

# MEMORIAL DA SOCIEDADE PORTUGUESA DE ESTATÍSTICA

*Editor*

Fernando Rosado



*Título:*

Memorial da Sociedade Portuguesa de Estatística

*Editor:*

Fernando Rosado

Edição parcialmente subsidiada pela Fundação para a Ciência e Tecnologia

Execução gráfica

Jmm-Artes Gráficas, Lda.

*ISBN 972-8890-02-8*

*Depósito Legal: 226817/05*

*Tiragem: 1000 exemplares*

## Agradecimentos

À Editorial Pendor por ter concedido autorização para a publicação do texto do Prof. Tiago de Oliveira.

À Universidade Lusíada, por ter autorizado a publicação do texto do Prof. Pedro Braumann.

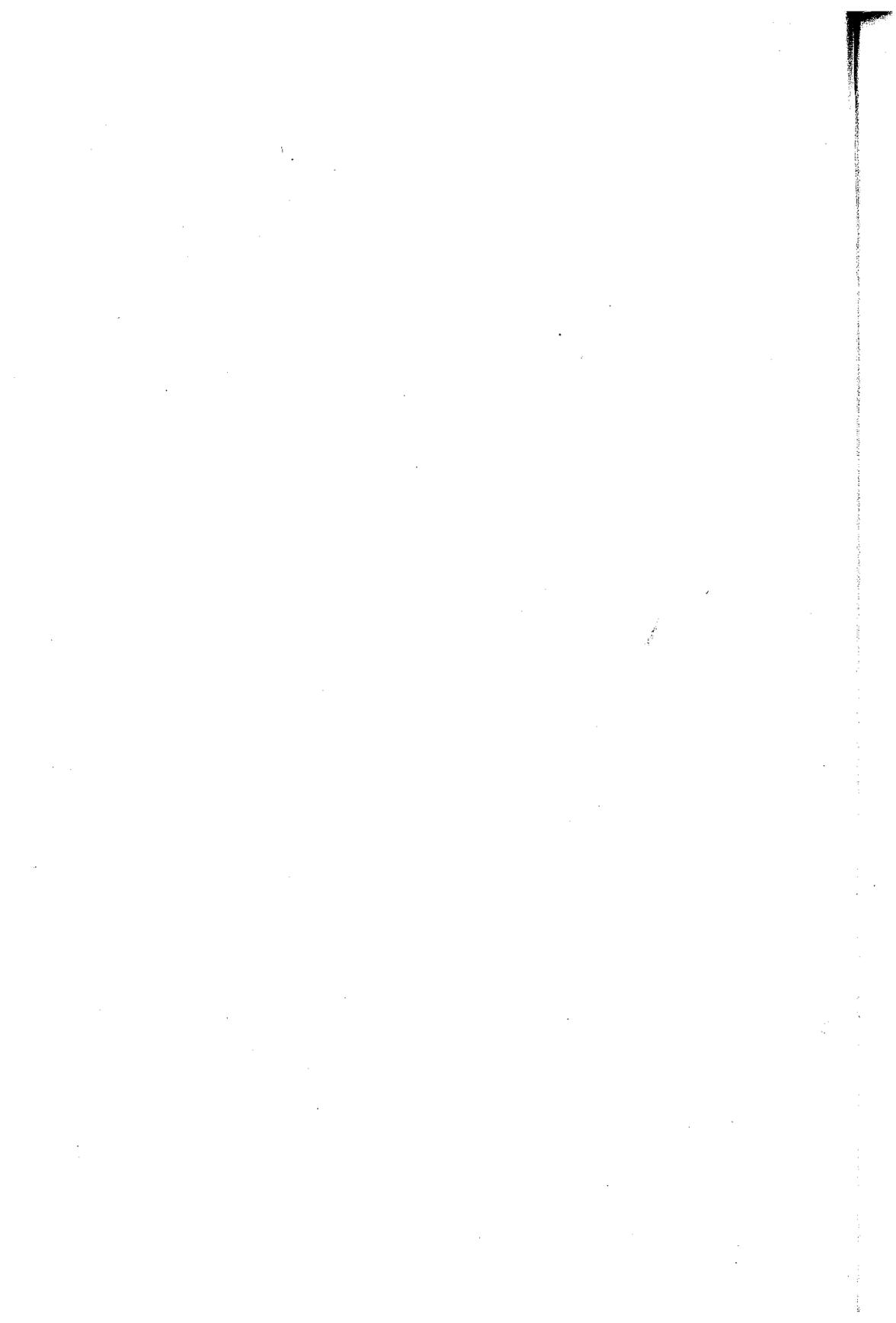
À Escolar Editora, por todo o apoio prestado na edição desta obra, em particular, aos Srs. Serafim Pereira e José Maria Marques pelo trabalho gráfico e de maquete, a concepção da capa e sugestões para a edição.

Aos Profs. José Carlos Tiago de Oliveira e Carlos Braumann na selecção de textos da autoria de seus pais – Prof. Tiago de Oliveira e Prof. Pedro Braumann – e que permitiu, *in memoriam*, incluí-los neste projecto.

Aos colegas, que desde o início encorajaram o projecto e se prontificaram a colaborar.

A todos, desejo manifestar o meu mais profundo reconhecimento.

*Fernando Rosado*



# Índice

Prefácio .....	xvii
Medida: Convergência de Estudos Dispersos para uma Teoria Unificada .....	1
<i>Pedro Bruno Teodoro Braumnann</i>	
A Estatística em Portugal nos últimos 25 anos .....	11
<i>Bento José Ferreira Murteira</i>	
A Educação em Portugal e o Século XX .....	23
<i>José Tiago da Fonseca Oliveira</i>	
O Ensino da Estatística em Portugal nos Últimos 150 Anos .....	31
<i>Maria de Fátima Fontes de Sousa</i>	
“Extremistas” num extremo da Europa .....	37
<i>Maria Ivette Gomes</i>	
Os Primeiros 25 Anos, Alicerces do Futuro .....	47
<i>Dinis Duarte Pestana</i>	
O Acaso e a Vida .....	53
<i>Carlos A. Braumann</i>	
Estatística Robusta: contribuição portuguesa .....	73
<i>João A. Branco</i>	
Estatística dos doutoramentos em Estatística .....	93
<i>Maria Antónia Amaral Turkman</i>	
Pluridisciplinaridade, Interdisciplinaridade e Convergências .....	117
<i>Helena Bacelar-Nicolau</i>	
O Sistema Estatístico Nacional .....	125
<i>Manuel José Vilares</i>	
“Outliers” em Português .....	139
<i>Fernando Rosado</i>	

Ciências Actuarias .....	151
<i>Maria de Lourdes Centeno</i>	
A Evolução do Sistema Estatístico Nacional após a Reestruturação de 1989 .....	163
Pontos Fortes e Pontos Fracos	
<i>Paulo Gomes</i>	
Modelos Normais Ortogonais: dos Anos 70 ao Presente .....	179
<i>João Tiago Mexia</i>	
Estabilidade e comportamento limite forte das estatísticas de ordem.....	199
<i>Margarida Brito</i>	
A Evolução da Estatística Bayesiana em Portugal.....	215
<i>Carlos Daniel Paulino</i>	
Estatística e <i>Agronomia</i> — Percurso e Ensino. Algumas Notas Históricas.....	219
<i>M. Manuela Neves</i>	
Sobre o Estudo de Séries Temporais e Previsão. Breve Nota Pessoal.....	231
<i>Nuno Crato</i>	

*À memória do Prof. Tiago de Oliveira*  
*À memória do Prof. Pedro Braumann*  
*Aos Fundadores da Sociedade Portuguesa de Estatística*



# Sociedade Portuguesa de Estatística

Fundada em 28 de Novembro de 1980

## Escritura de Formação - Outorgantes

José Tiago da Fonseca Oliveira

José Joaquim da Silva Dias Coelho

Pedro Bruno Teodoro Braumann

Bento José Ferreira Murteira

Maria Fernanda Neto Ramalhoto

Maria Ivette Leal de Carvalho Gomes

Dinis Duarte Ferreira Pestana

Maria Cristina de Sales Viana Serôdio Sernadas

Amílcar dos Santos Costa Sernadas

Maria Antónia da Conceição Abrantes Amaral Turkman

Maria de Fátima Fontes de Sousa

## Sócios Fundadores:

Todos os que aderiram à formação da Sociedade até 31 de Julho de 1981  
(artº 4 dos Estatutos)

## Sócios Honorários

Bento Murteira (eleito em 1993)

C. R. Rao (eleito em 2002)

David Cox (eleito em 2005)

# Presidentes da Sociedade Portuguesa de Estatística

J. Tiago de Oliveira (1981 - 1989)

M. Ivette Gomes (1990 - 1994)

João A. Branco (1994 - 2000)

Fernando Rosado (2000 - 2006)

# Órgãos Administrativos Estatutários (2003-2006)

## *Mesa da Assembleia Geral*

Presidente:

Carlos Braumann (Universidade de Évora)

Primeiro Vogal:

Salomé Cabral (Universidade de Lisboa)

Segundo Vogal:

Russell Alpizar Jara (Universidade de Évora)

## *Direcção*

Presidente:

Fernando Rosado (Universidade de Lisboa)

Vice-Presidente:

Lucília Carvalho (Universidade de Lisboa)

Secretário:

Manuela Neves (Universidade Técnica de Lisboa)

Secretário-Adjunto:

Paula Brito (Universidade do Porto)

Tesoureiro:

Zilda Mendes (Associação Nacional de Farmácias)

## *Conselho Fiscal*

Presidente:

Daniel Muller (Universidade Técnica de Lisboa)

Secretário:

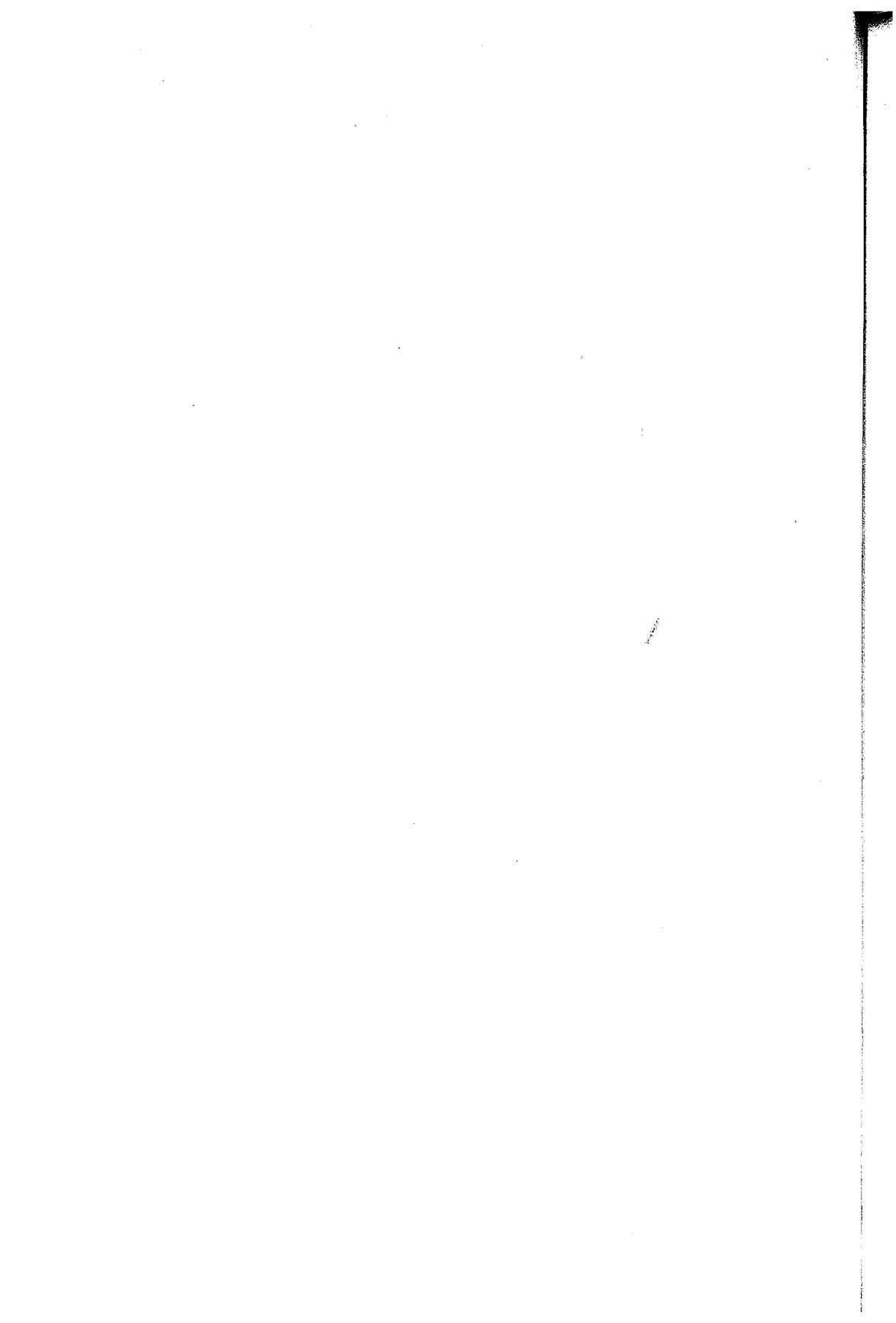
Ivone Figueiredo (Instituto de Investigação das Pescas e do Mar)

Relator:

Isabel Natário (Universidade Nova de Lisboa)

## Colóquios / Conferências / Congressos

- I Colóquio de Estatística e Investigação Operacional - 1975, Lisboa.
  - II Colóquio de Estatística e Investigação Operacional - 1981, Fundão/Covilhã.
  - III Colóquio de Estatística e Investigação Operacional - 1984, Lagos.
- 
- I Conferência em Estatística e Optimização promovida pelo Centro de Estatística e Aplicações da Universidade de Lisboa e pelo DEIOC/FCUL realizada de 3 a 5 de Dezembro de 1990 em Tróia. O elevado número de participantes, cerca de 200, mostrou como era necessário e urgente a realização de uma conferência cujos objectivos principais fossem o de promover o encontro e a troca de experiências entre os trabalhadores daqueles dois ramos do saber.
- 
- I Congresso 1993 – Vimeiro (Comissão Organizadora presidida por Dinis Pestana da Universidade de Lisboa)
  - II Congresso 1994 – Luso (Comissão Organizadora presidida por Nazaré Lopes da Universidade de Coimbra)
  - III Congresso 1995 – Guimarães (Comissão Organizadora presidida por Paulo Gomes do Instituto Nacional de Estatística)
  - IV Congresso 1996 – Funchal (Comissão Organizadora presidida por Rita Vasconcelos da Universidade da Madeira)
  - V Congresso 1997 – Curia (Comissão Organizadora presidida por Beatriz Matias da Universidade de Aveiro)
  - VI Congresso 1998 – Tomar (Comissão Organizadora presidida por Carlos Daniel Paulino da Universidade Técnica de Lisboa)
  - VII Congresso 1999 – Fão (Comissão Organizadora presidida por Pedro Oliveira da Universidade do Minho)
  - VIII Congresso 2000 – Peniche (Comissão Organizadora presidida por Manuela Neves da Universidade Técnica de Lisboa)
  - IX Congresso 2001 – Ponta Delgada (Comissão Organizadora presidida por Fátima Brilhante da Universidade dos Açores)
  - X Congresso 2002 – Porto (Comissão Organizadora presidida por Paula Brito da Universidade do Porto)
  - XI Congresso 2003 – Faro (Comissão Organizadora presidida por Efigénio Rebelo da Universidade do Algarve)
  - XII Congresso 2004 – Évora (Comissão Organizadora presidida por Carlos Braumann da Universidade de Évora)
  - XIII Congresso 2005 – Ericeira (Comissão Organizadora presidida por Luísa Canto e Castro da Universidade de Lisboa)
  - XIV Congresso 2006 – Covilhã (Comissão Organizadora presidida por Maria Eugénia Ferrão da Universidade da Beira Interior)



# Prefácio

A Sociedade Portuguesa de Estatística – SPE – tem 706 Sócios efectivos, 10 Sócios Colectivos e 3 Sócios Honorários.

Para viver é preciso esquecer; dizem alguns grandes pensadores. Mas, dizem mal! defendem outros. A cultura da memória - a não confundir com culto do passado - é filha da informação e do discernimento e mãe da lucidez. Dela, decerto, também se vive!

Em data jubilar – nos 25 anos – (também) é importante (e talvez fundamental!) explicar aos jovens (estatísticos) como vamos de investigação e encontrar a motivação para uma carreira de crucial importância nas mais diversificadas áreas do saber. Neste âmbito elevamos a Estatística ao lugar cimeiro, desejando contribuir para o maior e melhor esclarecimento de “como foi e como é em Portugal”. E como será!?

A SPE surgiu “na sequência” do 25 de Abril de 1974. Talvez por acaso, mas decerto por necessidade! Seguramente, foi o “sentimento de necessidade” mas também o despertar para novos projectos que, congregando valores científicos em redor do fundador pioneiro – o Professor Tiago de Oliveira – deram corpo à associação científica a que, por vontade dos sócios, presido e à qual desde a fundação pertencemos.

De facto, no final da década de 70 assistíamos ao regresso de jovens doutores que no estrangeiro tinham obtido a sua graduação. Portugal era parco nestas graduações e “em espírito de aventura” alguns pioneiros deram grandes passos para aproximar o ritmo científico português daquele que era marcado por algumas universidades de referência.

## **25 anos passados!**

Neste aniversário solene, todos vivemos “novos tempos”, outras leis e regras também (para e) na ciência, (para e) nas universidades, (para e) nas carreiras científicas e profissionais, (para e) no sistema de financiamento das actividades e das instituições... enfim, novos métodos e diferentes desafios!? Uma nova etapa a percorrer, talvez iniciando uma nova era...

Nesta nova era, que já vivemos, muito ligada a um desenvolvimento intensivo, ao controlo do rigor, à avaliação do resultado, da solidez das instituições, da necessidade dos investigadores e do apuramento de custos e benefícios dos investimentos.

A ciência em geral e a estatística em particular é uma nobre actividade, necessária ao corpo e ao espírito, indispensável ao bem-estar e à felicidade. Mas, a ciência é cara. Só os ricos a podem praticar ... e os pobres se a praticam ficam mais pobres. Embora exigindo

Ciências Actuarias .....	151
<i>Maria de Lourdes Centeno</i>	
A Evolução do Sistema Estatístico Nacional após a Reestruturação de 1989.....	163
Pontos Fortes e Pontos Fracos	
<i>Paulo Gomes</i>	
Modelos Normais Ortogonais:dos Anos 70 ao Presente.....	179
<i>João Tiago Mexia</i>	
Estabilidade e comportamento limite forte das estatísticas de ordem.....	199
<i>Margarida Brito</i>	
A Evolução da Estatística Bayesiana em Portugal.....	215
<i>Carlos Daniel Paulino</i>	
Estatística e <i>Agronomia</i> — Percurso e Ensino. Algumas Notas Históricas.....	219
<i>M. Manuela Neves</i>	
Sobre o Estudo de Séries Temporais e Previsão. Breve Nota Pessoal.....	231
<i>Nuno Crato</i>	

publicado no Boletim Informativo de Estatística e Investigação Operacional, Nº 1, Fevereiro de 1979. Este artigo do Prof. Tiago de Oliveira é, de facto, uma excelente síntese histórica das origens da SPE e por consequência também uma boa referência na história da estatística em Portugal.

“Talvez tenha algum interesse referir, em curso rápido, a pesquisa, com componentes teóricas, que recentemente se tem feito entre nós, após a II Guerra Mundial.

Essencialmente parece que os núcleos fundamentais de estudiosos têm estado ligados ao Instituto Superior de Ciências Económicas e Financeiras, à Faculdade de Ciências de Lisboa e, também, de índole bastante mais aplicada, ao Instituto Nacional de Estatística. Referimos, para breve notícia, que há alguns estudos individuais de investigação aplicada e ainda de outras organizações, como o Banco de Portugal e também grupos incipientes.

A actividade do Centro de Estudos de Estatística Económica (do IAC) ligado a gente do ISCEF, e que se estende dos anos 50 a meados dos anos 60, foi essencialmente votada ao estudo de metodologia ligada à pesquisa económica (em grande parte crono-séries e programação) e um pouco à pesquisa demográfica. Este núcleo, durante a sua experiência, à volta de uma dúzia de anos, trouxe a Portugal alguns investigadores de renome. Deve ainda salientar-se o estudo feito por 1963 pelo Centro relativo às necessidades educacionais para 1975, integrado no Projecto Regional do Mediterrâneo.

Virado mais à investigação teórica, desde meados dos anos 50, no Seminário de Matemática, depois no Centro de Matemáticas Aplicadas e, actualmente, no Centro de Estatística e Aplicações (do IAC e, agora do INIC) a gente da Faculdade de Ciências de Lisboa e outros têm prosseguido estudos, essencialmente em teoria das probabilidades, estatística de extremos, processos estocásticos e planeamento de experiências, além do apoio consultivo a investigadores em Biologia, Medicina, Psicologia, História, etc. O trazer de alguns investigadores e a realização de cursos têm sido actividades, também, dos sucessivos centros. Actualmente, no CEA, as linhas de pesquisa são a decisão estatística, em particular relativa a extremos estatísticos, processos estocásticos e aplicações, planeamento de experiências e programação e aplicações.

Constituídos em meados dos anos 40 os Centros de Estudos Económicos e Demográficos do Instituto Nacional de Estatística têm executado alguns estudos ligados à sua definição. Deverão recordar-se, entre outros, os estudos de demografia portuguesa realizados.

É bem pouco, quer na área de investigação teórica, quer na das aplicações o que entre nós se tem realizado, muitas vezes fruto do interesse e persistência individual. Países da dimensão do nosso têm, há muito, Institutos de Estatística nas Universidades em ligação com os Institutos de Aplicação. Estamos bem atrasados! Temos, pois, de avançar rapidamente para o futuro, para o que temos gente capaz. Não se pode perder mais tempo!”

A actual direcção vive o último ano do segundo mandato que, em assembleia eleitoral de 2002, lhe foi conferido.

Nunca poderei exprimir plenamente a minha gratidão a todos e a cada um dos sócios que sempre deram o maior apoio e todo o esforço para ajudar no sucesso das iniciativas a

grande esforço e dedicação, a solução deve estar em (apesar de tudo) fazer ciência para caminhar na saída daquele dilema.

É o que se exige aos estatísticos portugueses congregados em torno de um projecto líder – a SPE!

Aproveitemos a simbólica data jubilar para uma experiência de mudança no sentido de concretizar a consolidação que se sente no meio científico dos estatísticos portugueses – desde a política editorial até à excelência de investigação.

A SPE vive um momento de grande dinamismo que se deve registar. Num contexto simples de “cultura da memória”, desafiei vários seniores estatísticos portugueses para, em torno de um projecto literário, testemunharmos vivências e relatarmos factos científicos vividos de modo (a ajudar) a construir (um)a história da estatística em Portugal. Como em todos os projectos editoriais deste tipo, depois da congratulação pela (boa) ideia, surgiu a fase de concretização que requer, desde logo, sincronismo de etapas para todos os autores, exigindo trabalho redobrado pela diversidade de agendas e disponibilidades profissionais dos diversos intervenientes. O grande esforço por todos desenvolvido, deve ser salientado e permitiu que chegássemos à meta, transportando um documento único na história da estatística em Portugal.

O nosso trabalho é apresentado nos 19 capítulos desta obra, sequenciados de acordo com a antiguidade académica dos seus autores.

O texto produzido contém, de facto, a maior parte da recente história da Ciência Estatística em Portugal. Assim, são diversos os acontecimentos que, simultaneamente são relatados por diferentes autores. O cruzamento de toda a informação poderá conduzir o leitor a “alguma imprecisão”, por exemplo em datas – uma tese de doutoramento entregue no final do ano, decerto é discutida no ano seguinte e poderá proporcionar diferentes registos. A revisão desses “detalhes” implicaria uma morosidade de edição que, de todo, não se justificará. Salientada esta questão, resta desejar que todas essas (pequenas) incorrecções se possam esclarecer facilmente e que (elas também) possam estimular a pesquisa de modo a construir o maior rigor. O fundamental decerto está fiel e, por isso, estamos a elaborar memórias!

E porque, memorial é obra literária que relata factos históricos... em tempos de aniversário jubilar, o projecto inicial de “um livro sobre os 25 anos SPE” transformou-se no **Memorial da Sociedade Portuguesa de Estatística** que tenho o maior prazer em apresentar.

