendidas em pacotes de 50, germinam. Atendendo a que os totais de sementes germinadas foram 42, 40 e 43 numa experiência com sementes de 3 pacotes, teste a hipótese $H_0 : p = 0.9$ contra $H_1 : p < 0.9$. Obtenha o valor-p aproximado do teste. Para que níveis de significância se deve rejeitar H_0?

- **V.a. de interesse**

$$X = \begin{cases} 1, & \text{se a semente germina} \\ 0, & \text{c.c.} \end{cases}$$

- **Situação**

  $X \sim \text{Bernoulli}(p)$

  $p = P(\text{semente germinar})$ DESCONHECIDA

  $n = 3 \times 50 \geq 30$

- **Hipóteses**

  $H_0 : p = p_0 = 0.9$

  $H_1 : p < p_0$

- **Estatística de teste**

  $$T \quad = \quad \frac{\bar{X} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \overset{a}{\sim}_{H_0} \text{normal}(0,1)$$

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

  Lidamos com um teste unilateral inferior ($H_1 : p < p_0$), logo a região de rejeição de H_0 é do tipo $W = (-\infty, c)$.

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

  O valor observado da estatística de teste é

  $$t \quad = \quad \frac{\bar{x} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{\frac{42+40+43}{3\times 50} - 0.9}{\sqrt{\frac{0.9\times(1-0.9)}{3\times 50}}} \simeq -2.72.$$

  Como a região de rejeição deste teste é um intervalo à esquerda, temos:

  $$valor-p \quad = \quad P(T < t \mid H_0) \simeq \Phi(-2.72) = 1 - \Phi(2.72) \overset{tabela/calc.}{=} 1 - 0.9967 = 0.0033.$$

  Consequentemente, é suposto:

  – rejeitarmos H_0 a qualquer n.s. $\alpha_0 > 0.33\%$, designadamente aos n.u.s. de 1%, 5% e 10%[;

  – não rejeitarmos H_0 a qualquer n.s. $\alpha_0 \leq 0.33\%$.]

Pergunta 9 — 2 valores

Um engenheiro conjetura que uma certa variável observável X se distribui de acordo com a função de probabilidade dada por

$$P_0(X = x) = \frac{(x-1)(x-2)}{2} \times 0.6^3 \times 0.4^{x-3}, \quad x = 3, 4, \ldots$$

(hipótese H_0). Para avaliar esta conjetura, recolheu uma amostra casual de 100 valores de X que agrupou na tabela seguinte, a que juntou as frequências esperadas calculadas recorrendo a $P_0(X = x)$:

X	3	4	5	6	≥ 7
Frequência absoluta observada	28	22	23	18	9
Frequência absoluta esperada sob H_0	21.60	25.92	20.74	13.82	17.92

Com base no teste de ajustamento do qui-quadrado, o engenheiro deve

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**

  X

- **Hipóteses**

  $H_0 : P(X = x) = P_0(X = x) = \frac{(x-1)(x-2)}{2} \times 0.6^3 \times 0.4^{x-3}, \quad x = 3, 4, \ldots$

  $H_0 : P(X = x) \neq P_0(X = x), \quad$ para algum 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 aproximado)**

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	3	28	21.6	$\frac{(28-21.6)^2}{21.6} \simeq 1.896$
2	4	22	25.92	0.593
3	5	23	20.74	,0.246
4	6	18	13.82	1.264
5	$\{7,8,\ldots\}$	9	17.92	4.440
		$\sum_{i=1}^k o_i = n = 100$	$\sum_{i=1}^k E_i = n = 100$	$t = \sum_{i=1}^k \frac{(o_i-E_i)^2}{E_i} = 8.439$

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

$$valor - p \quad = \quad P(T > t \mid H_0) \simeq 1 - F_{\chi^2_{(5-1)}}(8.439) \overset{calc.}{\simeq} 0.076.$$

Assim sendo, devemos decidir pela:

- rejeição de H_0 a qualquer n.s. $\alpha_0 > 7.67579\%$, designadamente ao n.u.s. de 10%;
- não rejeição de H_0 a qualquer n.s. $\alpha_0 \leq 7.67579\%$, nomeadamente ao n.u.s. de 1% e 5%.

