

XV Congresso Anual da Sociedade Portuguesa de Estatística

Lisboa, 19 a 21 de Agosto de 2007

# **Introdução aos Métodos Estatísticos Robustos**

**Ana M. Pires e João A. Branco**

Edições SPE

## Edições SPE

### Manuais

- *Introdução à Probabilidade e à Estatística - com complementos de Excel*, por Maria Eugénia Graça Martins

### Minicursos

- *Tópicos de Sondagens*, por Paulo Gomes
- *Controlo Estatístico de Qualidade*, por Ivette Gomes e Isabel Barão
- *Modelos Lineares Generalizados*, por Antónia Turkman e Giovanni Silva
- *Inferência sobre Localização e Escala*, por Fátima Brilhante, Dinis Pestana, José Rocha e Sílvio Velosa
- *Modelos Heterocedásticos. Aplicações com o software Eviews*, por Daniel Muller
- *Séries Temporais - Modelações Lineares e Não Lineares*, por Esmeralda Gonçalves e Nazaré Mendes Lopes
- *Uma Introdução à Análise de Clusters*, por João Branco
- *Introdução às Equações Diferenciais Estocásticas e Aplicações*, por Carlos Braumann
- *Outliers em Dados Estatísticos*, por Fernando Rosado
- *Introdução aos Métodos Estatísticos Robustos*, por Ana Pires e João Branco

### Actas

- *Afirmar a Estatística. Um Desafio para o Século XXI - Actas do VI Congresso Anual da SPE*. C. Paulino, A. Pacheco, A. Pires e F. da Cunha (Ed.)
- *Um Olhar sobre a Estatística - Actas do VII Congresso Anual da SPE*. P. Oliveira e E. Athayde (Ed.)
- *A Estatística em Movimento - Actas do VIII Congresso Anual da SPE*. M. M. Neves, J. Cadima, M. J. Martins e F. Rosado (Ed.)

### FICHA TÉCNICA:

**Título:** Introdução aos Métodos Estatísticos Robustos

**Autores:** Ana Pires e João Branco

**Editora:** Sociedade Portuguesa de Estatística

**Concepção gráfica da capa:** ISCTE - Instituto Superior de Ciências do Trabalho e da Empresa

**Produção gráfica e impressão:** Instituto Nacional de Estatística

**Tiragem:** 550 exemplares

**ISBN:** 978-972-8890-10-0

**Depósito legal:** 260674/07

- *Novos Rumos em Estatística - Actas do IX Congresso Anual da SPE.* L. Carvalho, F. Brillhante e F. Rosado (Ed.)
- *Literacia e Estatística - Actas do X Congresso Anual da SPE.* P. Brito, A. Figueiredo, F. Sousa, P. Teles e F. Rosado (Ed.)
- *Estatística com Acaso e Necessidade - Actas do XI Congresso Anual da SPE.* P. Rodrigues, E. Rebelo, e F. Rosado (Ed.)
- *Estatística Jubilar - Actas do XII Congresso Anual da SPE.* C. A. Braumann, P. Infante, M. M. Oliveira, R. Alpizar-Jara e F. Rosado (Ed.)
- *Ciência Estatística - Actas do XIII Congresso Anual da SPE.* L. Canto e Castro, E. G. Martins, C. Rocha, M. F. Oliveira, M. M. Leal e F. Rosado (Ed.)
- *Estatística: Ciência Interdisciplinar - Actas do XIV Congresso Anual da SPE.* M. E. Ferrão, C. Nunes e C. A. Braumann (Ed.)

## História da Estatística

- *Memorial da Sociedade Portuguesa de Estatística.* F. Rosado (Ed.)

## Prefácio

Uma questão concreta, um conjunto de dados e um modelo são os principais ingredientes que fazem funcionar a inferência estatística. Esta estrutura tem-se revelado muito apelativa e extremamente útil pois oferece ao utilizador da estatística a possibilidade de atacar uma multitude de problemas da vida real. Porém, esta abordagem introduz uma dificuldade funcional, a que a inferência estatística tradicional não dá relevo, que pode prejudicar seriamente os resultados de uma análise estatística de dados reais. A dificuldade resulta de um desajustamento entre o que se passa na teoria e o que se passa na realidade e traduz-se no facto de as hipóteses necessárias para o modelo funcionar, de acordo com o previsto, raramente serem encontradas nos dados para os quais o modelo foi afinal concebido. Um exemplo claro é o modelo de regressão linear com todas as hipóteses que se lhe costumam associar para que o método dos mínimos quadrados possa realizar eficientemente o seu trabalho de estimação. Sabe-se que se os dados aos quais se pretende ajustar o modelo de regressão não respeitarem as hipóteses que o modelo exige, as estimativas dos mínimos quadrados poderão dar uma ideia falsa sobre a estrutura dos dados. E quantos conjuntos de dados existem que satisfaçam exactamente tais hipóteses?

Cientes desta realidade incontornável os especialistas inquietos lançaram novos olhares sobre a estatística e fizeram nascer a estatística robusta. Na visão da estatística robusta não há um modelo único, formulado em termos de hipóteses rígidas, que explique a estrutura dos dados, mas há sim um conjunto de modelos que se situam numa vizinhança estreita de um modelo ideal e são esses modelos que estão em condições de explicar a estrutura dos dados. Ao adoptar esta estratégia a estatística robusta fica habilitada a controlar os desvios que os dados reais apresentam em relação ao modelo ideal que se lhe quer impor e que a estatística clássica não é capaz de dominar. É, no fundo, a flexibilidade contra a rigidez da visão clássica.