Muita história (mais ou menos recente) da estatística e dos estatísticos portugueses pode ser lida em documentos que fazem parte do espólio da SPE – como, por exemplo, o Boletim Informativo – ou daquele que foi o seu antecedente – o Boletim Informativo de Estatística e Investigação Operacional. Em lugar de destaque, obviamente, devemos colocar as Actas dos Congressos Anuais publicadas desde 1993 e que, além da distribuição pelos congressistas, nos últimos anos temos enviado para cerca de 90 bibliotecas de todo o país.

Então, de Memorial se trata! .... e porque faz história transcrevo: “Alguns Núcleos Recentes de Investigação Estatística em Portugal” da autoria de J. Tiago de Oliveira, e

publicado no Boletim Informativo de Estatística e Investigação Operacional, Nº 1, Fevereiro de 1979. Este artigo do Prof. Tiago de Oliveira é, de facto, uma excelente síntese histórica das origens da SPE e por consequência também uma boa referência na história da estatística em Portugal.

“Talvez tenha algum interesse referir, em curso rápido, a pesquisa, com componentes teóricas, que recentemente se tem feito entre nós, após a II Guerra Mundial.

Essencialmente parece que os núcleos fundamentais de estudiosos têm estado ligados ao Instituto Superior de Ciências Económicas e Financeiras, à Faculdade de Ciências de Lisboa e, também, de índole bastante mais aplicada, ao Instituto Nacional de Estatística. Referimos, para breve notícia, que há alguns estudos individuais de investigação aplicada e ainda de outras organizações, como o Banco de Portugal e também grupos incipientes.

A actividade do Centro de Estudos de Estatística Económica (do IAC) ligado a gente do ISCEF, e que se estende dos anos 50 a meados dos anos 60, foi essencialmente votada ao estudo de metodologia ligada à pesquisa económica (em grande parte crono-séries e programação) e um pouco à pesquisa demográfica. Este núcleo, durante a sua experiência, à volta de uma dúzia de anos, trouxe a Portugal alguns investigadores de renome. Deve ainda salientar-se o estudo feito por 1963 pelo Centro relativo às necessidades educacionais para 1975, integrado no Projecto Regional do Mediterrâneo.

Virado mais à investigação teórica, desde meados dos anos 50, no Seminário de Matemática, depois no Centro de Matemáticas Aplicadas e, actualmente, no Centro de Estatística e Aplicações (do IAC e, agora do INIC) a gente da Faculdade de Ciências de Lisboa e outros têm prosseguido estudos, essencialmente em teoria das probabilidades, estatística de extremos, processos estocásticos e planeamento de experiências, além do apoio consultivo a investigadores em Biologia, Medicina, Psicologia, História, etc. O trazer de alguns investigadores e a realização de cursos têm sido actividades, também, dos sucessivos centros. Actualmente, no CEA, as linhas de pesquisa são a decisão estatística, em particular relativa a extremos estatísticos, processos estocásticos e aplicações, planeamento de experiências e programação e aplicações.

Constituídos em meados dos anos 40 os Centros de Estudos Económicos e Demográficos do Instituto Nacional de Estatística têm executado alguns estudos ligados à sua definição. Deverão recordar-se, entre outros, os estudos de demografia portuguesa realizados.

É bem pouco, quer na área de investigação teórica, quer na das aplicações o que entre nós se tem realizado, muitas vezes fruto do interesse e persistência individual. Países da dimensão do nosso têm, há muito, Institutos de Estatística nas Universidades em ligação com os Institutos de Aplicação. Estamos bem atrasados! Temos, pois, de avançar rapidamente para o futuro, para o que temos gente capaz. Não se pode perder mais tempo!”

A actual direcção vive o último ano do segundo mandato que, em assembleia eleitoral de 2002, lhe foi conferido.

Nunca poderei exprimir plenamente a minha gratidão a todos e a cada um dos sócios que sempre deram o maior apoio e todo o esforço para ajudar no sucesso das iniciativas a

que nos propusemos nos últimos seis anos em que estivemos à frente dos destinos da SPE.

Tivemos oportunidade de organizar seminários e congressos dinâmicos, por todo o país, com sucesso, editando anualmente em Actas o produto do trabalho científico, em todos os campos valorizando a SPE.

Este Memorial da Sociedade Portuguesa de Estatística é publicado em final de mandato da actual equipa directiva da SPE. É também momento oportuno para rejubilar; pelo grupo dedicado, que comigo, mais de perto, colaborou e com o qual tive o prazer da vivência da amizade acrescida de excelentes momentos profissionais e académicos. Desde o primeiro minuto, para nos lançarmos, aceitando o desafio de candidatura e de “submissão a votos”, foram seis anos de total dedicação à SPE que devo testemunhar. Com a Lucília, a Manuela, o Manuel João (no primeiro triénio), a Zilda e a Paula (no segundo triénio) trabalhei formando uma equipa que soube dar resposta a todos os projectos que avançámos. É também devido um agradecimento especial à Margarida pela prontidão, disponibilidade e dedicação sempre manifestadas em todas as tarefas administrativas que lhe foram confiadas.

Todas as épocas têm uma atmosfera. O dinamismo a que assistimos na SPE é consequência natural da resposta empenhada de todos os sócios. Com todos contámos desde 1999 no momento da apresentação do nosso programa de candidatura. Sempre eles estiveram presentes e a todos reconhecidamente agradecemos o apoio que sentimos.

De todos, muito bem dizemos!

De nós, tentámos que se diga que procurámos servir a SPE.

A Sociedade Portuguesa de Estatística e Investigação Operacional foi fundada em 1980 usando a sigla SPEIO. Inicialmente era uma organização com interesses científicos em Estatística e Investigação Operacional.

Durante os primeiros dez anos consolidou-se e melhor se definiram os diferentes rumos científicos conduzindo em termos práticos na separação do grupo de investigação operacional que se juntou à recém criada Associação Portuguesa de Investigação Operacional (APDIO).

A primeira Direcção da SPEIO - presidida por Tiago de Oliveira - foi eleita em 15 de Dezembro de 1980. Posteriormente, tendo sido alterados os estatutos em 21 de Fevereiro de 1982, foi realizada uma Assembleia Geral Extraordinária onde foram eleitos novos órgãos administrativos. A SPEIO/SPE desde muito cedo procurou projectar-se além fronteiras tendo começado por se filiar no ISI – International Statistical Institute no início de 1982. Desde 1980 até 1985 foram realizados 2 Colóquios. Desde 1985 até 1990 foi experimentada alguma inactividade. Em Junho de 1990 houve eleições. Ivette Gomes foi eleita Presidente e durante os tempos seguintes foi-se consolidando o projecto que conduziu à actual SPE, inclusive com a necessária alteração de estatutos. Em 15 de Fevereiro de 1991 em Assembleia Geral Extraordinária a SPE deu o grande passo na consolidação como associação científica. Desde 1993 que se têm realizado congressos anuais. Em Janeiro de 1994 iniciou funções uma nova equipa, liderada por João Branco e que tinha sido eleita em Junho de 1993, na Assembleia Geral realizada durante o I Congresso. Esta equipa continuou até 13 de Janeiro de 2000.

A Sociedade Portuguesa de Estatística não foi um objectivo inicial dos primeiros cientistas que, em Portugal nos anos 70, se congregaram em torno da Estatística e da Investigação Operacional. No Editorial do Boletim Informativo de Estatística e Investigação Operacional publicado em Novembro de 1980 pelo Centro de Estatística e Aplicações (INIC) e que actualmente está integrado na Universidade de Lisboa, escreve o editor: "Antes de concluir este editorial não queremos deixar de assinalar a fundação da "Sociedade Portuguesa de Estatística e Investigação Operacional" de que damos notícia mais detalhada na rubrica "Noticiário"...". De facto, essa notícia é apresentada na página 29 do referido Boletim. Embora com uma modesta notícia, de facto, tinha sido dado um grande passo científico nos domínios da estatística e da investigação operacional fruto do dinâmico grupo de cientistas daquelas áreas que integravam o Centro de Estatística e Aplicações da Universidade de Lisboa (aliás em cujas instalações se encontra ainda a sede da SPE – art 2 dos estatutos).

Resumidamente, este é o ponto da situação de um ciclo que começou há 25 anos. Sobre os anos passados, desde então, falam os autores que gentilmente acederam ao convite do Presidente da SPE para perpetuarem as suas vivências científicas no presente MEMORIAL da Sociedade Portuguesa de Estatística!



Um "simbólico dado" continua como bandeira desta associação científica – no início com a sigla (5 pintas) SPEIO ao alto e, desde 1991 "reequilibrado" com SPE.

Alea jacta est!

Os dados estão lançados! Mas, não apenas para registar os "sucessos"...

A SPE vale o que (no todo) conseguirem os seus sócios. Decerto, a Sociedade Portuguesa de Estatística, onde cada sócio é parte interveniente e fundamental em algum momento, continuará cada vez com mais sucesso, na certeza de que este é uma estatística "somatório" de pequenos acasos.

Esteja cada um atento ao seu momento de acção!

Fernando Rosado

*Presidente da Sociedade Portuguesa de Estatística*



# Medida: Convergência de Estudos Dispersos para uma Teoria Unificada

*Pedro Bruno Teodoro Braumann*

Oração de sapiência proferida na Universidade Lusíada  
em 22 de Março de 1993

1. Para começar, convém antecipar que a exposição subsequente contém asseverações a três níveis: umas que estão ao alcance de qualquer um, outras cujo entendimento pressupõe conhecimentos matemáticos administrados nos cursos liceais ou similares e, uma vez por outra, na primeira ou nas duas primeiras cadeiras de Análise Matemática dum curso superior, conforme as orientações dos respectivos programas; as terceiras de índole mais especializada, referidas sempre em termos genéricos que o leigo na matéria poderá aceitar sem relutância, apelando ao senso comum e confiando na capacidade dos matemáticos no que concerne à justificação das partes omissas. Não haverá nem fórmulas nem símbolos complicados.

2. Suponhamos (o famoso suponhamos dos matemáticos), talvez melhor não custa imaginar, que ouvimos falar pela primeira vez na existência dum ramo da Matemática intitulado “Teoria da Medida”, abreviadamente “Medida”. O que devemos pensar?

Bom, é natural que uma medida seja o valor numérico de algo que se mede, valor esse expresso numa unidade convenientemente escolhida. Por exemplo, o número de boletins de voto nulos que tenham sido depositados numa urna eleitoral; o valor do comprimento de certo segmento de recta, expresso digamos em centímetros; o valor do peso de um sólido, expresso digamos em gramas; etc.

Neste contexto o valor em causa pode obter-se, com precisão tolerável, recorrendo por ordem a: um processo de contagem; uma fita métrica apropriada; uma balança de tipo adequado; etc.

Em todos os exemplos citados e em muitos outros que se poderiam juntar, topamos com resultados numéricos que derivam de medições e a que parece justo atribuir a

designação de medidas. Será **premature** falar numa teoria matemática da medida; todavia **perspectivamos uma noção intuitiva de medida**, ao alcance de qualquer um, **digamos uma noção de medida popular** que pode apresentar-se sob formas bastante diversas e que retomaremos mais adiante. Passemos a palavra ao matemático.

3. Chegados a este ponto, é de esperar que o matemático, distanciando-se do leigo e atendendo porventura a razões ainda não apresentadas pretenda introduzir uma teoria matemática conforme. Para este efeito, é prudente e chega mesmo a ser indispensável para certos propósitos que se caracterize o **ambiente** ou o **enquadramento** no qual convém trabalhar. Assim se evitam ambiguidades, paradoxos e outros percalços pouco recomendáveis.

Neste contexto, começa-se por escolher um **conjunto pré-fixado** que seja não vazio, quer dizer, tenha elementos. Ele será o conjunto de trabalho e vamos chamar-lhe **universo** e também **espaço** onde a designação alternativa é preferida pelos **medidistas**, quer dizer os especialistas da medida, isso apesar de se prestar a certa discussão.

Claro que existem **conjuntos formados por elementos pertencentes ao espaço proposto**, incluindo o **conjunto máximo coincidente com o próprio espaço**, o **conjunto mínimo coincidente com o vazio** e em geral diversos conjuntos intermédios.

Tendo por elementos tais conjuntos no espaço dado podem formar-se **novos conjuntos**, de categoria superior, a que vamos chamar classes; isso no intuito de evitar confusões com os conjuntos primitivos, que passam a ser de categoria inferior <sup>1</sup>.

Encontram-se bastante divulgadas as operações (internas) que podem efectuar-se sobre conjuntos contidos num espaço dado: a **complementação** ou passagem dum conjunto para outro formado pelos elementos que não pertencem ao primeiro; a **união** ou seja a passagem de conjuntos propostos para o novo conjunto formado pelos elementos pertencentes a algum dos primeiros conjuntos; a **adição** ou seja o caso particular da união em que os conjuntos propostos são **disjuntos dois a dois**, quer dizer sem elementos comuns quando tomados dois a dois; a **intersecção** ou seja a passagem de conjuntos propostos para o novo conjunto formado pelos elementos comuns aos primeiros conjuntos; também aparecem **dois tipos de subtracção** entre dois conjuntos que facilmente se reduzem às operações já mencionadas.

Ora postos perante uma classe de conjuntos tirada de um espaço dado, pode acontecer que uma certa operação sobre conjuntos da classe considerada conduza a um novo conjunto a um novo conjunto **exterior** a essa classe. Pôr exemplo: se tomarmos o espaço formado pelos (números) dígitos e se aí introduzirmos a classe formada pelos conjuntos com um número ímpar de elementos, então a união do conjunto formado pelos 3 primeiros dígitos e do conjunto formado pelos 3 últimos dígitos, ambos pertencentes à classe tomada, conduz a um novo conjunto com 6 dígitos o qual é exterior à classe tomada.

<sup>1</sup> Citemos, a propósito, a expressão "o conjunto de todos os conjuntos" a qual se torna ambígua ou paradoxal por omissão – inconsciente ou deliberada – do facto de os conjuntos envolvidos não se inserirem todos no mesmo ambiente.

4. Fixado o seu espaço, o medidista prefere trabalhar com uma classe que seja **segura** (para os seus fins). O que se deve entender por uma classe segura?

O medidista, ao **contrário do topologista**, reconheceu poder cobrir as suas principais necessidades de associação de conjuntos contidos no espaço proposto, recorrendo a **operações intransnumeráveis**, quer dizer, sobre conjuntos – em número finito ou infinito – que possam referenciar-se por enumeração consecutiva. Frequentemente, basta-lhe considerar conjuntos em número finito; todavia, a **hipótese duma infinidade de conjuntos faz falta para conseguir uma generalidade aceitável e para poder recorrer aos préstimos da análise clássica** (que se esteia no conceito de infinito).

Ora bem, uma classe segura – denominada álgebra- $\sigma$  (a letra grega sigma) na linguagem típica usada pelos medidistas – dizíamos uma classe segura é uma classe da qual **não se pode sair ou descarrilar** quando se passa ao resultado de operações intransnumeráveis efectuadas sobre os seus conjuntos. Prova-se que uma tal classe **não pode deixar de fora nem o próprio espaço nem o conjunto vazio**. Por outro lado, é praticamente óbvio que é **segura a classe máxima** ou seja a classe formada por todos os conjuntos contidos no espaço dado, classe segura essa que volta e meia é **preterida** em virtude de se afigurar demasiado lata para os objectivos em vista.

Claro que, escolhida ao acaso uma classe, ela só raras vezes será uma classe segura. Embora uma situação facultadora de descarrilamento possa por vezes ser tratada directamente, é preferível **procurar ultrapassá-la recorrendo a um teorema** (aqui não demonstrado) que declara a **existência de uma e de uma só classe mínima que seja segura e que contenha a classe proposta**, classe mínima essa a que vamos chamar **classe segura gerada pela classe proposta**. Garantida assim a existência da classe segura gerada, **não é utópico** procurar determiná-la, tarefa essa que todavia pode deparar com dificuldades de execução.

Face ao exposto, afigura-se que o medidista encontrará um ambiente propício sempre que formar um **par ordenado** constituído por um **espaço dado** e por **uma classe segura nele instituída**. A um tal par é uso chamar **espaço mensurável**, com o adjectivo mensurável a aludir ao conceito de medida (*mensura* em latim), este ainda por definir. Neste contexto é uso chamar mensurável a todo o conjunto da classe segura.

Obviamente no meio de tantos ambientes matemáticos imagináveis não foi fácil reconhecer os espaços mensuráveis como entes (abstractos) **extremamente propícios ao objectivo primordial dos medidistas**, o de construir um conceito de medida **que tenha eficiência matemática, que seja satisfatoriamente lato, que se preste à aplicações e que se cubra a medida popular como caso particular**.<sup>2</sup>

Mas, feita a obra, nada nos impede de nos servirmos dela. Além disso fica explicado porque a chamada **Álgebra de Conjuntos**<sup>3</sup> desempenha um papel fundamental na Teoria da Medida.

<sup>2</sup> Para chegar lá foram necessários um trabalho afinado e uma visão admirável.

<sup>3</sup> Chama-se “álgebra de conjuntos” à parte do estudo dos conjuntos que trata das operações a efectuar sobre eles.

5. Tendo em mente o exposto nos números 3 e 4, afigura-se conveniente indagar o que **há de comum** aos exemplos de medida popular do nº 2 e a outros exemplos congéneres.

Vejam os.

- a) **Atribui-se um valor numérico real a um conjunto contido num espaço:** no 1º exemplo a um conjunto de elementos contido no espaço formado por todos os boletins de voto depositados na urna; no 2º exemplo a um conjunto de elementos contido no espaço formado por todos os pontos geométricos duma recta; no 3º exemplo a um conjunto de elementos contido no espaço formado por todos os pontos materiais do espaço físico, etc.

**O conjunto incidente pertence a uma classe não vazia:** no 1º exemplo a classe (máxima e) segura formada por todos os conjuntos de boletins de voto; no 2º exemplo à classe não-segura formada por todos os segmentos de uma recta; no 3º exemplo à classe não-segura formada por todos os sólidos, etc. **Caso a classe tomada não seja segura** podemos considerá-la imersa na classe segura por ela gerada.

- b) O valor atribuído a um conjunto da classe em causa é sempre um **número não-negativo, não ficando excluídos os valores extremos 0 e  $+\infty$** . Assim deve atribuir-se o valor 0 sempre que for vazio o respectivo conjunto (o dos boletins nulos no 1º exemplo e o dos pontos do segmento no 2º exemplo); no 2º exemplo deve atribuir-se o valor  $+\infty$  se for ilimitado o respectivo “segmento”.
- c) Caso se divida ou **fragmente** o conjunto proposto em partes disjuntas duas a duas, pertencentes à classe respectiva e formando uma colecção intransnumerável, então o **valor** do conjunto proposto resulta igual à soma dos valores das suas partes disjuntas. Esta asserção pode considerar-se óbvia no caso de uma colecção finita; no caso de uma colecção infinita ela impõe-se após uma pequena discussão de índole matemática.

O único óbice que pode surgir no meio disso tudo está relacionado com a parte final da alínea a), mais concretamente com a possibilidade da **existência de conjuntos que pertençam à classe segura sem pertencerem à classe inicial**. Tais conjuntos passam a ser apanhados pelo esquema e, por isso, devem ficar com valores, digamos, **extrapolados**. Cabe ao matemático providenciar por forma que as alíneas a), b) e c) não deixem de ser respeitadas pelos valores atribuídos, originais ou extrapolados, quando os conjuntos e suas partes pertencerem não apenas à classe inicial, mas mais geralmente à classe segura gerada por ela.

6. Acabamos de delinear um processo pelo qual o medidista, partindo de uma medida popular e apoiando-se nos préstimos da Matemática se tal for necessário, fica eventualmente capacitado a instituir um espaço mensurável (veja-se o nº 4) e, feito isso, a

organizar uma função matemática que atribua um valor numérico real a cada conjunto mensurável, isso de acordo com as 3 propriedades seguintes:

- 1ª é **não-negativo** (eventualmente igual a  $+\infty$ ) o valor atribuído a qualquer conjunto mensurável;
- 2ª é **nulo** o valor atribuído ao conjunto vazio;
- 3ª quando se fragmenta um conjunto mensurável em partes mensuráveis disjuntas duas a duas e formando uma coleção intransnumerável, então **o valor do conjunto proposto resulta igual à soma dos valores das suas partes.**

Por uma questão de simplificação da linguagem a empregar, vamos chamar **medida popular extrapolada** a uma função do tipo referido, quer ela seja uma medida popular quer ela resulte duma tal medida pelo processo de alargamento acima descrito. O matemático não pode deixar de pensar na vantagem que lhe advirá se ele suprimir a condição restritiva de o ponto de partida se situar numa medida popular.

Porque não se há-de considerar um **espaço mensurável absolutamente arbitrário** e de juntar a ele **uma função matemática qualquer contanto que satisfaça simultaneamente às 3 propriedades supracitadas?**

Ora bem, vamos chamar **função-medida**, abreviadamente **medida**, a toda a função do tipo acabado de referir, quer ela seja medida popular extrapolada quer não o seja. Dada uma tal função e escolhido arbitrariamente um conjunto mensurável, vamos chamar **medida do conjunto escolhido** ao valor numérico que a função lhe atribui.

Acabámos de chegar à noção mais geral de (função-) medida por via de **teorização ascendente**. Alcançada esta posição, podemos recorrer à via da **teorização descendente** para chegar a situações especiais variadíssimas entre as quais não podem deixar de figurar as medidas populares.

Talvez valha a pena acrescentar que até agora não se consideraram **conjuntos mensuráveis com valores negativos**, porventura uma limitação severa.

Sem entrarmos em pormenores do *modus faciendi*, assinalemos que a via a seguir é substituir todo o valor negativo pelo seu simétrico em moldes a esclarecer, isso sob a reserva expressa de não aparecer nenhuma “indeterminação” correspondente à subtracção de dois números ambos iguais a  $+\infty$ <sup>4</sup>.

Enfim, entrave comparativamente ligeiro. Aliás é pouco credível a existência de alguma teoria livre de entaves.

7. Em seguida vamos apresentar alguns **exemplos novos** de medidas definidas em conformidade com o nº 6. Em cada exemplo terão de ser verificadas as propriedades típicas 1ª, 2ª e 3ª, uma tarefa que é fácil de executar nos casos mais acessíveis e que deve

<sup>4</sup> Uma tal subtracção é em princípio indeterminada porque, salvo eventuais esclarecimentos ulteriores, não se pode discriminar se prevalece o  $+\infty$  do diminuendo ou se prevalece o  $+\infty$  do diminuidor ou se ocorre um “empate” nalgum número finito.

ser confiada a um perito nos demais casos. O ouvinte (respectivamente leitor) poderá ficar agradavelmente surpreendido ao constatar que ele já conhecia, pelo menos aproximadamente, a maioria se não a totalidade das medidas apresentadas, inseridas em sectores *a priori* desligados uns dos outros.

- 1° *exemplo*. Escolhido arbitrariamente um espaço mensurável, a “medida minimalista”, ou seja, a medida que atribui o valor mais baixo 0 a qualquer conjunto mensurável.
- 2° *exemplo*. Escolhido arbitrariamente um espaço mensurável, a “medida maximalista” ou seja a medida que atribui o valor mais alto  $+\infty$  a qualquer conjunto mensurável e não-vazio.
- 3° *exemplo*. Escolhido arbitrariamente um espaço mensurável, a “medida de contagem” ou seja a medida que atribui a qualquer conjunto mensurável um valor igual ao número dos seus elementos.
- 4° *exemplo*. Dado o espaço formado pelos alunos que se sujeitaram a um determinado exame, a “medida classificativa” ou seja a medida que atribui a qualquer conjunto formado por examinandos um valor igual à soma dos valores atribuídos aos seus elementos.
- 5° *exemplo*. Dado o espaço formado pelas equipas que participam num campeonato desportivo, digamos de xadrez ou de futebol, etc., a “medida pontualista” ou seja a medida que atribui a cada conjunto de equipas um valor igual à soma das pontuações alcançadas pelos seus elementos.
- 6° *exemplo*. Dão o espaço formado pelas equipas que participam no campeonato nacional de futebol, a “medida idolátrica” ou seja a medida – medida subjectiva – que atribui a qualquer conjunto de equipas ou o valor 0 ou o valor  $+\infty$ , 0 se o conjunto excluir uma determinada equipa e  $+\infty$  se o conjunto incluir essa equipa.
- 7° *exemplo*. Uma **distribuição de massas** tomada na classe segura gerada pela classe dos intervalos ou paralelepípedos do espaço real a 3 dimensões
- 8° *exemplo*. Uma **distribuição de cargas eléctricas não-negativas** tomada na classe segura referida no 7° exemplo.
- 9° *exemplo*. A **função volume**, abreviadamente o **volume**, com a mesma classe segura do 7° exemplo.<sup>5</sup>
- 10° *exemplo*. O **integral de Riemann indefinido** dada uma função contínua e não-negativa, integral esse tomado na classe segura gerada pela classe dos intervalos ou paralelepípedos do espaço real a N dimensões, com N finito.<sup>6</sup>

<sup>5</sup> A classe segura do texto é a álgebra de Borel tridimensional na linguagem usada pelos medidistas. Tem a propriedade de não excluir nenhuma região tal que a análise clássica ensine como determinar o respectivo volume.

