Probabilidades e Estatística / Introd. às Probabilidades e Estatística

TODOS OS CURSOS

Exame Época Especial
2020/2021
27/07/2021 – **09:00**

Duração: 120 + 30 minutos

Justifique convenientemente todas as respostas

Pergunta 1 2 valores

Num inquérito sobre a qualidade das instalações de uma dada residência universitária, 50% dos alunos residentes declarou estar satisfeita com as instalações desta residência. Entre os alunos que declararam estar satisfeitos 55% são *caloiros* e de entre os alunos insatisfeitos 10% são *caloiros*.

Suponha que um aluno é escolhido ao acaso entre os residentes. Qual é a probabilidade de estar satisfeito com as instalações da residência, sabendo que o aluno não é *caloiro*?

- **Quadro de acontecimentos e probabilidades**

Acontecimento	Probabilidade
S = aluno residente satisfeito com as instalações da residência	$P(S) = 0.50$
$\bar{C}$ = Aluno não caloiro	$P(\bar{C}) = ?$
	$P(C \mid S) = 0.55$
	$P(C \mid \bar{S}) = 0.10$

- **Prob. pedida**

$$
\begin{aligned}
P(S \mid \bar{C}) &\overset{teo.\ Bayes}{=} \frac{P(\bar{C} \mid S) \times P(S)}{P(\bar{C})} \\
&= \frac{P(\bar{C} \mid S) \times P(S)}{P(\bar{C} \mid S) \times P(S) + P(\bar{C} \mid \bar{S}) \times P(\bar{S})} \\
&= \frac{(1-0.55) \times 0.50}{(1-0.55) \times 0.50 + (1-0.10) \times (1-0.50)} \\
&= \frac{0.225}{0.675} \\
&= \frac{1}{3}.
\end{aligned}
$$

Pergunta 2 2 valores

Uma engenheira biomédica admite que 5% dos testes de despistagem da *doença de Hansen* são positivos.

Calcule a probabilidade de a engenheira biomédica ter de consultar os resultados de pelo menos 14 testes, selecionados aleatoriamente e de forma independente, até encontrar um que seja positivo.

- **V.a. de interesse**

  X = resultados consultados até encontrar um teste que seja positivo

- **Distribuição e f.p. de X**

  $X \sim \text{geométrica}(p), \quad \text{onde } p = 0.05$

  $P(X = x) = (1-p)^{x-1}\, p, \quad x \in \mathbb{N}$

- **Prob. pedida**

$$\begin{aligned}
P(X \geq 14) &= 1 - P(X \leq 14-1) \\
&= \sum_{x=1}^{14-1} (1-p)^{x-1}\, p \\
&= 1 - p\,\frac{1-(1-p)^{14-1}}{1-(1-p)} \\
&= (1-p)^{14-1} \\
&= (1-0.05)^{13} \simeq 0.513342.
\end{aligned}$$

Pergunta 3	2 valores

Um engenheiro de telecomunicações admite que o tempo (em horas) entre duas recepções consecutivas de um sinal de determinado tipo é uma variável aleatória X com função de densidade de probabilidade dada por $f_X(x) = \frac{2x}{\lambda^2}\, e^{-\left(\frac{x}{\lambda}\right)^2}$, para $x > 0$, onde λ é uma constante positiva tal que a mediana de X é igual 1.

Calcule a probabilidade de este engenheiro ter de aguardar adicionalmente mais de 24 minutos, sabendo que não foi recebido qualquer sinal nos primeiros 36 minutos.

- **V.a. de interesse**

  X = tempo (em horas) entre duas recepções consecutivas de um sinal

- **F.d.p. e f.d. de** X

$$f_X(x) = \begin{cases} \frac{2x}{\lambda^2}\, e^{-\left(\frac{x}{\lambda}\right)^2}, & x > 0 \\ 0, & \text{c.c.} \end{cases}$$

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)\,dt = \begin{cases} \int_0^x \frac{2t}{\lambda^2}\, e^{-\left(\frac{t}{\lambda}\right)^2} = -e^{-\left(\frac{t}{\lambda}\right)^2}\Big|_0^x = 1 - e^{-\left(\frac{x}{\lambda}\right)^2}, & x > 0 \\ 0, & \text{c.c.} \end{cases}$$

