DM DEPARTAMENTO DE MATEMÁTICA TÉCNICO LISBOA

Probabilidade e Estatística

LEBiol, LEIC-A, LEQ21; LEIC-T

2º Semestre – 2023/2024
18/06/2024
13:00–15:00

Duração: **120** minutos

Exame Época Normal – 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 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

A germinação de uma semente depende de dois fatores principais: temperatura ambiente e humidade do solo adequadas. Uma certa semente germina: com probabilidade 0.9 se ambos os fatores forem adequados; com probabilidade 0.3 se apenas um deles for adequado; e com probabilidade 0.1 se nenhum dos fatores for adequado. Considere que os dois fatores são independentes e que qualquer um deles é adequado com probabilidade 0.3.

Calcule a probabilidade de ambos os fatores terem sido adequados para uma semente que germinou.

- **Quadro de acontecimentos e probabilidades**

Acontecimento	Probabilidade
G = uma semente germinar	$P(G) = ?$
T = temperatura adequada	$P(T) = 0.3$
H = humidade adequada	$P(H) = 0.3$
$C_1 = T \cap H$	$P(C_1) \overset{T \perp\!\!\!\perp H}{=} P(T) \times P(H) = 0.3^2 = 0.09$
$C_2 = (\bar{T} \cap H) \cup (T \cap \bar{H})$	$P(C_2) = P(\bar{T} \cap H) + P(T \cap \bar{H}) = 2 \times 0.3 \times 0.7 = 0.42$
$C_3 = \bar{T} \cap \bar{H}$	$P(C_3) = P(\bar{T}) \times P(\bar{H}) = 0.7^2 = 0.49$
	$P(G \mid C_1) = 0.9$
	$P(G \mid C_2) = 0.3$
	$P(G \mid C_3) = 0.1$

- **Probabilidade pedida**

Uma vez que os cenários C_1, C_2 e C_3 são disjuntos e constituem uma partição do espaço de resultados Ω, temos, pelo teorema de Bayes,

$$
\begin{aligned}
P(C_1 \mid G) &= \frac{P(G \mid C_1) \times P(C_1)}{P(G)} = \frac{P(G \mid C_1) \times P(C_1)}{\sum_{i=1}^{3} P(G \mid C_i) \times P(C_i)} \\
&= \frac{0.9 \times 0.09}{0.9 \times 0.09 + 0.3 \times 0.42 + 0.1 \times 0.49} \\
&\simeq 0.316406.
\end{aligned}
$$

Pergunta 2 2 valores

Considere X uma variável aleatória uniforme discreta em $\{1, 2, \ldots, n\}$, onde n é um inteiro positivo.

Obtenha a variância de X, atendendo a que: $\sum_{x=1}^{n} x = n(n+1)/2$; $\sum_{x=1}^{n} x^2 = n(n+1)(2n+1)/6$.

- **V.a., sua distribuição e f.p.**

  $X =$ uniforme discreta$(\{1,2,\dots,n\})$

  $P(X=x) = \frac{1}{n}, \quad x = 1,2,\dots,n$

- **Valor esperado, segundo momento e variância de** X

$$\begin{aligned}
E(X) &= \sum_{x=1}^{n} x \times P(X=x) = \frac{\sum_{x=1}^{n} x}{n} = \frac{n(n+1)/2}{n} = \frac{n+1}{2} \\
E(X^2) &= \sum_{x=1}^{n} x^2\, P(X=x) = \frac{\sum_{x=1}^{n} x^2}{n} = \frac{n(n+1)(2n+1)/6}{n} = \frac{(n+1)(2n+1)}{6} \\
V(X) &= E(X^2) - E^2(X) = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \frac{n+1}{2}\left(\frac{2n+1}{3} - \frac{n+1}{2}\right) = \frac{n+1}{2}\,\frac{4n+2-3n-3}{6} \\
&= \frac{n^2-1}{12}.
\end{aligned}$$

Pergunta 3	2 valores

Seja X uma variável aleatória que representa o volume padronizado da descarga total anual de uma estação de tratamento de águas. Suponha que X possui função de distribuição

$$F_X(x) = \frac{1}{1+e^{-x}}, \quad x \in \mathbb{R}.$$

Calcule $P(-2 \le X \le 2 \mid -3 \le X \le 3)$.