<sup>6</sup> A classe segura do texto é a *Álgebra de Borel a N dimensões* na linguagem usada pelos medidistas. Tem a propriedade de não excluir nenhuma região tal que a Análise Clássica ensine como determinar o respectivo volume a N dimensões (comprimento se  $N=1$ , área se  $N=2$ , volume tradicional se  $N=3$  e hipervolume se  $N>3$ ).

Talvez valha a pena acrescentar que o nosso conceito de medida permite variantes de saída desde as mais triviais, relacionadas com a vida quotidiana, até às mais avançadas, relacionadas com assuntos matemáticos difíceis. Entre as últimas, citemos, a título informativo, uma **(teoria da) integração que engloba e ultrapassa** todos os tipos de integração anteriormente estudados. A este propósito, torna-se interessante comparar os avanços conseguidos através da medida, com os conseguidos através da topologia. Esta comparação foi efectuada por Zaanen no seu livro *Integration*.

8. No nº 2 abreviámos “Teoria da Medida” para “Medida”. Mas há quem diga “Medida e Probabilidade” em lugar de “Medida”. Pergunta-se: com que fundamento?

Todos nós já ouvimos falar em probabilidades: no mínimo em probabilidades mais ou menos populares; talvez em probabilidades mais organizadas, como as discretas e as contínuas; excepcionalmente em probabilidades mais gerais.

Do ponto de vista *histórico*, a noção de probabilidade começou a introduzir-se através de jogos de azar (com moedas, dados, apostas, etc.) e depois evoluiu estendendo-se a outros aspectos da actividade humana e mais geralmente da própria natureza. Curiosamente, até princípios deste século era mais ou menos “crença” que um ou outro “sector” da natureza albergava um ente algo misterioso, denominado probabilidade (sectorial ou privativa), que o matemático devia espiar e eventualmente filiar no seu mundo a fim de sacar resultados úteis por meio de processamentos devidamente ajustados. Era um posicionamento de índole **empírica** que levava à resolução satisfatória de muitos problemas sugeridos pela prática, que se revelava incerto ou mesmo incoerente perante outros problemas delicados (por vezes decorrentes dos anteriores) e que falhava sem apelo quando a conjuntura impunha ampliações significativas do quadro primitivamente traçado. Em resumo, **havia necessidade premente duma doutrina rigorosa**.

Coube a Kolmogorov e a outros o mérito de reconhecer que todos os tipos de probabilidade até então considerados, **a priori separados** uns dos outros, se sujeitavam, ou directamente ou através da sua extrapolação, à classe segura gerada pela classe inicialmente tomada, diríamos se sujeitavam às propriedades  $1^a$ ,  $2^a$  e  $3^a$ , referidas no nº 6, com a particularidade de o respectivo espaço ficar com o valor 1. Assim retirada a restrição de se partir alguma probabilidade de tipo já conhecido, podemos calcar o caminho seguido na passagem das medidas populares extrapoladas para as medidas arbitrárias.

Deste modo atingimos um conceito genérico de **(função - ) probabilidade** que se identifica com o conceito de (função -) medida dotada da particularidade de atribuir o valor 1 ao respectivo espaço. Neste contexto, fixados uma (função -) probabilidade e um conjunto mensurável, chamaremos **probabilidade** do conjunto fixado ao valor numérico que a função lhe atribui.

9. Posto isto, vamos chamar **medida finitizável** – medida finita- $\sigma$  (a letra grega sigma) na linguagem típica dos medidistas – a toda a medida que permita fragmentar qualquer conjunto mensurável em partes mensuráveis disjuntas duas a duas e formando

uma colecção intransnumerável de modo que cada uma das partes fique com uma medida finita.

Conhecem-se exemplos de medidas que não são finitizáveis; está neste caso a medida do 2º exemplo do nº 7. Todavia as **medidas finitizáveis são**, tirando uma ou outra excepção, as **medidas mais importantes** do ponto de vista quer teórico quer prático. Para o conhecedor podemos mencionar, a propósito, que as **medidas de Lebesgue-Stieltjes** são medidas finitizáveis especiais.

Por outro lado, vamos chamar **combinação linear e não-negativa formada por probabilidades** a qualquer soma cujas parcelas perfaçam uma colecção intransnumerável e sejam produtos de constantes finitas e não-negativas por funções-probabilidade.

Face ao exposto no nº 8, **os interessados primordialmente na noção geral de medida** interpretarão a noção de probabilidade como trivialmente incluída e os **interessados primordialmente na noção específica de probabilidade** limitar-se-ão a considerar a noção de medida apenas nos casos específicos em que ela possa ser atingida por um processo de generalização de tratabilidade mais ou menos imediata. **Os dois pontos de vista acabados de referir podem conciliar-se notoriamente se limitarmos o nosso estudo às medidas finitizáveis.**

Com efeito, escolhida arbitrariamente uma medida finitizável, existe um processo de construção, não pormenorizado aqui e não necessariamente único, que permite igualar a medida escolhida a uma combinação linear e não-negativa formada por probabilidades.

Não é aqui o lugar para referir o uso da probabilidade agora citada em questões de integração e noutras questões medidistas. Apenas uma das consequências possíveis é a seguinte: seja qual for o número natural  $N$ , a **função volume a  $N$  dimensões**<sup>7</sup>, definida na classe segura do 10º exemplo, pode igualar-se a uma soma de probabilidades. Eis um exemplo de unificação de estudos, talvez um pouco inesperado para o principiante.

---

<sup>7</sup> Esta função ou este volume não é senão o extrapolado do volume a  $N$  dimensões clássico.

**Pedro Bruno Teodoro Braumann (1919.01.31-2003.04.30)**

Doutorou-se em Ciências Matemáticas em 1951 na Faculdade de Ciências da Universidade de Lisboa, onde foi Professor grande parte da sua vida, participando na criação do Departamento de Estatística, Investigação Operacional e Computação. Foi ainda Professor da Universidade de Luanda (onde inaugurou cursos de Matemática e chefiou as Secções de Matemática de Luanda e de Sá da Bandeira, tendo também leccionado em Nova Lisboa), da Universidade de Aveiro (onde dirigiu o Departamento de Matemática e presidiu à organização das V Jornadas Luso-Espanholas de Matemática), da Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa (onde se jubilou em 1989) e, após a jubilação, da Universidade Lusíada (onde proferiu a oração de sapiência do ano lectivo de 1992/93).

A sua obra científica e didáctica inclui vários livros e dezenas de artigos (além de conferências e comunicações nacionais e internacionais) em áreas variadas com predomínio para a Teoria da Medida e Probabilidade, de que foi pioneiro e divulgador em Portugal. O seu trabalho de investigação mais importante (um dos seus resultados ficou conhecido por “Braumann’s theorem”) refere-se ao estudo das distribuições infinitamente divisíveis e limites de somas de variáveis aleatórias independentes.

Foi um dos fundadores da Sociedade Portuguesa de Estatística.



# A Estatística em Portugal nos últimos 25 anos

*Bento José Ferreira Murteira*

Numa óptica pessoal, portanto em termos subjectivos, versando naturalmente os temas preferidos do Autor<sup>1</sup>, há muitos anos jubilado, a presente nota pretende esboçar alguns dos aspectos que caracterizam a evolução da estatística em Portugal nos últimos 25 anos, isto é, desde os fins da década de 70 até hoje. As referências dirigem-se, sobretudo, a métodos estatísticos. Não são abordados, apesar de muito importantes, quaisquer aspectos associados com as estatísticas oficiais e o INE<sup>2</sup>.

Os fins da década de 70 são particularmente gratos ao Autor pois foi nessa data que publicou a 1ª edição dos dois volumes de «Probabilidades e Estatística», texto revisto e ampliado no princípio dos anos 80 e 90. Tratava-se de uma época em que não abundavam os livros de estatística em língua portuguesa e por isso, com um pouco de imodéstia, julgava-se que aqueles textos se tornaram de facto úteis. Felizmente, na década de 90 apareceram os livros de J. Tiago de Oliveira<sup>3</sup> e de F. Galvão de Mello<sup>4</sup> e mais recentemente a *Introdução à Probabilidade e à Estatística* de Dinis Duarte Pestana e Sílvio Filipe Velosa<sup>5</sup> e os dois volumes *Probabilidades e Estatística* de Esmeralda Gonçalves e Nazaré Mendes Lopes<sup>6</sup>.

Sem entrar em pormenores sobre o ensino da Estatística, que deixo a colegas menos afastados das salas de aula, convém dizer que, por exemplo, numa grande Escola de Economia a necessidade de textos bem mais elementares<sup>7</sup> que os do Autor acima referidos

---

<sup>1</sup> A presente nota não seria possível sem a preciosa e enorme ajuda dos colegas Maria Antónia Amaral Turkman, Maria Ivette Gomes, Kamil Feridun Turkman, Dinis Duarte Pestana e Carlos Alberto da Silva Ribeiro, a quem se agradece calorosamente. Ao Autor não foi possível focar tantos pontos quanto gostaria. Espera-se, assim, que ninguém se sinta excluído ou tratado com menos justiça. Se houver lugar a “queixas”, sobretudo dos mais novos, a única responsabilidade é do Autor.

<sup>2</sup> Leia-se, por exemplo, de Carlos Corrêa Gago, «O Sistema Estatístico Português e o seu enquadramento europeu», *Actas do VI Congresso da SPE*, 1998. (Sobre este tema e outros, diversos capítulos deste Memorial são “Leituras Complementares” na sequência das sugestões propostas pelo autor em notas de rodapé - Nota do editor).

<sup>3</sup> *Probabilidades e Estatística, Conceitos, Métodos e Aplicações*, McGraw-Hill, Lisboa, Vol.I, 1990, Vol.II, 1991.

<sup>4</sup> *Probabilidades e Estatística, Conceitos e Métodos Fundamentais*, Escolar Editora, Lisboa, Vol.I, 1993, Vol.II, 1997.

<sup>5</sup> Vol.1, Fundação Calouste Gulbenkian, Lisboa, 2002.

<sup>6</sup> Escolar Editora, Lisboa, 2000, 2003.

mostra que ao nível da licenciatura a deficiente preparação matemática, aliás verificada em muitas outras escolas, não tem permitido grandes voos. Por outro lado, no aspecto positivo, tem-se a introdução, nos anos 80, do ensino da estatística no secundário, facto a que J. Tiago de Oliveira deu uma ajuda mas que se deve em grande parte a António St. Aubyn que foi quem mais colaborou com o grupo que propôs os programas. De notar também o envolvimento nesse campo de Eugénia Graça Martins, Fernando Rosado e João Branco e de destacar ainda as acções, ao abrigo do Programa FOCO, em que participaram Fernanda Oliveira, Margarida Mendes Leal, Helena Iglésias Pereira, Salomé Cabral, Maria Antónia Amaral Turkman e Dinis Pestana.

Aspectos positivos a realçar noutros campos são, essencialmente:

- A fundação em 1975, por J. Tiago de Oliveira, do Centro de Estatística e Aplicações da Universidade de Lisboa (CEAUL) – berço da investigação em Probabilidades e Estatística em Portugal – e também a criação em 1981, na Faculdade de Ciências de Lisboa, do Departamento de Estatística e Investigação Operacional com duas licenciaturas: “Probabilidades e Estatística” e “Estatística e Investigação Operacional”.
- A fundação em 1980 da Sociedade Portuguesa de Estatística e Investigação Operacional, que em 1991 foi profundamente reestruturada, passando a ter a designação actual, **Sociedade Portuguesa de Estatística**. A SPE tem por objectivo promover, cultivar e desenvolver, em Portugal, o estudo da Estatística, suas aplicações e ciências afins e a sua acção tem sido decisiva para o progresso verificado no País. Além da realização regular de congressos e publicação das respectivas actas, depois de um trabalho de arbitragem que tem tido bastante mérito no desenvolvimento da investigação, a SPE publica também um «Boletim Informativo», promove anualmente um curso breve em áreas de interesse actual e concede prémios que têm incentivado jovens estatísticos na fase inicial da carreira.
- A instalação em 1989, na Universidade Nova de Lisboa, do Instituto Superior de Estatística e Gestão da Informação onde, no quadro de licenciaturas e mestrados, são ministrados cursos nas várias áreas da Estatística. A instalação do ISEGI deve muito ao esforço e à iniciativa de Manuel José Vilares que à data presidia ao INE.
- A instituição em 1990/91, no Instituto Superior de Economia e Gestão, de uma licenciatura em matemática aplicada à economia e à gestão com uma forte componente estatística<sup>8</sup>.

<sup>7</sup> Bento Murteira, Carlos da Silva Ribeiro, João Andrade e Silva e Carlos Pimenta, *Introdução à Estatística*, McGraw-Hill, Lisboa, 2002.

<sup>8</sup> As disciplinas obrigatórias são: Estatística Descritiva, Probabilidades, Estatística I, Econometria I e Econometria II. As opcionais são: Estatística II, Processos Estocásticos e Aplicações, Sucessões Cronológicas e Previsão, Tópicos de Econometria, Teoria do Risco, Estatística Multivariada e Teoria e Técnicas de Amostragem

- O início da actividade em 1991, na Universidade Católica, do Centro de Estudos e Sondagem de Opinião que muito tem contribuído para a visibilidade da Estatística junto do público em geral.
- A criação em 1994, na Universidade de Évora, do Centro de Investigação em Matemática e Aplicações em que a segunda linha de investigação se ocupa, actualmente, de Processos Estocásticos, Estatística e Investigação Operacional.
- A publicação a partir de 1996 da «Revista de Estatística» editada pelo INE.

Abordar alguns aspectos relevantes da evolução da Estatística em Portugal nos últimos 25 anos é tarefa difícil. No entanto, como já foi dito, prevalecem as preferências do Autor que muito fica a dever, também, à informação recolhida nas Actas dos Congressos da SPE dada a dificuldade que tem, em muitos casos, de acesso a outras fontes relevantes.

A Teoria da Probabilidade é a base da Inferência Estatística. Nos seus aspectos puros, além do tratamento elementar encontrado nos livros de texto acima indicados, tem sido ultimamente objecto de alguma atenção por parte dos investigadores portugueses. Algumas referências devem fazer-se. Começa-se pela obra de Pedro Bruno Teodoro Braumann<sup>9</sup>, infelizmente incompleta devido ao seu falecimento em 2003, e os seus estudos anteriores sobre limites de somas de variáveis aleatórias independentes. Pelo seu lado, João Pedro Faria sentiu necessidade de se afastar da “popular” axiomática de Kolmogorov<sup>10</sup> conforme acentua na sua tese sobre probabilidade comparativa<sup>11</sup>. Mais clássicos são, a título de exemplo, os estudos<sup>12</sup> de Eugénia Graça Martins e Dinis Pestana, sobre correlação e independência e sobre leis limites em contextos não clássicos, de J. Tiago Mexia e Sandra Cristina Nunes, sobre densidades simétricas, de C. M. Martins, sobre *moment inequalities*, de Paulo Eduardo Oliveira, sobre um teorema limite central para variáveis associadas, de Dinis Pestana e Sílvio Filipe Velosa, sobre classes de leis N-finitamente divisíveis, de Sílvio Filipe Velosa, sobre novas classes de leis infinitamente divisíveis discretas, de Helena Iglésias Pereira, sobre domínios de atracção e velocidades de convergência para estáveis multivariadas, de Andreia Hall, sobre processos extremos com falhas e sobre filtragem e extremos de variáveis discretas, de Maria Ivette Gomes e Dinis Pestana, sobre convergência e *penultimate approximations*, e de Agra Coelho, sobre distribuições e aproximações em estatística multivariada. Para terminar, apesar dos muitos nomes que ficam por citar, é justo referir o trabalho de Fátima Brilhante, Dinis Pestana, José Rocha e Sílvio Velosa<sup>13</sup>, sobre populações gaussianas, homocedasticidade, populações gaussianas, heterocedasticidade e studentização em populações não-gaussianas.

<sup>9</sup> *Teoria da Medida e da Probabilidade*, Parte I, Fundação Calouste Gulbenkian, Lisboa, 1987

<sup>10</sup> «A discussion with Kolmogorov», *Notas e Comunicações*, INIC, N.º 1, 1989.

<sup>11</sup> *Compatibilidade Crisscross em Estruturas de Probabilidade Comparativa*, Lisboa, 1991.

<sup>12</sup> Que se mencionam sem notas bibliográficas pela dificuldade em identificar o local de publicação de alguns estudos, muitos deles em revistas estrangeiras.

<sup>13</sup> *Inferência sobre Localização e Escala*, Mini-Curso SPE, Ponta Delgada, 2001.

Acompanhando o desenvolvimento dos métodos estatísticos, um dos factos mais marcantes das últimas décadas foi o extraordinário impulso dado pela informática. Sabe-se que em algumas faculdades e centros de investigação havia *main-frames* [por exemplo, no ISEG, o primeiro grande computador, um VAX da Digital, foi instalado em 1985] e que há naturalmente, ainda hoje, problemas que exigem o recurso a computadores potentes e a programas específicos. Não resta dúvida, porém, que a maior acessibilidade a *packages* de estatística cada vez mais completos, como SAS, SPSS, S-plus, Statistica<sup>14</sup>, etc., tem permitido que, dentro de certos limites, o trabalho dos estatísticos, sobretudo daqueles que não têm tempo ou não conseguem escrever os seus próprios programas, seja feito comodamente no seu gabinete ou em casa. Lembro, a propósito, que nos anos 50 passei muitas semanas com uma máquina FACIT a calcular autocorrelações (10 *lags* no máximo) de sucessões cronológicas não muito longas, enquanto hoje em décimos de segundo qualquer desses *packages* faz autocorrelações para muitos *lags* de sucessões praticamente tão grandes quanto se quiser.

Mas o impacto dos crescentes meios de computação não se fez sentir apenas nas aplicações da estatística ao torná-las possíveis ou menos trabalhosas, mais rápidas e, certamente, mais precisas. O impacto é muito mais profundo já que os métodos de computação intensiva, como a reamostragem, revolucionaram a estatística e já foram considerados por alguns como o casamento entre a estatística e os computadores. Quando a análise matemática (ou a incapacidade do investigador) não permite obter a resposta desejada, a reamostragem baseada em técnicas de simulação, como *bootstrap*, permutação (ou casualização quando se quer ter consciência de que nem sempre é viável efectuar todas as permutações dos dados da amostra), etc., permite ladear a complexidade analítica e chegar a resultados muito satisfatórios em domínios tais como o ensaio de hipóteses e os limites de confiança. Aliás não é difícil perceber a razão pela qual os testes de casualização, ao contrário dos testes paramétricos, escapam aos impasses analíticos: não é necessário trabalhar com amostras casuais (que muitas vezes não existem), raramente se pensa em termos de populações ou na estimação dos seus parâmetros, conseqüentemente não se depende das hipóteses de normalidade ou de homocedasticidade, etc.. Em particular, no que diz respeito ao *bootstrap* é interessante recordar que foi praticamente há 25 anos que a ideia foi lançada por Bradley Efron<sup>15</sup> [«Bootstrap Methods: Another Look at the Jackknife», *Annals of Statistics*, 1979] que tivemos o privilégio de ter entre nós no XII Congresso da SPE.

O Autor realizou alguns estudos sobre o *bootstrap* e orientou o trabalho de vários alunos<sup>16</sup>.

---

<sup>14</sup> Para não falar em *packages* especializados como Autobox, TSP, etc.. O Autor teve recentemente conhecimento de uma obra de Joaquim P. Marques de Sá, da Universidade do Porto – *Applied Statistics using SPSS, STATISTICA and MATLAB*, Springer-Verlag, 2003 – de que pôde consultar apenas um resumo mas que parece ser recomendável.

<sup>15</sup> Que mais recentemente, em colaboração com R. J. Tibshirani, publicou *An introduction to bootstrap* (Chapman & Hall, Nova Iorque, 1993), obra fundamental para o estudo da matéria.

<sup>16</sup> Por exemplo, Isabel Maria Dias Proença, «Método Bootstrap» e José Manuel Zorro Mendes, «Bootstrap Iterativo: Aplicação ao Índice de Gini», ISEG, 1990

Mais recentemente o interesse pelo *bootstrap* tem-se acentuado no nosso País<sup>17</sup>.

Outro método de computação intensiva – o algoritmo de Expectação-Maximização ou algoritmo EM<sup>18</sup> – tem sido, até onde sei, pouco aplicado entre nós. A excepção parece ser os *papers* de Paulo Gomes e Adelaide Figueiredo<sup>19</sup>, de Carlos Sousa, Carlos Braumann e Margarida Castro<sup>20</sup> e de M. S. E. Cabral<sup>21</sup>.

Sob o impulso de John W. Tukey e da sua originalidade<sup>22</sup> outra autêntica revolução foi a introdução da análise exploratória de dados, análise que, para aquele estatístico, é trabalho de detective com números, com contagens e com gráficos. A estatística descritiva foi, por assim dizer, refrescada, tendo ganho novas perspectivas e novos termos. Hoje em dia já não se hesita em falar de caixas-de-bigodes e em gráficos de caule-e-folhas termos que já aparecem, por exemplo, nas lições da Universidade Aberta<sup>23</sup>. A difusão em Portugal da análise exploratória de dados deve muito à iniciativa de Dinis Pestana de fazer traduzir para português a excelente obra de David C. Hoaglin, Frederik Mosteller e John W. Tukey<sup>24</sup>. Nesta obra se inspirou o presente Autor no seu texto sobre a matéria<sup>25</sup>. Como o título da obra daqueles três autores indica, a análise exploratória de dados é efectuada com a preocupação de os métodos utilizados serem robustos – válidos para um largo leque de hipóteses ou modelos teóricos – e resistentes – insensíveis à presença de erros grosseiros (*outliers*) ou de um grande número de pequenos erros. Foi nessa linha de preocupação que em boa hora foi publicada em 1993, pela Editorial Salamandra, a tradução da obra de Frederik Mosteller e R. E. K. Rourke<sup>26</sup>, designada em português «Estatísticas Firmes»

<sup>17</sup> Maria Ivette Gomes, «Metodologias Jackknife e Bootstrap em Estatísticas de Extremos», *Actas do II Congresso da SPE*, 1994. Teresa Themido Pereira, «Seleção Ótima da Fracção da Amostra na Estimação do Índice de Valores Extremos: Um Método Baseado no Bootstrap», *Actas do V Congresso da SPE*, 1997. Conceição Amado e Ana M. Pires, «A utilização da função de influência para melhorar o desempenho de reamostragem bootstrap», *Actas do VI Congresso da SPE*, 1998. Margarida Brito e Alexandre Ramos, «Método de Bootstrap Generalizado: caudas e centro de processos empíricos», *Actas do VI Congresso da SPE*, 1998. M. Manuela Neves, M. João Martins e M. Ivette Gomes, «Comparação de abordagens da metodologia bootstrap na estimação semi-paramétrica do índice de cauda», *Actas do VII Congresso da SPE*, 1999. Margarida Brito, «A Metodologia Bootstrap. Validade Assintótica e alguns casos de êxito», *Actas do VII Congresso da SPE*, 2000. Dora Prata Gomes e Manuela Neves, «A metodologia bootstrap: relação com o Jackknife», *Actas do VII Congresso da SPE*, 2000. M. Ivette Gomes e Cristina Miranda, «A metodologia Jackknife na estimação do índice extremal», *Actas do X Congresso da SPE*, 2002. Javier Roca Pardiñas, «Intervalos de confiança Bootstrap vs. assintóticos em modelos aditivos generalizados (GAM) para resposta binária», *Actas do X Congresso da SPE*, 2002.

<sup>18</sup> A matéria pode estudar-se na excelente monografia de G. J. McLachlan e T. Krishnan, «The EM Algorithm and Extensions», Wiley, Nova Iorque, 1997.

<sup>19</sup> «Identificação de uma Mistura de Leis de Bingham Através de um Algoritmo do tipo EM», *Actas do III Congresso da SPE*, 1995.

<sup>20</sup> «Uma aplicação do algoritmo EM a dados de captura de peixe», *Actas do IX Congresso da SPE*, 2001.

<sup>21</sup> «Misturas Finitas de Distribuições», Nota n.º 7 do CEAUL, 1989; «Teste de razão de verosimilhanças para uma mistura de normais. Um estudo de simulação», *Actas do II Congresso Anual da SPE*, 1994; «O modelo de misturas de normais na análise e interpretação de ANOVA's Significativas», *Actas do IV Congresso Anual da SPE*, 1996.