- **Obtenção de** λ

$$\begin{aligned}
\lambda \quad : \quad & F_X(1) = 0.5 \\
& 1 - e^{-\left(\frac{1}{\lambda}\right)^2} = 0.5 \\
& \frac{1}{\lambda^2} = -\ln(0.5) \\
& \lambda = \frac{1}{\sqrt{-\ln(0.5)}} \simeq 1.201122
\end{aligned}$$

- **Prob. pedida**

  Como 24 (resp. 36) minutos correspondem a 0.4 (resp. 0.6) horas, obtemos

$$\begin{aligned}
P(X > 0.6 + 0.4 \mid X > 0.6) &= \frac{P(X > 0.6, X > 0.6 + 0.4)}{P(X > 0.6)} \\
&= \frac{P(X > 0.6 + 0.4)}{P(X > 0.6)} \\
&= \frac{1 - F_X(0.6 + 0.4)}{1 - F_X(0.6)} \\
&\simeq \frac{e^{-\left(\frac{0.6+0.4}{1.201122}\right)^2}}{e^{-\left(\frac{0.6}{1.201122}\right)^2}} \\
&= e^{-\frac{1}{1.201122^2}\left[(0.6+0.4)^2 - 0.6^2\right]}
\end{aligned}$$

$$P(X > 0.6 + 0.4 \mid X > 0.6) \quad \simeq \quad 0.641713.$$

Pergunta 4	2 valores

Considere que: X é a variável aleatória que indica o número de dias de semana em que ocorrem pedidos de reparação a uma oficina mecânica; e Y a variável aleatória que representa o número total semanal de pedidos de reparação que foram aceites pela oficina. Admita que o par aleatório (X, Y) possui função de probabilidade conjunta dada por

	Y		
X	4	5	6
5	$\frac{1}{10}$	$\frac{2}{15}$	$\frac{1}{10}$
6	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{2}{15}$
7	$\frac{2}{15}$	$\frac{1}{10}$	$\frac{1}{10}$

Calcule $E(Y \mid X = 7)$, isto é, o valor esperado do número total semanal de pedidos de reparação aceites pela oficina, sabendo que ocorreram pedidos de reparação em 7 dias da semana.

- **V.a. de interesse**

  $Y \mid X = 7$

- **F.p. marginal de** X

  $P(X = x) = \sum_y P(X = x, Y = y) = \begin{cases} \frac{1}{10} + \frac{2}{15} + \frac{1}{10} = \frac{1}{3}, & x = 5, 6, 7 \\ 0, & \text{outros valores de } x \end{cases}$

  [I.e., $X \sim \text{uniforme}(\{5, 6, 7\})$.]

- **F.p. de** $Y \mid X = 7$

  É sabido que $P(Y = y \mid X = x) = \frac{P(X=x, Y=y)}{P(X=x)}$. Logo, para $x = 7$, temos

$$P(Y = y \mid X = 7) \quad = \quad \begin{cases} \frac{\frac{2}{15}}{\frac{1}{3}} = 0.4, & y = 4 \\ \frac{\frac{1}{10}}{\frac{1}{3}} = 0.3, & y = 5, 6 \\ 0, & \text{outros valores de y.} \end{cases}$$

- **Valor esperado de** $Y \mid X = 7$

$$\begin{aligned} E(Y \mid X = 7) &= \sum_y y \times P(Y = y \mid X = 7) \\ &= 4 \times 0.4 + 5 \times 0.3 + 6 \times 0.3 \\ &= 4.9. \end{aligned}$$

Pergunta 5	2 valores

O tempo (em minutos) que uma viatura espera em determinado cruzamento é uma variável aleatória com distribuição exponencial com valor esperado igual a 2 minutos.

Calcule a probabilidade aproximada de o tempo total de espera de $n = 81$ viaturas seja superior a 172 minutos. Assuma que os tempos de espera das n viaturas são variáveis aleatórias independentes e identicamente distribuídas a X.

