Probabilidades e Estatística / Introd. às Probabilidades e Estatística

TODOS OS CURSOS

Exame Época Especial
2017/2018
23/07/2017 – **09:00**

Duração: 3 horas

Justifique convenientemente todas as respostas

Grupo I	5 valores

1. Uma fábrica recebe componentes eletrónicas dos fornecedores A e B. De acordo com os registos desta fábrica, 80% das componentes são fornecidas pelo fornecedor A e 20% pelo fornecedor B. Sabe-se que, das componentes fornecidas por A (respetivamente por B), 5% são defeituosas (respetivamente 20% são defeituosas). Admitindo que se selecionou ao acaso uma componente eletrónica:

(a) Obtenha a probabilidade de ela ser defeituosa. (1.5)

- **Quadro de acontecimentos e probabilidades**

Acontecimento	**Probabilidade**
$A = \{$componente eletrónica do fornececedor$A\}$	$P(A) = 0.8$
$B = \{$componente eletrónica do fornececedor $B\}$	$P(B) = 0.2$
$D = \{$componente eletrónica selecionada é defeituosa$\}$	$P(D) = ?$
	$P(D \mid A) = 0.05$
	$P(D \mid B) = 0.2$

- **Probabilidade pedida**
  Tirando partido da lei da probabilidade total, segue-se

$$\begin{aligned} P(D) &= P(D \mid A) \times P(A) + P(D \mid B) \times P(B) \\ &= 0.05 \times 0.8 + 0.2 \times 0.2 \\ &= 0.08. \end{aligned}$$

(b) Determine a probabilidade de ela ter sido fornecida por B, sabendo que não é defeituosa. (1.0)

- **Probabilidade pedida**
  Invocando o teorema de Bayes, tem-se

$$\begin{aligned} P(B \mid \overline{D}) &= \frac{P(\overline{D} \mid B) \times P(B)}{P(\overline{D})} \\ &= \frac{(1-0.2) \times 0.2}{1-0.08} \\ &\simeq 0.173913. \end{aligned}$$

2. A ocorrência de falhas de determinado transmissor rege-se de acordo com um processo de Poisson com taxa igual a 2 falhas por hora.

(a) Calcule a probabilidade de se verificarem mais de 29 falhas, deste transmissor, em 12.5 horas. (1.0)

- **V.a. de interesse**
  $X_t =$ número de falhas do transmissor em t horas de funcionamento $\quad (t > 0)$

- **Distribuição da variável aleatória X_t**
  Dado que lidamos com um processo de Poisson com taxa igual a 2 falhas por hora, temos $X_t \sim \text{Poisson}(2 \times t)$.

- **F.p. de $X_{12.5}$**
  $P(X_{12.5} = x) = \frac{e^{-2\times 12.5}(2\times 12.5)^x}{x!} = \frac{e^{-25}\,25^x}{x!}, \quad x = 0, 1, 2, \ldots$

- **Probabilidade pedida**

$$\begin{aligned} P(X_{12.5} > 29) &= 1 - P(X_{12.5} \le 29) \\ &= 1 - F_{Poisson(25)}(29) \\ &\overset{tabela/calc.}{\simeq} 1 - 0.8179 \\ &\simeq 0.1821. \end{aligned}$$

(b) Qual é a probabilidade de o tempo entre os instantes de ocorrência de duas falhas consecutivas deste transmissor ser superior a 20 minutos? (1.5)

- **Variável aleatória de interesse**
  T = tempo (em horas) entre duas falhas consecutivas do transmissor

- **Distribuição de T**
  Lidamos com um processo de Poisson com taxa de 2 falhas do transmissor por hora, logo $T \sim \text{Exponencial}(\lambda)$ com $\lambda = 2$.

- **F.d.p. de T**
  $$f_T(t) = \begin{cases} 0, & t < 0 \\ 2 \times e^{-2t}, & t \ge 0 \end{cases}$$

- **Probabilidade pedida**

$$\begin{aligned} P(T > 1/3) &= \int_{1/3}^{+\infty} 2 \times e^{-2t}\,dt \\ &= -e^{-2t}\Big|_{1/3}^{+\infty} \\ &= e^{-2/3} \\ &\simeq 0.5134. \end{aligned}$$

[Em alternativa, $P(T > 1/3) = P(X_{1/3} = 0) = \frac{e^{-2\times 1/3}(2\times 1/3)^0}{0!} = e^{-2/3} \simeq 0.5134$ com $X_{1/3}$ é a v.a. que descreve o número de falhas do transmissor num período de $1/3$ hora e $X_{1/3} \sim \text{Poisson}(2 \times 1/3)$.]

