Probabilidade e Estatística

LEAer, LEBiom, LEMec; LEE, LERC

2º Semestre – 2023/2024
02/07/2024
15:30–17:30

Duração: **120** minutos **Exame Época Recurso – B**

- 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 múltipla 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

Num concurso para uma agência espacial os candidatos são aprovados se conseguirem passar três provas eliminatórias, que se realizam sequencialmente. Mais, cada candidato só participa na eliminatória seguinte, caso tenha passado a eliminatória anterior. Sabe-se de edições anteriores que: na primeira prova, para cada candidato que não passa, há três que passam; na segunda prova são eliminados 80% dos candidatos que nela participam; a probabilidade de sucesso de um candidato que participa na terceira prova é de 10%.

Qual é a probabilidade de um candidato, escolhido ao acaso entre os que vão agora iniciar as provas, vir a ser aprovado?

- **Quadro de acontecimentos e probabilidades**

Acontecimento	Probabilidade
A_1 = candidato passa na primeira prova eliminatória	$P(A_1) = 3\,P(\bar{A}_1)$
A_2 = candidato passa na segunda prova eliminatória	$P(A_2) = ?$
A_3 = candidato passa na terceira prova eliminatória	$P(A_3) = ?$
	$P(\bar{A}_2 \mid A_1) = 0.8$
	$P[A_3 \mid (A_1 \cap A_2)] = 0.1$

- **Cálculo auxiliar**

$$P(A_1) + P(\bar{A}_1) = 1 \Leftrightarrow 3P(\bar{A}_1) + P(\bar{A}_1) = 1 \Leftrightarrow P(\bar{A}_1) = 1/4 \Leftrightarrow P(A_1) = 0.75$$

- **Prob. pedida**

Usando a lei da probabilidade composta, temos

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P(A_1) \times P(A_2 \mid A_1) \times P[A_3 \mid (A_1 \cap A_2)] \\ &= 0.75 \times [1 - P(\bar{A}_2 \mid A_1)] \times 0.1 \\ &= 0.75 \times 0.2 \times 0.1 \\ &= 0.015. \end{aligned}$$

Pergunta 2 2 valores

Registos recentes indicam que somente 20% dos clientes contactados decidem comprar um jacto privado produzido por determinada empresa. Seja X a variável aleatória que representa o número total de contactos efectuados por uma gestora dessa empresa até à venda de 5 jactos privados. Então a função de probabilidade de X é dada por $P(X = x) = \binom{x-1}{5-1}(1-0.2)^{x-5}\,0.2^5$, para $x = 5, 6, \ldots$

Indique a probabilidade de tal gestora ter de efectuar mais de 8 contactos até à venda de 5 jactos privados, sabendo que já efectuou mais de 5 contactos sem que tenha atingido tal objectivo.

A: 0.0992 **B**: 0.9899 **C**: 0.9012 **D**: 0.0104

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

  X = no. total de contactos efectuados até à venda de 5 jactos privados

  $P(X=x) = \binom{x-1}{5-1}(1-0.2)^{x-5}\,0.2^5, \quad x=5,6,\ldots$

- **Probabilidade pedida**

$$\begin{aligned} P(X>8 \mid X>5) &= \frac{P(X>8, X>5)}{P(X>5)} = \frac{P(X>8)}{P(X>5)} = \frac{1-P(5\le X\le 8)}{1-P(X\le 5)} \\ &= \frac{1-\sum_{x=5}^{8}\binom{x-1}{5-1}(1-0.2)^{x-5}\,0.2^5}{1-\binom{5-1}{5-1}(1-0.2)^{5-5}\,0.2^5} \overset{calc.}{\simeq} \frac{1-0.010406}{1-0.00032} \simeq 0.9899. \end{aligned}$$

Pergunta 3	2 valores

O erro de medição associado a um certo método para medir a concentração de azoto no solo é uma variável aleatória X com a seguinte função de densidade de probabilidade $f_X(x) = \frac{1}{2}e^{-|x|}$, para $-\infty < x < +\infty$. Calcule o primeiro e o terceiro quartis de X.

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

  X = erro de medição...