- **V.a.**

  $X =$ volume padronizado da descarga total anual de uma estação de tratamento de águas

- **F.d.**

  $F_X(x) = P(X \le x) = \frac{1}{1+e^{-x}}, \quad x \in \mathbb{R}$

- **Probabilidade pedida**

$$\begin{aligned}
P(-2 \le X \le 2 \mid -3 \le X \le 3) &= \frac{P(-2 \le X \le 2, -3 \le X \le 3)}{P(-3 \le X \le 3)} = \frac{P(-2 \le X \le 2)}{P(-3 \le X \le 3)} = \frac{F_X(2) - F_X(-2)}{F_X(3) - F_X(-3)} \\
&= \frac{\frac{1}{1+e^{-2}} - \frac{1}{1+e^{2}}}{\frac{1}{1+e^{-3}} - \frac{1}{1+e^{3}}} \qquad \left[= \frac{1+e+e^3+e^4}{1+e+2e^2+e^3+e^4}\right] \\
&\simeq 0.841403.
\end{aligned}$$

Pergunta 4	2 valores

Sejam X e Y as variáveis aleatórias que representam o atraso à partida e à chegada (em horas e arredondado para valores inteiros) de certo voo de longo curso de uma companhia aérea. Os registos da companhia mostram que a função de probabilidade conjunta de (X,Y) é dada pela tabela à direita.

	Y		
X	0	1	2
0	0.1	0.2	0.1
1	0.1	0.0	0.3
2	0.1	0.1	0.0

Calcule a variância de Y condicional a $X=0$, $V(Y \mid X=0)$.

- **Par aleatório** (X,Y)

  $X =$ atraso à partida

  $Y =$ atraso à chegada

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

  $Y \mid X = 0$

$$P(Y=y \mid X=0) = \frac{P(X=0,Y=y)}{P(X=0)} = \frac{P(X=0,Y=y)}{\sum_{y=0}^{2} P(X=0,Y=y)} = \begin{cases} \frac{0.1}{0.1+0.2+0.1} = \frac{1}{4}, & y=0 \\ \frac{0.2}{0.4} = \frac{1}{2}, & y=1 \\ \frac{0.1}{0.4} = \frac{1}{4}, & y=2 \\ 0, & \text{outros valores de } y \end{cases}$$

- **Valor esperado, segundo momento e variância de** $Y \mid X = 0$

$$\begin{aligned} E(Y \mid X=0) &= \sum_{y=0}^{2} y\, P(Y=y \mid X=0) = 0 \times \frac{1}{4} + 1 \times \frac{1}{2} + 2 \times \frac{1}{4} = 1 \\ E(Y^2 \mid X=0) &= \sum_{y=0}^{2} y^2\, P(Y=y \mid X=0) = 0^2 \times \frac{1}{4} + 1^2 \times \frac{1}{2} + 2^{\text{"}} \times \frac{1}{4} = \frac{3}{2} \\ V(Y \mid X=0) &= E(Y^2 \mid X=0) - E^2(Y \mid X=0) = \frac{3}{2} - 1^2 = 1/2. \end{aligned}$$

Pergunta 5 2 valores

O responsável por um laboratório químico admite que qualquer peça de material de vidro que é encomendada tem uma probabilidade de 0.05 de se partir durante o transporte e que qualquer peça se pode partir independentemente das restantes. São aguardadas três encomendas de dimensões 20, 30 e 50 de três fornecedores distintos que operam independentemente.

Indique o número mais provável de peças partidas no conjunto de peças que está para chegar?