Este texto constitui o material de suporte ao mini-curso que antecede o XV Congresso Anual da Sociedade Portuguesa de Estatística. Depois de um panorama geral da estatística robusta que é dado no Capítulo 1, segue-se, no Capítulo 2, o conjunto de conceitos básicos da estatística robusta e, no Capítulo 3, aborda-se o problema geral da estimação e apresentam-se os estimadores mais conhecidos. No Capítulo 4 pode apreciar-se a actuação da estatística robusta no estudo da regressão, um dos métodos mais correntemente usados em estatística. O Capítulo 5 mostra o papel das ferramentas da estatística robusta e a sua interligação com a estatística clássica na análise de uma aplicação complexa.

O texto contém abundantes ilustrações gráficas e exemplos com dados simulados e dados reais que permitem compreender melhor os conceitos e mostrar a utilidade dos métodos robustos.

Em face do objectivo deste projecto não foram incluídos vários tópicos de muito interesse para a estatística robusta, em particular a análise multivariada onde se têm verificado avanços de grande relevo.

Este livro é em grande medida o resultado do estudo sobre o tema, iniciado pelos autores em 1989 com a participação no segundo curso ECAS (*European Courses in Advanced Statistics*), intitulado “*Robustness in Statistics – Theory and Applications*”, que teve lugar em Schloss Reisensburg, Alemanha, de 2 a 7 de Outubro, e do trabalho de investigação desenvolvido desde essa altura (integrado nas actividades do CEMAT – Centro de Matemática e Aplicações do Instituto Superior Técnico, financiado pela Fundação para a Ciência e a Tecnologia, entidade a quem se agradece o apoio).

Gostaríamos de deixar aqui um agradecimento às nossas colegas Conceição Amado, pela leitura atenta dos Capítulos 2 e 3, e Isabel Rodrigues, co-autora do trabalho em que se baseia o Capítulo 5. É evidente que quaisquer gralhas ou incorrecções são da nossa inteira responsabilidade e desde já agradecemos a todos os que queiram comunicar falhas que encontrarem ou contribuir com comentários e sugestões.

Instituto Superior Técnico,  
Lisboa, Junho de 2007

João A. Branco  
Ana M. Pires

## Índice

<b>1</b>	<b>Um panorama da estatística robusta</b>	<b>1</b>
1.1	Introdução	1
1.2	Motivação e nota histórica	3
1.3	Relações com outros métodos	12
1.4	A estatística robusta na prática	14
1.5	O presente e o futuro da estatística robusta	17
<b>2</b>	<b>Conceitos básicos</b>	<b>21</b>
2.1	Introdução	21
2.2	Curva de sensibilidade	22
2.3	Função de influência	28
2.3.1	Conceito de funcional	30
2.3.2	Definição e propriedades	35
2.3.3	Medidas de robustez baseadas na função de influência	42
2.3.4	Exemplos	45
2.3.5	Problemas multivariados e multiparamétricos	75
2.3.6	Generalizações	85
2.4	Robustez qualitativa	86
2.5	Ponto de rotura	88

2.6	Síntese das propriedades mais relevantes	95	4.3.5	Mínimos quadrados aparados (LTS)	189
<b>3</b>	<b>Estimação</b>	<b>97</b>	4.3.6	Estimadores-S	190
3.1	Introdução	97	4.3.7	Estimadores-MM	192
3.2	Os estimadores-M	98	4.3.8	Regressão mais profunda	193
3.2.1	Definição geral	98	4.4	Comparação de estimadores	197
3.2.2	Modelo de localização	101	4.5	Análise dos resultados de uma regressão robusta	197
3.2.3	Modelo de escala	125	4.5.1	Estimação de $\sigma$ e coeficiente de determinação	197
3.2.4	Situações multivariadas e multiparamétricas	144	4.5.2	Intervalos de confiança e testes de hipóteses	199
3.2.5	Modelo de localização e escala	145	4.5.3	Exemplos	203
3.3	Breve referência a outras classes de estimadores	148	4.6	Apreciação geral	216
3.4	Para além da estimação pontual	152	<b>5</b>	<b>Uma aplicação</b>	<b>219</b>
3.4.1	Distribuições assintóticas	153	5.1	Introdução	219
3.4.2	Correcções e outras aproximações	156	5.2	Enquadramento e descrição dos dados	221
3.4.3	<i>Jackknife</i> e <i>bootstrap</i>	157	5.3	Métodos	226
3.4.4	Robustez de intervalos de confiança e testes de hipóteses	158	5.3.1	Diagnóstico	228
<b>4</b>	<b>Regressão</b>	<b>161</b>	5.3.2	Estimação com erros correlacionados	232
4.1	Introdução	161	5.3.3	Estimação robusta com erros correlacionados	238
4.2	Méritos e defeitos do método dos mínimos quadrados	162	5.4	Resultados	245
4.2.1	Heterocedasticidade e não normalidade	165	5.5	Discussão e conclusões	248
4.2.2	Presença de <i>outliers</i>	173	<b>Referências Bibliográficas</b>	<b>249</b>	
4.3	Métodos robustos de regressão	178			
4.3.1	Mínimos desvios absolutos	182			
4.3.2	Estimadores-M	186			
4.3.3	Estimadores-M generalizados	187			
4.3.4	Mínima mediana dos quadrados (LMS)	188			