

# Estatística e Delineamento

Jorge Cadima

Secção de Matemática (DCEB)  
Instituto Superior de Agronomia (ULisboa)

2021-22

## 1 Professores:

- ▶ Jorge Cadima (Responsável)
- ▶ Elsa Gonçalves
- ▶ Maria João Martins
- ▶ Rita Neres

2 Webpage no Sistema Fenix ← Indispensável

## 3 Software:

- ▶ Homepage: [www.r-project.org](http://www.r-project.org)
- ▶ Repositório (para descarregar): [cran.r-project.org](http://cran.r-project.org)

# Pressupostos

Admite-se que houve frequência numa disciplina introdutória de Estatística no primeiro ciclo (semelhante à existente no ISA) e que são conhecidos:

- principais indicadores descritivos (média, variância, covariância, coeficiente de correlação linear, etc.) e suas propriedades;
- conceitos básicos de probabilidades;
- variáveis aleatórias e sua caracterização;
- principais distribuições de probabilidades (Normal, t-Student,  $\chi^2$ , F, etc.);
- conceitos de intervalos de confiança e testes de hipóteses.

Breve revisão nos [Exercícios Introdutórios](#) (também com introdução ao R).

**AVISO:** Vejam-se os materiais de apoio da UC [Estatística](#) dos primeiros ciclos do ISA (ou equivalentes), disponíveis na [página web](#) dessa UC).

# Enquadramento

Nas disciplinas introdutórias de Estatística aborda-se sobretudo o estudo das observações de **uma** variável.

A UC Estatística e Delineamento é uma disciplina de aprofundamento, que procura **relacionar uma variável de interesse com outras variáveis**.

O programa da UC consiste no estudo do principal **modelo estatístico**: o **Modelo Linear**, que inclui como casos particulares:

- a **Regressão Linear** (Simples e Múltipla);
- e as **Análises de Variância** (ANOVAs).

No Calendário Escolar para 2021-22, decidido pelos órgãos de gestão do ISA, o semestre lectivo tem **12 semanas**. Os horários da disciplina são afectados por **4 feriados**. **As aulas Teóricas e seis das oito turmas práticas apenas terão 10,5 semanas de aulas**.

# Aulas e horários de dúvidas

**Todas as aulas** começam a partir de segunda-feira, dia 27 de Setembro (mesmo a turma prática com aula antes da Teórica).

- **Aulas teóricas (2 vezes por semana, 1h cada)** - Há dois blocos diferentes, ambos com aulas às 2<sup>as</sup> e 4<sup>as</sup>-feiras.
- **Aulas práticas (2 vezes por semana, 1h30 cada)** - Há oito turmas (inscrições via Fénix).
- Há **horários de dúvidas** (pormenores na webpage) **a partir da semana de 4 de Outubro. Todos os horários de dúvidas são para qualquer aluno da UC.**

# Material de Apoio

- Há [material de apoio às aulas](#) na página *web* da UC (ver secção lateral de nome *Materiais de Apoio*).
- Existem [apontamentos](#) de acompanhamento da matéria.
- Há apontadores para os [enunciados e resoluções de exames e testes](#) de anos anteriores. Atenção que os programas da disciplina nem sempre são iguais ao longo dos anos.
- Nalgumas salas de aulas Práticas (A2 e A3) há apenas o computador do professor (projectado em ecrã). [Quem puder, traga portátil](#).
- Quem trazer portátil deverá [ter o R previamente instalado](#). Nos [materiais de apoio às Aulas Práticas](#) há um [video](#) de apoio à instalação do R, bem como outros materiais (escritos e video) de apoio à sua utilização.

# Funcionamento das aulas

Nas aulas práticas devem ter disponíveis:

- os enunciados dos Exercícios;
- se utilizar um computador na sala de aula, uma *pen* para guardar a sessão de trabalho.

**AVISO:** Para ter acesso aos computadores do ISA, é preciso ter conta informática de aluno (em caso de problema contactar a Divisão de Informática).

# Avaliação de conhecimentos

Por exame final (nota de 0 a 20, aprovação a partir de 9,5 valores).

## AVISOS:

- Todas as avaliações serão presenciais.
- Qualquer aluno inscrito na UC pode apresentar-se a avaliação de conhecimentos.
- Nas avaliações **não são admitidas calculadoras gráficas, nem qualquer tipo de equipamento electrónico.**

# Programa

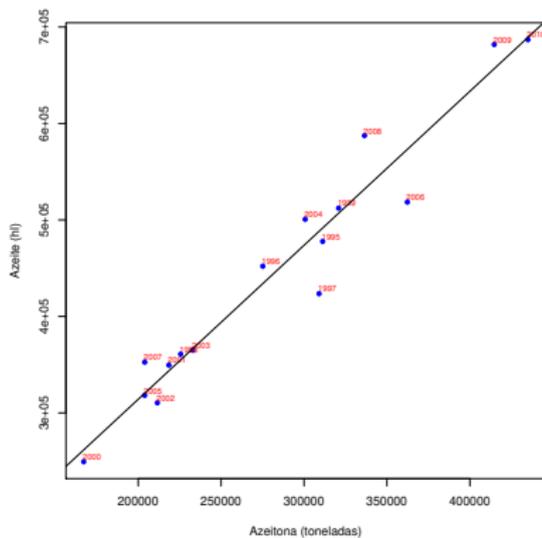
## ○ Modelo Linear.