<sup>22</sup> *Exploratory Data Analysis*, Limited Preliminary Edition, Addison-Wesley, Reading, 1970.

<sup>23</sup> Maria Eugénia Graça Martins e Alexandre Gomes Cerveira, *Introdução às probabilidades e à estatística*, Universidade Aberta, 1999.

<sup>24</sup> *Análise Exploratória de Dados, Técnicas Robustas*, Edições Salamandra, 1983, Lisboa.

<sup>25</sup> *Análise Exploratória de Dados, Estatística Descritiva*, McGraw-Hill, 1993, Lisboa.

<sup>26</sup> «Sturdy Statistics. Nonparametrics e Order Statistics», Addison-Wesley, 1973, Nova Iorque.

sendo o termo “firmes” escolhido pelo Editor para qualificar os métodos apresentados no livro de forma a abarcar o sentido dos termos técnicos “robustos” e “resistentes”<sup>27</sup>.

A Análise de Dados, na óptica de Tukey e de Benzécri, e o estudo dos *outliers* são capítulos da estatística que não têm sido negligenciados pelos investigadores portugueses. No primeiro caso é bom ler João Branco<sup>28</sup> e consultar os trabalhos de Helena Bacelar-Nicolau e Fernando Nicolau<sup>29</sup> e de Paulo Gomes. No caso dos *outliers* convém não esquecer que a caixa-de-bigodes é o primeiro e, muitas vezes, o último instrumento de selecção de *outliers* num conjunto de dados. Quem o diz é Fernando Rosado que entre nós tem sido pioneiro no estudo das várias formas de “acolhimento” a dados extravagantes. Da sua vasta obra refere-se apenas a tese de doutoramento<sup>30</sup> e o *survey* de 1999<sup>31</sup> a que se acrescentam dois importantes *papers* de M. M. Brauman<sup>32</sup>.

Não é possível abordar a estatística portuguesa nos últimos anos sem uma forte referência a J. Tiago de Oliveira em relação ao qual se tem um pesado débito. O seu desaparecimento precoce em 1992 foi uma perda dramática para a comunidade estatística portuguesa. Uma lista dos seus numerosos trabalhos e de outras actividades de natureza científica pode ver-se em *Tiago de Oliveira, o Homem e a Obra*<sup>33</sup>. Aqui refere-se apenas o seu último livro – *Statistical Analysis of Extremes*, Pendor, Évora, 1997 – que J. Tiago de Oliveira não pôde, infelizmente, ver publicado e que é fundamental para quem deseja estudar a matéria. Uma boa descrição das contribuições científicas de J. Tiago de Oliveira encontra-se no artigo de Maria Ivette Gomes<sup>34</sup> em que esta presta justiça ao papel notabilíssimo que teve no desenvolvimento da estatística em Portugal, quer como investigador internacionalmente reconhecido e que aqui fez escola, quer na imposição da estatística como ramo autónomo do conhecimento. Embora J. Tiago de Oliveira tenha tido no seu trabalho de investigador variados interesses (cartas de controle, *outliers*, estatística não paramétrica, teoria do risco, etc.) não parece errado identificar como suas áreas de eleição as seguintes:

- Extremos Multivariados em Contexto Paramétrico.
- Escolha Estatística de Modelos Extremais.
- Extremos Univariados em Sucessões não I.I.D.
- Inferência Estatística em Modelos Extremais.

<sup>27</sup> Sobre robustez veja-se também: Ana M. Pires e João Branco, «Estatísticas Robustas: Passado, Presente e Futuro», *Actas do I Congresso da SPE*, 1993.

<sup>28</sup> «O Fascínio de Coleccionar e Analisar Dados Estatísticos», *Actas do V Congresso da SPE*, 1997.

<sup>29</sup> Por exemplo, «Estatísticas e Análise de Dados Multivariados: Passado e Futuro», *Actas do I Congresso da SPE*, 1993.

<sup>30</sup> *Existência e Detecção de Outliers. Uma abordagem Metodológica*, Faculdade de Ciências, Lisboa, 1984.

<sup>31</sup> «*Outliers* em dados estatísticos – O Passado e o Presente. E o Futuro?», *Actas do VII Congresso da SPE*, 1999.

<sup>32</sup> «Medidas de Performance para Testes de Detecção de *Outliers* em Populações Exponenciais com Parâmetros Conhecidos», *Actas do V Congresso da SPE*, 1997, e «Efeito do uso de hipóteses gerais sobre as medidas de performance para testes de detecção de *outliers* em populações com parâmetros conhecidos», *Actas do VI Congresso da SPE*, 1998.

<sup>33</sup> Edições Colibri, Lisboa, 1993.

<sup>34</sup> «A Obra Científica de J. Tiago de Oliveira» em *Estatísticas Robustas, Extremos e Mais Alguns Temas*, Edições Salamandra, Lisboa, 1993.

Em todas estas áreas J. Tiago de Oliveira tem tido seguidores, com suficiente qualidade para enfrentar a permanente evolução que se verifica na Teoria dos Extremos como, aliás, em todos os ramos do saber. No entanto, se J. Tiago de Oliveira iniciou a Escola de Extremos existente em Portugal (para isso foi fundamental o Congresso da NATO do Vimeiro, em 1983) o desenvolvimento e reconhecimento internacional da mesma Escola muito deve a Maria Ivette Gomes de que se referem alguns trabalhos mais recentes<sup>35</sup>. Esse reconhecimento contribuiu, naturalmente, para que se realizasse na Universidade de Aveiro, em Julho de 2004, o «*Third International Symposium on Extreme Value Analysis: Theory and Practice*»<sup>36</sup>.

Voltando aos outros seguidores de J. Tiago de Oliveira além de Maria Ivette Gomes, correndo o risco de pecar por omissão, podem indicar-se, nomeadamente, Alexandra Dias, Alexandra Ramos, Ana Ferreira, Andreia Hall, Cláudia Neves, Emília Athayde, Helena Ferreira, Isabel Barão, Isabel Fraga Alves, Kamil Feridun Turkman, Luísa Canto e Castro Duarte, Margarido Brito, Maria João Martins, Manuel Scotto, Manuela Neves, Teresa Alpuim e Teresa Themido Pereira. Entre os seguidores, uns têm-se mantido em contexto paramétrico, outros têm procedido à análise em contextos semi-paramétricos e não paramétricos, outros, enfim, têm tratado o tema mais recente dos Extremos Espaciais A título indicativo, sem de modo nenhum dizer que são os mais importantes ou com mais valia científica, convém fazer referência a mais alguns *papers*<sup>36</sup>.

Antes de deixar a estatística clássica convém mencionar alguns trabalhos recentes de J. Tiago Mexia<sup>37</sup>.

<sup>35</sup> «A new class of semi-parametric estimators of the second order parameter», *Portugaliae Mathematica* 60:2, 2003. «A class of asymptotically unbiased semi-parametric estimators of the tail index» (com F. Caeiro), *Test* 11:2, 2002. «Semi-parametric estimation of the second order parameter - asymptotic and finite sample behaviour» (com L. de Haan and L. Peng), *Extremes* 5:4, 2002. «Asymptotically unbiased» estimators of the tail index based on external estimation of the second order parameter», *Extremes* 5:1, 2002. «Generalizations of the Hill estimator asymptotic versus finite sample behaviour» (com M. J. Martins), *J. Statist. Planning and Inference*, 93, 2001. «Alternatives to a semi-parametric estimator of parameters of rare events - the Jackknife methodology» (com M. J. Martins and M. Neves), *Extremes* 3:3, 2000. «Approximations by extreme value distributions» (com L. de Haan), *Extremes* 2:1, 1999. «Statistical choice of extreme value domains of attraction» (com M. I. Fraga Alves), *Communications in Statistics*, 25, 1996. «Penultimate limiting forms in extreme value theory», *Ann. Inst. Statist. Math.* 36A, 1984.

<sup>36</sup> Maria Isabel Barão e J. A. Tawn, «Extremal analysis of short series with outliers: sea-levels and athletics records», *Applied Statistics*, Vol.48, 1999. Jorge Oliveira e Kamil Feridun Turkman, «A Study of the Air Quality in Lisbon: A Spatio-temporal Model for Prediction of NO2 Levels», *Journal of Chemometrics*, 16, 2002. Ana Luísa Correia e Manuela Neves, «Escolha Estatística em Modelos Extremos - Testes de Ajustamento», *Actas do III Congresso da SPE*, 1995. Maria Isabel Fraga Alves e Maria Ivette Gomes, «Statistical choice of extreme value domains of attraction», *Communications in Statistics, Theory and Methods*, 1996. Cláudia Pedrosa Neves e Maria Isabel Fraga Alves «Escolha automática de Reiss-Thomas para topo da amostra - heurístico versus assintótico», *Actas do VIII Congresso da SPE*, 2000. Maria Teresa Alpuim, «An Extremal Markovian Sequence», *J. Appl. Probability*, 1989. Luísa Canto e Castro Duarte, «Distribuição generalizada de valores extremos, domínio de atracção e velocidade de convergência», *Actas do I Congresso da SPE*, 1993. Andreia Hall, «Comportamento Extremal em Séries de Sucessões Estacionárias Sujeitas a Falhas», *Actas do IV Congresso da SPE*, 1996. Andreia Hall e Manuel G. Scotto, «Comportamento Extremal em Agregações Associadas a Modelos de Médias Móveis de Variáveis Aleatórias Inteiras», *Actas do X Congresso da SPE*, 2002. Manuel G. Scotto e K. F. Turkman, «On the extremal behaviour of sub-sampled solutions of stochastic difference equations», *Portugaliae Mathematica*, Vol.50, 2002.

<sup>37</sup> «Selective F tests for sub-normal models» (com Célia Nunes), *Discussiones Mathematicae, Probability and Statistics*, 22, 2002. «Exact Distribution for the generalized F tests» (com M. Fonseca e R. Zmyslony), *Discussiones Mathematicae, Probability and Statistics*, 22, 2002.

A chamada estatística clássica repousa no princípio da “amostragem repetida” segundo o qual os procedimentos estatísticos devem apreciar-se através do respectivo comportamento num grande número de repetições nas mesmas condições. Uma das faces do princípio reside na interpretação frequencista de probabilidade, a outra face reside na avaliação dos procedimentos estatísticos em termos das frequências com que fornecem respostas correctas ou “bons” resultados.

Apesar de existirem várias escolas alternativas, a grande rival da estatística clássica é a estatística bayesiana a qual reconhece que anterior à informação dada pela amostra – dado fixo que não se repete – existe informação inicial que deve ser formalizada em termos probabilísticos, normalmente do tipo subjectivista, através da chamada distribuição *a priori*. Combinando os dois tipos de informação através do Teorema de Bayes chega-se à distribuição a posteriori que é, por excelência, a base instrumental da inferência bayesiana.

Se o Autor pode talvez ser responsável pelo lançamento da semente Bayesiana no nosso País não há dúvida que tal paradigma ganhou, entre nós, grande balanço com a plena adesão de Maria Antónia Amaral Turkman e a publicação dos seus trabalhos de 1980<sup>38</sup> sobre análise preditiva e do artigo de 1981<sup>39</sup>. Na impossibilidade de citar os numerosos trabalhos de Maria Antónia Amaral Turkman e dos seus seguidores<sup>40</sup>, citam-se alguns dos seus mais recentes estudos<sup>41</sup>.

Outro “praticante” bayesiano é Carlos Daniel Paulino com vasta obra da qual se destacam alguns escritos ultimamente publicados<sup>42</sup>.

É um facto que desde meados dos anos 80 se tem verificado no domínio da Estatística e suas aplicações um enorme desenvolvimento da metodologia bayesiana a que não é estranho, conforme já se assinalou a propósito da estatística clássica, o avanço em capacidade de computação e o aparecimento de *software* específico. No entanto, foi na década de 90 que se deu uma autêntica reviravolta<sup>43</sup> quando o impasse verificado no cálculo de alguns integrais exigidos no processo de inferência bayesiana foi ultrapassado através da adaptação e aperfeiçoamento de algoritmos aplicados em Física Estatística e

<sup>38</sup> *Some Theoretical and Practical Applications of Predictive Distributions*, Ph.D. Thesis, Universidade de Sheffield e, em colaboração com I. R. Dunsmore, «Optimal Estimates of Predictive Distributions», *Biometrika*, 67.

<sup>39</sup> «Inferência Bayesiana», *Actas do II Colóquio de Estatística e Investigação Operacional*, Fundação/Covilhã.

<sup>40</sup> Lisete Sousa, Luzia Gonçalves, Giovanni Loiola da Silva, Isabel Simões Pereira e Fernando Magalhães, entre outros.

<sup>41</sup> P. de Zea Bermudez, M. A. Amaral Turkman e K. F. Turkman, «A Predictive Approach to Tail Probability Estimation», *Extremes*, 4, 2001. P. de Zea Bermudez e M. A. Amaral Turkman, «Bayesian Approach to Parameter Estimation of the Generalized Pareto Distribution», *Test*, 12, 2003. M. Antunes, M. A. Amaral Turkman e K. F. Turkman, «A Bayesian Approach to Event Prediction», *Journal of Time Series Analysis*, 24, 2003. I. Pereira e M. A. Amaral Turkman, «Bayesian Prediction in Threshold Autoregressive Models with Exponential White Noise», *Test*, 12, 2003. G. L. Silva e M. A. Amaral Turkman, «Bayesian Analysis of an additive survival model with frailty», *Communications in Statistics – Theory and Methods*, 30, 2004.

<sup>42</sup> Carlos Daniel Paulino (com C. A. B. Pereira), «Bayesian Methods for Categorical Data under Informative General Censoring», *Biometrika*, 82, 1995. Carlos Daniel Paulino (com P. Soares), «Incomplete categorical data analysis. A Bayesian perspective», *Journal of Statistical Computation and Simulation*, 69, 2001. Carlos Daniel Paulino (com P. Soares), «Analysis of rates in incomplete Poisson data», *The Statistician*, 52, 2003. Carlos Daniel Paulino (com P. Soares e J. Neuhaus), «Binomial Regression with Misclassification», *Biometrics*, 59, 2003.

conhecidos pela designação de Métodos de Monte Carlo via Cadeias de Markov (MCMC). Pois bem, na recentemente aparecida *Estatística Bayesiana*<sup>44</sup>, em que o Autor teve a honra de colaborar com Maria Antónia Amaral Turkman e Carlos Daniel Paulino, além de um tratamento actual da metodologia bayesiana e das suas aplicações, faz-se justiça à importância das modernas técnicas de computação bayesiana.

A Decisão Estatística trabalha com a informação inicial, a informação dada pela amostra e, ainda – o que faz a diferença – com a informação sobre as consequências (em geral, económicas) dos procedimentos decisoriais (que são, em regra, procedimentos estatísticos – estimação, ensaio de hipóteses, etc.) formalizada através de uma função perda (custos/benefícios) ou, mais correctamente, de uma função utilidade. Com alguma surpresa verifica-se que nos artigos incluídos nos 10 Volumes das Actas dos Congressos da SPE não há, salvo erro, uma única referência explícita à decisão estatística, a funções perda ou a funções utilidade<sup>45</sup>. Pensa-se que o desinteresse tem, talvez, explicação no facto de enquanto a inferência bayesiana ultrapassou o impasse das distribuições *a priori* subjectivas, a decisão estatística não ladeou as dificuldades da construção de funções perda. Semelhante facto não é fácil de entender quando se pensa que no ensaio de hipóteses devem ser bem ponderadas as consequências de um erro de 1.<sup>a</sup> espécie (rejeitar um bom medicamento, por exemplo) e de um erro de 2.<sup>a</sup> espécie (aceitar um mau medicamento) e que nos problemas de estimação os custos de um pequeno erro são diferentes dos custos de um grande erro. Esta reflexão parece incontornável por mais que contrarie R. A. Fisher quando mesmo da prática de Neyman-Pearson – de aceitar ou rejeitar hipóteses – diz ser um procedimento próprio de comerciantes, i.e., pouco digno de um cientista. Em qualquer caso, o texto do Autor<sup>46</sup> tem tido pouco eco.

A Teoria dos Processos Estocásticos, fundamento indispensável para Análise das Sucessões Cronológicas, tem aplicações bastante mais vastas. Uma linha de investigação<sup>47</sup> que se considera interessante é a desenvolvida por Carlos A. Braumann no estudo do crescimento de populações em ambiente aleatório<sup>48</sup> com importantes aplicações no domínio das pescas. De referir também «O Acaso, a Bolsa ou a Vida»<sup>49</sup> em que Carlos A. Braumann oferece uma claríssima introdução ao estudo das equações diferenciais estocásticas. Há muitos outros trabalhos no quadro dos processos estocásticos e aplicações de que se cita apenas uma pequena amostra<sup>50</sup>.

<sup>44</sup> Fundação Calouste Gulbenkian, Lisboa, 2003.

<sup>45</sup> Se não errei na contagem, *papers* Bayesianos aparecem, em média, dois por Acta do Congresso.

<sup>46</sup> *Estatística: Inferência e Decisão*, IN-CM, Lisboa, 1988.

<sup>47</sup> Segundo me informam parte da investigação em Processos Estocásticos tem sido feita no âmbito da matemática pura, quer pelo grupo liderado por Ana Bela Cruzeiro, quer pelo grupo liderado por Streit.

<sup>48</sup> De que se cita apenas a sua Tese Doutoramento, *Population Growth in Random Environment*, State University of New York at Stony Brook, 1979 e um dos seus últimos artigos – «O uso de diferentes cálculos estocásticos na modelação do crescimento populacional em ambiente aleatório» – publicado nas *Actas do X Congresso da SPE*, 2002.

<sup>49</sup> *Actas do V Congresso da SPE*, 1997

<sup>50</sup> Daniel Assunção Muller, *Inference Statistique pour la Reproduction Moyenne des Processus de Galton-Watson*, Université Pierre et Marie Curie, Paris, 1982. José António Serôdio e Maria Lucília Carvalho, «Processos de nascimento e morte em ambiente oscilante: escalas de tempo e resposta de populações de microalgas em ambiente intermareal», *Actas do I Congresso da SPE*, 1993. Maria Lucília Carvalho, «Processos de ramificação e ensaios clínicos sequenciais», *Actas*

Tem sido no domínio da Análise das Sucessões Cronológicas<sup>51</sup> que se tem processado investigação mais intensa. Desde 1970, data em que foi publicado o estudo seminal de Box e Jenkins sobre os modelos ARMA e ARIMA, o *stock* de modelos de sucessões cronológicas, lineares e não lineares, não tem parado de aumentar não esquecendo as sucessões de valores inteiros<sup>52</sup>.

Para ilustrar a actividade dos investigadores portugueses no domínio das sucessões cronológicas citam-se alguns trabalhos de Kamil Feridun Turkman<sup>53</sup>, de Esmeralda Gonçalves e Nazaré Mendes Lopes<sup>54</sup>, de Nuno Crato<sup>55</sup>, de Daniel Assunção Muller e outros autores<sup>56</sup>.

---

*Congresso da SPE*, 1996. Cláudia Nunes e António Pacheco, «Martingalas. Uma ferramenta em inferência estatística», *Actas do VI Congresso da SPE*, 1997. João Tiago Mexia, «Differential Equations for Moment Generating Functions of Diffusions», *Actas do VII Congresso da SPE*, 1999. João Figueira e Manuel Alberto Ferreira, «Aplicação de uma equação diferencial estocástica linear ao estudo dum fundo de pensões», *Actas do VII Congresso da SPE*, 1999. Helena Guerra, Manuela Oliveira e Sónia Inácio e João Mexia, «Inferência para o processo de Ornstein-Uhlenbeck», *Actas do VII Congresso da SPE*, 1999. Rui M. Almeida, «Modelos Financeiros com Saltos», *Actas do VII Congresso da SPE*, 1999. Rita Sousa e Sónia Torres, «Uma aplicação das cadeias de Markov ao estudo da mobilidade do mercado de trabalho na região Norte», *Actas do X Congresso da SPE*, 2002. J. Nicolau, «Stationary processes that look like random walks – The bounded random walk process in discrete and continuous time», *Econometric Theory*, 18, 2002.

<sup>51</sup> Existe um texto em português que resultou da colaboração do Autor com Daniel Muller e Kamil Feridun Turkman: *Análise de Sucessões Cronológicas*, McGraw-Hill, Lisboa, 1993

<sup>52</sup> Por exemplo, ver, Vera Lúcia Oliveira e Maria Eduarda Silva, «Processos para Séries Temporais de Contagem», *Actas do VII Congresso da SPE*, 1999 e Sara Morgado Nunes e Maria Eduarda Silva, «Subamostragem em processos autoregressivos de valores inteiros», *Actas do X Congresso da SPE*, 20002.

<sup>53</sup> K. F. Turkman (com M. Walker), «A Stability Result for the Periodogram», *Annals of Probability*, 18, 1990. K. F. Turkman (com Subba Rao), «On extremes of bilinear and random coefficient autoregressive processes», *Notas e Comunicações*, CEAUL, 1997. K. F. Turkman e M. A. Amaral Turkman, «Extremes of bilinear time series models», *Journal of Time Series*, 18, 1997. Kamil Feridun Turkman (com Marília Reis), «Modelos de regime para séries temporais», *Actas do X Congresso do SPE*, 2002. Kamil Feridun Turkman (com Manuel G. Scotto), «Métodos de estimação em modelos bilineares», *Actas do V Congresso do SPE*, 1997. Kamil Feridun Turkman (com Marília Reis), «Um modelo regido pelas observações para precipitação», *Actas do X Congresso do SPE*, 2002. P. Teles (com R. M. Heiberger), «Display for direct comparison of ARIMA models», *American Statistician*, 56, 2002. K. F. Turkman, M. A. Amaral Turkman e M. Antunes, «A Bayesian approach to event prediction», *Journal of Time Series*, 24, 2003.

<sup>54</sup> (com P. Jacob) «A new technique for testing general ARMA models», *Actas do I Congresso da SPE*, 1993 e «A decision procedure for bilinear time series based on the asymptotic separation», *Statistics*, 2000. «Series temporais com erros condicionalmente heteroscedásticos», *Actas do III Congresso da SPE*, 1995; «Ensaio estatístico entre processos de erro e modelos auto-regressivos vectoriais», *Actas do VIII Congresso da SPE*, 2000.

<sup>55</sup> Nuno Crato (com Bonnie. K. Ray), «Some problems in the overspecification of ARMA and ARIMA processes using ARFIMA models», *Actas do III Congresso da SPE*, 1995. Nuno Crato, «Séries temporais de memória longa. Da teoria às aplicações», *Actas do X Congresso da SPE*, 2002. Nuno Crato (com R. Ramjee e B. K. Ray), «A note on moving average forecasts of long memory processes with an application to quality control», *International Journal of Forecasting*, 18, 2002.

<sup>56</sup> Daniel Muller (com W. W. S. Wei), «Iterative Least Squares and Estimation and Identification of the transfer function model», *Journal of Time Series Analysis*, 18, 1997. Daniel Muller (com Cristina Oliveira), «O sobreajustamento na identificação de um modelo de função de transferência bivariado através da função de autocorrelação estendida generalizada amostral», *Actas do X Congresso do SPE*, 2002. Anabela Afonso, «Uma aplicação dos modelos arma-garch às taxas de rentabilidade do índice PSI-20», *Actas do X Congresso do SPE*, 2002. Marília Reis, *Some Problems in Non-Linear Prediction*, DEIO, 2002. F. F. R. Ramos, «Forecasts of market shares from VAR e BVAR models: a comparison of their accuracy», *International Journal of Forecasting*, 19, 2003. P. M. M. Rodrigues (com U. Hassler), «Seasonal Unit Roots tests under structural breaks», *Journal of Time Series Analysis*, 25, 2004. P. M. M. Rodrigues (com A. M. R. Taylor), «Alternative estimators and unit roots for seasonal autoregressive processes», *Journal of Econometrics*, 120, 2004. P. M. M. Rodrigues e P. M. D. C. B. Gouveia, «Threshold cointegration and the PPP Hypothesis», *Journal of Applied Statistics*, 31, 2004.

A Econometria é de criação relativamente recente (praticamente em 1930 com a fundação da *Econometric Society* em que teve importante papel Ragnar Frisch) e pode dizer-se, em termos gerais, que envolve uma fecunda combinação da Economia, da Matemática e da Estatística. O seu desenvolvimento tem sido notável e hoje está-se muito longe da Econometria dessa data ou da matéria ensinada no 1.º Curso ministrado nos ISCEF (hoje ISEG) em 1952/53 e que foi também o 1.º efectuado no País<sup>57</sup>. Na presente nota tem interesse uma concepção menos vasta segundo a qual a Econometria está sobretudo relacionada com os problemas de especificação e inferência em modelos estatísticos para fenómenos económicos. Quer dizer, o estudo moderno da Econometria tem que ter em conta os seguintes aspectos:

- a natureza dos dados – dados seccionais, dados temporais, dados de painel;
- os processos de geração dos dados – tipos de amostragem, processos estocásticos subjacentes;
- as hipóteses que delimitam os modelos econométricos – comportamento e relacionamento das variáveis, exogeneidade ou pré-determinação versus endogeneidade, estacionaridade versus não estacionaridade, ergodicidade, identificação e, enfim, as premissas que permitem obter as distribuições exactas ou assintóticas dos estimadores dos parâmetros.