- **V.a.**

  $X_i =$ tempo de espera da viatura i, $\quad i = 1,\ldots,n$

  $n = 81$

- **Distribuição, valor esperado e variância comuns**

  $X_i \stackrel{i.i.d.}{\sim} X \sim \text{exponencial}(\frac{1}{2}), \quad i = 1,\ldots,n$

  $E(X_i) = E(X) = \mu \stackrel{form.}{=} 2, \quad i = 1,\ldots,n$

  $V(X_i) = V(X) = \sigma^2 \stackrel{form.}{=} 2^2 = 4, \quad i = 1,\ldots,n$

- **V.a. de interesse**

  $S_n = \sum_{i=1}^n X_i =$ tempo total de espera de n viaturas

- **Valor esperado e variância de S_n**

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

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

- **Distribuição aproximada de S_n**

  De acordo com o teorema do limite central (TLC), temos

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

- **Valor aproximado da probabilidade pedida**

  $$\begin{aligned} P(S_n > 172) &= P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} > \frac{172 - n\mu}{\sqrt{n\sigma^2}}\right) \\ &\stackrel{TLC}{\simeq} 1 - \Phi\left(\frac{172 - 81 \times 2}{\sqrt{81 \times 2^2}}\right) \\ &= 1 - \Phi(5/9) \\ &\simeq 1 - \Phi(0.56) \\ &\stackrel{tabelas,calc.}{=} 1 - 0.7123 \\ &= 0.2877. \end{aligned}$$

Pergunta 6 — 2 valores

Seja X a variável aleatória que contabiliza o tempo (em minutos) que um aluno leva a responder a uma pergunta de uma prova de avaliação. Admita que X segue uma distribuição exponencial com parâmetro desconhecido λ ($\lambda > 0$).

Calcule a estimativa de máxima verosimilhança de $P(X > 15 \mid X > 12.5)$, atendendo a que a concretização de uma amostra aleatória de 10 perguntas conduziu a um total de 115.2 minutos.

- **V.a. de interesse**

  $X =$ tempo que um aluno leva a responder a uma pergunta (em minutos)

- **Distribuição de X**

  $X \sim$ exponencial(λ)

- **Parâmetro desconhecido; espaço paramétrico**

  λ

  $\Theta = \mathbb{R}^+$

- **F.d.p.**

  $$f_X(x) \stackrel{form.}{=} \begin{cases} \lambda\, e^{-\lambda\, x}, & x \geq 0 \\ 0, & \text{c.c.} \end{cases}$$

- **Amostra**

  $$\underline{x} = (x_1, \ldots, x_n) \quad : \quad \begin{aligned} & n = 10 \\ & \sum_{i=1}^{n} x_i = 115.2 \end{aligned}$$

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

  Passo 1 — Função de verosimilhança

  $$\begin{aligned} L(\lambda \mid \underline{x}) &= f_{\underline{X}}(\underline{x}) \\ &\stackrel{X_i\ indep}{=} \prod_{i=1}^{n} f_{X_i}(x_i) \\ &\stackrel{X_i \sim X}{=} \prod_{i=1}^{n} f_X(x_i) \\ &= \prod_{i=1}^{n} \left(\lambda\, e^{-\lambda\, x_i}\right) \\ &= \lambda^n\, e^{-\lambda \sum_{i=1}^{n} x_i}, \quad \lambda \in \Theta = \mathbb{R}^+ \end{aligned}$$

  Passo 2 — Função de log-verosimilhança

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

  Passo 3 — Maximização

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

  $$\hat{\lambda} \quad : \quad \begin{cases} \left.\frac{d \ln L(\lambda|\underline{x})}{d\lambda}\right|_{\lambda=\hat{\lambda}} = 0 & \text{(ponto de estacionaridade)} \\ \left.\frac{d^2 \ln L(\lambda|\underline{x})}{d\lambda^2}\right|_{\lambda=\hat{\lambda}} < 0 & \text{(ponto de máximo).} \end{cases}$$

  $$\begin{cases} \frac{n}{\hat{\lambda}} - \sum_{i=1}^{n} x_i = 0 \\ -\frac{n}{\hat{\lambda}^2} < 0 \end{cases}$$

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

  Passo 4 — Concretização

  Para esta amostra tem-se

$$\begin{aligned}\hat{\lambda} &= \frac{10}{115.2}\\ &\simeq 0.086806.\end{aligned}$$

- **Outro parâmetro desconhecido**

  Invocando a propriedade de falta de memória da distribuição exponencial, temos

$$\begin{aligned}h(\lambda) &= P(X > 15 \mid X > 12.5)\\ &= P(X > 15 - 2.5)\\ &= e^{-\lambda\,(15-12.5)}\\ &= e^{-2.5\,\lambda}.\end{aligned}$$

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

  A propriedade de invariância dos estimadores de MV permite-nos concluir que a estimativa de MV de $h(\lambda)$ é

$$\begin{aligned}\widehat{h(\lambda)} &= h(\hat{\lambda})\\ &= e^{-2.5\,\hat{\lambda}}\\ &= e^{-2.5\times 0.086806}\\ &\simeq 0.804918.\end{aligned}$$

Pergunta 7	2 valores

Admita que a distribuição da altura X (em centímetros) de mulheres em determinado país é uma variável aleatória com distribuição normal com valor esperado desconhecido μ e variância desconhecida σ^2. O resultado de uma amostragem casual de $n = 11$ mulheres desse país conduziu à média amostral $\bar{x} = 166.7$ e à variância amostral corrigida $s^2 = 44.15$.

