

Estadística e Delineamento

Jorge Cadima

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

2015-16

1 Professores:

- ▶ Jorge Cadima
- ▶ Elsa Gonçalves
- ▶ Manuel Campagnolo
- ▶ Maria João Martins

2 Webpage: Sistema Fénix

3 Software:



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

Objectivos

- **Introduzir alguns dos principais modelos estatísticos:**
 - ▶ Regressão Linear (Simple e Múltipla)
 - ▶ Análises de Variância
- **Introduzir ferramentas para estudar dados de contagens:**
 - ▶ Testes χ^2 baseados na estatística de Pearson

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

Nas disciplinas introdutórias de Estatística aborda-se o estudo das observações numa variável. Nesta disciplina procura-se estudar variáveis explicativas dessas observações.

Enquadramento

Na UC **Estadística e Delineamento** admite-se que 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);
- Conceitos de intervalos de confiança e testes de hipóteses.

AVISO: Quem precisa de (re)ver estes conhecimentos deve utilizar também os materiais de apoio da UC **Estadística dos primeiros ciclos do ISA** (ou equivalentes), que estão disponíveis na página *web* dessa UC (seguir o apontador na página *web* de ED).

Aulas e horários de dúvidas

- **Aulas teóricas (2 vezes 1h por semana)** - Há dois blocos diferentes, ambos com aulas 3as. e 5as.-feiras.
- **Aulas práticas (2 vezes 1h30 por semana)** - Há oito turmas (inscrições via Fénix). Esta semana só funciona a segunda aula de cada turma.

AVISOS:

- O material de apoio para as aulas é disponibilizado na página web da UC (ver secção lateral de nome *Materiais de Apoio*).
- Antes da primeira aula prática confirmar que tem conta (informática) de aluno (em caso de problema contactar CIISA)
- Para a primeira aula prática levar (i) **Enunciados** dos Exercícios Introdutórios; e (ii) uma *pen* para guardar a sessão de trabalho.

AVISO: Existem **horários de dúvidas** 4 dias por semana (ver *webpage*). Todos os horários de dúvidas são para qualquer aluno da UC.

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:

- **Aprovação por testes:** classificação média igual ou superior a 9,5 valores no conjunto de dois testes, desde que em nenhum dos testes a classificação seja inferior a 8,0 valores.
- **Aprovação em exame:** classificação igual ou superior a 9,5 valores.
- Qualquer aluno inscrito na UC pode apresentar-se a avaliação de conhecimentos, desde que regularmente inscrito.
- Nas avaliações **não são admitidas calculadoras gráficas, nem qualquer tipo de equipamento electrónico.**
- **Proposta data primeiro teste:** Segunda-feira, 2 de Novembro, 18h.

Programa

- 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 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 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**, havendo porta-enxertos melhores do que outros?

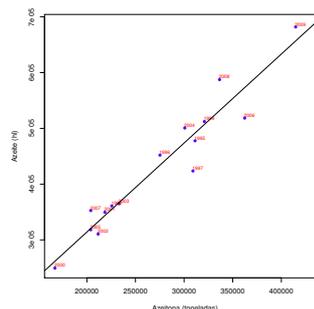
Testes χ^2 para dados de contagem (cont.)

- Breve revisão da teoria de testes de hipóteses.
- Teste de **ajustamento dum distribuição unidimensional** (as contagens dum 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 acima).
 - ▶ Aplicações à teoria genética (probabilidade conhecidas).
 - ▶ Teste de homogeneidade
 - ▶ Teste de independência

2.1. 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)
 - ▶ estimação dos parâmetros da recta;
 - ▶ Propriedades da recta ajustada;
 - ▶ Relações não-lineares e transformações linearizantes.
- **Inferência** (sobre a recta populacional)
 - ▶ 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.

2.2. Modelo linear: Regressão Linear Múltipla

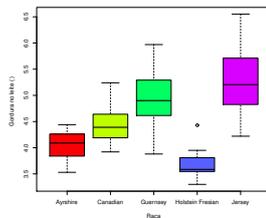
Objectivo: Relacionar uma variável resposta numérica com dois ou mais preditores numéricos, através dum relação linear.

- **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.

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)

3 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]