- **F.d. de X**

$$F_X(x) = P(X\le x) = \int_{-\infty}^{x} f_X(t)\,dt = \begin{cases} \int_{-\infty}^{x}\frac{1}{2}e^t\,dt = \frac{e^t}{2}\Big|_{-\infty}^{x} = \frac{e^x}{2}, & x<0 \\ \int_{-\infty}^{0}\frac{1}{2}e^t\,du + \int_0^x \frac{1}{2}e^{-t}\,dt = \frac{e^t}{2}\Big|_{-\infty}^{0} - \frac{e^{-t}}{2}\Big|_0^x = 1-\frac{e^{-x}}{2}, & x\ge 0 \end{cases}$$

- **Primeiro quartil de X**

$$q_{0.25}<0 \;:\; F_X(q_{0.25}) = 0.25 \quad \Leftrightarrow \quad \frac{1}{2}e^{q_{0.25}} = 0.25 \quad \Leftrightarrow \quad q_{0.25} = \ln 0.5 \quad [\Leftrightarrow \quad q_{0.25} \simeq -0.6931].$$

- **Terceiro quartil de X**

$$q_{0.75}>0 \;:\; F_X(q_{0.75}) = 0.5 \quad \Leftrightarrow \quad 1-\frac{1}{2}e^{-q_{0.75}} = 0.75 \quad \Leftrightarrow \quad q_{0.75} = -\ln 0.5 \quad [\Leftrightarrow \quad q_{0.75} \simeq 0.6931].$$

  [Em alternativa, podia ser invocada a simetria da f.d.p. e concluir que $q_{0.75} = -q_{0.25}$.]

Pergunta 4	2 valores

O número de pulsos, por segundo, de ondas de choque radiais, antes e depois de tomar um anti-inflamatório, é representado pelo par aleatório (X,Y) com função de probabilidade conjunta dada pela tabela à direita.

	Y		
X	0	1	2
0	0.2	0.1	0.1
1	0.1	0.2	0
2	0	0.1	0.2

Calcule $E(Y \mid X=2)$ e $V(Y \mid X=2)$.

- **Par aleatório**

  X = # de pulsos, por segundo, de ondas de choque radiais, antes de tomar um anti-inflamatório

  Y = # de pulsos, por segundo, de ondas de choque radiais, depois de tomar um anti-inflamatório

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

  $Y \mid X = 2$

$$P(Y = y \mid X = 2) = \frac{P(Y = y, X = 2)}{P(X = 2)} = \frac{P(Y = y, X = 2)}{\sum_{y=0}^{2} P(X = 2, Y = y)} = \frac{P(Y = y, X = 2)}{0 + 0.1 + 0.2} = \begin{cases} \frac{0.1}{0.3} = \frac{1}{3}, & y = 1 \\ \frac{0.2}{0.3} = \frac{2}{3}, & y = 2 \\ 0, & \text{c.c.} \end{cases}$$

- **Valor esperado e variância condicionais pedidos**

$$\begin{aligned} E(Y \mid X = 2) &= \sum_{y=0}^{2} y \times P(Y = y \mid X = 2) = 1 \times \frac{1}{3} + 2 \times \frac{2}{3} = \frac{5}{3} \\ E(Y^2 \mid X = 2) &= \sum_{y=0}^{2} y^2 \times P(Y = y \mid X = 2) = 1^2 \times \frac{1}{3} + 2^2 \times \frac{2}{3} = 3 \\ V(Y \mid X = 2) &= E(Y^2 \mid X = 2) - E^2(Y \mid X = 2) = 3 - \left(\frac{5}{3}\right)^2 = \frac{2}{9}. \end{aligned}$$

Pergunta 5 2 valores

Uma empresa farmacêutica vai lançar um novo medicamento, anunciando que cada dose tem uma concentração do princípio ativo (X), com valor esperado 50 mg/mL e desvio-padrão 2 mg/mL. Serão analisadas 100 doses do medicamento e supõe-se que as respetivas concentrações do princípio ativo são variáveis aleatórias independentes e identicamente distribuídas a X.

Qual é o valor (aproximado) para a probabilidade da concentração média do princípio ativo nestas 100 doses ser inferior a 49.95 mg/mL?