Indicam-se na nota abaixo alguns estudos importantes de Econometria publicados por investigadores portugueses<sup>58</sup>.

Enfim, é escusado frisar que muitíssimo ficou por dizer sobre a evolução dos métodos estatísticos no nosso País nos últimos 25 anos. O Autor lamenta, por exemplo, não ter feito referência à impressionante lista de aplicações da Estatística à medicina e aos problemas de sobrevivência. No entanto, as muitas omissões devem ser consideradas uma evidência da explosão do interesse pela Estatística em Portugal, sendo muito difícil para qualquer pessoa acompanhar, mesmo superficialmente, como foi o caso, os inúmeros campos em que se desdobra a nossa disciplina. Espera-se, pois, que outros colegas chamados a dar a sua contribuição preencham as inevitáveis lacunas e os inevitáveis desequilíbrios ou enviesamentos do presente escrito de que o Autor tem consciência.

<sup>57</sup> Veja-se o excelente estudo de José A. F. Machado e J. M. C. Santos Silva, «50 Anos de Ensino de Econometria em Portugal», *Economia*, 26, 2002.

<sup>58</sup> Pedro Portugal (com J. T. Addison), «On the Distributional Shape of Unemployment Duration», *The Review of Economics and Statistics*, 69, 1987. Isabel Cristina Teixeira Andrade, *Three Essays in Applied Multivariate Econometrics*, Ph.D., University of Southampton, 1993. J. M. C. Santos Silva (com F. A. G. Windmeijer), «Endogeneity in Count Data Models: An Application for Demand of Health Care», *Econometrics*, 12, 1997. Nuno Crato e Pedro Lima (com F. Jay Breidt), «The detection and estimation of long memory in stochastic volatility», *Journal of Econometrics*, 83, 1998. J. A. F. Machado e J. M. C. Santos Silva, «Glejser's test revisited», *Journal of Econometrics*, 97, 2000. J. A. F. Machado e J. Mata, «Counterfactual Decomposition of Changes in Wage Distributions using Quantile Regression», *Journal of Applied Econometrics*, a publicar em 2004.

## **Bento José Ferreira Murteira**

Licenciado em Finanças pelo Instituto Superior de Ciências Económicas e Financeiras.

Doutor em Finanças pela Universidade Técnica de Lisboa (1953).

Docente no Instituto Superior de Economia e Gestão desde 1947. Em 1954 passou de assistente a 1.º assistente, em 1961 passou a Professor Extraordinário e em 1963 a Professor Catedrático categoria em que se Jubilou em Agosto de 1994. Nos últimos anos de Catedrático (inclusivamente no ano lectivo de 1994-95) colaborou activamente no Mestrado em Matemática Aplicada à Economia e Gestão, de que foi um dos fundadores.

No âmbito da Fundação das Universidades Portuguesas, fez parte em 1997 e 1998 das Comissões de Avaliação Externa das Licenciaturas em Economia e das Licenciaturas em Matemática; em 2001 fez parte da Subcomissão de Avaliação Externa das Licenciaturas em Estatística

Tem uma extensa lista de publicações nas áreas de estatística, econometria e investigação operacional embora recentemente se dedique apenas à estatística. O volume editado pelo ISEG, por ocasião do Jubileu (Junho de 1994), contém uma boa parte dos artigos publicados. Impossível é também reproduzir a infindável lista de participações em júris do mais diverso tipo e as conferências realizadas.

É sócio honorário da Sociedade Portuguesa de Estatística, desde 1993.

# A Educação em Portugal e o Século XX<sup>1</sup>

*José Tiago da Fonseca Oliveira*

Fazer um esforço rápido do que tem sido a Educação em Portugal no nosso século é como que uma maratona em face das diversas viragens, positivas e negativas, que tem tido. Vamos tentá-lo, de modo breve, procurando chegar ao dia de hoje, visando já o século XXI e a Comunidade Europeia que agora tem 320 milhões de cidadãos e, com a evolução histórica acelerada a que e está assistindo, a médio prazo poderá ser muito maior.

## I

Em 1910 proclama-se em Portugal a I República. A grandeza e generosidade do ideário republicano, em 3 curtos anos, traça as linhas de desenvolvimento da Educação no País e Colónias. Poderemos considerar como marcos fundamentais a reforma das Universidades, em que as do Porto e a de Lisboa se vem juntar à de Coimbra, a criação da Faculdade de Direito de Lisboa, a passagem das Escolas Médico-Cirúrgicas do Porto e Lisboa a Faculdades de Medicina, a autonomia dos cursos de Farmácia, a transformação da Academia Politécnica do Porto e da Escola Politécnica de Lisboa em Faculdades de Ciências, a criação das Faculdades de Letras por transformação da Faculdade de Teologia e do Curso Superior de Letras de Lisboa, a criação das Escolas Normais Superiores de Coimbra e Lisboa, a organização do ensino técnico superior que vem para Instrução Pública do Ministério do Fomentos, as Escolas que vêm do Ministério das Colónias, etc. Fora do sistema ficaram as Escolas Militares. Embora com erros, em particular no modo de escolha de professores, erros de que o ensino universitário se virá, mais tarde, a ressentir, a abertura que essa reforma mostrava, abria caminhos ao fechado ensino universitário. Na Constituição Universitária de 1911 institucionaliza-se a investigação como parte essencial. É o eco do ensaio de 1841 de Herculano intitulado “Da Escola Politécnica e do Colégio dos Nobres”, em defesa da primeira e crítica acérrima ao segundo onde diz “A Universidade deve ser o padrão por onde se afina tudo o que diz respeito à

---

<sup>1</sup> Conferência proferida na Academia das Ciências de Lisboa, em Novembro de 1990 e publicada em *Collected Works. Obras. J. Tiago de Oliveira. Volume I*, p. 487-496. Editorial Pendor, 1994.

inteligência: a sua missão é duplicada – de escola quando provê de novos sujeitos as classes excepcionais; de academia, quando os seus membros, vivendo pela ciência e para a ciência, tem na mão o facho que alumia as escolas de aplicação, o facho dos progressos puramente especulativos” (Opúsculos, tomo VIII).

Houve depois alterações, apenas de pormenor, até ao Estado Novo.

Não tem a mesma envergadura a reorganização do ensino secundário: só a criação dos trabalhos práticos quando possível (de 1914), a preparação para trabalhos domésticos das jovens (também de 1914) e pequenas reformas de curta duração, ou mesmo nula, entre as quais avulta talvez a criação de 3 liceus femininos, com autorização para que as jovens estudantes, fora do Porto, Coimbra e Lisboa onde se localizavam os liceus femininos, se pudessem matricular nos liceus masculinos, um por cada capital dos outros distritos. Todavia um ponto é importante e transmite aqui a ideia normativa do ideário republicano: “A instrução cívica deve ser dada pelos professores durante o ensino das disciplinas que para tal ofereçam ensejo, e nomeadamente das de Português e História.”, o que era passo em frente à data (1918).

O ensino técnico-profissional, não universitário, é principalmente obra de 1919 com as Escolas de Artes e Ofícios – que tão grande papel virão a desempenhar com os Altos Comissários de Norton de Matos, em Angola e de Brito Camacho em Moçambique, e com o governo Álvaro de Castro também em Moçambique -, as Escolas Industriais e as Escolas Preparatórias (depois Institutos Industriais) e as Aulas de Comércio, Escolas Comerciais e Institutos Comerciais. O ensino agrícola vai essencialmente apoiar-se nas escolas móveis.

Mas era essencialmente o ensino primário e também o infantil, que mais falavam ao coração dos republicanos, perfeitamente conscientes da nossa elevadíssima taxa de analfabetismo; cerca de 70% em 1911 e reduzida apenas para cerca de 65% em 1920. A 29 de Março de 1911 sai uma notável reforma do ensino primário, embora criticada em certos pontos por João de Barros e João de Deus Ramos, que, se tivesse sido executada a passo rápido, nos poria a par da Europa. O mesmo decreto legisla ainda sobre o ensino infantil e ensino normal. O ensino infantil, facultativo, desenvolve-se pouco na órbita do Estado, mas permite ensino particular dentro do qual temos a rede dos Jardins-Escolas João de Deus que ainda hoje subsistem.

O ensino primário com 3 escalões – elementar (obrigatório) complementar e superior – é, em princípio, gratuito e dá as noções básicas da língua, das ciências exactas e naturais, da geografia, da história e tem ainda “moral prática” e “educação social, económica e civil”. São criadas escolas primárias em todas as freguesias do País e Ilhas, se possível com separação dos sexos. É o retomar, para os dois sexos, das Escolas Menores (gratuitas), com os Mestres Régios, que o Marquês quis criar em cada bairro de Lisboa e uma ou duas em cada vila do Continente, Ilhas e Ultramar. Então não foi exequível mas todavia entre nós e também no Brasil produziu efeitos positivos. As Escolas Primárias Superiores, curiosamente em regimes de coeducação, derivativo semi-profissionalizado para quem não seguisse os outros ensinos de continuação, não chegam a exercer grande peso embora algumas tivessem tido aspectos de notar. As Escolas Normais Primárias que

recebiam alunos vindos ou do ensino primário superior ou de classes análogas dos liceus, em princípio uma por distrito, iriam formar os professores primários com cuja acção a I República esperava manter o ideal republicano, dominar o analfabetismo e alargar os horizontes. Deve notar-se que este decreto, em parte executado, é muito influenciado pela estrutura educacional francesa coeva.

Porém 1923 poderia ter sido o segundo marco na organização educacional do País, se fosse aprovado o projecto de João Camoesas da “Lei sobre a reorganização da educação nacional”, de 2 de Julho, em grande parte influenciado pela “Seara Nova”. A sua dúzia e meia de páginas tem a fundamentação e o esquema organizativo do ensino e do ministério que o regeria (a chamar-se já Ministério da Educação Nacional). Muito se teria ganho se ele tem sido aprovado e executado. António Sérgio di-lo claramente “Quem conspira contra a reforma medite bem no que vai fazer: porque assume perante o povo a mais tremenda das responsabilidades. *Um dia a nação nos há-de julgar.*” E Jaime Cortezão, em sintonia, mas detalhando outros aspectos, diz: “A reforma constitue, na crítica dos métodos de ensino e nos fins gerais que tacitamente se propõe, não só o mais sério documento político emanado de um governo, dentro da República, como a primeira tentativa de reforma nacional, orientada por um espírito democrático”. O ministério a que pertencia João Camoesas cai em Novembro e o documento é hoje, quase só, uma peça histórica. Quase só peça histórica, pois embora necessitando evidentemente de ajustamentos, há ainda hoje pontos a desenvolver e a aplicar, como por exemplo a cultura moral, social e artística na instrução primária, a disciplina com o mesmo nome no curso geral dos liceus, esta com trabalhos práticos no complementar. A cultura física está sempre presente. No ensino técnico elementar (agrícola, comercial, doméstico, industrial e marítimo) há, além da formação básica geral, as aplicações à veterinária, tecnologia agrícola, economia rural, prática comercial, contabilidade, economia doméstica, trabalhos de ferro e madeira, economia industrial, navegação costeira, direito marítimo, etc., consoante os casos. As escolas técnicas complementares, em que se convertiam os liceus femininos, fariam ensino comercial, industrial, artístico e de enfermagem continuando a mesma linha. É curiosa, e do tempo, a preocupação de separação dos sexos.

As Universidades proporião ao governo a sua própria reforma e as Escolas Normais Primárias e Superiores passariam a Faculdades de Ciências da Educação, só recentemente criadas, que se integrariam nas Universidades. E vem ainda a Inspeção médico-escolar, a Inspeção técnica do ensino, as Universidades Populares, uma Junta de Fomento de Actividades Sociais e Investigações Científicas, a estruturação das construções escolares, o Fundo de Educação Nacional, a reforma do Ministério, a criação de escolas-modelo (experimentais e para deficientes) e até um Museu Pedagógico. A 29 de Dezembro de 1923 António Sérgio, num efémero Ministério em que a “Seara Nova” participou em convite directo, criou a Junta de Orientação de Estudos – correspondente à Junta de Fomento das Actividades Sociais e Investigações Científicas – com uma largueza de vistas ainda não retomada 66 anos depois e o Instituto do Cancro, que após vicissitudes várias se mantém hoje, tendo a Junta o nome actual de INIC.

## II

E, assim, se chega ao Estado Novo, perdido que foi o projecto de João Camoesas, com a Educação seguindo essencialmente o modelo dos primeiros anos da República.

O Estado Novo, até ao ministério de Veiga Simão, representa quase sempre um retrocesso ou pelo menos paragem de evolução educacional. Como pontos positivos temos o retomar, sob outro nome, da ideia da Junta de Orientação de Estudos de Sérgio e Camoesas, a criação da Universidade Técnica de Lisboa por federação de Escolas existentes e o envio de bolsheiros para o estrangeiro, em maior ou menor número, que vai permitindo uma remodelação das Universidades, lenta e insuficiente. E, há que admiti-lo, certas admissões precipitadas de professores, esquecendo o pensamento de Herculano consignado na Constituição Universitária, causaram problemas durante os anos 30.

Outros aspectos são a criação dos Estudos Gerais Universitários em Angola e Moçambique, a imposição progressiva dos doutoramentos, a lateralmente útil Campanha Nacional de Educação de Adultos que os levou a enviar os filhos para a escola, reformas do ensino liceal e técnico ambas de 1947 com alguns aspectos positivos, permitindo a introdução da Escola Activa e os estudos para o Fomento Cultural ligados à OCDE/Projecto Regional do Mediterrâneo (1959/64). Mas que dizer das demissões políticas de 35 e 47, da extinção da autonomia parcial das Universidades e do encerramento de Faculdades em 1933, da redução da escolaridade obrigatória quando a taxa de analfabetismo era cerca de 63% em 1930 e 50% em 1940, da criação em 1936 dos regentes escolares que ensinaram o melhor possível aquele pouco que sabiam, da suspensão das matrículas nas Escolas Normais (ditas de Magistério Primário) que vai até 1939, da lei do casamento das professoras primárias, do decair do ensino infantil oficial, da criação da Mocidade Portuguesa afastando o movimento escuteiro livre de Sir Baden Pwell, do lento, injusto escorrer de rejeições de candidatos por actividades ditas contra o Estado, da sucessão de crises estudantis dos anos 60, etc.?

Veiga Simão, no fim do Estado Novo, representa um grande arejamento do sistema (que já começara com certas reformas anteriores da Universidade). É uma forte remodelação global do ensino, com aspectos controversos e manchas tecnocráticas, mas que representa uma séria adaptação aos tempos de então: fim dos anos 60, início dos anos 70.

O “Projecto do Sistema Escolar” e as “Linhas Gerais de Reforma do Ensino Superior” que não vale a pena descrever com detalhe, pois é o sistema vigente quase completamente até há pouco e parcialmente ainda hoje, foram como se disse um salto em frente, que urge reformular e prolongar, ajustando às condições de hoje em que não há Império Colonial e a 3 anos de prazo estaremos totalmente integrados na Comunidade Europeia.

## III

É, agora, a altura de entrarmos no tema Educação na III República, após o 25 de Abril. Em seguida à confusão no sistema educacional, em que a educação cívica politécnica foi um passo feliz, importante, mas com erros e falhado, houve que estabilizar o sistema com

os diplomas sobre a gestão do ensino que deveriam ter duração curta e ainda subsistem. Depois, efectuaram-se alguns progressos como o alargamento da escolaridade obrigatória, agora atingindo praticamente 100% de escolaridade primária e com uma taxa de analfabetismo em morte lenta – embora com algum semi-analfabetismo –, a reintrodução do ensino profissionalizante, a reorganização dos centros de estudo do INIC, parada todavia há tantos anos!, certa expansão da escolaridade portuguesa no estrangeiro, os diplomas dos departamentos universitários, dos cursos pós-graduados e dos créditos, a perda de oportunidade de adequar os Institutos de pós-graduação universitários que não existem!, a semi-eleição dos Reitores Universitários (tantas vezes partidarizada, para o que basta ler o “Expresso” em cada instância), a institucionalização do ensino politécnico, oscilante e meandrizada, a melhoria de certas disciplinas e de manuais escolares, eis o que parece poder dizer-se como modificações, de maior ou menor vulto, do projecto Veiga Simão! Pelo caminho ficaram afastamentos injustos e promoções infelizes.

É pouco, é um País adiado na Educação, 15 anos após.

#### IV

E agora e aqui, Portugal, 1989, olhos no futuro?

Regressemos ao princípio dos tempos perguntando-nos o que é a Educação? Benda responde-nos dizendo que educar é propor uma hierarquia de valores à juventude. E eu ajuntaria “de saberes”, que são consentâneos com os valores, os fundamentam e explicitam. E é, nesta perspectiva pessoal, como cidadão, que assumo totalmente a responsabilidade do que vai ser dito, sem consultar nem co-responsabilizar ninguém.

Em que medida as perspectivas actuais, supostas voltadas ao futuro, expressam uma hierarquia de valores e de saberes apresentada aos jovens?

Permitam-me faça extractos da carta que o Presidente François Mitterrand, em 13 Fev. 84, dirigiu ao “Collège de France” dizendo:

“Votre haute institution est, depuis sa naissance, au centre des principales avancées du savoir et de l’enseignement. Aussi, j’attacherai le plus grand prix à ce que, sans entrer dans le détail des programmes, le collège de France veuille bien réfléchir a ce que pourraient être, selon lui, les principes fondamentaux de *l’enseignement de l’avenir*, intégrant la culture littéraire et artistique la plus universelle, avec les savoirs et les méthodes des sciences les plus récentes.

Il conviendrait pour cela de prendre en compte les changements technologiques et sociaux, de favoriser la connaissance de toutes les cultures, enfin et surtout de doter les jeunes français d’instruments modernes de pensée, d’expression et d’action». A ideia dos valores e dos saberes está pois presente nesta carta.

De modo análogo agiram os Senhores Reis de Portugal e dos Algarves, no século passado, em que a Academia deu o seu parecer fundamentado sobre a introdução do sistema métrico decimal de pesos e medidas, sobre a Comissão Geológica – hoje Serviços Geológicos de Portugal –, sobre o Sistema estatístico nacional, criou a instituição vacínica,

que generalizou a vacinação e realizando a sua missão se tornou inútil, e o curso superior de Letras. Curiosamente, a actual “Lei de bases do Sistema Estatístico Nacional”, aprovada por unanimidade no Parlamento, é muito inferior à lei de Armindo Monteiro, de 1935, sobre o “Instituto Nacional de Estatística” que de resto vem seguir-se, após certa evolução, à Comissão Central de Estatística do Reino de 1857. Para ela foi importante o notável parecer de Marrecas Ferreira em 1854, actual ao tempo, em sequência a uma intervenção de Herculano nesta Academia. Não foi agora consultada a Academia – quem foi mesmo consultado? - e, de repente, surge um “engano” de 130 milhões de contos, superior ao montante do capital social de 80 novas seguradoras!

Mas regressemos a Paris e vejamos qual a resposta da Assembleia do “Collège de France” ao Presidente Mitterrand, cerca de um ano depois. As “Propositions pour l’enseignement de l’avenir” de 1985, enunciam 9 princípios cujos títulos citamos:

1. L’unité dès sciences et la pluralité dès cultures;
2. La diversification des formes d’excellence;
3. La multiplication des chances;
4. L’unité dans et par le pluralisme;
5. La revision périodique des savoirs enseignés;
6. L’unification des savoirs transmis;
7. Une éducation ininterrompue et alternée;
8. L’usage des techniques modernes de diffusion;
9. L’ouverture dans et par l’autonomie.

São estes os princípios básicos que o «Collège de France» designou e que, tanto quanto creio, a Academia reiteraria na sua linha geral se tivesse sido consultada, como devia sê-lo.

De resto, quem foi consultado sobre a “Lei de bases do Sistema Estatístico Nacional”, a “Lei da autonomia das Universidades”, a “Lei de bases do sistema educativo”?

Ele há assuntos demasiado sérios!

A comparação da “Lei de bases do sistema educativo”, de 1986, no seu artº 2 (de título “Princípios gerais”), difere bem das “Propositions” do “Collège de France” e é pena!

Deixemo-la para trás e falemos, um pouco em corrida, sobre algumas das leis já em vigor, tanto havendo para dizer e mais ainda para fazer.

No decreto-lei 286-89 que organiza os currículos do ensino ante-superior embora com aspectos válidos podem, entre outras, revelar-se as seguintes situações:

1. No 3º ciclo do ensino básico (7º, 8º e 9º anos) há uma opção de 3h/semana entre uma 2ª língua estrangeira, a educação tecnológica e a educação musical. Como a 2ª língua estrangeira é obrigatoriamente oferecida, e a Comunidade Europeia está à porta, é natural que seja em grande parte escolhida, o que reduz a formação tecnológica num Mundo em permanente evolução e coloca os jovens portugueses em clara situação de subalternidade.

Todavia a I República já a começava com o livrinho “Lições de Coisas” que estudei na escola primária, com proveito, e agora recorro com saudade;

2. Não há educação cívica como disciplina autónoma. Irá aparecer em parte das cerca de 100 h/ano do 3º ciclo na multiforme área-escola. Todavia ela é essencial para os cidadãos; a alternativa Desenvolvimento Pessoal e Social/Educação Religiosa é insuficiente e desigual. A I República dava-lhe a importância e dignidade necessária e o próprio Estado Novo manteve a Educação Moral e Cívica.

Civismo/cidadania e Tecnologia eis duas faltas graves da reforma curricular em curso. E esquecemos a discutida solução do insucesso escolar, o estatuto dos professores com as áreas científicas em desertificação, o desenvolvimento do calendário das chamadas “inovações”, etc.

A lei de autonomia universitária desagradou a todos os colegas que conheço o que nos leva a tentar contorná-la nos estatutos das Universidades e Faculdades, visando o século XXI.

Qual das filosofias que “The Ideal of the University”, de Robert Paul Wolff, é a seguida no artº 1 (Missão da Universidade): a do “Sanctuary of scholarship”, a do “Training camp for the professions”, a da “Social service station” ou a da “Assembly line for the establishment”? Talvez a escolha venha a ser feita por razões tecno-burocráticas, dados os recursos humanos, laboratoriais, bibliográficos e financeiros necessários em cada caso. E onde se reflectem as 9 “Propositions” do “Collège de France”?

As condições de escolaridade são francamente más com 40 ou mais alunos nas aulas práticas e mais de 100 nas teóricas, tantas vezes; quando e se há verbas a contratação leva meses.

O programa Ciência irá mobilizar avultados recursos financeiros para as Universidades ainda em 1989. Hoje ainda nada se sabe sobre ele.

Uma Universidade tem de ter uma componente de artes e humanidades e outra de ciências e técnicas, o que o Estado, de resto, pratica. Porém no Estatuto do ensino superior particular e cooperativo, de 1989, não se exige ensino tecnológico, pelo que continuamos a poder ter as Universidades privadas – tão necessárias de facto – como Universidades de “lápiz e papel”, com eventualmente um pequeno parque de microcomputadores!

Estamos bem longe da concepção de Multiversidade de que falava Kerr, Presidente da Universidade de Columbia, em New York, nos anos 60.

Analogamente podemos perguntar-nos o que se passa com os cursos de formação, sobre alguns dos quais o F. S. E. se está já a debruçar. Continuam a faltar os cursos de reciclagem e especialização bem como o seu ordenamento no mercado de trabalho.

*Quo Vadis*, Educação em Portugal nos finais do século XX?

**José Tiago da Fonseca Oliveira** (22.12.1928 – 23.6.1992)

Professor Universitário e Cientista de renome mundial.

Autor de duas centenas de artigos e de uma dezena de livros. Pioneiro do ensino das Probabilidades e Estatística em Portugal. Criador, com E.J. Gumbel, da Teoria Matemática dos Valores Extremos.

Secretário de Estado da Investigação Científica do I Governo Constitucional.

Fundador e principal impulsionador do primeiro Projecto de Matemáticas Aplicadas em Portugal, na sequência do qual foi criado o Departamento de Estatística e Investigação Operacional da Faculdade de Ciências da Universidade de Lisboa.

Fundador e principal impulsionador da Sociedade Portuguesa de Estatística de que foi o primeiro Presidente.