A: 6 **B**: 5 **C**: 3 **D**: 4

- **V.a. e distribuições**

  X_i = número de peças partidas na encomenda ao fornecedor i, $i = 1,2,3$

  $X_i \overset{indep.}{\sim}$ binomial(n_i, p), $n_1 = 20$, $n_2 = 30$, $n_3 = 50$, $p = 0.05$

- **V.a. de interesse e distribuição exacta**

  $Y = \sum_{i=1}^{3} X_i$ = número total de peças partidas nas três encomendas

  $Y \sim$ binomial$\left(n = \sum_{i=1}^{3} n_i = 100, p = 0.05\right)$, por admitirmos independência entre os diferentes fornecedores.

- **Moda de** Y

  É sabido que $mo = mo(Y) = \text{argmax}_y P(Y=y) = \text{argmax}_{y \in \{0,1,\dots,n\}} \frac{n!}{y!(n-y)!}\, p^y\, (1-p)^{n-y}$. Logo,

$$mo = mo(X) \in \{0,1,\dots,n\} \quad : \quad \begin{cases} P(Y=mo) \geq P(Y=mo-1) \\ P(Y=mo) \geq P(Y=mo+1) \end{cases} \quad \begin{cases} \frac{P(Y=mo)}{P(Y=mo-1)} \geq 1 \\ \frac{P(Y=mo+1)}{P(Y=mo)} \leq 1 \end{cases}$$

$$\begin{cases} \dfrac{\frac{n!}{mo!(n-mo)!}\, p^{mo}\, (1-p)^{n-mo}}{\frac{n!}{(mo-1)!(n-(mo-1))!}\, p^{mo-1}\, (1-p)^{n-(mo-1)}} \geq 1 \\ \dfrac{\frac{n!}{(m0+1)!(n-(mo+1))!}\, p^{mo+1}\, (1-p)^{n-(mo+1)}}{\frac{n!}{mo!(n-mo)!}\, p^{mo}\, (1-p)^{n-mo}} \leq 1 \end{cases} \quad \begin{cases} \frac{n-(mo-1)}{mo} \times \frac{p}{1-p} \geq 1 \\ \frac{n-mo}{mo+1} \times \frac{p}{1-p} \leq 1 \end{cases}$$

$$\begin{cases} mo \leq (n+1)\, p = 5.05 \\ mo \geq (n+1)\, p - 1 = 4.05, \end{cases}$$

  i.e., $mo = mo(Y) = 5$.

Pergunta 6	2 valores

Suponha que X representa o tempo (em segundos) entre duas chegadas consecutivas de pedidos a um servidor. Admita que a função de densidade de probabilidade de X é dada por $f_X(x) = \lambda^2\, x\, e^{-\lambda x}$, para $x > 0$, onde λ é uma constante positiva desconhecida.

Obtenha a estimativa de máxima verosimilhança de $P(X > 2) = e^{-2\lambda} + 2\lambda\, e^{-2\lambda}$ com base numa amostra casual $(x_1, \ldots, x_n)$ de dimensão $n = 20$ e média $\bar{x} = 4.58$.

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

  X = tempo (em segundos) entre duas chegadas consecutivas de pedidos a um servidor

  $$f_X(x) = \begin{cases} \lambda^2\, x\, e^{-\lambda x}, & x > 0 \\ 0, & \text{outros valores de } x \end{cases}$$

- **Amostra**

  $\underline{x} = (x_1, \ldots, x_n)$, amostra de dimensão $n = 20$ proveniente de X e tal que $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i = 4.58$.

- **Parâmetro desconhecido**

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

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

  Passo 1 — Função de verosimilhança

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

  Passo 2 — Função de log-verosimilhança

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

  Passo 3 — Maximização

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

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

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

  Passo 4 — Estimativa de MV de λ

  $$\hat{\lambda} = \frac{2n}{\sum_{i=1}^n x_i} = \frac{2}{\bar{x}} = \frac{2}{4.58} \simeq 0.436681$$