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

  X_i = concentração do princípio ativo na dose i, $\quad i = 1, \ldots, n$

  $n = 100$

  $X_i \overset{i.i.d.}{\sim} X$

  $E(X_i) = E(X) = \mu = 50$

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

- **V.a. de interesse**

  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ = concentração média do princípio ativo em n doses

- **Valor esperado e variância de $\bar{X}_n$**

  $E(\bar{X}_n) = \cdots = \mu = 50$

  $V(\bar{X}_n) = \cdots = \frac{\sigma^2}{n} = \frac{4}{100} = 0.04$

- **Distribuição aproximada de $\bar{X}_n$**

  Segundo o teorema do limite central (TLC),

$$\frac{\bar{X}_n - E(\bar{X}_n)}{\sqrt{V(\bar{X}_n)}} = \frac{\bar{X}_n - E(X)}{\sqrt{V(X)/n}} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \overset{a}{\sim} \text{normal}(0,1).$$

- **Probabilidade pedida (valor aproximado)**

$$\begin{aligned} P(\bar{X}_n < 49.95) &= P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq \frac{49.95 - \mu}{\sigma/\sqrt{n}}\right) \overset{TLC}{\simeq} \Phi\left(\frac{49.95 - \mu}{\sigma/\sqrt{n}}\right) = \Phi\left(\frac{49.95 - 50}{2/\sqrt{100}}\right) = \Phi(-0.25) \\ &= 1 - \Phi(0.25) \overset{tabelas/calc.}{=} 1 - 0.5987 = 0.4013. \end{aligned}$$

Pergunta 6	2 valores

Quando um supermercado abre, apenas uma caixa é aberta; a segunda caixa só é aberta aquando da chegada do quarto cliente. Se X representar o tempo (em minutos) que decorre da abertura do supermercado até ao instante em que a segunda caixa é aberta, então a função de densidade de probabilidade de X é dada por $f_X(x) = \frac{1}{6}\lambda^4 x^3 e^{-\lambda x}$, para $x > 0$, onde λ é um parâmetro desconhecido real positivo.

Deduza a estimativa de máxima verosimilhança de $E(X) = \frac{4}{\lambda}$ com base na seguinte amostra casual: $(41, 15, 22, 11, 15)$.

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

  X = tempo que decorre da abertura do supermercado ao instante em que a segunda caixa é aberta

  $f_X(x) = \frac{1}{6}\lambda^4 x^3 e^{-\lambda x}, \quad x > 0$

- **Parâmetro desconhecido**

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

- **Amostra**

  $\underline{x} = (x_1, \ldots, x_5) = (41, 15, 22, 11, 15)$ é uma amostra de dimensão $n = 5$, proveniente da população X.

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

  Passo 1 — Função de verosimilhança

$$
\begin{aligned}
L(\lambda \mid \underline{x}) &= f_{\underline{X}}(\underline{x}) \stackrel{X_i\, indep}{=} \prod_{i=1}^{n} f_{X_i}(x_i) \stackrel{X_i \sim X}{=} \prod_{i=1}^{n} f_X(x_i) = \prod_{i=1}^{n} \left(\frac{1}{6}\lambda^4 x_i^3 e^{-\lambda x_i}\right) \\
&= 6^{-n}\lambda^{4n} \left(\prod_{i=1}^{n} x_i\right)^3 \times e^{-\lambda \sum_{i=1}^{n} x_i}, \quad \lambda > 0
\end{aligned}
$$

  Passo 2 — Função de log-verosimilhança

$$
\ln L(\lambda \mid \underline{x}) = -n\ln(6) + 4n\ln(\lambda) + 3\sum_{i=1}^{n}\log(x_i) - \lambda\sum_{i=1}^{n} x_i, \quad \lambda > 0
$$

  Passo 3 — Maximização

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

$$
\hat{\lambda} : \begin{cases} \left.\frac{d\ln L(\lambda|\underline{x})}{d\lambda}\right|_{\lambda=\hat{\lambda}} = 0 & \text{(ponto de estacionaridade)} \\ \left.\frac{d^2\ln L(\lambda|\underline{x})}{d\lambda^2}\right|_{\lambda=\hat{\lambda}} < 0 & \text{(ponto de máximo)} \end{cases} \quad \Leftrightarrow \quad \begin{cases} \frac{4n}{\hat{\lambda}} - \sum_{i=1}^{n} x_i = 0 \\ -\frac{4n}{\hat{\lambda}^2} < 0 \quad \text{(prop. verdadeira)} \end{cases}
$$

  Passo 4 — Estimativa de MV de λ

$$
\hat{\lambda} = \frac{4n}{\sum_{i=1}^{n} x_i} = \frac{4 \times 5}{104} = \frac{5}{26}. \qquad [\simeq \hat{\lambda} \simeq 0.192308]
$$

- **Outro parâmetro desconhecido**

  $h(\lambda) = E(X) = \frac{4}{\lambda}$

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

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

$$
\widehat{h(\lambda)} = h(\hat{\lambda}) = \frac{4}{\hat{\lambda}} = \frac{4}{5/26} = 20.8.
$$

Pergunta 7	2 valores

Uma engenheira biomédica supõe que o custo de produção de um *kit* de certo teste de despistagem é uma variável aleatória (X, em euros) com distribuição normal, com valor esperado e desvio padrão desconhecidos. Tendo-se recolhido uma amostra casual de $n = 18$ unidades de tal artigo, obteve-se uma média e variância corrigida amostrais iguais a $\bar{x} = 3.92$ e $s^2 = 0.64^2$.