Professor da Faculdade de Ciências de Lisboa, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa, Universidade da Baía, Technion, University of California at Santa Barbara e Universidade Livre de Bruxelas, entre outras.

Membro da Academia das Ciências de Lisboa.

# O Ensino da Estatística em Portugal nos Últimos 150 Anos

*Maria de Fátima Fontes de Sousa*

A reforma de Passos Manuel, iniciada com o decreto de 15 de Novembro de 1836, estende-se a todos os níveis de ensino nomeadamente ao universitário, procurando por este meio introduzir na formação das novas gerações os ideais liberais que se firmaram em Portugal sobretudo a partir da Regeneração.

Assim, são organizadas as escolas médico-cirúrgicas, criam-se a Escola Politécnica de Lisboa, a Academia Politécnica do Porto, o Conservatório de Arte Dramática e as Academias de Belas Artes do Porto e Lisboa Faculdade de Direito em Coimbra por fusão de duas.

Em 5 de Dezembro do mesmo ano constitui a faculdade de Direito de Coimbra a partir da de Leis e Cânones, então aí existentes.

Para além destas inovações, alarga o quadro das cadeiras até então leccionadas a novas disciplinas como, por exemplo, Álgebra, Mineralogia, Química e Economia Política.

A nova Faculdade de Direito de Coimbra -certamente incentivada por este espírito renovador - passou a incluir um curso de Economia Política regido por Adrião Forjaz de Sampaio que, atento ao que se passava no ensino superior na Europa, introduziu como complemento deste curso os primeiros ensinamentos em Estatística.

Deste modo, em 1841, foi publicada por aquele professor uma obra com o título “Primeiros elementos da ciência da estadística”, extraídos principalmente do Jornal dos Trabalhos da Sociedade Francesa de Statistique Universelle, para auxílio das prelecções suplementares do curso de Economia Política da Faculdade de Direito de Coimbra.

Esta obra, que surgiu como compêndio, foi traduzida, ainda em 1841, para castelhano por Vicente Diez Canseco e impressa em Madrid sob o título de “Elementos de la Ciencia de la Estadística” tornando-se o 1º manual de estatística no país vizinho, pois que só em 1844 irá surgir outra obra de estatística agora de autor espanhol, José Maria Ibanez, com o título “Tratado Elemental de Estadística”. Este compêndio de Forjaz de Sampaio teve, posteriormente, sucessivas edições com algumas variações de conteúdo e título

Assim, por exemplo, em 1855, saíu o compêndio “Estudos sobre os primeiros elementos da theoria da estadística” e “Novos elementos de Economia Política e Estadística” 5ª edição em 1859.

Nessas obras Forjaz de Sampaio inspira-se sobretudo na escola francesa e refere-se muito a Quetelet, cientista que, em 1834, publicou “Essai de physique sociale” onde já trata de problemas do foro da Estatística e que foi o organizador dos primeiros Congressos Internacionais de Estatística iniciados, em 1853, em Bruxelas. Mais uma vez Coimbra inspirada na Escola Francesa é pioneira na introdução de novos estudos.

Até prova em contrário, tudo indica, pois, que data de 1841 a primeira iniciativa em Portugal de introduzir na Universidade o ensino de métodos estatísticos, embora, ao tempo, de forma ainda muito retórica e pouco formalizada, mas onde já se pode reconhecer o discurso da estatística descritiva no que diz respeito a apresentação de tabelas e gráficos bem como na introdução de algumas noções importantes como a de média e dispersão de séries estatísticas.

Mais tarde, em Lisboa, na Escola Politécnica, recentemente criada a partir do Real Colégio dos Nobres, é inserido, a partir de 1865, na cadeira de Cálculo Infinitesimal o ensino de rudimentos de Cálculo das Probabilidade.

E assim continuou até 1911.

Com a implantação da Republica, foi promulgada uma reforma criando, na recente Faculdade de Ciências de Lisboa, uma cadeira trimestral de Cálculo das Probabilidades a qual passou, em 1932, a anual, inserindo-se no 3º ano do curso de Matemática de que foi professor o Prof. Doutor Victor Hugo de Lemos e assistente o Doutor Pedro Brauman.

Em 1949 ainda eram estes professores que davam esta disciplina e de quem fui aluna. Só posteriormente, em 1957, é criada uma outra cadeira semestral com o nome de “Probabilidades, Erros e Estatística” a qual dizia respeito aos preparatórios de Engenharias que ao tempo podiam ser feitos na Faculdade de Ciências até ao 3º ano.

Mas, em 1952/53, existiam na Secção de Matemática da Faculdade de Ciências de Lisboa um *Centro de Álgebra* e um *Centro de Matemática Aplicada ao Estudo da Energia Nuclear* que congregavam vários docentes e investigadores em matemática e que eram orientados pelo Prof. A. Almeida e Costa. Estes centros visavam o estudo e investigação mais avançados quer em Álgebra quer em Aplicações da Matemática a outros domínios.

Foi nesta data e nestes Centros que conheci o Tiago de Oliveira o qual, levado pelo Prof. Almeida e Costa, tinha ingressado na F.C.U.L como 2º assistente da Secção de Matemática -Grupo de Análise e Geometria. Na época eu também fazia parte destes Centros como elemento externo da Faculdade, auferindo, a partir de 1955, uma pequena bolsa -400\$00 mensais - para participar nos seus seminários e efectuar trabalho de investigação.

Os seminários realizavam-se uma vez por semana e os trabalhos de investigação dos vários investigadores eram publicados em separatas relativas aos respectivos centros e com os títulos dados pelos investigadores seus autores. Desta forma foram assim publicados mais de cem títulos.

Nessa altura já o Tiago de Oliveira se interessava pela Estatística, talvez devido à sua passagem anterior pelo Instituto de Biologia Marítima onde tinha sido Assistente de Investigação em Biometria e Bioestatística e publicára já em 1952/3 vários artigos científicos sobre problemas estatísticos.

Nas longas conversas formais e informais que tivemos entre 1953 e 1957 era muitas vezes lamentado o facto da Estatística não fazer parte dos curricula de qualquer curso de Matemática em Portugal, tendo assim progressivamente nascido na mente do Tiago de Oliveira a ideia de introduzir essa disciplina no Ensino Universitário português, à qual ideia devo ter sido a primeira aderente entusiástica.

Assim, quando em 1958 surgiu a oportunidade de ter uma bolsa do Governo Francês - através do Instituto Francês em Portugal- decidi, com o apoio do Tiago, aproveitá-la para fazer um curso de Estatística no "Institut de Statistique de l'Université de Paris" e, simultaneamente, preparar uma tese sobre um tema em Estatística para posterior doutoramento em Portugal.

Durante os três anos que lá permaneci tirei o curso e comecei sob a orientação do Prof. Dugué a preparação da tese com que mais tarde me doutorei.

Circunstancias adversas da minha vida particular bem como um parecer desfavorável da P.I.D.E. que levantou dificuldades em relação ao meu ingresso num lugar no Instituto de Biologia Marítima e que só foram resolvidos 4 anos mais tarde, fizeram com que só em 1965 eu regressasse à F.C.L. para dar aulas práticas na disciplina de Probabilidades e Estatística para os cursos de Engenharia que continuavam a existir na F.C.L. O Prof. Tiago dava as aulas teóricas e eu as práticas.

O meu regresso às lides universitárias foi calorosamente apoiado pelo Tiago de Oliveira que na época já tinha conseguido convencer o Conselho Escolar da Faculdade a propor uma licenciatura em Matemática com duas opções a partir do 3º ano, respectivamente "Matemática Pura" e "Matemática Aplicada" iniciando-se em 1966 pela primeira vez o 3º ano das novas opções.

Com o meu ingresso como assistente para a Matemática Aplicada foi possível oferecer, a partir de 1968, algumas cadeiras de opção nos 3º e 4º anos no domínio da Estatística que o Tiago considerava fundamentais para implementar definitivamente o Ensino daquela disciplina em Portugal.

Com o seu espírito agudo e empreendedor o Tiago viu, nesta ampliação e flexibilização dos curricula, a oportunidade para introduzir novas disciplinas em dois campos afins da Matemática e Estatística : Investigação Operacional e Computação.

Estes domínios, na altura de cultivo recente (três dezenas de anos) nos países mais avançados, eram totalmente inexistentes no panorama do ensino em Portugal e passaram a estar incluídos no seu projecto de modernizar e mudar as Matemáticas Aplicadas em Portugal.

Foi assim que nas cadeiras de opção dos 3º e 4º anos da Matemática Aplicada surgiram sucessivamente disciplinas como Métodos Matemáticos de Investigação Operacional, Programação Matemática, Cálculo Automático, Computadores, Grafos, Simulação, Programação Dinâmica, para só citar algumas, a maior parte dadas por mim.

Todas as diversas opções implementadas bem como as disciplinas obrigatórias Teoria da Probabilidade e Estatística Matemática, foram a partir de 1966 até finais da década de 1970 a base de formação dos novos licenciados na opção em Matemática Aplicada. Foi dos licenciados formados com este espírito, sobretudo dos anos 70/71/72 que nasceu um “núcleo duro” de novos discípulos que o Tiago iniciou na investigação em Estatística e que quer cá quer no estrangeiro fizeram doutoramentos na década de 70.

Orgulho-me de ter sido a primeira a iniciar este núcleo, mesmo antes do projecto ter dado frutos, e de ter trabalhado duramente a seu lado para a sua realização.

Ter projectos é relativamente fácil, aliciante e agora até está na moda. *O difícil é realizá-los ao longo de anos.* É preciso ter *criatividade, perseverança e muita dedicação apaixonada.* O Tiago tinha as três qualidades com abundância suficiente para arrastar nesta senda de pioneirismo científico alguns jovens licenciados na época referida.

Todo este trabalho culminou em 1981 com a criação formal do novo Departamento de Estatística, Investigação Operacional e Computação -o DEIOC- de que ele foi o 1º presidente- e de duas novas licenciaturas “Probabilidades e Estatística” e “Estatística e Investigação Operacional” que ainda hoje subsistem praticamente com os mesmos currícula. Logo de seguida criou dois cursos de pós-graduação com os mesmos nomes das licenciaturas e que começaram a funcionar a partir de 1983.

Contou também o Prof. Tiago com as colaborações de mais dois professores: o Prof. Doutor Pedro Brauman na área da Teoria da Medida e da Probabilidade e o Prof. Doutor Bento Murteira que o apoiou na Estatística Matemática quando o Tiago se encontrava de serviço no Ministério da Educação, bem como em inúmeros doutoramentos e concursos.

Não se limitou o Tiago a criar apenas cadeiras e cursos, mas preocupou-se em criar também *meios indispensáveis* para haver *investigação e aperfeiçoamento* constante nos diversos domínios abrangidos pelos cursos, pois ele considerava absolutamente fundamental que tal acontecesse para haver qualidade de ensino.

Deste modo, teve o cuidado de fundar uma *Biblioteca* de livros e revistas em Probabilidade, Estatística Investigação Operacional e Computação -que hoje tem mais de 10.000 obras da especialidade.

Formou também um Centro de Matemática Aplicada ainda na década de 60, que posteriormente passou a ter o nome de Centro de Estatística e Aplicações e que tinha sobretudo a missão de incentivar a investigação em Probabilidades e Estatística bem como nas suas Aplicações, mediante a realização de seminários, trabalhos de orientação de doutoramentos e actividades de intercambio científico.

Muito haveria a dizer sobre a influência do Centro de Estatística na investigação em Estatística em Portugal, mas tal alongaria demasiado esta síntese sobre o tema proposto.

Basta dizer que ainda hoje existe esse Centro que dispõe já de cerca de 50 investigadores e que tem obtido as melhores classificações na qualidade dos serviços de formação e investigação prestados.

Finalmente foi o grande obreiro da fundação em 1981 da Sociedade Portuguesa de Estatística e Investigação Operacional e seu primeiro Presidente, tendo organizado os

primeiros colóquios e congressos a nível nacional, mas alguns já com participação de cientistas estrangeiros.

Hoje esta Instituição deu lugar a duas Sociedades: a Sociedade Portuguesa de Estatística e a Sociedade Portuguesa de Investigação Operacional, que realizam os seus congressos anuais em Portugal e com projecção internacional. X

Resta dizer que a Faculdade de Ciências de Lisboa é, ainda hoje, a única instituição de ensino e investigação onde existe um Departamento de Estatística e Investigação Operacional com quadro próprio, embora haja quem procure transformá-lo em mais um Departamento de Matemática Aplicada. Com efeito nas outras Universidades do País existem diversas disciplinas de Estatística e Investigação Operacional mas integradas em Departamentos de Matemática ou de Matemática Aplicada. X

Podemos dizer que praticamente todos os que hoje estão à frente do Departamento de Estatística e Investigação Operacional da F.C.U.L foram seus alunos e mais não têm feito que continuar a sua labuta em prol do ensino e investigação da Estatística e da Investigação Operacional. X

O DEIO é autónomo, com duas secções: a de Estatística e a de Investigação Operacional. Também deu origem a um outro Departamento da FCUL, o Departamento de Informática, por separação da componente de computação do anterior DEIOC.

O D.E.I.O dispõe hoje de cerca de 40 docentes e presta serviço aos outros departamentos da F.C.U.L nas diversas disciplinas do foro da Estatística, Probabilidade e Investigação Operacional que fazem parte dos curricula dos seus diversos cursos.

## Referências Bibliográficas

- “Faculdade de Ciências da Universidade de Lisboa: Passado, Presente e Perspectivas Futuras” 150 anos da Escola Politécnica e os 75 anos da Faculdade de Ciências de Lisboa. 1986, Lisboa Museu da Ciência
- “Aspectos Históricos da Estatística” 1981 Instituto Nacional de Estatística
- “História da Estatística em Portugal” Fernando Pereira Sousa, 1996, Instituto Nacional de Estatística

### **Maria de Fátima Fontes de Sousa**

Licenciada em Matemática e Doutorada em Matemática Aplicada pela Universidade de Lisboa. Foi Professora Catedrática na Faculdade de Ciências de Lisboa, onde fez toda a sua carreira universitária. Actualmente encontra-se jubilada. No decurso dessa carreira foi co-fundadora, juntamente com o Prof. J. Tiago de Oliveira, do 1º Departamento de Estatística e Investigação Operacional em Portugal, em 1981, do 1º Centro de Estatística e Aplicações da F.C.U.L. (1979) e da Sociedade Portuguesa de Estatística.

Inaugurou, em Portugal, o ensino regular de disciplinas do âmbito da estatística, tais como Processos Estocásticos, Simulação, Planeamento de Experiências, Programação Matemática e Amostragem entre outras.

Escreveu quatro livros - "Análise Numérica e Cálculo Automático", "Probabilidades e Estatística: um curso básico", "Amostragem: uma introdução" e "Processos Estocásticos Aplicados" sendo os dois primeiros publicados pela Escolar Editora e os dois últimos pela Universidade Aberta.

## “Extremistas” num extremo da Europa

*Maria Ivette Gomes*

Quando terminei a minha licenciatura em *Matemática Pura*, especialidade de *Álgebra*, na Faculdade de Ciências de Lisboa, e decidi enveredar pela área das “*Probabilidades, Estatística e Processos Estocásticos*”, comecei por trabalhar em *Testes de Aleatoriedade e Metodologias Não-paramétricas*, sob a orientação do Professor J. Tiago de Oliveira, já na altura referência internacional na área da *Estatística de Extremos*. Durante o período que permaneci em Portugal como assistente, e apesar de não ter nessa fase trabalhado directamente na *Teoria de Valores Extremos*, apercebi-me de algumas das potencialidades e beleza desta área. Tomada a decisão de ir trabalhar para Doutoramento em Inglaterra, em *Metodologias Não Paramétricas*, proporcionou o acaso ser professor em Sheffield Clive Anderson, o qual conhecia bem Tiago de Oliveira, uma vez que tinha defendido a sua tese de Doutoramento em 1971, sobre “*Contributions to the Asymptotic Theory of Extreme Values*”. Foi assim que acabei por enveredar, quase em “full-time”, nesta área. A trabalhar em Sheffield para Doutoramento, também numa área muito próxima, mas sob a orientação de Morris Walker, encontrava-se na altura Feridun Turkman, actualmente Professor Catedrático no Departamento de Estatística e Investigação Operacional (DEIO), tal como eu.

Terminei a minha tese de Doutoramento, “*Some Probabilistic and Statistical Problems in Extreme Value Theory*”, em 1978. Em 1980 terminava a sua tese de Doutoramento, “*Limiting Distributions of Maxima of Certain Types of Non-Stationary Stochastic Processes*”, Feridun Turkman. Os dois, em conjunto com Tiago de Oliveira, propusémos a organização de um “*NATO Advanced Study Institute*” (ASI) em “*Statistical Extremes and Applications*”, que se realizou no Vimeiro, em Setembro de 1983. Participaram nesse ASI reconhecidos nomes da área de Extremos, de que recordo alguns: Clive Anderson (Sheffield University, UK), Paul Deheuvels (Université Paris VI, France), Benjamin Epstein (Technion, Israel), Janos Galambos (Temple University, USA), Laurens de Haan (Erasmus University of Rotterdam, The Netherlands), A.M. Hasofer (University of New South Wales, Australia), Ross Leadbetter (University of North Carolina, USA), Jef Teugels (Katholieke Universiteit Leuven, Belgium), Ishay Weissman (Technion, Israel), entre outros. Tiveram também a oportunidade de participar neste ASI alguns alunos do primeiro Mestrado ministrado pelo então *DEIOC* (Departamento de Estatística,

Investigação Operacional e Computação), de que destaco a Teresa Alpuim, a Isabel Barão e a Fátima Miguéns.

Até meados de 80 a *Estatística de Extremos* que eu conhecia era de índole puramente paramétrica, fundamentada no teorema limite extremal. Nessa altura, em Portugal, os trabalhos de investigação na área eram quase exclusivamente os trabalhos de Tiago de Oliveira, que já tinha, desde o início dos anos 70, uma obra vasta, reconhecida internacionalmente. Os temas em que desenvolvi trabalho, numa fase inicial de carreira pós-doutoramento, foram:

1. Comportamento pré-assintótico em Teoria de Valores Extremos.
2. Concomitantes de estatísticas ordinais.
3. Inferência estatística em modelos de extremos multivariados e multi-dimensionais.

Também pioneira nesta área, com uma tese de Doutoramento defendida na Universidade de Paris VI, em 1987, sobre o tema “*Encadrement Presque Sûr des Statistiques d’Ordre*” e sob a orientação de Paul Deheuvels, é a Margarida Brito, actualmente Professora Associada da Universidade do Porto. A primeira aluna a doutorar-se em Portugal, na área de Extremos, penso ter sido a Teresa Alpuim, que se doutorou em 1989, sob minha orientação, com uma tese intitulada “*Contribuições à Teoria de Extremos em Sucessões Dependentes*”, defendida na Universidade de Lisboa. A sua tese deu origem a trabalhos seminais na área de *Extremos em Sucessões Dependentes*, de que não consigo deixar de referir dois, publicados em 1988, um intitulado “An extremal Markovian sequence”, *J. Applied Probab.* **26**, 219-232, e outro intitulado “High level exceedences in stationary sequences with extremal index”, *Stochastic Processes* **30**, 1-16. Quase em simultâneo, em 1990, doutorou-se a Manuela Neves, actualmente Professora Catedrática da Universidade Técnica de Lisboa. A sua tese, intitulada “*Estimação por Blocos dos Parâmetros da Distribuição de Fréchet — Comparação de Métodos Expeditos*”, foi defendida na Universidade Nova de Lisboa, sob a orientação de Tiago de Oliveira.

Foram também minhas alunas de Doutoramento a Luísa Canto e Castro e a Isabel Fraga Alves, que concluíram as suas teses em 1992, sobre “*Velocidades de Convergência em Teoria de Valores Extremos*” e “*Inferência Estatística em Modelos Extremais*”, respectivamente. Concluiu também douramento em 1992 a Fernanda Oliveira, com uma tese sobre “*Leis Limites em Sucessões Dependentes de uma Cadeia*”, orientada por Feridun Turkman.

O Professor Tiago de Oliveira só tardiamente decidiu investir na orientação de alunos de Doutoramento na sua área de eleição, e penso que, infelizmente, devido à sua morte prematura em 1992, só chegou a presenciar o Doutoramento da Manuela Neves. Duas outras alunas de Doutoramento de Tiago Oliveira, nesta área, a Isabel Barão e a Teresa Themido Pereira, terminaram as suas teses de Douramento sobre “*Comparação de Populações de Gumbel*” e “*Contribuições à Teoria de Valores Extremos*”, já sob minha orientação, em 1993 e em 1994, respectivamente. O quarto aluno de Tiago de Oliveira na

área de *Extremos*, a M. Fátima Miguéns, trabalhava numa área muito perto da área de investigação por excelência de Tiago de Oliveira, a área de *Extremos Multivariados* e, com um percurso de investigação mais atribulado, só muito recentemente, já em 2004, concluiu a sua Tese de Doutoramento, sob a orientação de M. Fátima Fontes de Sousa.

A partir de 1994 e até finais de 2003, data a que reporto este documento, os doutoramentos na área de *Extremos*, e temas relacionados, continuaram a bom ritmo:

1. Helena M. Simões Ferreira obteve o grau de Doutor (Universidade de Coimbra), em Junho de 1994, com uma tese intitulada "*Condições de Dependência Local em Teoria de Valores Extremos*", por mim orientada.
2. M. Emília Athayde obteve o grau de Doutor (Universidade de Lisboa), em Outubro de 1994, com uma tese intitulada "*Estudo de Algumas Sucessões Markovianas com Relevância para a Teoria de Extremos*", também sob minha orientação.
3. João Gomes obteve o grau de Doutor (Universidade de Lisboa), em Julho de 1996, com uma tese intitulada "*Valores Extremos e Alguns Processos Markovianos*", sob a orientação de M. Teresa Alpuim. Trata-se pois do meu primeiro "neto" científico.
4. M. Adelaide Valente de Freitas obteve o grau de Doutor (Universidade de Aveiro), em Julho de 1998, com uma tese intitulada "*Nova Classe de Aproximações em Teoria de Valores Extremos*", sob minha orientação.
5. Andreia Oliveira Hall obteve o grau de Doutor (Universidade de Lisboa), em Julho de 1998, com uma tese intitulada "*Extremos de Sucessões de Contagem — Do Outro Lado do Espelho*", sob minha orientação.
6. M. Fátima Brilhante obteve o grau de Doutor (Universidade dos Açores), em Dezembro de 1999, com uma tese intitulada "*Inferência Estatística em Modelos Não-Gaussianos com Recurso a Spacings e Outras Funções de Estatísticas Ordinais*", sob orientação de Dinis Pestana.
7. Sandra Mendonça obteve o grau de Doutor (Universidade da Madeira) em 2000, com uma tese intitulada "*Tópicos Sobre a Convergência Fraca de Sucessões de Variáveis Aleatórias*", sob a orientação de Dinis Pestana.
8. M. Graça Temido obteve o grau de Doutor (Universidade de Coimbra) em Novembro de 2000 com uma tese intitulada "*Classes de Leis Limites em Teoria de Valores Extremos — Estabilidade e Semi-estabilidade*", sob minha orientação, conjunta com Luísa Canto e Castro.
9. M. João Martins obteve o grau de Doutor (Universidade de Lisboa) em Março de 2001 com uma tese intitulada "*Estimação de Caudas Pesadas — Variantes ao Estimador de Hill*", sob minha orientação, conjunta com M. Manuela Neves.
10. Manuel Scotto obteve o grau de Doutor (Universidade de Lisboa) em 2001, com uma tese intitulada "*On the Extremes of Certain Transformations of Time Series*", sob a orientação de Feridun Turkman.

11. Bruno Cecílio de Sousa obteve o grau de Doutor (The University of Michigan), em Outubro de 2002, com uma tese intitulada “*A Contribution to the Estimation of the Tail Index of Heavy-Tailed Distributions*”, sob a orientação de Bruce Hill e G. Michailidis.
12. M. Luísa Pereira obteve o grau de Doutor (Universidade da Beira Interior), em Julho de 2002, com uma tese intitulada “*Valores Extremos Multidimensionais de Variáveis Dependentes*”, sob a orientação de Helena Ferreira.
13. Ana Ferreira obteve o grau de Doutor (Universidade de Tilburg), em Novembro de 2002, com uma tese intitulada “*Statistics of Extremes. Estimation and Optimality*”, sob a orientação de Laurens de Haan e John Einmahl.
14. Patrícia Bermudez obteve o grau de Doutor (Universidade de Lisboa), em Fevereiro de 2003, com uma tese intitulada “*Bayesian Approach to Extreme Quantile Estimation*”, sob a orientação de Feridun Turkman.
15. Fernanda Otilia Figueiredo obteve o grau de Doutor (Universidade de Lisboa), em Março de 2003, com uma tese intitulada “*Controlo Estatístico da Qualidade — Métodos Robustos*”, sob minha orientação, e em que também surgem resultados da área das *Estatísticas Ordiniais e Extremos*.
16. Orlando Oliveira (falecido em Junho de 2003) obteve o grau de Doutor (Universidade de Lisboa), em Maio de 2003, com uma tese intitulada “*In Extremis*”, sob minha orientação.
17. Alexandra Ramos obteve o grau de Doutor (Universidade de Surrey), em Fevereiro de 2003, com uma tese intitulada “*Multivariate Joint Tail Modelling and Score Tests of Independence*”, sob a orientação de Anthony Ledford.
18. Alexandra Dias obteve o grau de Doutor (ETH, Zurique), em Dezembro de 2003, com uma tese intitulada “*Copula Inference for Finance and Insurance*”, sob a orientação de Paul Embrechts.

