

Estatística e Delineamento

Jorge Cadima

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

2020-21

1 Professores:

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

2 **Webpage:** Sistema Fenix ← Indispensável

3 **Software:** 

- ▶ *Homepage:* www.r-project.org
- ▶ Repositório (para descarregar): cran.r-project.org

Pressupostos

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

Na UC **Estatística e Delineamento** admite-se 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, χ^2 , t-Student, F, Binomial, Poisson, etc.);
- conceitos de **intervalos de confiança** e **testes de hipóteses**.

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

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, ou com hipóteses explicativas.**

O fundamental do programa da UC diz respeito ao principal **modelo estatístico**: o **Modelo Linear**.

Existem apontamentos desta parte da matéria.

Aulas e horários de dúvidas

O semestre lectivo tem 14 semanas. **As aulas práticas começam na segunda-feira, dia 14 de Setembro** (antes das Teóricas).

- **Aulas teóricas** (2 vezes por semana, 1h15 cada), **online** - Há dois blocos diferentes, ambos com aulas 5as. e 6as.-feiras.
- **Aulas práticas** (1 vez por semana, 2h30), **presenciais** - Há oito turmas (inscrições via Fénix).
- **Material de apoio às aulas** na página *web* da UC (ver secção lateral de nome *Materiais de Apoio*).
- Haverá **horários de dúvidas** 4 dias por semana (pormenores na *webpage*). **Todos os horários de dúvidas são para qualquer aluno da UC.**

Funcionamento das aulas práticas

Por decisão do Conselho de Gestão do ISA, as aulas práticas presenciais decorrem uma única vez por semana, durante 2h30, em salas que assegurem o distanciamento físico.

Não há PCs suficientes para todos os alunos.
Nalgumas salas há apenas um PC do professor (projectado em ecrã).

Quem puder, traga portátil (*laptop*) para as aulas.

AVISO: Não é avaliado o conhecimento do *software* R, mas apenas a interpretação estatística dos respectivos *outputs*.

Material para as aulas práticas

- enunciados dos Exercícios;
- computadores portáteis, onde tenha previamente sido instalado o *software* R, que pode ser livremente descarregado em:

<https://cran.r-project.org>

- ou, se utilizar um PC na sala de aula, uma *pen* para guardar a sessão de trabalho.

AVISO: Se nas aulas práticas usar PC do ISA, é preciso ter conta informática de aluno do ISA (em caso de problema contactar a Divisão de Informática).

AVISO: Se preferir usar o R Studio pode obter versões gratuitas em [rstudio.com](https://www.rstudio.com).

Avaliação de conhecimentos

- Por testes (dois testes: um a meio do semestre e outro na data da primeira chamada de exame); ou
- Por exame final.

AVISOS:

- Todas as avaliações serão presenciais.
- Aprovação por testes: classificação média $\geq 9,5$ valores, desde que em nenhum teste a classificação seja inferior a 8,0 valores.
- Aprovação em exame: classificação $\geq 9,5$ valores.
- 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.
- Proposta primeiro teste: Quarta-feira, 4 Novembro, fim da tarde.

Programa

1 Modelo Linear.

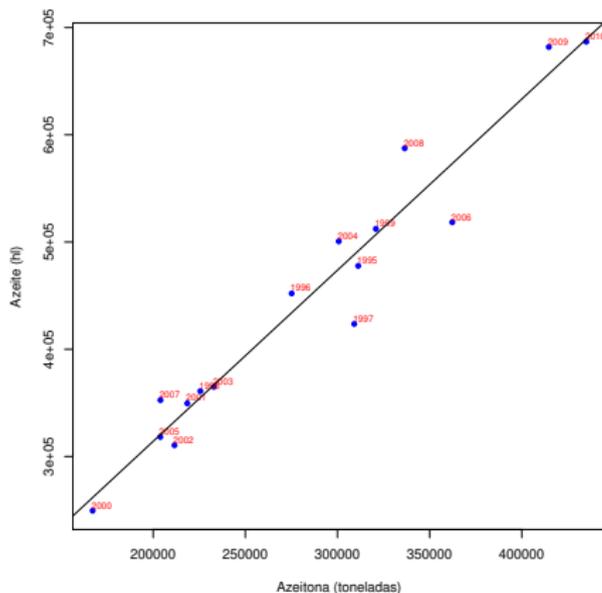
- 1 Regressão Linear Simples
- 2 Regressão Linear Múltipla
- 3 Análises de Variância (ANOVA)

