Departamento de Matemática da Universidade de Coimbra

# Estatística – LEI, LECD

1 de julho de 2025 **Exame de recurso** Duração: 2h30

**Observação**: A resolução completa das questões inclui a justificação do raciocínio utilizado e a apresentação dos cálculos efetuados.

## I

1. Determinada empresa tem 5% dos seus computadores infetados com um certo vírus. Foi decidido instalar, em todos os computadores da empresa, o mesmo antivírus que, logo após a instalação, realiza um teste que sinaliza (ou não) a presença do vírus no computador. Sabe-se que o antivírus sinaliza a presença do vírus em 95% dos computadores infetados, mas também sinaliza *erradamente* a presença do vírus em 2% dos computadores não infetados. Foi escolhido aleatoriamente um computador daquela empresa.

   (a) Calcule a probabilidade do antivírus sinalizar a presença do vírus nesse computador.

   (b) Sabendo que o teste do antivírus deu positivo (para a presença do vírus), qual é a probabilidade do computador estar realmente infetado?

   (c) Em 15 computadores retirados ao acaso dos mais de 1500 computadores da empresa, qual é a probabilidade de haver pelo menos 2 infetados?

   (d) Sabe-se que o número de novos ficheiros corrompidos por dia num computador infetado segue uma lei de Poisson com variância 4. Determine a probabilidade de que, o número de novos ficheiros infetados por dia, seja superior a 6.

2. As variáveis aleatórias $X$ e $Y$ que representam a duração (em anos) de dois tipos de baterias de iões de lítio, A e B, são modeladas pelas leis $\mathcal{N}(8,1)$ e $\mathcal{N}(10,\sqrt{5})$, respetivamente, sendo também independentes.

   (a) Calcule $P(X > 6)$.

   (b) Calcule $P(2X - 4 < Y)$.

3. A variável aleatória $X$ tem densidade dada por $f_X(x) = \begin{cases} 2(1-x) & 0 \leq x \leq 1, \\ 0 & \text{se não,} \end{cases}$ , com $E(X) = \frac{1}{3}$.

   (a) Calcule $P(X \leq 0.3)$.

   (b) Considere outra variável aleatória $Y$ tal que $Y = \theta X + 1$, onde $\theta$ é um parâmetro real desconhecido.

      i. Construa um estimador de $\theta$.

      ii. Mostre que $Y$ não tem lei normal.

      iii. Dispomos de uma amostra de $Y$, de dimensão 51, que apresenta: $\overline{y} = 1.684$ e $s = 0.249$. Construa um intervalo que contenha a média de $Y$ com 95% de confiança.

      iv. Usando a alínea anterior, construa um intervalo que contenha $\theta$ com 95% de confiança.

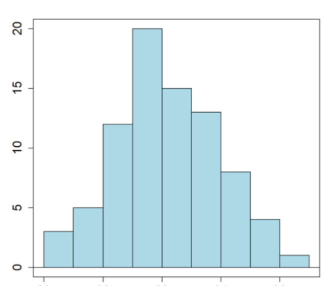
## II

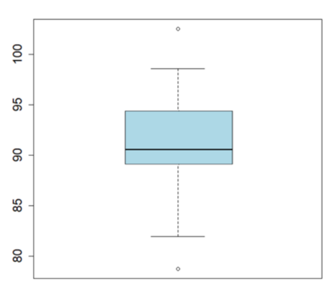
O tempo, em minutos, de reparação de um certo tipo de avaria numa linha de montagem é uma variável aleatória que representamos por $X$. Dispomos de uma amostra de $X$, de dimensão 81, que apresenta os seguintes valores:

| $\overline{x}$ | $s$ | $q_1$ | $q_2$ | $q_3$ |
|---|---|---|---|---|
| 91.613 | 4.104 | 89.147 | 91.778 | 94.428 |

1. Averigue se o valor amostral 102.513 é um outlier desta amostra.

2. Mostre que nenhum dos seguintes gráficos pode corresponder a esta amostra.




3. Foi realizado um teste estatístico para averiguar a compatibilidade desta amostra com a hipótese de normalidade da variável aleatória $X$ que representa o referido tempo de reparação. Com base na informação apresentada ao lado e gerada em R, o que pode concluir?

```
        Anderson-Darling normality test

data:  x
A = 0.39704, p-value = 0.3609
```

**v.p.f.**

4. Realize um teste que lhe permita averiguar se a média de $X$ deve ser considerada inferior a 92. Use o nível de significância 0.05.