Foram também cerca de meia centena os Mestrados e Provas de Aptidão Pedagógica e Científica na área de extremos ou relacionada.

A partir de um núcleo duro de alguns dos doutores já referidos, o grupo de investigação em *Extremos* tem crescido de forma salutar. Também outros probabilistas portugueses, como D. Pestana, M. E. Graça Martins, H. Iglésias Pereira, entre outros, contribuíram para o desenvolvimento de extensões do modelo clássico, sem que esta área seja central no seu trabalho. Até finais de 2003, e incluindo a minha e a de Kamil Feridun Turkman, foram 28 as teses de Doutoramento na área (13 das quais nos últimos 5 anos — de 1999 a 2003), escritas por investigadores portugueses e a investigar actualmente em Portugal, embora não necessariamente na área de Extremos. A dinâmica de publicação tem sido elevada, e considero-a nitidamente acima dos padrões médios em Portugal. Embora a quantidade não signifique obviamente qualidade, atrevo-me a avançar com algumas das publicações em revistas internacionais, no período 1999-2003:

1999

1. Barão, M. I. and Tawn, J. A. "Extremal analysis of short series with outliers: sea-levels and athletic records". *Applied Statistics* **48**, 469-488.
2. Ferreira, H. "Limit distributions for point processes of exceedances of random levels". *Test* **8**:1, 191-200.
3. Fraga Alves, M. I. "Asymptotic distribution of Gumbel statistic in a semi-parametric approach". *Portugaliae Mathematica* **56**:3, 282-298.
4. Gomes, M. I. "Generalized Jackknife Moment estimator of the tail index". *Bull. of the International Statistical Institute* **58** (1), 401-402.
5. Gomes, M. I. "Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields, by R.-D. Reiss and M. Thomas. Birkhauser, 1977 (Book Review)". *Extremes* **2**:1, 111-113.
6. Gomes, M. I. and de Haan, L. "Approximations by extreme value distributions". *Extremes* **2**:1, 71-85.
7. Martins, M. J., Gomes, M. I. and Neves, M. "Some results on the behaviour of Hill's estimator". *J. Statist. Comp. and Data Analysis* **63**, 283-297.

2000

1. Canto e Castro, L., de Haan, L. and Temido, M. G. "Rarely observed sample maxima". *Th. Probability and Applications* **45**:4, 658-662.
2. Draisma, G., de Haan, L., Peng, L. and Pereira, T. T. "A Bootstrap-based method to achieve optimality in estimating the extreme-value index". *Extremes* **2**:4, 367-404 (1999).
3. Ferreira, H. "A note on extremes of concomitants of order statistics". *Extremes* **3**:4, 385-392.
4. Gomes, M. I., Martins, M. J. and Neves, M. "Alternatives to a semi-parametric estimator of parameters of rare events — the Jackknife methodology". *Extremes* **3**:3, 207-229.
5. Pereira, L. and Ferreira, H. "Limit distribution for point processes of high local maxima". *J. Statistical Planning and Inference* **97**, 227-233.
6. Scotto, M. G. and Soares, C. G. "Modelling the long-term time series of significant wave height with non-linear threshold models". *Coastal Engineering* **51** (5-6), 387-394.
7. Temido, M. G. "Mixture results for extremal behaviour of strongly dependent nonstationary Gaussian sequences". *Test* **9**:2, 439-453.

## 2001

1. Barne-Delcroix, M. F. and Brito, M. "Multidimensional stability and strong limiting behaviour of intermediate order statistics". *Multivariate Analysis* **79**, 157-170.
2. Brannas, K. and Hall, A. "Estimation in integer-valued moving average models". *Applied Stochastic Models in Business and Industry* **17**: 3, 277-291.
3. Fraga Alves, M. I. "Weiss-Hill Estimator". *Test* **10**:1, 203-224.
4. Gomes, M. I. and Martins, M. J. "Generalizations of the Hill estimator — asymptotic versus finite sample behaviour". *J. Statist. Planning and Inference* **93**, 161-180.
5. Pereira, L. and Ferreira, H. "The asymptotic locations of the maximum and minimum of stationary sequences". *J. Statistical Planning and Inference* **104**, 287-295.
6. Turkman, K. F. "Extremal Behaviour of Trigonometric Polynomials with Random Coefficients". *Nonlinear Analysis* **47**, 3113-3124.

## 2002

1. Bermudez, P., Amaral Turkman, A. and Turkman, K. "A Predictive Approach to Tail Probability Estimation". *Extremes* **4**:4, 295-314 (2001).
2. Caeiro, C. and Gomes, M. I. "A class of "asymptotically unbiased" semi-parametric estimators of the tail index". *Test* **11**:2, 345-364.
3. Caeiro, F. and Gomes, M. I. "Bias reduction in the estimation of parameters of rare events". *Theory of Stochastic Processes* **8** (24), 1-2, 67-76.
4. Ferreira, A. "Optimal asymptotic estimation of small exceedance probabilities". *J. Statist. Planning and Inference* **104**, 83-102.
5. Ferreira, H. and Scotto, M. G. "On the asymptotic location of high values of a stationary sequence". *Statistics and Probability Letters* **60**, 475-482.
6. Fraga Alves, M. I. "A location invariant Hill-Type estimator". *Extremes* **4**:3, 199-217 (2001).
7. Gomes, M. I. and Martins, M. J. "Asymptotically unbiased estimators of the tail index based on external estimation of the second order parameter". *Extremes* **5**:1, 5-31.
8. Gomes, M. I., Martins, M. J. and Neves, M. "Generalized Jackknife semi-parametric estimators of the tail index". *Portugaliae Mathematica* **59**:4, 393-408.
9. Gomes, M. I. and Oliveira, O. "The bootstrap methodology in Statistics of Extremes — choice of the optimal sample fraction". *Extremes* **4**:4, 331-358 (2001).
10. Gomes, M. I., de Haan, L. and Peng, L. "Semi-Parametric Estimation of the Second Order Parameter — Asymptotic and Finite Sample Behaviour". *Extremes* **5**:4, 387-414.
11. Hall, A. "Extremes of moving average models with regularly varying tails". *Extremes* **4**:3, 219-239 (2001).

12. Scotto, M. G. and Turkman, K. F. "On the Extremal Behaviour of Sub-Sampled Solutions of Stochastic Difference Equations". *Portugaliae Mathematica* **59**, 267-282.

2003

1. Bermudez, P. de Zea and Amaral Turkman, M. A. "Bayesian Approach to Parameter Estimation of the Generalized Pareto Distribution". *Test* **12**:1, 259-277.
2. Brito, M. and Moreira Freitas, A. C. "Limiting behaviour of a geometric-type estimator for tail indices". *Insurance: Mathematics and Economics* **33**, 211-216.
3. Dias, A. and Embrechts, P. "Change point analysis for dependence". *Insurance Mathematics and Economics* **32** (1), 152-152.
4. Ferreira, H. "Extremes of associated variables". *Statistics and Probability Letters* **63**, 333-338.
5. Ferreira, H. and Martins, A. "The extremal index of sub-sampled periodic sequences with strong local dependence". *RevStat* **1**, 15-24.
6. Ferreira, A., de Haan, L. and Peng, L. "On optimizing the estimation of high quantiles of a probability distribution". *Statistics* **37**, 401-434.
7. Fraga Alves, M. I. "Estimation of first and second order parameters in heavy tails". *Insurance Mathematics and Economics* **32** (1), 153-153.
8. Fraga Alves, M. I., Gomes, M. I. and de Haan, L. "A new class of semi-parametric estimators of the second order parameter". *Portugaliae Mathematica* **60**:2, 193-213.
9. Fraga Alves, M. I., de Haan, L. and Tao Lin. "Estimation of the parameter controlling the speed of convergence in Extreme Value Theory". *Mathematical Methods of Statistics* **12** (2), 155-176.
10. Gomes, M. I. and Oliveira, O. "Censoring estimators of the tail index". *Statistics and Probability Letters* **65**, 147-159.
11. Gomes, M. I. "Bias reduction in financial risk modelling". *Bull. Intern. Statist. Inst.* **LX**, 1:1, 165-168.
12. Gomes, M. I. and Oliveira, O. "How can non-invariant statistics work in our benefit in the semi-parametric estimation of parameters of rare events". *Comm. in Statist.— Simulation and Computation* **32**:4, 1005-1028.
13. Gomes, M. I. and Oliveira, O. "Maximum likelihood revisited under a semi-parametric context — estimation of the tail index". *J. Statist. Computation and Simulation* **73**:4, 285-301.
14. Hall, A. and Scotto, M. G. "Extremes of sub-sampled integer-valued moving average models with heavy-tailed innovations". *Statistics and Probab. Letters* **63** (1), 97-105.
15. Scotto, M. G. "A note on the asymptotic distribution of the maxima in disaggregated time-series models". *Statistics and Probab. Letters* **65** (2), 127-137.
16. Scotto, M., Anderson, C. and Turkman, K. F. "Extremes of Sub-sampled Series". *Journal of Time Series Analysis* **24**:5, 505-512.

17. Scotto, M. G. and Ferreira, H. "Extremes of deterministic sub-sampled moving averages with heavy-tailed innovations". *Applied Stochastic Models in Business and Industry* **19**, 303-313.
18. Temido, M. G. "New limiting distribution of maxima of independent random variables". *Portugaliae Mathematica* **60**:4, 379-388.
19. Temido, M. G. and Canto e Castro, L. "Max-semistable laws in extremes of stationary random sequences". *Theory Probab. and its Applications* **47**:2, 365-374.
20. Themido Pereira, T. "Spacial extremes: the stationary case". *Bull. Intern. Statist. Inst.* **LX**, 2:2, 249.

Laurens de Haan aparece como co-autor de algumas das mais significativas publicações acima arroladas. A permanência prolongada deste eminente "extremista" entre nós — com uma "Gulbenkian Professorship" em 2005, que elegeu a FCUL como local principal de trabalho, desde que se retirou da docência activa na Universidade Erasmus de Roterdão — teve decerto uma influência profunda no desenvolvimento deste campo de investigação em Portugal.

Convém ainda registar que o *Grupo de Extremos*, apesar da elevada contribuição a nível internacional, não tem descurado a publicação a nível nacional, como pode ser atestado pelas publicações de artigos em português, nas diferentes colectâneas de texto editadas pela *Sociedade Portuguesa de Estatística*, desde 1992. E refiro apenas as Actas dos onze Congressos Anuais da Sociedade Portuguesa de Estatística, onde a produção na área de *Extremos* tem sido, em média, de 17% (8 trabalhos) ao ano.

Existem actualmente, tanto quanto sei, nove alunos de Doutoramento na área, que apresentamos por pretensa ordem de obtenção do grau. Três deles deverão defender as suas teses em 2005: M. Cristina Miranda, sobre "*Estimação dos Índices Extremal e de Cauda*", Ana Paula Martins, sobre "*Coefficientes Extremais*" e Ana Cristina Moreira de Freitas, sobre "*Estimação do Coeficiente de Cauda Exponencial. Aplicações à Teoria de Risco*". Os outros seis estudantes de Doutoramento na área de *Extremos* são a Cláudia Neves, a Sandra Dias, o Frederico Caeiro, a Dora Prata Gomes, a Marta Ferreira e a Lígia Henriques. Consigo ainda contabilizar 6 alunos de Mestrado, a iniciar trabalho na área de *Extremos*, o que promete o alargamento do grupo num futuro próximo.

Actualmente, são muito diversificados os temas da área de extremos investigados em Portugal. Para além de um grupo forte a trabalhar na área de *Estimação Semi-paramétrica de Parâmetros de Acontecimentos Raros*, com trabalhos pioneiros na área de estimadores invariantes, redução de viés com a utilização da metodologia Jackknife, alisamentos através de combinações lineares adequadas, utilização da metodologia "bootstrap" na escolha adaptativa da fracção óptima da amostra, temos ainda grupos fortes nas vertentes de *Extremos e Modelação de Risco*, *Extremos e Ambiente*, *Extremos de Sucessões Dependentes Univariadas*, *Multivariadas e Multi-Dimensionais* e *Extremos Espaciais*. É pois de crer que o dinamismo do grupo proporcionará um desenvolvimento salutar da área, com reconhecimento internacional de uma *Escola de Extremos* em Portugal.

## **Maria Ivette Gomes**

É Professora Catedrática no Departamento de Estatística e Investigação Operacional (DEIO), Faculdade de Ciências da Universidade de Lisboa, desde 1985.

Foi Presidente da Sociedade Portuguesa de Estatística (SPE) de 1990 até 1994, Presidente do DEIO de 1995 a 1997 e tem sido Directora do Centro de Estatística e Aplicações (CEAUL) desde Julho de 1999.

Foi Editora Associada da revista *Portugaliae Mathematica* de 1994 até 2001 e é Editora-Principal da revista *RevStat*, editada pelo INE, e com início em 2003.



# Os Primeiros 25 Anos, Alicerces do Futuro

*Dinis Duarte Pestana*

Há alguns anos, o saudoso António Simões Neto, com mais alguns carolas, procuraram insuflar vida nova numa sociedade de estatística que tinha vindo a apagar-se, depois de alguma força inicial, encarregando-se nomeadamente de algumas questões estatutárias desajustadas.

Do anterior formato da Sociedade mal me lembro; houvera congressos no Fundão, em Lagos, em Tróia, mas nesse tempo eu era ainda novo de mais para arrecadar muitas coisas na memória. Lembro-me que foi no Fundão que conheci o João Branco, em Lagos que conheci a Nazaré Mendes Lopes, em Tróia que conheci a Corália Vicente e o José Maia, quando uma nova fase da Estatística, ou das aplicações da Estatística, começava a surgir em diversas universidades. O motor dos congressos do Fundão e de Lagos foi naturalmente J. Tiago de Oliveira, e o de Tróia foi uma organização liderada por Antónia Amaral Turkman e Lucília Carvalho, dessa vez no âmbito das actividades do CEAUL e do DEIO, e não da SPEIO. Neles a Investigação Operacional teve um papel interessante, mas que possivelmente evidenciou que os interesses comuns eram poucos. Entre os congressos a actividade da Sociedade era quase invisível.

Mesmo assim, aquela Sociedade (SPEIO) desempenhou um papel importante. Até 1980, tirando os colegas mais próximos, da FCUL ou colaboradores do projecto ML4 que esteve na origem do Centro de Estatística e Aplicações da Universidade de Lisboa, sabia vagamente da actividade do Doutor Murta, na Universidade de Coimbra. Em Lisboa, o Professor Tiago de Oliveira tinha uma personalidade dominante, e o Professor Bento Murteira uma simpatia irradiante, acompanhada de uma curiosidade sadia por tantos campos e aplicações, que depressa o tornava, a par de António Simões Neto, um interlocutor irresistível. E havia a obra notável, em profundidade e erudição, de Pedro Braumann. Não eram os únicos mas, sendo de outra geração, eram os modelos.

Por razões que têm uma quota de mistério, um dia foi necessário abandonar os modelos, e a geração seguinte, a que pertença, assumir transitoriamente o seu protagonismo. Foi nessa fase que a Sociedade se recriou como SPE, augurando alguma esperança num futuro mais pacífico.

Mas não se deve omitir que os congressos da anterior Sociedade mista de Estatística e Investigação Operacional proporcionaram um ambiente amistoso em que alargámos o

círculo de conhecidos e de amigos; sem isso, provavelmente a SPE não poderia ter existido.

Passaram vinte e cinco anos desde a primeira fundação (ena, quase parece o clássico de Asimov, colocando um desenvolvimento da Probabilidade — a Psico-história — no cerne da organização do futuro), e o Presidente da SPE convidou-me a particiar num balanço do que foi feito. Começarei por referir o papel da SPE na evolução dos campos por que me interesseo.

Depois, como SPE evoca espectáculo, deter-me-ei no que actualmente dá maior visibilidade à Sociedade Portuguesa de Estatística. Mas SPE aponta também esperança, e num balanço também há que referir o que foi adiado, ou ficou por fazer, na esperança que daqui a vinte e cinco anos esses aspectos façam parte do activo do balanço. Os alicerces estão prontos, há que construir o futuro — o que só será possível com uma participação mais activa de todos, porque os órgãos estatutários da SPE são o motor, o agenciador de facilidades para o desenvolvimento, que é, este, responsabilidade de todos nós.

### **Desenvolvimento de algumas áreas da Estatística nos últimos 25 anos**

Nos 25 anos que decorreram desde a fundação da primeira Sociedade até hoje, o desenvolvimento da Estatística em Portugal foi notável. Provavelmente sou um dos sócios fundadores só porque todos os estatísticos (e não só) ligados ao ensino superior foram sócios fundadores. Nessa altura, se os dedos das mãos não chegavam para contar os cultores da Estatística entre nós, os dedos das mãos e dos pés, para quem lá chegue, já deviam bastar. Hoje, centenas de participantes trocam ideias nos congressos anuais da SPE, sendo necessárias várias sessões paralelas para acomodar as muitas comunicações interessantes que anualmente são submetidas.

### **Que evolução houve, nestes 25 anos, nas áreas que me interessam?**

A *Probabilidade*, cujo ensino nas universidades portuguesas tem declinado, porque raros são os alunos com interesse genuíno por questões mais abstractas, tem tido um papel interessante nas actividades da SPE. Muitos trabalhos publicados nas Actas dos Congressos inscrevem-se, de facto, na área da Probabilidade, embora visem em geral o desenvolvimento das ideias em outras áreas da Estatística. Alguns deles têm sido premiados; e assim a SPE tem contribuído, como lhe compete, para incentivar o desenvolvimento desta área fundamental. Mas seria porventura de recomendar um investimento mais visível nesta área, por exemplo um curso breve na área de Probabilidade num dos próximos congressos, ou um melhor aproveitamento de um dos maiores especialistas do campo, o Professor Yurinski, se ter radicado em Portugal.

A *Amostragem* foi objecto de um curso breve, e no Congresso de Guimarães o Professor Paulo Gomes organizou um debate sobre sondagens que teve grande êxito. Mas em minha opinião os estatísticos portugueses continuam a ter um interesse marginal pela questão da recolha da informação, e a Amostragem e o Planeamento de Experiências têm

tido um papel muito menor do que o que lhes seria devido. A SPE poderia ter uma intervenção importante nesta questão, nomeadamente pedindo aos *referees* dos trabalhos na área da utilização da Estatística que avaliassem cuidadosamente a questão do protocolo da recolha dos dados — sendo, naturalmente, previamente divulgada esta política editorial.

Poderia também a SPE ter um papel cultural importante, agindo junto de outras sociedades científicas e de autoridades académicas, chamando a atenção para o papel ímpar desempenhado pela Estatística na metodologia da investigação científica, em que as áreas de Amostragem e de Planeamento de Experiências têm que passar a gozar de uma autoridade que lhes tem vindo a ser negada por incultura.

O bom senso dos dois últimos presidentes da SPE desencadeou uma aproximação com os produtores de *Estatísticas Oficiais* que a meu ver deve ser aprofundada: a semente foi lançada, mas poucos membros da SPE têm cultivado este terreno fértil — recorde-se que D. R. Cox, no seu lúcido trabalho quando deixou a presidência do *ISI*, indicava a aproximação entre a estatística matemática e as estatísticas oficiais como um dos investimentos importantes por fazer na investigação estatística.

A *Bioestatística* em Portugal continua incipiente. Há muitos trabalhos nas áreas de Biologia e Biomedicina em que se recorre a argumentos estatísticos e a análise de dados, mas a produção de trabalho de fundo em Bioestatística — a própria percepção do que distingue esta área da Estatística *tout court* — é insuficiente.

Esta análise necessariamente rápida parece pouco optimista. De facto, a evolução da Estatística em Portugal nos últimos 25 anos foi notável, mas até certo ponto decepcionante. Um estatístico que queira fazer investigação deve ser um bom Matemático, e não tem sido fácil captar para a Estatística muitos jovens de talento naquela área, podendo mesmo ser questionada a clarividência de, no plano do ensino universitário, em algumas instituições se ter procedido à separação das duas áreas (claro que numa sociedade saudável os investigadores são uma parte diminuta dos profissionais de uma determinada área, e para a formação de estatísticos profissionais aquela separação pode ter sido uma política correcta). Um estatístico, hoje, deve dominar razoavelmente bem várias disciplinas de Informática, nomeadamente o que se prende às novas áreas de exploração e análise de grandes bancos de dados. E deve ter uma formação ecuménica, ser capaz de se interessar pelos assuntos mais diversos, ter um entendimento razoável dos problemas que são o motor de novos desenvolvimentos da Estatística. Ora novos campos de investigação, concorrentes da Estatística, ganharam ímpeto nos últimos 25 anos, e nós, os da SPE, não tivemos as iniciativas necessárias para continuar a captar para a Estatística os novos talentos, com aquela multimodalidade de valências, que são necessários para o seu progresso. Assim, sendo embora o desenvolvimento da Estatística em Portugal florescente relativamente ao que era há 25 anos, a meu ver estiolou face ao que se anunciava.

O papel da SPE no desenvolvimento das áreas referidas é comparável ao que operou em todas as outras áreas de Estatística: apoiou todos os projectos que lhe foram submetidos, e se mais não houve foi por falta de iniciativa da comunidade estatística, em que me incluo. Estimulou equilibradamente todos os campos, procedendo a uma escolha

judiciosa de oradores convidados nos congressos, e proporcionando cursos breves no período ante-congresso. Poderá, eventualmente, no futuro, haver um diagnóstico mais fino de áreas a apoiar e desenvolver, e acções mais concertadas — e sobretudo menos episódicas do que a realização do congresso anual — para atingir os objectivos propostos.

### Visibilidade da SPE

Para início oficial das suas actividades, a Sociedade Portuguesa de Estatística decidiu dar continuidade aos congressos da SPEIO. Uma alteração de monta (que na altura foi polémica): decidiu-se que passariam a ter periodicidade anual. Afinal foi possível, e a Sociedade Portuguesa de Estatística tem assim desempenhado um papel motor na criação da cultura estatística, e mesmo de vocabulário científico próprio, em Portugal. A decisão de apostar na edição de volumes, que não são propriamente actas no sentido em que há uma filtragem decorrente de um processo cada vez mais exigente de avaliação por pares, tem reforçado esse papel cultural, ainda que actualmente se possa pôr em causa se esse esforço de publicação não terá prejudicado a procura de publicação em foruns internacionais mais apropriados.

No primeiro desses congressos, realizado no Vimeiro, a SPE honrou-se elegendo como membro honorário o Professor Bento Murteira. Estas distinções não devem ser banalizadas, mas lamento que não tenha sido posteriormente concedido igual distinção aos Presidentes que serviram a SPE com tanta generosidade.

A actividade da SPE tem-se diversificado, e a Sociedade apoia muitos seminários em diversas instituições. No que respeita o incentivo aos jovens, para além da criação de um prémio científico cujo recipiente é desvendado no último dia do congresso, a SPE tem vindo a divulgar, em séries de seminários, os contributos dos recém-doutorados em Estatística. O *Boletim da Sociedade Portuguesa de Estatística* não tem tido a regularidade esperada, mostrando que a comunidade estatística portuguesa está menos activa do que seria desejável.

De qualquer modo, a visibilidade da SPE está actualmente ligada ao espectáculo anual do congresso. O “ano estatístico” divide-se em dois períodos: aquele em que se prepara a participação no congresso que se avizinha, e aquele em que se aprimora a apresentação escrita do trabalho, se espera a reacção dos *referees*, e se atende aos seus comentários.

É óbvio o que pode haver de perverso neste estado de coisas, sobretudo se atendermos à escassez de publicações internacionais em revistas reputadas por parte de muitos estatísticos portugueses que na SPE se revelam prolíficos.

