

• Responsáveis: Profs. Manuel Scottó (geral)  
António Pacheco (LEEC)

• Docente: Miguel Abreu (eu)

• Página da cadeira: Fé'nix.

• Programa

- ① Análise Exploratória de Dados + R/RStudio [3]
- ② Noções básicas de probabilidade [3]
- ③ Variáveis aleatórias discretas e contínuas [4]
- ④ Pares aleatórios [4]
- ⑤ Combinações lineares de variáveis aleatórias e teorema do limite central [2]
- ⑥ Estimação pontual [2]
- ⑦ Estimação intervalar [2]
- ⑧ Testes de Hipóteses [4]
- ⑨ Introdução à regressão linear simples [4]

• Bibliografia:

→ Probabilidades e Estatística: teoria, exemplos & exercícios  
Manuel Cabral Morais, IST Press

→ Exercícios (página de PE → Material Didático Base)

→ Formulário ( " " " → " " " )

→ Tabelas Estatísticas ( " " " → " " " )

- Avaliação tem 3 componentes

1) Séries de Problemas: (SP, 20%)

6 séries de 5 problemas, via Feñix, nas semanas 4, 6, 11, 12, 14 e 16 (primeira na semana 28.mar - 03.abr)

SP = média das 5 melhores

2) Projeto computacional (P, 30%)

Resolução de exercícios feita com R/RStudio, via plataforma Moodle, distribuído na semana 4 (28.mar - 1.abr) e com data final de entrega a 12.junho. Horários de dúvidas da semana 8 (25-29.abril) serão dedicados ao projeto.

3) Exame Final (E1 e E2, 50%)

6 de Julho e 22 de Julho, duração 2h.

Nota mínima: 7,5.

Nota Final:  $0,5 * \max\{E1, E2\} + 0,3 * P + 0,2 * SP$

com prova adicional para  $NF \geq 17,5$ .

Importante: para acesso ao Moodle (projeto) é necessário que no Feñix tenham como email institucional o vosso email do IST, i.e. xxx.yyy@tecnico.ulisboa.pt

- Meu horário de dúvidas: 3<sup>as</sup> e 5<sup>as</sup>, 14:00 - 16:00, via Zoom (por defeito) ou presencial (a combinar).

# ① Análise exploratória de dados + R/RStudio

- Recolha de dados → Análise Descritiva -  
Estatística Descritiva
- Conceitos básicos
  - População (ou universo): conjunto de todos os ele<sup>tos</sup> com determinada característica em comum (e.g. todos os alunos do IST)
  - Unidade estatística: ele<sup>to</sup> da população (aluno do IST)
  - Variável: característica de interesse em estudo (ex.  $X$  - altura dos alunos do IST  
 $x$  - altura observada de um aluno do IST)  
↑ dado)
  - Amostra: subconjunto da população observado,  
 $x = (x_1, \dots, x_n)$
- Classificação da variável:
  - Qualitativa ou Categórica: identificam uma qualidade ou atributo dos ele<sup>tos</sup> da população
    - Nominal: sexo, cor dos olhos
    - Ordinal: classe social, grau de instrução
  - Quantitativa ou Numérica: identificam números resultantes de uma contagem ou medição

→ Contínua: peso, altura

→ Discreta: idade, média final do curso

• Variáveis quantitativas - indicadores numéricos ou amostrais de  $x = (x_1, \dots, x_n)$

→ Medidas de localização amostrais:

média, moda, quantis, mediana e quartis.

→ Medidas de dispersão amostrais:

variância, desvio padrão e coeficiente de variação

• Medidas de localização

→ Média:  $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$

→ Moda (mo): valor que ocorre com maior frequência

→ Quantil- $\alpha$  ( $q_\alpha$ ,  $0 < \alpha < 1$ ): valor  $q_\alpha$  t.s. há uma proporção  $\alpha$  de observações  $\leq q_\alpha$  e há uma proporção  $1-\alpha$  de observações  $\geq q_\alpha$ .

Mais precisamente:

$x = (x_1, \dots, x_n) \rightsquigarrow x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$   
" " " " " "  
mín  $x$  " máx  $x$

$\rightsquigarrow q_\alpha = \begin{cases} \frac{x_{(n\alpha)} + x_{(n\alpha+1)}}{2}, & \text{se } n\alpha \in \mathbb{N} \\ x_{(\lfloor n\alpha \rfloor + 1)}, & \text{se } n\alpha \notin \mathbb{N} \end{cases}$

→ Mediana =  $q_{1/2}$

→ Quartis: 1º quartil =  $q_{1/4}$

2º quartil =  $q_{1/2}$  = mediana

3º quartil =  $q_{3/4}$

• Medidas de dispersão

→ Amplitude:  $R = \max(x) - \min(x) = x_{(n)} - x_{(1)}$

→ Amplitude inter-quartis:  $IQR = q_{3/4} - q_{1/4}$

→ Variância: 
$$s_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \frac{\left(\sum_{i=1}^n x_i^2\right) - n\bar{x}^2}{n-1}$$

→ Desvio padrão:  $s_x = \sqrt{s_x^2}$

(mesma unidade de medida de variável e de média)

→ Coefficiente de variação:  $cv = \frac{s}{\bar{x}}$

• adimensional e permite comparar a dispersão amostral de variáveis distintas

• só se calcula qd a variável tome valores de um só sinal.

• R / RStudio : <http://www.r-project.org>

Eg. R Programming Tutorial by Barton Paulson on Youtube (2 horas).