Grupo II	5 valores

1. O tempo de execução de um algoritmo, em minutos, é representado pela variável aleatória X com função densidade de probabilidade

$$f_X(x) = \begin{cases} 2x^2, & 0 \le x \le 1 \\ 1, & 1 < x \le 4/3 \\ 0, & \text{caso contrário.} \end{cases}$$

(a) Qual é a probabilidade de X não exceder 1 minuto? (0.5)

- **V.a. de interesse**
  X = tempo de execução de um algoritmo

- **F.d.p. de X**
  $$f_X(x) = \begin{cases} 2x^2, & 0 \le x \le 1 \\ 1, & 1 < x \le 4/3 \\ 0, & \text{caso contrário.} \end{cases}$$

- **Prob. pedida**

$$\begin{aligned} P(X \le 1) &= \int_{-\infty}^{1} f_X(x)\,dx \\ &= \int_{0}^{1} 2x^2\,dx \end{aligned}$$

$$\begin{aligned} P(X \le 1) &= \left. \frac{2x^3}{3} \right|_0^1 \\ &= \frac{2}{3}. \end{aligned}$$

(b) Obtenha a mediana de X. (1.0)

- **Mediana de X**
  A mediana de X será representada por $me(X) = me$ e verifica a condição $F_X(me) = \frac{1}{2}$.
  Na alínea anterior, constatou-se que $P(X \le 1) = F_X(1) = \frac{2}{3}$. Mais ainda, como $F_X(me) = 1/2 < 2/3 = F_X(1)$, conclui-se que $me < 1$. Logo,

$$\begin{aligned} me = me(X) \in (0,1) \quad &: \quad F_X(me) = \frac{1}{2} \\ & \int_{-\infty}^{me} f_X(x)\,dx = \frac{1}{2} \\ & \int_0^{me} 2x^2\,dx = \frac{1}{2} \\ & \frac{2}{3} \times (me)^3 = \frac{1}{2} \\ & me = \left(\frac{3}{4}\right)^{\frac{1}{3}} \\ & me \simeq 0.908561. \end{aligned}$$

(c) Sabendo que $E(X) = \frac{8}{9}$ e $V(X) = \frac{1}{15}$, determine um valor aproximado para a probabilidade de o tempo médio de 100 execuções do algoritmo exceder 0.9 minuto. Suponha que os tempos de execução são variáveis aleatórias independentes e identicamente distribuídas a X. (1.5)

- **V.a.**
  X_i = tempo da i-ésima execução do algoritmo, onde $i = 1, \ldots, n$
  $n = 100$

- **Distribuição, valor esperado e variância comuns**
  $X_i \stackrel{i.i.d.}{\sim} X, \quad i = 1, \ldots, n$
  $E(X_i) = E(X) = \mu = \frac{8}{9}, \quad i = 1, \ldots, n$
  $V(X_i) = V(X) = \sigma^2 = \frac{1}{15}, \quad i = 1, \ldots, n$

- **V.a. de interesse**
  $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ = tempo médio de n execuções do algoritmo

- **Valor esperado e variância de $\bar{X}$**
  $E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n}\sum_{i=1}^n E(X_i) \stackrel{X_i \sim X}{=} \frac{1}{n} \times nE(X) = E(X) = \mu$

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

- **Distribuição aproximada de $\bar{X}$**
  Pelo teorema do limite central (TLC) pode escrever-se

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

- **Valor aproximado da probabilidade pedida**

$$\begin{aligned} P(\bar{X} > 0.9) &= 1 - P(\bar{X} \le 0.9) \\ &= 1 - P\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \le \frac{0.9 - \mu}{\frac{\sigma}{\sqrt{n}}}\right) \end{aligned}$$

$$
\begin{aligned}
P(\bar{X} > 0.9) & \overset{TLC}{\simeq} 1 - \Phi\left(\frac{0.9 - \frac{8}{9}}{\sqrt{\frac{1}{1500}}}\right) \\
& \simeq 1 - \Phi(0.43) \\
& \overset{tabela/calc}{=} 1 - 0.6664 \\
& = 0.3336.
\end{aligned}
$$

2. Seja (X, Y) um par aleatório, em que X e Y representam o diâmetro interior e o comprimento de um pino-guia, respetivamente. Admita que a função de densidade de probabilidade conjunta de (X, Y) é dada por

$$
f_{X,Y}(x,y) = \begin{cases} \frac{6}{7}\left(x^2 + \frac{xy}{2}\right), & 0 < x < 1, \quad 0 < y < 2 \\ 0, & \text{caso contrário.} \end{cases}
$$