Calcule um intervalo de confiança a 90% para a variância de X.

A: [0.2679, 0.6409] **B**: [0.5024, 0.8961] **C**: [0.2524, 0.8030] **D**: [0.2412, 0.7416]

- **V.a. de interesse**

  X = custo de produção de um *kit* de teste

- **Situação**

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

  $\mu = E(X)$ desconhecido

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

- **Obtenção do IC para σ^2**

  Passo 1 — Variável aleatória fulcral para σ^2

$$Z = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

  Passo 2 — Quantis de probabilidade

  Dado que $n = 18$ e $(1-\alpha) \times 100\% = 90\%$, usaremos os quantis

$$(a_\alpha, b_\alpha) : \begin{cases} P(a_\alpha \leq Z \leq b_\alpha) = 1-\alpha \\ P(Z < a_\alpha) = P(Z > b_\alpha) = \alpha/2 \end{cases} \Leftrightarrow \begin{cases} a_\alpha = F^{-1}_{\chi^2_{(n-1)}}(\alpha/2) = F^{-1}_{\chi^2_{(17)}}(0.05) \overset{tabela/calc.}{=} 8.672 \\ b_\alpha = F^{-1}_{\chi^2_{(n-1)}}(1-\alpha/2) = F^{-1}_{\chi^2_{(17)}}(0.95) \overset{tabela/calc.}{=} 27.59. \end{cases}$$

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

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

$$P\left[a_\alpha \leq \frac{(n-1)S^2}{\sigma^2} \leq b_\alpha\right] = 1-\alpha$$

$$P\left[\frac{1}{b_\alpha} \leq \frac{\sigma^2}{(n-1)S^2} \leq \frac{1}{a_\alpha}\right] = 1-\alpha$$

$$P\left[\frac{(n-1)S^2}{b_\alpha} \leq \sigma^2 \leq \frac{(n-1)S^2}{a_\alpha}\right] = 1-\alpha$$

  Passo 4 — Concretização

  Atendendo à expressão geral do IC para σ^2,

$$IC_{(1-\alpha)\times 100\%}(\sigma) = \left[\frac{(n-1)\,s^2}{F^{-1}_{\chi^2_{(n-1)}}(1-\alpha/2)}, \frac{(n-1)\,s^2}{F^{-1}_{\chi^2_{(n-1)}}(\alpha/2)}\right],$$

  ao par de quantis acima e ao valor de s^2, temos

$$\begin{aligned} IC_{90\%}(\sigma^2) &= \left[\sqrt{\frac{(18-1)\times 0.64^2}{27.59}}, \sqrt{\frac{(18-1)\times 0.64^2}{8.672}}\right] \\ &\simeq [0.2524, 0.8030]. \end{aligned}$$

Pergunta 8	2 valores

Pretende-se testar a resistência à tração de um novo material compósito projetado para uso na indústria aeronáutica. As especificações exigem que o valor esperado da resistência à tração seja igual a 500 MPa, sendo o compósito declarado como inadequado caso a tração esperada seja inferior 500 MPa. Para tal, foram recolhidos 25 espécimes do material, tendo-se observado uma média da resistência à tração igual a 478 MPa.

Teste a hipótese $H_0 : \mu = 500$ contra $H_1 : \mu < 500$, assumindo que a resistência à tração segue uma distribuição normal, com desvio padrão igual a 10 MPa, e considerando um nível de significância de 6%.