Determine um intervalo de confiança a 90% para σ^2.

- **V.a. de interesse**

  X = altura de mulher em determinado país

- **Situação**

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

  μ desconhecido

  σ^2 DESCONHECIDO

- **Obtenção de IC para σ^2**

  Passo 1 — Seleção da v.a. fulcral para σ^2

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

  Passo 2 — Obtenção dos quantis de probabilidade

$$\begin{cases} a_\alpha = F^{-1}_{\chi^2_{(n-1)}}(\alpha/2) = F^{-1}_{\chi^2_{(11-1)}}(0.1/2) = F^{-1}_{\chi^2_{(10)}}(0.05) \overset{tabela/calc.}{=} 3.940 \\ b_\alpha = F^{-1}_{\chi^2_{(n-1)}}(1-\alpha/2) = F^{-1}_{\chi^2_{(11-1)}}(1-0.1/2) = F^{-1}_{\chi^2_{(10)}}(0.95) \overset{tabela/calc.}{=} 18.31. \end{cases}$$

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

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

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

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

Passo 4 — Concretização

A expressão geral do IC para σ^2 é

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

Ao termos em conta a dimensão da amostra e os quantis e $s^2 = 44.15$ o IC pedido é

$$\begin{aligned} IC_{90\%}(\sigma^2) &\simeq \left[\frac{(11-1)\times 44.15}{18.31}, \frac{(11-1)\times 44.15}{3.940}\right] \\ &\simeq [24.1125,\ 112.0558]. \end{aligned}$$

Pergunta 8	2 valores

Um laboratório farmacêutico afirma que a vacina que produz garante imunidade a certa doença, dois meses após a administração da vacina, com probabilidade $p_0 = 0.7$. Para verificar esta afirmação, foram vacinadas $n = 400$ pessoas, sendo que dois meses depois 138 delas não estavam imunes a essa doença.

Confronte as hipóteses $H_0: p = p_0$ e $H_1: p < p_0$. Decida com base no valor-p aproximado.

- **V.a. de interesse**

  X = indicador de imunidade após vacinação

- **Situação**

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

  p DESCONHECIDO

- **Hipóteses**

  $H_0: p = p_0 = 0.7$

  $H_1: p < p_0$

- **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

  Teste unilateral inferior, logo a região de rejeição de H_0 é do tipo $W = (-\infty, c)$.

- **Decisão (com base no valor-p)**

  Atendendo a que

$$
\begin{aligned}
t &= \frac{\bar{x}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \\
&= \frac{\frac{400-138}{400}-0.7}{\sqrt{\frac{0.7\times(1-0.7)}{400}}} \\
&\simeq -1.96 \\
valor-p &= P\left(T < t \mid H_0\right) \\
&\simeq \Phi(t) \\
&\simeq \Phi(-1.96) \\
&= 1-\Phi(1.96) \\
&\overset{tabelas,calc.}{=} 1-0.9750 \\
&= 0.025,
\end{aligned}
$$

  devemos:

  – não rejeitar H_0 a qualquer n.s. $\alpha_0 \leq 2.5\%$, designadamente ao n.u.s. de 1%;

  – rejeitar H_0 a qualquer n.s. $\alpha_0 > 2.5\%$, por exemplo, aos n.u.s. de 5% e 10%.

Pergunta 9 2 valores

Numa fábrica de material eletrónico, o engenheiro de produção defende a hipótese H_0 de que a vida útil (em milhares de horas) das componentes produzidas é modelada pela variável aleatória X que segue uma distribuição exponencial com parâmetro $\lambda = 0.5$.