- **Outro parâmetro desconhecido**

  $h(\lambda) = P(X > 2) = e^{-2\lambda} + 2\lambda\, e^{-2\lambda}$

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

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

  $$\widehat{h(\lambda)} = h(\hat{\lambda}) = e^{-2\hat{\lambda}} + 2\hat{\lambda}\, e^{-2\hat{\lambda}} \simeq e^{-2\times 0.436681} + 2 \times 0.436681\, e^{-2\times 0.436681} \simeq 0.782214.$$

Pergunta 7	2 valores

Uma engenheira pretende estimar o valor esperado μ da duração (em horas) de determinada reação química. Uma amostra casual de 40 dessas reacções resultou numa duração total de 804 (horas) e numa variância amostral $s^2 \simeq 355.938$ (horas^2).

Deduza um intervalo de confiança aproximado a 99% para μ.

- **V.a. de interesse**

  X = duração de uma reação química

- **Situação**

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

- **Obtenção do IC aproximado para μ**

  Passo 1 — Variável aleatória fulcral para μ

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

  Passo 2 — Quantis de probabilidade

  Dado que $(1-\alpha) \times 100\% = 99\%$, usaremos os quantis

$$(a_\alpha, b_\alpha) \quad : \quad \begin{cases} P(a_\alpha \leq Z \leq b_\alpha) = 1 - \alpha \\ P(Z < a_\alpha) = P(Z > b_\alpha) = \alpha/2 \end{cases}$$

$$\begin{cases} a_\alpha = \Phi^{-1}(\alpha/2) = -\Phi^{-1}(1-\alpha/2) \overset{tab.,calc.}{=} -2.5758 \\ b_\alpha = \Phi^{-1}(1-\alpha/2) = \Phi^{-1}(0.995) \overset{tab.,calc.}{=} 2.5758. \end{cases}$$

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

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

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

$$P\left[\bar{X} - b_\alpha \times \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} - a_\alpha \times \frac{S}{\sqrt{n}}\right] \simeq 1 - \alpha$$

  Passo 4 — Concretização

  Atendendo à expressão geral do IC para μ,

$$IC_{(1-\alpha)\times 100\%}(\mu) \quad = \quad \left[\bar{x} \pm \Phi^{-1}(1-\alpha/2) \times \frac{s}{\sqrt{n}}\right],$$

  ao par de quantis acima e ao facto de $\bar{x} = \frac{804}{40} = 20.1$ e $s^2 \simeq 355.938$, temos:

$$\begin{aligned} IC_{99\%}(\mu) &\simeq \left[20.1 - 2.5758 \times \frac{\sqrt{355.938}}{\sqrt{40}}, 20.1 + 2.5758 \times \frac{\sqrt{355.938}}{\sqrt{40}}\right] \\ &\simeq [12.4163, 27.7837]. \end{aligned}$$

Pergunta 8	2 valores

O fabricante de uma marca de silenciadores de automóveis supõe que a vida útil dos seus silenciadores (X, em anos) segue uma distribuição normal com valor esperado e variância, μ e σ^2, desconhecidos. Uma amostra casual de 16 desses silenciadores possui um desvio padrão de $s = 1$ ano.

Teste a hipótese $H_0 : \sigma^2 = 0.81$ contra $H_1 : \sigma^2 > 0.81$. Obtenha o valor-p (ou um intervalo para o valor-p) do teste. Para que níveis de significância não se deve rejeitar H_0?

- **V.a. de interesse**

  X = vida útil dos seus silenciadores

- **Situação**

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

  μ desconhecido

  σ DESCONHECIDO

- **Hipóteses**

  $H_0 : \sigma^2 = \sigma_0^2 = 0.81$ vs. $H_1 : \sigma^2 > \sigma_0^2 = 0.81$

- **Estatística de teste**

  $$\frac{(n-1)\,S^2}{\sigma_0^2} \sim_{H_0} \chi^2_{(n-1)}$$

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

  O teste é unilateral superior ($H_1 : \sigma^2 > \sigma_0^2$), logo a região de rejeição de H_0 é do tipo $W = (c, +\infty)$.

