

Estatística e Delineamento

Jorge Cadima

Seção de Matemática (DCEB), Instituto Superior de Agronomia (UL)

2014-15

1 Professores:

- ▶ Jorge Cadima
- ▶ Fernanda Valente
- ▶ Manuel Campagnolo
- ▶ Elsa Gonçalves

2 Webpage: Sistema Fénix

3 Software:

- ▶ <http://www.r-project.org>
- ▶ <http://cran.r-project.org>

Enquadramento

Na UC **Estatística e Delineamento** admite-se que:

- houve frequência duma disciplina introdutória de Estatística no primeiro ciclo (semelhante à existente no ISA).
- são conhecidos conceitos como:
 - ▶ principais indicadores descritivos (média, variância, covariância, 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);
 - ▶ Intervalos de confiança e testes de hipóteses.

Quem não tem estes conhecimentos deve utilizar também os materiais de apoio da UC **Estatística** dos primeiros ciclos do ISA (ou equivalentes).

Programa

Nas disciplinas introdutórias de Estatística aborda-se o estudo das observações duma variável.

Nesta disciplina procura-se aprofundar o estudo de **possíveis causas explicativas da variabilidade associadas a essas observações**.

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

2 Modelo Linear.

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

1. 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 uma técnica de enxertia chamada “enxertos prontos”. O sucesso depende de se verificar, ou não, o chamado *pegamento* (formação do calo de enxertia). Deseja-se comparar o comportamento de três porta-enxertos específicos (1103P, SO4 e R99) para a casta Castelhão. 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** e há um porta-enxertos preferível?

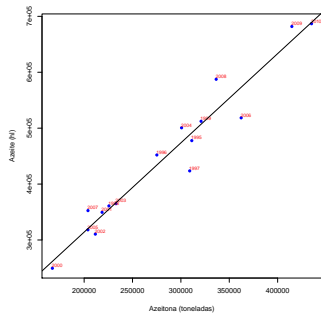
Testes χ^2 para dados de contagem (cont.)

- Breve **revisão de testes de hipóteses**.
- Teste de **ajustamento duma distribuição unidimensional** (as contagens duma experiência são compatíveis com uma distribuição Binomial? Ou uma Poisson?).
 - ▶ Probabilidades totalmente pré-especificadas
 - ▶ Probabilidades estimadas
- Testes para **tabelas de contingência** (contagens em **tabela de dupla entrada**, como no exemplo).
 - ▶ Aplicações à teoria genética.
 - ▶ Teste de homogeneidade
 - ▶ Teste de independência

2.1. Modelo Linear: regressão linear simples

Objectivo: Relacionar uma **variável resposta numérica** com um **preditor numérico**, através duma **relação linear**.

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



Modelo Linear: Regressão Linear Simples

- Contexto descritivo e estimação dos parâmetros da recta
- Relações não-lineares e transformações linearizantes
- O modelo para o contexto inferencial
- 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

2.2. Modelo linear: regressão linear múltipla

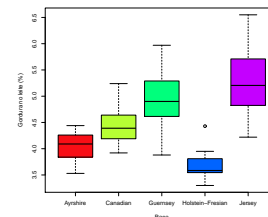
Objectivo: Relacionar uma **variável resposta numérica** com **dois ou mais preditores numéricos**, através duma **relação linear**.

- Uma ferramenta: a notação matricial
- Contexto descritivo e estimação dos parâmetros
- O modelo para o contexto inferencial
- Propriedades distribucionais dos estimadores
- Inferência
- Submodelos e selecção de submodelos
- Análise dos resíduos

2.3. 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-Friesian, 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)
- **Extensão:** O modelo a um factor, com **efeitos aleatórios**

Bibliografia

1 Referências Base:

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

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]**

Bibliografia (continuação)

Referências de apoio à utilização do R

- ▶ **Docentes da disciplina de Estatística (2008/09)**, *Introdução à Aplicação R*, www.isa.utl.pt/dm/estat/estat/AulasR.pdf
- ▶ **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]