(a) Determine $P(X > Y)$. (1.0)

- **Par aleatório** (X, Y)
  X = diâmetro interior de um pino-guia
  Y = comprimento de um pino-guia

- **F.d.p. do par aleatório** (X, Y)
  $$
  f_{X,Y}(x,y) = \begin{cases} \frac{6}{7}\left(x^2 + \frac{xy}{2}\right), & 0 < x < 1, \quad 0 < y < 2 \\ 0, & \text{caso contrário.} \end{cases}
  $$

- **Prob. pedida**
  $$
  \begin{aligned}
  P(X > Y) & = \int_0^1 \int_0^x f_{X,Y}(x,y)\, dy\, dx \\
  & = \int_0^1 \int_0^x \frac{6}{7}\left(x^2 + \frac{xy}{2}\right) dy\, dx \\
  & = \int_0^1 \left(\frac{6x^2}{7} \times y + \frac{6x}{14} \times \frac{y^2}{2}\right)\Bigg|_0^x dx \\
  & = \int_0^1 \left(\frac{6x^2}{7} \times x + \frac{6x}{14} \times \frac{x^2}{2}\right) dx \\
  & = \int_0^1 \frac{15x^3}{14}\, dx \\
  & = \frac{15}{14} \times \frac{x^4}{4}\Bigg|_0^1 \\
  & = \frac{15}{56}.
  \end{aligned}
  $$

(b) Comprove que as variáveis aleatórias X e Y são dependentes. (1.0)

- **F.d.p. marginal de** X
  Para $0 < x < 1$, tem-se
  $$
  \begin{aligned}
  f_X(x) & = \int_{\mathbb{R}} f_{X,Y}(x,y)\, dy \\
  & = \int_0^2 \frac{6}{7}\left(x^2 + \frac{xy}{2}\right) dy \\
  & = \left(\frac{6x^2}{7} \times y + \frac{6x}{14} \times \frac{y^2}{2}\right)\Bigg|_0^2 \\
  & = \frac{12\,x^2 + 6\,x}{7}.
  \end{aligned}
  $$

- **F.d.p. marginal de Y**
  Para $0 < y < 2$, tem-se

$$\begin{aligned} f_Y(y) &= \int_{\mathbb{R}} f_{X,Y}(x,y)\,dx \\ &= \int_0^1 \frac{6}{7}\left(x^2 + \frac{xy}{2}\right) dx \\ &= \frac{6}{7} \times \left.\frac{x^3}{3}\right|_0^1 + \frac{6y}{14} \times \left.\frac{x^2}{2}\right|_0^1 \\ &= \frac{2}{7} + \frac{3y}{14}. \end{aligned}$$

- **(In)dependência entre as variáveis X e Y**
  X e Y são v.a. DEPENDENTES sse

$$f_{X,Y}(x,y) \neq f_X(x) \times f_Y(y), \quad \text{para algum } (x,y) \in \mathbb{R}^2.$$

  Ora, por um lado

$$\begin{aligned} f_{X,Y}\left(\frac{1}{2},\frac{1}{2}\right) &= \frac{6}{7} \times \left(\frac{1}{2}^2 + \frac{\frac{1}{2}\times\frac{1}{2}}{2}\right) \\ &= \frac{9}{28} \qquad [\simeq 0.321429]. \end{aligned}$$

  e por outro

$$\begin{aligned} f_X\left(\frac{1}{2}\right) \times f_Y\left(\frac{1}{2}\right) &= \frac{12 \times \frac{1}{2}^2 + 6 \times \frac{1}{2}}{7} \times \left(\frac{2}{7} + \frac{3 \times \frac{1}{2}}{14}\right) \\ &= \frac{6}{7} \times \frac{11}{28} \\ &= \frac{33}{98} \qquad [\simeq 0.336735]. \end{aligned}$$

  Assim, conclui-se que

$$f_{X,Y}(0.5, 0.5) \neq f_X(0.5) \times f_Y(0.5).$$

  pelo que X e Y são efectivamente v.a. DEPENDENTES.

Grupo III	5 valores

1. Admita que a leitura do rendimento de um processo químico (em ppm) é representado pela variável aleatória X com função de densidade de probabilidade igual a

$$f_X(x) = \begin{cases} \frac{1}{\theta}\, e^{-\frac{x-90}{\theta}}, & x \geq 90 \\ 0, & \text{caso contrário,} \end{cases}$$

onde θ é um parâmetro positivo desconhecido.