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

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

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

$$\begin{aligned} F^{-1}_{\chi^2_{(5-1)}}(0.9) = 7.779 &\quad < \quad 8.439 < 8.496 = F^{-1}_{\chi^2_{(5-1)}}(0.925) \\ 0.075 = 1 - 0.925 &\quad < \quad 1 - F_{\chi^2_{(5-1)}}(8.439) < 1 - 0.90 = 0.10. \end{aligned}$$

Logo, devemos:

- rejeitar H_0 a qualquer n.s. $\alpha_0 \geq 10\%$, designadamente ao n.u.s. de 10%;
- não rejeitar de H_0 a qualquer n.s. $\alpha_0 \leq 7.5\%$, nomeadamente aos n.u.s. de 1% e 5%.

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

Pergunta 10 — 2 valores

Numa experiência de laboratório, uma quantidade x_i de insulina foi injectada no indivíduo i, anotando-se a seguir a redução percentual de glicose no sangue Y_i. Considere um modelo de regressão linear com $Y_i \sim \text{normal}(\beta_0 + \beta_1 x_i, \sigma^2)$, $i = 1, \ldots, 10$, independentes. Um resumo dos dados obtidos nesta experiência é dado por:

$$\bar{x} = 2.34, \quad \sum_{i=1}^{10} x_i^2 = 66.80, \quad \hat{\beta}_0 = 1.696, \quad \hat{\sigma}^2 = 20.595$$

Use estas quantidades para testar $H_0 : \beta_0 = 0$ contra $H_1 : \beta_0 \neq 0$, ao nível de significância de 5%.

- **Hipóteses**

  $H_0 : \beta_0 = \beta_{0,0} = 0$

  $H_1 : \beta_0 \neq \beta_{0,0}$

- **Nível de significância**

  $\alpha_0 = 5\%$

- **Estatística de teste**

  Ao admitirmos que $Y_i \overset{indep.}{\sim} \text{normal}(0, \sigma^2)$, $i = 1, \ldots, n$, temos

$$T = \frac{\hat{\beta}_0 - \beta_{0,0}}{\sqrt{\hat{\sigma}^2 \times \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}\right)}} \sim_{H_0} t_{(n-2)}$$

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

  Como o teste é bilateral ($H_1 : \beta_0 \neq \beta_{0,0}$) e $f_{T|H_0}(z)$ é simétrica em relação à origem, a região de rejeição de H_0 é do tipo $W = (-\infty, -c) \cup (c, +\infty)$, onde $c : P(\text{Rejeitar} H_0 \mid H_0) = \alpha_0$, i.e.,

$$c = F^{-1}_{t_{(n-2)}}(1 - \alpha_0/2) = F^{-1}_{t_{(8)}}(0.975) \overset{tabelas, calc.}{=} 2.306.$$

- **Decisão**

  Uma vez que

  - $n = 10$
  - $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i = 2.34$
    $\sum_{i=1}^n x_i^2 = 66.80$
    $\sum_{i=1}^n x_i^2 - n\,\bar{x}^2 = 66.80 - 8 \times 2.34^2 = 12.044$
  - $\hat{\beta}_0 = 1.696$
  - $\hat{\sigma}^2 = 20.595$

  o valor observado da estatística de teste é dado por

$$t = \frac{\hat{\beta}_0 - \beta_{0,0}}{\sqrt{\hat{\sigma}^2 \times \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}\right)}} \simeq \frac{1.696 - 0}{\sqrt{20.595 \times \left(\frac{1}{10} + \frac{2.34^2}{12.044}\right)}} \simeq 0.501813.$$

  Como $t \simeq 0.501813 \not\in W = (-\infty, -2.306) \cup (2.306, +\infty)$ não devemos rejeitar H_0 ao n.s. $\alpha_0 = 5\%$ [ou a qualquer n.s. inferior a $\alpha_0 = 5\%$].