- **V.a. de interesse**

  X = resistência à tração

- **Situação**

  $X \sim \text{normal}(\mu, 10^2)$

  μ DESCONHECIDO

  $\sigma^2 = 10^2$ conhecido

- **N.s.**

  $\alpha_0 = 6\%$

- **Hipóteses**

  $H_0 : \mu = \mu_0 = 500$

  $H_1 : \mu < \mu_0 = 500$

- **Estatística de teste**

  $$T = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim_{H_0} \text{normal}(0,1)$$

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

  O teste é unilateral inferior ($H_1 : \mu < 500$), logo a região de rejeição de H_0 é do tipo $W = (-\infty, c)$, onde $c : P(\text{Rejeitar } H_0 \mid H_0) = \alpha_0$, i.e.,

  $$c = \Phi^{-1}(\alpha_0) = \Phi^{-1}(0.06) = -\Phi^{-1}(1-0.06) \overset{tabelas/calc.}{=} -1.5548.$$

- **Decisão**

  O valor observado da estatística de teste é

  $$t = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = (478-500)/(10/\sqrt{25}) = -11$$

  Como $t = -11 \in W = (-\infty, -1.5548)$, a hipótese H_0 deve ser rejeitada ao nível de significância de 6% [ou a qualquer n.s. superior a α_0].

Pergunta 9	2 valores

Uma equipa de engenheiros está a analisar a distribuição da velocidade (X, em m/s) do vento dominante em determinado local onde se pretende construir uma aerogare. Eles estão interessados em determinar se X possui função de distribuição dada por $F_0(x) = 1 - \exp[-(x/\beta)^\alpha]$, para $x > 0$, onde $\alpha = 1.85$ e $\beta = 3.9$ (hipótese H_0). Para tal, foram recolhidas 100 observações casuais, tendo-se organizado a seguinte tabela de frequências, sendo que as frequências absolutas esperadas sob H_0 foram arredondadas à unidade:

Classe	$(0, 1.734]$	$(1.734, 2.713]$	$(2.713, 3.720]$	$(3.720, 5.044]$	$(5.044, +\infty)$
Freq. abs. observada	17	23	14	30	16
Freq. abs. esperada sob H_0	20	20	20	20	20

Obtenha o valor-p aproximado (ou um intervalo para o valor-p aproximado) do teste de ajustamento do qui-quadrado. Para que níveis de significância deverá a equipa de engenheiros rejeitar H_0?

- **V.a. de interesse**

  X = velocidade do vento dominante (em m/s)

- **Hipóteses**

  $H_0 : F_X(x) = F_0(x) = 1 - \exp[-(x/\beta)^\alpha], \quad \forall x > 0 \; (\alpha = 1.85, \beta = 3.9)$

  $H_1 : F_X(x) \neq F_0(x), \quad$ para algum $x > 0$

- **Estatística de teste**

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

  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	$(0, 1.734]$	17	20.0	$\frac{(17-20.0)^2}{20.0} = 0.45$
2	$(1.734, 2.713]$	23	20.0	0.45
3	$(2.713, 3.720]$	14	20.0	1.8
4	$(3.720, 5.044]$	30	20.0	5
5	$(5.044, +\infty)$	16	20.0	0.8
		$\sum_{i=1}^k o_i = n = 100$	$\sum_{i=1}^k E_i = n = 100$	$t = \sum_{i=1}^k \frac{(o_i-E_i)^2}{E_i} = 8.5$

  Atendendo a que a região de rejeição de H_0 é um intervalo à direita num teste de ajustamento do qui-quadrado, temos

$$valor-p \quad = \quad P(T > t \mid H_0) \simeq 1 - F_{\chi^2_{(5-1)}}(8.5) \stackrel{calc.}{\simeq} 0.074887.$$

  Assim sendo, devemos decidir pela:

  – rejeição de H_0 a qualquer n.s. $\alpha_0 > 7.4887\%$, designadamente ao n.u.s. de 10%[;

  – não rejeição de H_0 a qualquer n.s. $\alpha_0 \leq 7.4887\%$, nomeadamente ao n.u.s. de 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

$$F^{-1}_{\chi^2_{(5-1)}}(0.925) = 8.496 \quad < \quad 9.488 <= F^{-1}_{\chi^2_{(5-1)}}(0.95)$$
$$0.05 = 1 - 0.95 \quad < \quad 1 - F_{\chi^2_{(5-1)}}(8.5) < 1 - 0.925 = 0.075.$$

  Logo, devemos:

  – rejeitar H_0 a qualquer n.s. $\alpha_0 \geq 7.5\%$, designadamente ao n.u.s. de 10%[;

  – não rejeitar de H_0 a qualquer n.s. $\alpha_0 \leq 5\%$, nomeadamente aos n.u.s. de 1% e 5%].