(a) Deduza o estimador de máxima verosimilhança de θ, com base numa amostra aleatória $(X_1, \ldots, X_n)$ associada à leitura do rendimento deste processo químico, em n dias consecutivos. (1.5)

- **V.a. de interesse**
  X = leitura do rendimento de um processo químico
- **F.d.p. de X**

$$f_X(x) = \begin{cases} \theta^{-1}\, \exp\left(\frac{90-x}{\theta}\right), & x \geq 90 \\ 0, & \text{caso contrário} \end{cases}$$

- **Parâmetro desconhecido**
  $\theta, \quad \theta > 0$

- **Amostra**
  $\underline{x} = (x_1, \ldots, x_n)$ é uma amostra de dimensão n proveniente da população X [para a qual se tem $x_i > 90,\ i = 1, \ldots, n \quad \Leftrightarrow \quad \sum_{i=1}^{n} x_i > 90\,n \quad \Leftrightarrow \quad \bar{x} > 90$].

- **Obtenção do estimador de MV de θ**

  Passo 1 — Função de verosimilhança

$$
\begin{aligned}
L(\theta \mid \underline{x}) &= f_{\underline{X}}(\underline{x}) \\
&\overset{X_i\, indep}{=} \prod_{i=1}^{n} f_{X_i}(x_i) \\
&\overset{X_i \sim X}{=} \prod_{i=1}^{n} f_X(x_i) \\
&= \prod_{i=1}^{n} \left[\theta^{-1} \exp\left(-\frac{x_i - 90}{\theta}\right)\right] \\
&= \theta^{-n} \exp\left[-\sum_{i=1}^{n}\left(\frac{x_i - 90}{\theta}\right)\right] \\
&= \theta^{-n} \exp\left(-\frac{\sum_{i=1}^{n} x_i - 90\,n}{\theta}\right) \\
&= \theta^{-n} \exp\left(-n \times \frac{\bar{x} - 90}{\theta}\right), \quad \theta > 0
\end{aligned}
$$

  Passo 2 — Função de log-verosimilhança

$$
\ln L(\theta \mid \underline{x}) = -n \ln(\theta) - n \times \frac{\bar{x} - 90}{\theta}
$$

  Passo 3 — Maximização
  A estimativa de MV de θ é doravante representada por $\hat{\theta}$ e

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

$$
\begin{cases} -\frac{n}{\hat{\theta}} + n \times \frac{\bar{x}-90}{\hat{\theta}^2} = 0 \quad \Leftrightarrow \quad \frac{\hat{\theta}+90-\bar{x}}{\hat{\theta}^2} = 0 \\ \frac{n}{\hat{\theta}^2} - 2n \times \frac{\bar{x}-90}{\hat{\theta}^3} < 0 \end{cases}
$$

$$
\begin{cases} \hat{\theta} = \bar{x} - 90 \\ \frac{n}{(\bar{x}-90)^2} - 2n\frac{\bar{x}-90}{(\bar{x}-90)^3} = -\frac{n}{(\bar{x}-90)^2} < 0 \quad \text{(proposição verdadeira já que } \bar{x} > 90\text{).} \end{cases}
$$

  Passo 4 — Estimador de MV de θ

$$
EMV(\theta) = \frac{\sum_{i=1}^{n}(X_i - 90)}{n} = \bar{X} - 90
$$

(b) Obtenha a estimativa de máxima verosimilhança de $P(X > 100) = \exp\left(-\frac{10}{\theta}\right)$, tendo em conta uma amostra $(x_1, \ldots, x_{10})$ para a qual $\sum_{i=1}^{10} x_i = 910$. (0.5)

- **Estimativa de MV de θ**

$$
\begin{aligned}
\hat{\theta} &= \bar{x} - 90 \\
&= \frac{910}{10} - 90 \\
&= 1
\end{aligned}
$$

- **Outro parâmetro desconhecido**

$$
h(\theta) = P(X > 100) = \exp\left(-\frac{10}{\theta}\right)
$$

- **Estimativa de MV de $h(\theta)$**
  Invocando a propriedade de invariância dos estimadores de máxima verosimilhança, pode concluir-se que a estimativa de MV de $h(\theta)$ é dada por

$$\begin{aligned}
\widehat{h(\theta)} &= h(\hat{\theta}) \\
&= \exp\left(-\frac{10}{\hat{\theta}}\right) \\
&= \exp\left(-\frac{10}{1}\right) \\
&\simeq 0.000045.
\end{aligned}$$