Uma amostra de $n = 400$ componentes foi selecionada aleatoriamente da produção da fábrica, tendo-se medido a vida útil de cada componente. A análise dos resultados obtidos conduziu à seguinte tabela de frequências:

Vida útil da componente	$]0, 0.5]$	$]0.5, 1.0]$	$]1.0, 1.5]$	$]1.5, 2.0]$	> 2.0
Frequência absoluta observada	98	78	58	36	130
Frequência absoluta esperada sob H_0	E_1	68.91	53.67	41.79	E_5

Após ter calculado as frequências absolutas esperadas E_1 e E_5 (aproximando-as às centésimas), averigue se H_0 é consistente com este conjunto de dados. Decida com base no valor-p aproximado.

- **V.a. de interesse**

  $X =$ vida útil (milhares de horas) da componente produzida

- **Hipóteses**

  $H_0 : X \sim \text{exponencial}(0.5)$

  $H_1 : X \not\sim \text{exponencial}(0.5)$

- **Estatística de teste**

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

onde:

- k = no. de classes = 5;
- O_i = freq. abs. observável da classe i;
- E_i = freq. abs. esperada sob H_0 da classe i;
- $\beta = 0$.

- **Frequência esperadas sob H_0 omissas**

$$\begin{aligned} E_1 &= n \times P(X \leq 0.5 \mid H_0) \\ &= 400 \times \left(1 - e^{\frac{-0.5}{2}}\right) \\ &\simeq 88.48 \\ E_5 &= n - \sum_{i=1}^{4} E_i \\ &= 400 - (88.48 + 68.91 + 53.67 + 41.79) \\ &= 147.15. \end{aligned}$$

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

  Ao lidarmos com 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 i	Freq. abs. obs. o_i	Freq. abs. esp. sob H_0 E_i	Parcelas valor obs. estat. teste $\frac{(o_i-E_i)^2}{E_i}$
1	$]0, 0.5]$	98	E_1	$\frac{(98-88.48)^2}{88.48} \simeq 1.0243$
2	$]0.5, 1.0]$	78	68.91	1.1991
3	$]1.5, 2.0]$	58	53.67	0.3493
4	$]2.0, 2.5]$	36	41.79	0.8022
5	> 2.5	130	147.15	1.9988
		$\sum_{i=1}^{k} o_i = n$ $= 400$	$\sum_{i=1}^{k} e_i = n$ $= 400$	$t = \sum_{i=1}^{k} \frac{(o_i-e_i)^2}{e_i}$ $\simeq 5.3737$

Dado que o valor observado da estatística de teste é $t = \sum_{i=1}^{k} \frac{(o_i-E_i)^2}{E_i} \simeq 5.3737$ e $W = (c, +\infty)$, obtemos

$$\begin{aligned} valor - p &= P(T > t \mid H_0) \\ &\simeq 1 - F_{\chi^2_{(k-1)}}(t) \\ &= 1 - F_{\chi^2_{(4)}}(5.3737) \\ &\simeq 0.251056 \end{aligned}$$

e devemos

- não rejeitar de H_0 a qualquer n.s. $\alpha_0 \leq valor - p = 25.1056\%$, designadamente aos n.u.s. $(1\%, 5\%, 10\%)$;
- rejeitar de H_0 a qualquer n.s. $\alpha_0 > valor - p = 25.1056\%$.

Alternativamente e recorrendo às tabelas de quantis da distribuição do qui-quadrado, podemos obter um intervalo para o valor-p deste teste:

$$
\begin{aligned}
F^{-1}_{\chi^2_{(4)}}(0.70) = 4.878 &< t = 5.3737 < 5.989 = F^{-1}_{\chi^2_{(4)}}(0.80)\\
0.70 &< F_{\chi^2_{(4)}}(5.3737) < 0.80\\
0.20 = 1-0.80 &< valor-p \simeq 1 - F_{\chi^2_{(4)}}(5.3737) < 1-0.70 = 0.30.
\end{aligned}
$$

Logo:

- não devemos rejeitar H_0 a qualquer n.s. $\alpha_0 \leq 20\%$, nomeadamente aos n.u.s. $(1\%, 5\%, 10\%)$;
- devemos rejeitar H_0 a qualquer n.s. $\alpha_0 \geq 30\%$.

Pergunta 10 2 valores

Considere os dados abaixo sobre diâmetro (x, em cm) e altura (Y, em m) de pinheiros da espécie *Pinus brutia*:

$$\sum_{i=1}^{10} x_i = 222, \quad \sum_{i=1}^{10} x_i^2 = 5498, \quad \sum_{i=1}^{10} y_i = 202, \quad \sum_{i=1}^{10} y_i^2 = 4135.82, \quad \sum_{i=1}^{10} x_i y_i = 4660.$$

Admita que as variáveis x e Y estão relacionadas de acordo com o modelo de regressão linear simples: $Y = \beta_0 + \beta_1 x + \varepsilon$.