Pergunta 10	2 valores

Admita que o modelo de regressão linear simples $Y = \beta_0 + \beta_1\, x + \epsilon$ foi usado para relacionar a velocidade de um veículo (x) e a eficiência de combustível $(Y$, em milhas por galão, mpg) observada durante os testes de estrada. Para tal, foram recolhidas 32 observações, tendo-se obtido os seguintes valores:

$$\sum_{i=1}^{32} x_i = 270, \quad \sum_{i=1}^{32} x_i^2 = 2453, \quad \sum_{i=1}^{32} y_i = 105, \quad \sum_{i=1}^{32} x_i\, y_i = 895, \quad \hat{\sigma}^2 \simeq 16.966637$$

Tendo em conta estes valores e as hipóteses de trabalho convenientes, obtenha um intervalo de confiança a 98% para β_1 e pronuncie-se sobre a significância do modelo de regressão.

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

  Y = eficiência de combustível (v.a. resposta)

  x = velocidade de um veículo (variável explicativa)

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

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

- **Estimativa de β_1**

  Temos $n = 32$ e

  - $\sum_{i=1}^n x_i = 270$

    $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i = \frac{270}{32} = 8.4375$

    $\sum_{i=1}^n x_i^2 = 2453$

    $\sum_{i=1}^n x_i^2 - n\,\bar{x}^2 = 2453 - 32 \times 8.4375^2 = 174.875$

  - $\sum_{i=1}^n y_i = 105$

    $\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i = \frac{105}{32} = 3.28125$

  - $\sum_{i=1}^n x_i\, y_i = 895$

    $\sum_{i=1}^n x_i\, y_i - n\,\bar{x}\,\bar{y} = 895 - 32 \times 8.4375 \times 3.28125 = 9.0625.$

  Logo, a estimativa de β_1 é dada por:

  $$\hat{\beta}_1 \quad = \quad \frac{\sum_{i=1}^n x_i\, y_i - n\,\bar{x}\,\bar{y}}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2} = \frac{9.0625}{174.875} \simeq 0.051823.$$

- **Obtenção do IC a** 98% **para β_1**

  Passo 1 — V.a. fulcral para β_1

  $$Z = \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} \sim t_{(n-2)}, \quad \text{onde} \quad se(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}}$$

  Passo 2 — Quantis de probabilidade

  Já que $(1-\alpha) \times 100\% = 98\%$, temos $\alpha = 0.02$ e lidaremos com

  $$\begin{cases} a_\alpha = -b_\alpha \\ b_\alpha = F^{-1}_{t_{(n-2)}}(1-\alpha/2) = F^{-1}_{t_{(32-2)}}(1-0.02/2) = F^{-1}_{t_{(30)}}(0.99) \overset{tabela/calc.}{=} 2.457 \end{cases}$$

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

  $$P(a_\alpha \le Z \le b_\alpha) = 1-\alpha \quad \Leftrightarrow \quad P\left[a_\alpha \le \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} \le b_\alpha\right] = 1-\alpha$$

  $$P\left[\hat{\beta}_1 - b_\alpha \times se(\hat{\beta}_1) \le \beta_1 \le \hat{\beta}_1 - a_\alpha \times se(\hat{\beta}_1)\right] = 1-\alpha$$

Passo 4 — Concretização

A expressão geral do $IC_{(1-\alpha)\times 100\%}\left(\beta_1\right)$ é dada por $\left[\hat{\beta}_1 \pm b_\alpha \times \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}}\right]$ e $\hat{\sigma}^2 \simeq 16.966637$. Donde,

$$IC_{95\%}(\beta_1) \quad = \quad \left[0.051823 \pm 2.457 \times \sqrt{\frac{16.966637}{174.875}}\right] \simeq [-0.713490, 0.817136].$$

- **Teste de significância do modelo de RLS**
  - **Hipóteses**

    $H_0 : \beta_1 = \beta_{1,0} = 0$

    $H_1 : \beta_1 \neq \beta_{1,0}$
  - **N.s.**

    $\alpha_0 = 100\% - 98\% = 2\%$
  - **Decisão**

    Invocando a analogia entre IC e testes de hipóteses (bilaterais), concluímos que o valor conjeturado para β_1 em H_0

    $$\beta_{1,0} \quad = \quad 0 \in IC_{98\%}(\beta_1) = [-0.713490, 0.817136],$$

    pelo que a hipótese $H_0 : \beta_1 = \beta_{1,0} = 0$ não deve ser rejeitada ao n.s. $\alpha_0 = 2\%$ [ou a qualquer n.s. superior 2%]. Assim sendo, a eficiência de combustível (Y) não parece ser influenciada pela velocidade do veículo a tal n.s.