2. Uma clínica pretende comparar dois tipos de dieta. Com esse objetivo, escolheram-se casualmente 15 pacientes que foram sujeitos à dieta 1 e outros 15 que foram sujeitos à dieta 2. Após 10 semanas, anotou-se o total de peso perdido (em kg) por cada paciente sujeito à dieta i, X_i, para $i = 1,2$. As amostras conduziram a: $\bar{x}_1 = 9.3$ kg e $s_1^2 = 5.76$ kg^2 para a dieta 1; $\bar{x}_2 = 8.2$ kg e $s_2^2 = 6.76$ kg^2 para a dieta 2. Admitindo que X_1 e X_2 possuem distribuições normais:

(a) Obtenha um intervalo de confiança a 95% para a variância do total de peso perdido por um paciente sujeito à dieta 1 durante 10 semanas. (1.5)

- **V.a. de interesse**
  X_1 = total de peso perdido (em kg) por paciente sujeito à dieta 1

- **Situação**
  $X_1 \sim \text{normal}(\mu_1, \sigma_1^2)$
  $\mu_1 = E(X_1)$ desconhecido
  $\sigma_1^2 = V(X_1)$ DESCONHECIDO
  $[X_{11}, X_{12}, \ldots, X_{1\,n_1}$ são v.a. i.i.d. a X_1, com $n_1 = 15$.]

- **Obtenção do IC para σ_1^2**
  Passo 1 — Selecção da v.a. fulcral para σ_1^2

$$Z = \frac{(n_1 - 1)\, S_1^2}{\sigma_1^2} \sim \chi^2_{(n_1-1)}$$

  [uma vez que é suposto determinar um IC para a variância de uma população normal, com valor esperado desconhecido.]

  Passo 2 — Obtenção dos quantis de probabilidade
  Ao ter-se em consideração que $n_1 = 15$ e $(1-\alpha)\times 100\% = 95\%$, far-se-á uso dos quantis

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

$$\begin{cases} a_\alpha = F^{-1}_{\chi^2_{(n_1-1)}}(\alpha/2) = F^{-1}_{\chi^2_{(14)}}(0.025) \stackrel{tabela/calc.}{=} 5.628 \\ b_\alpha = F^{-1}_{\chi^2_{(n_1-1)}}(1-\alpha/2) = F^{-1}_{\chi^2_{(14)}}(0.975) \stackrel{tabela/calc.}{=} 26.12. \end{cases}$$

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

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

$$P\left[a_\alpha \le \tfrac{(n_1-1)S_1^2}{\sigma_1^2} \le b_\alpha\right] = 1-\alpha$$

$$P\left[\tfrac{1}{b_\alpha} \le \tfrac{\sigma_1^2}{(n_1-1)S_1^2} \le \tfrac{1}{a_\alpha}\right] = 1-\alpha$$

$$P\left[\tfrac{(n_1-1)S_1^2}{b_\alpha} \le \sigma_1^2 \le \tfrac{(n_1-1)S_1^2}{a_\alpha}\right] = 1-\alpha$$

  Passo 4 — Concretização
  Atendendo ao par de quantis acima e ao facto de

$$
\begin{aligned}
s_1^2 &= 5.76\\
IC_{(1-\alpha)\times 100\%}(\sigma_1^2) &= \left[\frac{(n_1-1)\,s_1^2}{F^{-1}_{\chi^2_{(n_1-1)}}(1-\alpha/2)},\ \frac{(n_1-1)\,s_1^2}{F^{-1}_{\chi^2_{(n_1-1)}}(\alpha/2)}\right],
\end{aligned}
$$

segue-se:

$$
\begin{aligned}
IC_{95\%}(\sigma_1^2) &= \left[\frac{(15-1)\times 5.76}{26.12},\ \frac{(15-1)\times 5.76}{5.628}\right]\\
&\simeq [3.087289,\ 14.325813].
\end{aligned}
$$

(b) Admitindo que X_1 e X_2 possuem variâncias iguais, teste a hipótese de igualdade dos valores esperados dos totais de peso perdido para as duas dietas durante 10 semanas, ao nível de significância de 10%. (1.5)

- **V.a. de interesse**
  X_i = total de peso perdido (em kg) por paciente sujeito à dieta i, $i = 1,2$

- **Situação**
  $X_1 \sim \text{Normal}(\mu_1, \sigma_1^2) \perp\!\!\!\perp X_2 \sim \text{Normal}(\mu_1, \sigma_2^2)$
  $(\mu_1 - \mu_2)$ DESCONHECIDO
  σ_1^2 e σ_2^2 desconhecidos, no entanto, assume-se que são IGUAIS: $\sigma_1^2 = \sigma_2^2 = \sigma^2$
  [$X_{i1}, X_{i2}, \ldots, X_{i\,n_i}$ são v.a. i.i.d. a X_i, para $i = 1,2$, com]
  $n_1 = n_2 = 15 \leq 30$