Após ter enunciado as hipóteses de trabalho que entender convenientes, obtenha um intervalo de confiança a 99% para o valor esperado da diferença de altura de dois pinheiros cujo diâmetro difere de 10 cm.

- **Modelo de RLS**

  Y = diâmetro de pinheiro (v.a. resposta)

  x = altura de pinheiro (variável explicativa)

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

- **Hipóteses de trabalho**

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

- **Estimativas de MV de β_0 e β_1; estimativa de σ^2**

  Importa notar que

  - $n = 10$
  - $\sum_{i=1}^n x_i = 222$

    $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i = \frac{222}{10} = 22.2$

    $\sum_{i=1}^n x_i^2 = 5498$

    $\sum_{i=1}^n x_i^2 - n\,\bar{x}^2 = 5498 - 10 \times 22.2^2 = 569.6$
  - $\sum_{i=1}^n y_i = 202$

    $\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i = \frac{202}{10} \simeq 20.2$

    $\sum_{i=1}^n y_i^2 = 4135.82$

    $\sum_{i=1}^n y_i^2 - n\,\bar{y}^2 = 4135.82 - 10 \times 20.2^2 \simeq 55.42$
  - $\sum_{i=1}^n x_i y_i = 4660$

    $\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y} = 4660 - 10 \times 22.2 \times 20.2 = 175.6.$

Logo,

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y}}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2} \\ &= \frac{175.6}{569.6} \\ &\simeq 0.308287\end{aligned}$$

$$\begin{aligned}[\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1\,\bar{x} \\ &= 20.2 - 0.308287 \times 22.2 \\ &\simeq 13.356029]\end{aligned}$$

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{n-2}\left[\left(\sum_{i=1}^n y_i^2 - n\,\bar{y}^2\right) - \left(\hat{\beta}_1\right)^2\left(\sum_{i=1}^n x_i^2 - n\,\bar{x}^2\right)\right] \\ &\simeq \frac{1}{10-2}\left(55.42 - 0.308287^2 \times 569.6\right) \\ &\simeq 0.160590\end{aligned}$$

- **Obtenção do IC para β_1 e do IC pretendido**

Passo 1 — Seleção da v.a. fulcral

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

Passo 2 — Obtenção dos quantis de probabilidade

$$\begin{cases} a_\alpha = F_{t_{(n-2)}}(\alpha/2) = -F_{t_{(10-2)}}(1-0.01/2) = -F_{t_{(8)}}(0.995) \stackrel{tabelas,calc.}{=} -3.355 \\ b_\alpha = F_{t_{(n-2)}}(1-\alpha/2) = F_{t_{(8)}}(0.995) \stackrel{tabelas,calc.}{=} 3.355 \end{cases}$$

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

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

$$P\left[a_\alpha \le \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}}} \le b_\alpha\right] = 1-\alpha$$

$$P\left[\hat{\beta}_1 - b_\alpha \times \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}} \le \beta_1 \le \hat{\beta}_1 - a_\alpha \times \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}}\right] = 1-\alpha$$

Passo 4 — Concretização

Tendo em conta a expressão geral do IC para β_1,

$$IC_{(1-\alpha)\times 100\%}(\beta_1) = \left[\hat{\beta}_1 \pm F^{-1}_{t_{(n-2)}}(1-\alpha/2) \times \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}}\right],$$

e os resultados anteriores, o IC pretendido para

$$\begin{aligned}E(Y \mid x = x_0 + 10) - E(Y \mid x = x_0) &= \beta_0 + \beta_1 \times (x_0 + 10) - \beta_0 + \beta_1 \times x_0 \\ &= 10 \times \beta_1\end{aligned}$$

é dado por

$$\begin{aligned}
IC_{99\%}(10\times\beta_1) &= 10\times IC_{(1-\alpha)\times 100\%}(\beta_1)\\
&\simeq 10\times\left[0.308287\pm 3.355\times\sqrt{\frac{0.160590}{569.6}}\right]\\
&\simeq 10\times[0.308287-0.05633,\, 0.308287+0.05633]\\
&= 10\times[0.251953,\, 0.364621]\\
&= [2.51953,\, 3.64621].
\end{aligned}$$