5. Que tipo de erro pode estar a cometer na decisão tomada na alínea anterior?

6. O teste da alínea 4. foi realizado via R. Qual das seguintes figuras correspondende ao output gerado? Justifique.

```
        One Sample t-test

data:  x
t = -0.84909, df = 80, p-value = 0.1992
alternative hypothesis: true mean is less than 92
95 percent confidence interval:
     -Inf 92.37168
sample estimates:
mean of x
 91.61278
```

```
        One Sample t-test

data:  x
t = 0.84909,  df = 80, p-value = 0.1992
alternative hypothesis: true mean is less than 92
95 percent confidence interval:
     -Inf 92.37168
sample estimates:
mean of x
 91.61278
```

7. Pretendemos testar $H_0 : \sigma^2 = 16$ contra $H_1 : \sigma^2 > 16$. Indique a região crítica deste teste, considerando $\alpha = 0.02$.

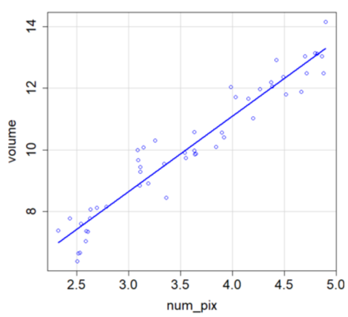
## III

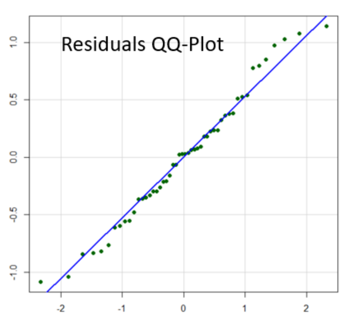
1. A descendência originada pelo cruzamento de dois tipos de plantas pode ser um dos quatro genótipos A, B, C e D que, segundo o modelo mendeliano teórico de sucessão genética, devem ocorrer nas proporções 0.1, 0.2, 0.3 e 0.4 respetivamente. Na sequência do cruzamento daqueles dois tipos de plantas, procedeu-se à classificação genética de 100 plantas, tendo sido registadaa frequência absoluta de cada um dos genótipos. É nosso objetivo realizar um teste de ajustamento do qui-quadrado para averiguar se estes resultados estão de acordo com o referido modelo teórico. A tabela seguinte contém parte da informação necessária à realização do teste.

| Genótipos | A | B | C | D |
|---|---|---|---|---|
| Número de plantas | 9 | 25 | 28 | 38 |
| Número de plantas esperado, sob $H_0$ | 10 | | 30 | |
| $\frac{(n_i - e_i)^2}{e_i}$ | 0.1 | 1.25 | | 0.1 |

Complete a tabela e realize o teste ao nível de significância 0.05.

2. Uma empresa produtora de ostras planeia implementar um sistema de classificação automática das ostras segundo o seu volume e avalia uma proposta de estimação do volume ($Y$, em cm$^3$) em função do número de píxeis ($x$, em milhões) numa representação tridimensional da ostra. Foi medido o volume de 50 ostras e registado o número de píxeis de uma imagem tridimensional de cada uma delas. Com estes registos, pretende-se testar a existência de associação entre $x$ e $Y$, descrita por um modelo de regressão linear simples, na forma $Y = ax + b + \varepsilon$. Apresenta-se a seguir parte do output, obtido via R, correspondente à regressão efetuada sobre os dados.




```
        Shapiro-Wilk normality test

data:  residuals.RegModel.1
W = 0.98051, p-value = 0.5738
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.29602    0.36642   3.537 0.000909 ***
num_pix      2.44982    0.09947  24.628  < 2e-16 ***
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5672 on 48 degrees of freedom
Multiple R-squared:  0.9267,    Adjusted R-squared:  0.9251
F-statistic: 606.5 on 1 and 48 DF,  p-value: < 2.2e-16
```

(a) Indique o valor do coeficiente de determinação e interprete-o.

(b) Com a informação disponível, podemos assumir que a variável erro ($\varepsilon$) é normalmente distribuída? Se sim, indique estimativas para os seus parâmetros.

(c) Qual o número de píxeis que permite obter uma estimativa de 11 cm$^3$ para o volume das ostras?

Cotação: I - **1.** 1.5+0.75+1.25+1.0 **2.** 1.0+1.5 **3.** 1.0+0.75+0.5+1.25+1.0
II - **1.** 1.0 **2.** 0.5 **3.** 0.5 **4.** 1.5 **5.** 0.25 **6.** 0.5 **7.** 0.5 III - **1.** 1.5 **2.** 0.75+1.0+0.5