- **Hipóteses**
  $H_0 : \mu_1 - \mu_2 = \mu_0 = 0$
  $H_1 : \mu_1 - \mu_2 \neq \mu_0$

- **Nível de significância**
  $\alpha_0 = 10\%$

- **Estatística de teste**

$$
T = \frac{(\bar{X}_1 - \bar{X}_2) - \mu_0}{\sqrt{\frac{(n_1-1)\,S_1^2 + (n_2-1)\,S_2^2}{n_1+n_2-2} \times \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim_{H_0} t_{(n_1+n_2-2)}
$$

  [dado que se pretende efectuar um teste sobre a diferença de valores esperados de duas populações normais independentes, com variâncias desconhecidas mas que se assume serem iguais.]

- **Região de rejeição de H_0** (para valores da estatística de teste)
  Estamos a lidar com um teste bilateral ($H_1 : \mu_1 - \mu_2 \neq \mu_0$), logo a região de rejeição de H_0 (para valores da estatística de teste) é $W = (-\infty, -c) \cup (c, +\infty)$, onde $c : P(\text{Rejeitar } H_0 | H_0) = \alpha_0$, i.e.,

$$
\begin{aligned}
c \quad : \quad & P(T \in W \mid H_0) = \alpha_0\\
& 2 \times \left[1 - F_{t_{(n_1+n_2-2)}}(c)\right] = \alpha_0\\
& c = F^{-1}_{t_{(n_1+n_2-2)}}(1 - \alpha_0/2)\\
& c = F^{-1}_{t_{(28)}}(0.95)\\
& c \overset{tabela/calc.}{=} 1.701.
\end{aligned}
$$

- **Decisão**
  Uma vez que

  $n_1 = 15 \quad \bar{x}_1 = 9.3 \quad s_1^2 = 5.76$

  $n_2 = 15 \quad \bar{x}_2 = 8.2 \quad s_2^2 = 6.76$

  o valor observado da estatística de teste é igual a

$$
\begin{aligned}
t &= \frac{(\bar{x}_1 - \bar{x}_2) - \mu_0}{\sqrt{\frac{(n_1-1)\, s_1^2 + (n_2-1)\, s_2^2}{n_1+n_2-2} \times \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \\
&= \frac{(9.3 - 8.2) - 0}{\sqrt{\frac{14\times 5.76 + 14\times 6.76}{15+15-2}\left(\frac{1}{15} + \frac{1}{15}\right)}} \\
&= \frac{1.1}{\sqrt{\frac{5.76+6.76}{15}}} \\
&\simeq 1.204027
\end{aligned}
$$

Como $t \simeq 1.204027 \notin W = (-\infty, -1.701) \cup (1.701, +\infty)$, não devemos rejeitar H_0 ao n.s. $\alpha_0 =$ 10% [ou a qualquer n.s. inferior a $\alpha_0 = 10\%$].

Grupo IV	5 valores

1. Conjetura-se que o número de defeitos de um circuito eletrónico, X, segue uma distribuição de Poisson com valor esperado 0.8. Ao inspecionarem-se 60 circuitos selecionados ao acaso, obteve-se a seguinte tabela de frequências:

Nº defeitos por circuito	0	1	≥ 2
Frequência absoluta observada	33	14	13
Frequência absoluta esperada sob H_0	26.96	E_2	E_3

(a) Calcule as frequências absolutas esperadas sob $H_0 : X \sim \text{Poisson}(0.8)$ que se representam na tabela por E_2 e E_3 (aproximando-as às centésimas). (0.5)

- **V.a. de interesse**
  $X =$ número de defeitos de um circuito eletrónico
- **Distribuição e f.p. conjecturadas**
  $X \sim \text{Poisson}(0.8)$
  $P(X = x) = \frac{e^{-0.8}\, 0.8^x}{x!}, \quad x = 0, 1, 2, \ldots$
- **Frequências absolutas esperadas omissas**
  Tendo em conta a dimensão da amostra $n = 60$ e a f.p. conjecturada, temos:

$$
\begin{aligned}
E_2 &= n \times P(X = 1) \\
&= 60 \times \frac{e^{-0.8}\, 0.8^1}{1!} \\
&\simeq 21.57; \\
E_3 &= n - \sum_{i=1}^{2} E_i \\
&\simeq 60 - (26.96 + 21.57) \\
&= 11.47.
\end{aligned}
$$

(b) Teste H_0, ao nível de significância de 5%. (1.5)

- **Hipóteses**
  $H_0 : X \sim \text{Poisson}(0.8)$
  $H_1 : X \not\sim \text{Poisson}(0.8)$
- **Nível de significância**
  $\alpha_0 = 5\%$

- **Estatística de Teste**

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

onde:

k = No. de classes = 3

O_i = Frequência absoluta observável da classe i

E_i = Frequência absoluta esperada, sob H_0, da classe i

β = No. de parâmetros a estimar = 0 [dado que em H_0 se conjectura uma distribuição em particular.]