- **Decisão (com base no valor-p)**

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

  $$\begin{aligned} t &= \frac{(n-1)\,s^2}{\sigma_0^2} = \frac{(16-1)\times 1}{0.81} \simeq 18.5185 \\ valor-p &= P(T > t \mid H_0) = 1 - F_{\chi^2_{(n-1)}}(t) = 1 - F_{\chi^2_{(15)}}(18.5185) \overset{calc.}{=} 0.2364, \end{aligned}$$

  é suposto:

  - não rejeitar H_0 a qualquer n.s. $\alpha_0 \leq valor-p = 23.64\%$, designadamente aos níveis de significância habituais (1%, 5%, 10%), [ou seja, há forte evidência a favor de H_0];
  - rejeitar H_0 a qualquer n.s. $\alpha_0 > valor-p = 0.35\%$].

  Alternativamente, poderíamos recorrer às tabelas de quantis da distribuição do qui-quadrado e adiantar um intervalo para o valor-p:

  $$\begin{aligned} F^{-1}_{\chi^2_{(15)}}(0.7) = 17.32 &< t = 18.5185 < 19.31 = F^{-1}_{\chi^2_{(15)}}(0.8) \\ 0.7 &< F_{\chi^2_{(15)}}(18.5185) < 0.8 \\ 1 - 0.8 &< 1 - F_{\chi^2_{(15)}}(18.5185) < 1 - 0.7 \\ 0.2 &< valor-p < 0.3. \end{aligned}$$

  Assim, é suposto:

  - não rejeitar H_0 a qualquer n.s. $\alpha_0 \leq 20\%$, nomeadamente aos n.u.s. (1%, 5%, 10%);
  - [rejeitar H_0 a qualquer n.s. $\alpha_0 \geq 0.5\%$, designadamente aos n.u.s. (1%, 5%, 10%)].

Pergunta 9	2 valores

Suponha que uma região comercial foi dividida em cinco áreas. Um gestor pretende testar a hipótese H_0 de os clientes se distribuirem uniformemente por estas cinco áreas. Tendo 500 clientes sido selecionados casualmente e inquiridos sobre a área de proveniência, obtiveram-se os valores que constam da tabela abaixo.

Área	A	B	C	D	E
Frequência absoluta observada	110	130	70	90	100

Com base no teste de ajustamento do qui-quadrado, que decisão deverá tomar o gestor?

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 = área de proveniência do cliente

- **Hipóteses**

  $H_0 : X \sim$ uniforme discreta$(\{A, B, C, D, E\})$ vs. $H_0 : X \not\sim$ uniforme discreta$(\{A, B, C, D, E\})$

- **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)**

  Ao con

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	A	110	100	$\frac{(110-100)^2}{100} = 1$
2	B	130	100	$\frac{(130-100)^2}{100} = 9$
3	C	70	100	$\frac{(70-100)^2}{100} = 9$
4	D	90	100	$\frac{(90-100)^2}{100} = 1$
5	E	100	100	$\frac{(100-100)^2}{100} = 0$
		$\sum_{i=1}^k o_i = n = 500$	$\sum_{i=1}^k E_i = n = 500$	$t = \sum_{i=1}^k \frac{(o_i-E_i)^2}{E_i} = 20$

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

$$valor-p \quad = \quad P(T > t =| H_0) \simeq 1 - F_{\chi^2_{(5-1)}}(20) \overset{tabelas, calc.}{\simeq} 1 - 0.9995 = 0.0005.$$

  Assim sendo, devemos decidir pela:

  - rejeição de H_0 a qualquer n.s. $\alpha_0 > 0.05\%$, designadamente a todos n.u.s. de 1%, 5%e 10%;
  - [não rejeição de H_0 a qualquer n.s. $\alpha_0 \leq 0.05\%$].

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

  - **A**: Rejeitar H_0 a 1%, 5% e 10%.