2 Testes de hipóteses para dados de contagem (baseados na estatística χ^2 de Pearson).

1.1. 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;
 - ▶ **Propriedades** distribucionais **dos estimadores** do modelo;
 - ▶ **Intervalos de confiança** para os parâmetros;
 - ▶ **Testes de hipóteses** para os parâmetros;
 - ▶ **Análise dos resíduos** para validação do modelo e identificação de observações especiais.

1.2. 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.

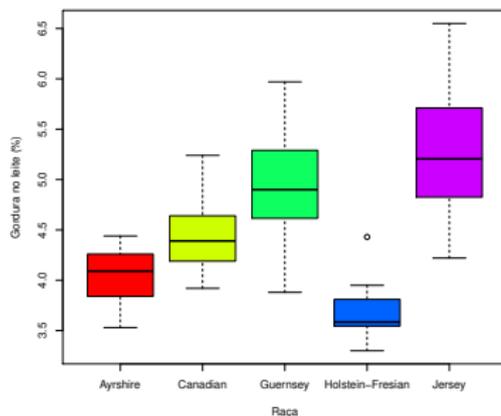
- **A inferência**

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

1.3. 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:



Há raças com leite mais gordo, ou as diferenças 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)

2. Testes de Hipóteses para dados de contagem (estatística χ^2 de Pearson)

Objectivo: Testar se dados de contagem são compatíveis com uma dada hipótese explicativa.

Exemplo: Em viticultura há técnicas de enxertia chamadas “enxertos prontos”. O sucesso depende de se verificar, ou não, o *pegamento* (formação do calo de enxertia). Deseja-se comparar o comportamento de três porta-enxertos específicos (1103P, SO4 e R99) para a casta Castelão. Eis os resultados dum estudo:

	1103P	SO4	R99
não pegamento	8	12	32
pegamento	954	943	939

Estes resultados são compatíveis com a hipótese de os porta-enxertos serem equivalentes, em matéria de pegamento? Ou a diferença nos resultados observados pode ser considerada **estatisticamente significativa**, havendo porta-enxertos melhores do que outros?

Testes χ^2 para dados de contagem (cont.)

- Teste de **ajustamento dum distribuição unidimensional** das contagens (as contagens dum experiência são compatíveis com uma distribuição Binomial? Ou uma Poisson?).
 - ▶ Probabilidades totalmente conhecidas
 - ▶ Probabilidades estimadas
- Testes para **tabelas de contingência** (contagens em **tabela de dupla entrada**, como no exemplo acima).
 - ▶ Probabilidades conhecidas (aplicações à teoria genética).
 - ▶ Teste de homogeneidade
 - ▶ Teste de independência

1 Referências Base:

- ▶ **Slides e apontamentos das aulas teóricas** (disponibilizados na página *web* da UC)
- ▶ **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 da disciplina de Estatística (2008/09)**, *Introdução à Aplicação R*, ver apontador na página web da disciplina - Materiais de Apoio - Aulas Práticas
- ▶ **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*, cran.r-project.org/doc/contrib/Torgo-ProgrammingIntro.pdf
- ▶ **Venables, W.N. e Ripley, B.D. (2002)**, *Modern Applied Statistics with S (fourth edition)*, Springer-Verlag [**BISA: U10-733**]