- **Frequências absolutas esperadas sob H_0**

De acordo com a tabela facultada e a alínea (a), os valores das frequências absolutas esperadas sob H_0 aproximados às centésimas são: $E_1 \simeq 26.96$; $E_2 \simeq 21.57$; $E_3 \simeq 11.47$.
[Não é necessário fazer qualquer agrupamento de classes uma vez que em pelo menos 80% das classes se verifica $E_i \geq 5$ e que $E_i \geq 1$ para todo o i. Caso fosse preciso efectuar agrupamento de classes, os valores de k e $c = F^{-1}_{\chi^2_{(k-\beta-1)}}(1-\alpha_0)$ teriam que ser recalculados...]

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

Tratando-se de um teste de ajustamento, a região de rejeição de H_0 escrita para valores de T é o intervalo à direita $W=(c,+\infty)$, onde

$$\begin{aligned} c &= F^{-1}_{\chi^2_{(k-\beta-1)}}(1-\alpha_0) \\ &= F^{-1}_{\chi^2_{(3-0-1)}}(1-0.05) \\ &= F^{-1}_{\chi^2_{(2)}}(0.95) \\ &\overset{tabela/calc.}{=} 5.991. \end{aligned}$$

- **Decisão**

No cálculo do valor observado da estatística de teste convém recorrer à seguinte tabela auxiliar:

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	$\{0\}$	33	26.96	$\frac{(33-26.96)^2}{26.96} \simeq 1.353$
2	$\{1\}$	14	21.57	2.657
3	$\{2,3,\ldots\}$	13	11.47	0.204
		$\sum_{i=1}^{k} o_i = n$ $= 60$	$\sum_{i=1}^{k} E_i = n$ $= 60$	$t=\sum_{i=1}^{k}\frac{(o_i-E_i)^2}{E_i}$ $\simeq 4.214$

Uma vez que $t \simeq 4.214 \not\in W=(5.991,+\infty)$, não devemos rejeitar H_0 ao n.s. de $\alpha_0 = 5\%$ [nem a qualquer outro n.s. inferior a 5%].

2. Um conjunto de dados relativos a cinco indivíduos forneceu os seguintes valores relativos ao logaritmo (de base e) da contagem bacteriana (Y) x dias após a inoculação de uma vacina:

$$\textstyle\sum_{i=1}^5 x_i = 33, \quad \sum_{i=1}^5 x_i^2 = 239, \quad \sum_{i=1}^5 y_i = 60.4, \quad \sum_{i=1}^5 y_i^2 = 730.32, \quad \sum_{i=1}^5 x_i y_i = 401.9,$$

onde $\left[\min_{i=1,\ldots,5} x_i,\ \max_{i=1,\ldots,5} x_i\right] = [3.0,\ 9.0]$.

Admitindo que os erros aleatórios associados ao modelo de regressão linear simples de Y em x satisfazem $\epsilon_i \overset{i.i.d.}{\sim} \text{Normal}(0,\sigma^2)$, $i=1,\ldots,5$:

(a) Obtenha a estimativa de máxima verosimilhança do valor esperado do logaritmo (de base e) da contagem bacteriana 5 dias após a inoculação da vacina. (1.0)

- **Estimativas de MV de β_1, β_1 e $E(Y \mid x=5)$**
  Dado que

  $n = 5$

  $\sum_{i=1}^n x_i = 33$
  $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i = \frac{33}{5} = 6.6$
  $\sum_{i=1}^n x_i^2 = 239$
  $\sum_{i=1}^n x_i^2 - n\,\bar{x}^2 = 239 - 5 \times 6.6^2 = 21.2$

  $\sum_{i=1}^n y_i = 60.4$
  $\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i = \frac{60.4}{5} = 12.08$
  $\sum_{i=1}^n y_i^2 = 730.32$
  $\sum_{i=1}^n y_i^2 - n\,\bar{y}^2 = 730.32 - 5 \times 12.08^2 = 0.6880$