Pergunta 10 2 valores

Uma empresa que produz fertilizantes realizou uma experiência visando estudar a relação entre a quantidade de fertilizante (x) e um índice de produção de milho (Y), tendo obtido os resultados seguintes:

$$\sum_{i=1}^{8} x_i = 10.8; \quad \sum_{i=1}^{8} x_i^2 = 18.36; \quad \sum_{i=1}^{8} y_i = 240; \quad \sum_{i=1}^{8} y_i^2 = 8\,500; \quad \sum_{i=1}^{8} x_i y_i = 385.5.$$

Admitindo a validade do modelo de regressão linear simples, $Y = \beta_0 + \beta_1 x + \epsilon$, e tendo em conta estas quantidades e as hipóteses de trabalho convenientes, teste a hipótese $H_0 : \beta_1 = 0$ contra $H_1 : \beta_1 \neq 0$, ao nível de significância de 1%. Interprete o resultado do teste.

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

  Y = índice de produção de milho (v.a. resposta)

  x = quantidade de fertilizante (variável explicativa)

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

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

- **Hipóteses**

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

  $H_1 : \beta_1 \neq \beta_{1,0}$

- **Nível de significância**

  $\alpha_0 = 1\%$

- **Estatística de teste**

$$T = \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}}} \sim_{H_0} t_{(n-2)}$$

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

  Estamos a lidar com um teste bilateral ($H_1 : \beta_1 \neq \beta_{1,0}$), logo a região de rejeição de H_0 é do tipo $W = (-\infty, -c) \cup (c, +\infty)$, onde $c : P(\text{Rejeitar } H_0 \mid H_0) = \alpha_0$, i.e.,

$$c = F^{-1}_{t_{(n-2)}}(1-\alpha_0/2) = F^{-1}_{t_{(8-2)}}(1-0.01/2) = F^{-1}_{t_{(6)}}(0.995) \overset{tabela/calc.}{\simeq} 3.707.$$

- **Cálculos auxiliares**

  Importa notar que temos

  - $n = 8$
  - $\sum_{i=1}^n x_i = 10.8$

    $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i = \frac{10.8}{8} = 1.35$

    $\sum_{i=1}^n x_i^2 = 18.36$

    $\sum_{i=1}^n x_i^2 - n\,\bar{x}^2 = 18.36 - 8 \times 1.35^2 = 3.78$
  - $\sum_{i=1}^n y_i = 240$

    $\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i = \frac{240}{8} = 30$

    $\sum_{i=1}^n y_i^2 = 8\,500$

    $\sum_{i=1}^n y_i^2 - n\,\bar{y}^2 = 8\,500 - 8 \times 30^2 = 1\,300$
  - $\sum_{i=1}^n x_i y_i = 385.5$

    $\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y} = 385.5 - 8 \times 1.35 \times 30 = 61.5.$

  Consequentemente, as estimativas de β_1 e σ^2 são dadas por:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{p}}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2} = \frac{61.5}{3.78} \simeq 16.269841;$$

$$\hat{\sigma}^2 = \frac{1}{n-2}\left[\left(\sum_{i=1}^n y_i^2 - n\,\bar{y}^2\right) - (\hat{\beta}_1)^2\left(\sum_{i=1}^n x_i^2 - n\,\bar{x}^2\right)\right] \simeq \frac{1}{8-2}\left[1\,300 - (16.269841)^2 \times 3.78\right]$$

$$\simeq 49.900799.$$

- **Decisão**

  O valor observado da estatística de teste é igual a

$$
\begin{aligned}
t &= \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}}} \\
&\simeq \frac{16.269841 - 0}{\sqrt{\frac{49.900799}{3.78}}} \\
&\simeq 4.477911.
\end{aligned}
$$

  Como $t \simeq 4.477911 \in W = (-\infty, -3.707) \cup (3.707, +\infty)$ devemos rejeitar H_0 ao n.s. de $\alpha_0 = 5\%$[, assim como a qualquer n.s. superior a 5%.]

- **Interpretação pedida**

  Há forte evidência a favor de H_1, i.e., da hipótese de a quantidade de fertilizante influenciar o índice de produção de milho.