- 1 Regressão Linear Simples
- 2 Regressão Linear Múltipla
- 3 Análises de Variância (ANOVA)

# 1. Modelo Linear: regressão linear simples

**Objectivo:** Relacionar linearmente duas variáveis numéricas.

**Exemplo:** Produção de azeitona e de azeite em Portugal, entre os anos 1995 e 2010 (Fonte:INE)



# Modelo Linear: Regressão Linear Simples

- **Contexto descritivo** (o ajustamento da recta na amostra)
  - ▶ Equação da **recta ajustada** (método dos mínimos quadrados);
  - ▶ **Propriedades da recta ajustada**;
  - ▶ Relações não-lineares e **transformações linearizantes**.
- **Inferência** (sobre a recta populacional, com base numa amostra)
  - ▶ O **modelo**;
  - ▶ **Estimadores** dos parâmetros da recta populacional e suas **propriedades**;
  - ▶ **Intervalos de confiança** para os parâmetros;
  - ▶ **Testes de hipóteses** para os parâmetros;
  - ▶ **Análise dos resíduos** para validação do modelo e identificação de observações especiais.

## 2. Modelo linear: Regressão Linear Múltipla

**Objectivo:** Relacionar uma **variável resposta numérica** com **dois ou mais preditores numéricos**, através duma relação **linear**.

- **Contexto descritivo**

- ▶ Uma ferramenta: a notação matricial;
- ▶ A **estimação dos parâmetros** do (hiper)plano ajustado;
- ▶ Propriedades do (hiper)plano ajustado.

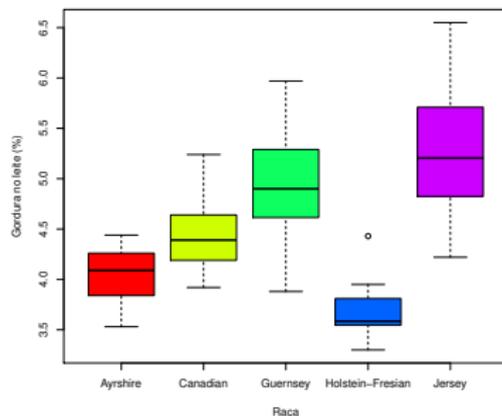
- **A inferência**

- ▶ O modelo;
- ▶ **Propriedades distribucionais dos estimadores**;
- ▶ **Intervalos de confiança e testes de hipóteses para os parâmetros**;
- ▶ **Submodelos e selecção de submodelos**;
- ▶ **Análise dos resíduos**.

### 3. Modelo Linear: Análise de Variância (ANOVA)

**Objectivo:** Relacionar uma **variável resposta numérica** com **um ou mais preditores categóricos (factores)**.

**Exemplo:** Comparar a % de gordura no leite de 5 raças de vacas (*Ayrshire*, *Canadian*, *Guernsey*, *Holstein-Fresian*, *Jersey*). Eis os diagramas de extremos e quartis para 20 vacas de cada raça:



As diferenças observadas são **significativas** ou são obra do acaso?

## Modelo Linear: ANOVAs (cont.)

- Introdução ao **delineamento experimental**
- Delineamento a um factor totalmente casualizado e o modelo correspondente (efeitos fixos)
- Delineamento factorial a dois factores. O modelo sem interacção e o modelo **com interacção** (efeitos fixos).
- Delineamento a dois factores hierarquizados e respectivo modelo (efeitos fixos)

## 1 Referências Base:

- ▶ Slides e apontamentos das aulas teóricas;
- ▶ **Kutner, M.H.; Nachtsheim, C.J.; Neter, J. e Li, W. (2005)**, *Applied Linear Statistical Models*, Irwin [**BISA: U10-727 e CD-236**]

## 2 Outras referências:

- ▶ **Draper, N.R. e Smith, H. (1998)**, *Applied Regression Analysis*, 3a. edição, John Wiley & Sons [**BISA: U10-734**] + [**SI-78**] ([**BISA: U10-412**] a primeira edição de 1981).
- ▶ **Montgomery, D.C. e Peck, E.A. (1982)**, *Introduction to Linear Regression Analysis*, John Wiley & Sons [**BISA: U10-329**]
- ▶ **Seber, G.A.F. (1977)**, *Linear Regression Analysis*, John Wiley & Sons [**BISA: U10-416**]

## 3 Referências de apoio à utilização do R

- ▶ **Docentes de Estatística (2008/09)**, Introdução à Aplicação R.
- ▶ **Maindonald, J. e Brown, W.J. (2003)**, *Data Analysis and Graphics using R*, Cambridge University Press [**BISA: U10-722**]
- ▶ **Torgo, L. (2006)**, Introdução à Programação em R
- ▶ **Venables, W.N. e Ripley, B.D. (2002)**, *Modern Applied Statistics with S (fourth edition)*, Springer-Verlag [**BISA: U10-733**]