  $\sum_{i=1}^n x_i y_i = 401.9$
  $\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y} = 401.9 - 5 \times 6.6 \times 12.08 = 3.26,$

  as estimativas de MV de β_1, β_0 e $E(Y \mid x=5)$ são, para este modelo de RLS, iguais a:

$$
\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^n x_i y_i - n\,\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2} \\
&= \frac{3.26}{21.2} \\
&\simeq 0.153774 \\
\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \times \bar{x} \\
&\simeq 12.08 - 0.153774 \times 6.6 \\
&\simeq 11.065092 \\
\hat{E}(Y \mid x=5) &= \hat{\beta}_0 + \hat{\beta}_1 \times 5 \\
&\simeq 11.065092 + 0.153774 \times 5 \\
&\simeq 11.833962.
\end{aligned}
$$

(b) Teste a significância do modelo de regressão. Decida com base no valor-p. (1.5)

- **Hipóteses**
  $H_0 : \beta_1 = \beta_{1,0} = 0$
  $H_1 : \beta_1 \neq 0$

- **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 da estatística de teste)
  Estamos a lidar com um teste bilateral ($H_1 : \beta_1 \neq 0$), pelo que a região de rejeição de H_0 é uma reunião de intervalos do tipo $W = (-\infty, -c) \cup (c, +\infty)$.

- **Decisão (com base no valor-p)**
  Tendo em conta os valores obtidos em (a), bem como o valor de

$$
\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n-2}\left[\left(\sum_{i=1}^n y_i^2 - n\,\bar{y}^2\right) - \left(\hat{\beta}_1\right)^2 \left(\sum_{i=1}^n x_i^2 - n\,\bar{x}^2\right)\right] \\
&\simeq \frac{1}{5-2}\left(0.6880 - 0.153774^2 \times 21.2\right) \\
&\simeq 0.062232,
\end{aligned}
$$

concluímos que o valor observado da estatística de teste é igual a

$$\begin{aligned} t &= \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}}} \\ &= \frac{0.153774 - 0}{\sqrt{\frac{0.062232}{21.2}}} \\ &= 2.838206. \end{aligned}$$

Uma vez que a região de rejeição deste teste é uma reunião de intervalos simétricos [e a distribuição da estatística e teste sob H_0 é simétrica em relação à origem], temos:

$$\begin{aligned} valor-p &= P(T > |t| \mid H_0) \\ &= 2 \times [1 - F_{t_{(n-2)}}(|t|)] \\ &= 2 \times [1 - F_{t_{(3)}}(2.838206)] \\ &\stackrel{calc}{=} 2 \times (1 - 0.967128) \\ &= 0.0657436. \end{aligned}$$

Deste modo é suposto:

- não rejeitar H_0 a qualquer n.s. $\alpha_0 \leq 6.57436\%$, pelo que H_0 não é contrariada pelos dados aos n.u.s. de 1% e 5%;
- rejeitar H_0 a qualquer n.s. $\alpha_0 > 6.57436\%$, por exemplo, ao n.u.s. de 10%.

[Alternativamente, poderíamos recorrer às tabelas de quantis da distribuição t-student com 3 graus de liberdade e adiantar um intervalo para o valor-p:

$$\begin{aligned} F^{-1}_{t_{(3)}}(0.95) = 2.353 &< t = 2.838206 < 3.182 = F^{-1}_{t_{(3)}}(0.975) \\ 0.95 &< F_{t_{(3)}}(2.838206) < 0.975 \\ 2 \times (1 - 0.975) &< 2 \times [1 - F_{t_{(3)}}(2.838206)] < 2 \times (1 - 0.95) \\ 0.05 &< valor - p < 0.10. \end{aligned}$$

Assim, é suposto:

- não rejeitar H_0 a qualquer n.s. $\alpha_0 \leq 5\%$, pelo que H_0 não é contrariada pelos dados aos n.u.s. de 1% e 5%;
- rejeitar H_0 a qualquer n.s. $\alpha_0 \geq 10\%$, nomeadamente, ao n.u.s. de 10%.

(c) Calcule e comente o valor do coeficiente de determinação do modelo ajustado. (0.5)

- **Cálculo do coeficiente de determinação**

$$\begin{aligned} r^2 &= \frac{\left(\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y}\right)^2}{\left(\sum_{i=1}^n x_i^2 - n\,\bar{x}^2\right) \times \left(\sum_{i=1}^n y_i^2 - n\,\bar{y}^2\right)} \\ &= \frac{3.26^2}{21.2 \times 0.6880} \\ &\simeq 0.728636. \end{aligned}$$

- **Interpretação coeficiente de determinação**

  Cerca de 73% da variação total do logaritmo (de base e) da contagem bacteriana é explicada pelo número de dias decorridos após a inoculação da vacina, através do modelo de regressão linear simples considerado. Consequentemente, podemos afirmar que a recta estimada parece ajustar-se bem ao conjunto de dados.