Mas globalmente o saldo é positivo, porque tem sido estimulados a sensação de comunidade e os laços de empatia tão necessários para nos apreciarmos uns aos outros e para lançarmos ideias de colaboração. Num plano mais profundo, os congressos anuais têm contribuído de forma idendelével para a formação de cientistas, ensinando-os a apresentar as suas ideias, e também a avaliar as ideias dos outros. Os congressos anuais da SPE têm sido uma escola de criatividade, de exigência, e de avaliadores da criatividade. Se mais não fizesse a SPE, só por isto merecia o nosso reconhecimento.

## SPE e esperança no futuro

Os dirigentes da Sociedade Portuguesa de Estatística tiveram uma acção notável na afirmação da SPE junto da comunidade estatística. No que se refere á sociedade em geral, houve acções incipientes, a que convém dar continuidade, e porventura prioridade — para bem da própria sociedade portuguesa.

Em primeiro lugar, há que estruturar acções continuadas de divulgação da Estatística, e do seu papel na cidadania. Nunca é demais recordar as sábias palavras de Carlyle, sobre o papel desempenhado pela Estatística como travão às ideias feitas, aos preconceitos, e à tentativa de fazer passar mentiras repetidas por verdades. Mais do que da direcção da SPE, isto pede o empenhamento de um punhado de entusiastas que mantenham um *site* actualizado, incentivando a intervenção de todos nós para a discussão estatística de questões que preocupam a sociedade em que nos inserimos.

Há, por outro lado, que ir alterando paulatinamente os hábitos dos empresários, que não se preocupam suficientemente com o controle de qualidade, com a imagem, e que continuam a não se apreçoer que entregar tarefas nobres a amadores pode ter ocasionalmente excelentes resultados, mas não é o padrão em que se deve apostar se se pretender dar o salto qualitativo de que a nossa indústria e os nossos serviços precisam.

Há ainda que formar a opinião dos cientistas de outras áreas, que em geral pensam que um conhecimento pela rama e a acessibilidade de *software* permite prescindir de uma Estatística forte em Portugal. Uma acção continuada junto de outras sociedades científicas e de instituições financiadoras e de tutela tem que insistir no papel formativo da Estatística nas metodologias da investigação científica, e na mais-valia que uma formação mais aprofundada em Amostragem e Planeamento de Experiências, Estatística, e Análise de Dados, confere à formação pós-graduada em qualquer área.

A par disto, a SPE devia influir na normalização do ensino da Estatística a vários níveis, e apoiar a inflexão para áreas que se mantêm atrasadas entre nós, nomeadamente a Estatística Aplicada, e a nova área de *Intelligent Data Analysis*, que não traduzo por falta de uma expressão adequada<sup>1</sup>. Naturalmente, o apoio à formação de pequenas e médias empresas de consultores de Estatística é uma das acções a recomendar.

Cativar novos talentos é um investimento prioritário, porque dele depende o futuro. Pede por isso uma acção continuada, para que decerto não será difícil conseguir o empenhamento dos mais dotados para este tipo de tarefa, e financiamento de instituições que têm uma visão estratégica sobre o que é necessário para um salto qualitativo da competitividade das nossas empresas. Mas não é só a juventude que deve estar na mira da SPE, as alterações profundas da sociedade impõem o dever de preparar acções de educação contínua, que a SPE poderia vantajosamente coordenar.

Finalmente (felizmente ninguém me requer que seja exaustivo), sou favorável a uma divulgação pública pela SPE de uma carta deontológica da prática da Estatística, que possa defender quer os produtores de estatística quer os seus utilizadores.

<sup>1</sup> Uma das razões de ser das academias, e por extensão das sociedades científicas, é o de curar de questões lexicais. Já indicámos o papel de relevo que a SPE tem tido a este respeito, há que continuar e ir mais longe.

## **Dinis Duarte Pestana**

Licenciou-se em Matemática, depois de uma breve passagem pela Faculdade de Letras.

Doutorou-se em Probabilidade na Universidade de Sheffield, em 1978.

É Professor Catedrático do Departamento de Estatística e Investigação Operacional da FCUL.

Publicou cerca de quarenta artigos, nas editoras Wiley, Springer e North Holland, e em revistas internacionais; e tem extensa obra de divulgação nas áreas da Probabilidade e da Estatística.

É co-autor do manual universitário, *Introdução à Probabilidade e à Estatística*, publicado pela Fundação Calouste Gulbenkian em 2002.

Foi representante do Conselho de Reitores das Universidades Portuguesas no Conselho Superior de Estatística e integrou as Comissões de Avaliação das Universidades Portuguesas nas áreas de Matemática e de Informática/Matemática.

# O Acaso e a Vida

*Carlos A. Braumann*

## 1. Os 25 anos da Sociedade Portuguesa de Estatística

A Sociedade Portuguesa de Estatística (SPE) faz 25 anos. Parabéns! E bem merecidos porque hoje, graças em boa parte à sua acção, a Probabilidade e a Estatística em Portugal estão bem implantadas e com grande pujança. Os Congressos anuais, e assisti a todos, têm demonstrado um crescimento impressionante do trabalho de investigação e de aplicação de qualidade e dos seus cultores. E aqui a presença feminina tem sido, ao contrário da quase totalidade dos outros países, não apenas apreciável, mas maioritária. Contudo, há um perigo que nos espreita a prazo, o da redução dos jovens universitários interessados em prosseguir estudos nesta área do saber, fruto da demografia e, principalmente, do trauma da Matemática, que adquiriu obesidade patológica alimentada pelo clima social e décadas de pedagogia familiar e escolar facilitista. Daí a necessidade de, sem descurmarmos os nossos outros deveres, nos preocuparmos em levar aos jovens a mensagem da importância e da beleza da Probabilidade e da Estatística e, mais prosaicamente, da empregabilidade dos seus licenciados.

Mas, naturalmente, os parabéns são para a visão dos fundadores, para o labor intenso dos seus Presidentes, das suas Direcções e dos restantes órgãos sociais e, “last but not the least”, para o empenho dos sócios na vida da sociedade e no trabalho diário de cultores, construtores e divulgadores destes saberes.

## 2. Objectivos

Quando o Presidente da SPE, Professor Doutor Fernando Rosado, me convidou para escrever um texto para este volume comemorativo, convite que muito agradeço, aceitei entusiasmado. O trabalho que tinha já não era pouco mas não imaginava então a dimensão do acréscimo que ele iria sofrer pelo facto de presidir à Comissão Organizadora do XII Congresso Anual da SPE (Évora, 2004) e integrar a Comissão Editorial das respectivas Actas (além de outras tarefas que assumi). Devia imaginar, pois já tinha experiência de organização deste tipo de eventos, mas a memória é selectiva... O certo é que as matérias

inadiáveis foram postergando a escrita do texto e que fui falhando, contra o meu feitio de cumprir compromissos custe o que custar, os sucessivos prazos até que, perante o inevitável ultimato que o tornou agora inadiável, aqui vai o texto que foi possível fazer em poucos dias. Ele não é o que eu queria fazer porque o que eu queria fazer exigia muito mais tempo do que o que dispunha. Ele não é o que os leitores mereciam mas, neste momento, só tinha duas opções: ou fazer o que me era possível ou não fazer. A última estava-me vedada pelo compromisso assumido com o Presidente da SPE de que só ele me poderia libertar. O que não fez, porque compreensivelmente ... não podia fazer. Só espero que o leitor seja mais generoso para comigo do que eu próprio e me saiba perdoar.

Pelas razões acima referidas vou-me limitar, por uma questão de exequibilidade, a analisar o que foram estes 25 anos em Portugal na área das aplicações das equações diferenciais estocásticas (EDE) à dinâmica de populações de seres vivos. É a minha principal área de investigação e é um nicho especializado com um número razoável (embora não muito grande) de cultores a nível internacional mas com poucos a nível nacional, onde só nos últimos anos foi possível atrair alguns jovens investigadores<sup>1</sup>. Deste modo, necessariamente, o meu trabalho pessoal vai ter aqui uma presença preponderante, do que peço desculpa aos leitores. A minha intenção inicial pretendia abranger a área mais vasta dos Processos Estocásticos, o que exigia uma pesquisa intensa e o consequente tempo de que agora não disponho. Talvez noutra ocasião...

O meu propósito é dar uma visão panorâmica não pormenorizada, dirigida ao não especialista, do que se tem feito nesta área. Assim, não irei descrever em pormenor a actividade de investigação desenvolvida, o que seria fastidioso, mas apenas os resultados mais marcantes. Tentarei ser poupado nas referências bibliográficas, dando preferência às de língua portuguesa (particularmente às Actas dos Congressos da SPE), normalmente em detrimento das publicações iniciais dos resultados (frequentemente em língua inglesa). O leitor pode encontrar estas últimas referenciadas nas primeiras, que servem melhor o objectivo divulgatório deste texto.

Como é óbvio, as equações diferenciais estocásticas têm inúmeras outras aplicações nos mais variados ramos da Ciência e da Tecnologia. Elas nasceram da Física, mais concretamente na modelação do chamado movimento browniano de partículas suspensas em fluidos<sup>2</sup>, mas o seu tratamento matemático iniciou-se com Itô<sup>3</sup> em 1951. Também é óbvio que a influência do “acaso” na dinâmica de populações de seres vivos pode ser modelada, conforme as circunstâncias e os objectivos, de variadas maneiras (processos de ramificação, processos de nascimento e morte, processos semi-markovianos, etc.). A área em que me vou mover é, assim, a da intersecção de um tipo de modelos (as equações diferenciais estocásticas) com um tipo de aplicações (a dinâmica de populações de seres vivos que vivem num ambiente que inevitavelmente sofre flutuações aleatórias).

<sup>1</sup> Felizmente noutras áreas de aplicação das EDE conseguiu-se maior capacidade de atracção.

<sup>2</sup> Uhlenbeck, G. E. e Ornstein, L. S. (1930). On the theory of Brownian motion. *Physical Review* 36: 823-841. NOTA. A solução da EDE usada no modelo é conhecida por processo de Ornstein-Uhlenbeck.

<sup>3</sup> O trabalho pioneiro foi: Itô, K. (1951). On stochastic differential equations. *Mem. Amer. Math. Soc.* 4: 1-51.

No final farei algumas referências à divulgação em Portugal das equações diferenciais estocásticas.

### 3. Os acasos na vida de um jovem investigador

Como é que me cheguei a interessar por estas matérias? É uma longa história fruto de vários acasos que ocorreram na minha vida, e aí também o título deste texto se aplica com propriedade. Tinha chegado de Angola em Dezembro de 1974 e quis o “acaso” que arranjasse colocação na Universidade de Évora, onde iniciei funções em Maio de 1975 e que se viria a tornar a minha casa. Na Universidade de Évora predominavam então os ensinamentos de matérias agrícolas e ambientais e o Reitor, Professor Ário Lobo Azevedo, manifestou interesse em que eu trabalhasse em aplicações biológicas da Matemática, tema que também me atraía bastante. Na altura, o Professor José Tiago de Oliveira organizava na Faculdade de Ciências de Lisboa um seminário muito dinâmico de Probabilidades e Estatística. Reunia numa sala do piso inferior do anexo ao Observatório na Rua da Escola Politécnica, um edifício um pouco degradado mas com o enquadramento paisagístico do Jardim Botânico. Aí aprendi que as Probabilidades e Estatística eram um mundo muito mais vasto do que me tinha apercebido na licenciatura em Matemática Aplicada, apesar de esta ter para a época uma razoável componente desta área. E aprendi com os melhores como se fazia e apresentava o trabalho de investigação. Estavam lá os mais seniores da altura, todos aliás jovens de espírito, como o Professor Tiago de Oliveira (que já conhecera como Professor em Angola), o meu Pai ou a Professora Maria de Fátima Fontes de Sousa. Estavam lá muitos dos que são hoje os seniores da nossa comunidade científica, embora entre nós nos consideremos ainda os mesmos jovens de então. O contacto com o Professor Tiago de Oliveira (que já leccionara e produzira um bom texto didáctico sobre modelos matemáticos em dinâmica de populações) e a Professora Fátima de Sousa levou-me a pedir-lhes conselho sobre leituras a fazer e eles amavelmente foram acompanhando o meu progresso. As suas cartas de recomendação, aliadas a uma boa nota de licenciatura, ajudaram a conseguir a bolsa do INIC (instituição antecessora da FCT) para poder fazer o doutoramento na State University of New York at Stony Brook, precisamente em aplicações biológicas. Fui aliciado para essa aventura pelos Professores R. R. Sokal e J. R. Rohlf, que conheci num curso de Verão por eles ministrado em Oeiras (organizado pela Fundação Calouste Gulbenkian). Chegado a Stony Brook em Agosto de 1976, fui aprender mais Matemática e Biologia (de que só tinha conhecimentos rudimentares), mas na altura nem sequer tinha ouvido falar de EDE. Nem sequer havia lá, como não havia cá e julgo que não haveria na generalidade das Universidades do mundo, nenhum curso sobre essa matéria, que na altura estava pouco divulgada.

Foi em 1977 numa disciplina de Seminário (sobre temas de Ecologia e Evolução) que tive de fazer uma apresentação aos meus colegas de um artigo da minha escolha. Procurei algo que utilizasse métodos “originais” e deparei-me<sup>4</sup> com um artigo interessantíssimo

<sup>4</sup> Não havia motores de busca para ajudar e talvez não tivessem ajudado porque não sabia o que buscava.

inadiáveis foram postergando a escrita do texto e que fui falhando, contra o meu feitio de cumprir compromissos custe o que custar, os sucessivos prazos até que, perante o inevitável ultimato que o tornou agora inadiável, aqui vai o texto que foi possível fazer em poucos dias. Ele não é o que eu queria fazer porque o que eu queria fazer exigia muito mais tempo do que o que dispunha. Ele não é o que os leitores mereciam mas, neste momento, só tinha duas opções: ou fazer o que me era possível ou não fazer. A última estava-me vedada pelo compromisso assumido com o Presidente da SPE de que só ele me poderia libertar. O que não fez, porque compreensivelmente ... não podia fazer. Só espero que o leitor seja mais generoso para comigo do que eu próprio e me saiba perdoar.

Pelas razões acima referidas vou-me limitar, por uma questão de exequibilidade, a analisar o que foram estes 25 anos em Portugal na área das aplicações das equações diferenciais estocásticas (EDE) à dinâmica de populações de seres vivos. É a minha principal área de investigação e é um nicho especializado com um número razoável (embora não muito grande) de cultores a nível internacional mas com poucos a nível nacional, onde só nos últimos anos foi possível atrair alguns jovens investigadores<sup>1</sup>. Deste modo, necessariamente, o meu trabalho pessoal vai ter aqui uma presença preponderante, do que peço desculpa aos leitores. A minha intenção inicial pretendia abranger a área mais vasta dos Processos Estocásticos, o que exigia uma pesquisa intensa e o conseqüente tempo de que agora não disponho. Talvez noutra ocasião...

O meu propósito é dar uma visão panorâmica não pormenorizada, dirigida ao não especialista, do que se tem feito nesta área. Assim, não irei descrever em pormenor a actividade de investigação desenvolvida, o que seria fastidioso, mas apenas os resultados mais marcantes. Tentarei ser poupado nas referências bibliográficas, dando preferência às de língua portuguesa (particularmente às Actas dos Congressos da SPE), normalmente em detrimento das publicações iniciais dos resultados (frequentemente em língua inglesa). O leitor pode encontrar estas últimas referenciadas nas primeiras, que servem melhor o objectivo divulgatório deste texto.

Como é óbvio, as equações diferenciais estocásticas têm inúmeras outras aplicações nos mais variados ramos da Ciência e da Tecnologia. Elas nasceram da Física, mais concretamente na modelação do chamado movimento browniano de partículas suspensas em fluidos<sup>2</sup>, mas o seu tratamento matemático iniciou-se com Itô<sup>3</sup> em 1951. Também é óbvio que a influência do “acaso” na dinâmica de populações de seres vivos pode ser modelada, conforme as circunstâncias e os objectivos, de variadas maneiras (processos de ramificação, processos de nascimento e morte, processos semi-markovianos, etc.). A área em que me vou mover é, assim, a da intersecção de um tipo de modelos (as equações diferenciais estocásticas) com um tipo de aplicações (a dinâmica de populações de seres vivos que vivem num ambiente que inevitavelmente sofre flutuações aleatórias).

<sup>1</sup> Felizmente noutras áreas de aplicação das EDE conseguiu-se maior capacidade de atracção.

<sup>2</sup> Uhlenbeck, G. E. e Ornstein, L. S. (1930). On the theory of Brownian motion. *Physical Review* 36: 823-841. NOTA. A solução da EDE usada no modelo é conhecida por processo de Ornstein-Uhlenbeck.

<sup>3</sup> O trabalho pioneiro foi: Itô, K. (1951). On stochastic differential equations. *Mem. Amer. Math. Soc.* 4: 1-51.

No final farei algumas referências à divulgação em Portugal das equações diferenciais estocásticas.

### 3. Os acasos na vida de um jovem investigador

Como é que me cheguei a interessar por estas matérias? É uma longa história fruto de vários acasos que ocorreram na minha vida, e aí também o título deste texto se aplica com propriedade. Tinha chegado de Angola em Dezembro de 1974 e quis o “acaso” que arranjasse colocação na Universidade de Évora, onde iniciei funções em Maio de 1975 e que se viria a tornar a minha casa. Na Universidade de Évora predominavam então os ensinamentos de matérias agrícolas e ambientais e o Reitor, Professor Ário Lobo Azevedo, manifestou interesse em que eu trabalhasse em aplicações biológicas da Matemática, tema que também me atraía bastante. Na altura, o Professor José Tiago de Oliveira organizava na Faculdade de Ciências de Lisboa um seminário muito dinâmico de Probabilidades e Estatística. Reunia numa sala do piso inferior do anexo ao Observatório na Rua da Escola Politécnica, um edifício um pouco degradado mas com o enquadramento paisagístico do Jardim Botânico. Aí aprendi que as Probabilidades e Estatística eram um mundo muito mais vasto do que me tinha apercebido na licenciatura em Matemática Aplicada, apesar de esta ter para a época uma razoável componente desta área. E aprendi com os melhores como se fazia e apresentava o trabalho de investigação. Estavam lá os mais seniores da altura, todos aliás jovens de espírito, como o Professor Tiago de Oliveira (que já conhecera como Professor em Angola), o meu Pai ou a Professora Maria de Fátima Fontes de Sousa. Estavam lá muitos dos que são hoje os seniores da nossa comunidade científica, embora entre nós nos consideremos ainda os mesmos jovens de então. O contacto com o Professor Tiago de Oliveira (que já leccionara e produzira um bom texto didáctico sobre modelos matemáticos em dinâmica de populações) e a Professora Fátima de Sousa levou-me a pedir-lhes conselho sobre leituras a fazer e eles amavelmente foram acompanhando o meu progresso. As suas cartas de recomendação, aliadas a uma boa nota de licenciatura, ajudaram a conseguir a bolsa do INIC (instituição antecessora da FCT) para poder fazer o doutoramento na State University of New York at Stony Brook, precisamente em aplicações biológicas. Fui aliciado para essa aventura pelos Professores R. R. Sokal e J. R. Rohlf, que conheci num curso de Verão por eles ministrado em Oeiras (organizado pela Fundação Calouste Gulbenkian). Chegado a Stony Brook em Agosto de 1976, fui aprender mais Matemática e Biologia (de que só tinha conhecimentos rudimentares), mas na altura nem sequer tinha ouvido falar de EDE. Nem sequer havia lá, como não havia cá e julgo que não haveria na generalidade das Universidades do mundo, nenhum curso sobre essa matéria, que na altura estava pouco divulgada.

Foi em 1977 numa disciplina de Seminário (sobre temas de Ecologia e Evolução) que tive de fazer uma apresentação aos meus colegas de um artigo da minha escolha. Procurei algo que utilizasse métodos “originais” e deparei-me<sup>4</sup> com um artigo interessantíssimo

<sup>4</sup> Não havia motores de busca para ajudar e talvez não tivessem ajudado porque não sabia o que buscava.

(embora com alguns problemas e um ou outro erro) de Levins (1969)<sup>5</sup>. Ele pegava nalguns modelos clássicos de crescimento de populações (equações diferenciais ordinárias) e introduzia um termo adicional aleatório para dar conta das perturbações que as naturais flutuações aleatórias das condições ambientais provocavam no crescimento da população. Eram EDE que ele adaptara das aplicações físicas para as aplicações biológicas. Ao contrário de outros modelos que estudávamos em Processos Estocásticos (como processos de ramificação ou processos de nascimento e morte), aqui não se estudava a aleatoriedade demográfica (que assume taxas médias determinísticas de natalidade e mortalidade mas tem em conta que os nascimentos e mortes ocorrem ao acaso) mas a aleatoriedade ambiental, em que são as próprias taxas que flutuam aleatoriamente conforme as condições ambientais são mais ou menos favoráveis à sobrevivência e à reprodução. Tive sorte pois era o primeiro artigo alguma vez publicado sobre o tema. Pesquisando mais, encontrei alguns<sup>6</sup> (muito poucos) que desenvolviam esse estudo pioneiro. E, naturalmente, fui à procura dos compêndios sobre equações diferenciais estocásticas para estudar os fundamentos matemáticos. Havia poucos ainda mas lá encontrei o Arnold (1974)<sup>7</sup>, o Gihman e Skorohod (1972)<sup>8</sup>, o Soong (1973)<sup>9</sup>, o McKean Jr. (1969)<sup>10</sup>. E com isso me tive de governar para em poucas semanas preparar um seminário que fosse inteligível a colegas que na sua maioria nunca haviam estudado Processos Estocásticos. O seminário teve algum impacto (tanto que no ano seguinte um colega fez a tese de mestrado na área). Entusiasmei-me e falei ao meu orientador, Professor L. R. Ginzburg, em trabalhar nisso para a minha tese, o que ele aceitou apesar de tal o obrigar a aprender comigo uma matéria que também para ele era nova. Acho que foi uma boa escolha pois era uma área ainda muito nova que as minhas contribuições podiam ajudar a construir, como veio a suceder. Aliás, a investigação para a tese<sup>11</sup>, que defendi em final de 1979 (Braumann, 1979) foi um diálogo constante entre o meu próprio trabalho e novos artigos que iam saindo sobre esta temática. Na altura, ainda havia a ideia de a tese dever ter resultados inéditos (o que deixava sempre o receio de alguém aparecer antes com algum resultado que tanto trabalho

<sup>5</sup> Levins, R. (1969). The effect of random variations of different types on population growth. *Proc. Natl. Acad. Sci. USA* 62: 1061-1065.

<sup>6</sup> Os que mais utilizei na altura (havia alguns outros) foram:

-Capocelli, R. M. e Ricciardi, L. M. (1974). A diffusion model for population growth in random environments. *Theoret. Popul. Biol.* 5: 28-41.

-Goel, N. S. e Richter-Dyn, N. (1974). *Stochastic Models in Biology*. Academic Press, N. Y.

-Kiester, A. R. e Barakat, R. (1974). Exact solutions to certain stochastic differential equation models of population growth. *Theoret. Popul. Biol.* 6: 199-216.

-May, R. M. (1973). Stability in randomly fluctuating versus deterministic environments. *Amer. Natur.* 107: 621-650.

-Roughgarden, J. (1975). A simple model for population dynamics in stochastic environments. *Amer. Natur.* 109: 713-736.

-Tuckwell, H. C. (1974). A study of some diffusion models of population growth. *Theoret. Popul. Biol.* 5: 345-357.

<sup>7</sup> Arnold, L. (1974). *Stochastic Differential Equations. Theory and Applications*. Wiley, N. Y.

<sup>8</sup> Gihman, I. I. e Skorohod, A. V. (1972). *Stochastic Differential Equations*. Springer, N. Y.

<sup>9</sup> Soong, T. T. (1973). *Random Differential Equations in Science and Engineering*. Academic Press, N. Y.

<sup>10</sup> McKean Jr., H. P. (1969). *Stochastic Integrals*. Academic Press, N. Y.

<sup>11</sup> Braumann, C. A. (1979). *Population Growth in Random Environments*. Ph. D. Thesis, State University of New York at Stony Brook.

nos levara a obter), pelo que as primeiras publicações minhas sobre o assunto só começam em 1981.

De volta a Portugal, é criada no ano seguinte a SPE (na altura Sociedade Portuguesa de Estatística e Investigação Operacional), a que aderi e a que me orgulho de pertencer, tendo participado activamente quer nos Colóquios de Estatística e Investigação Operacional, quer posteriormente, nos Congressos Anuais que organizou.

#### 4. Brevíssima introdução às equações diferenciais estocásticas

Mas o que são equações diferenciais estocásticas? São equações da forma

$$dX(t) = f(t, X(t))dt + g(t, X(t))dw(t), \quad X(0) = X_0, \quad (1)$$

onde  $w(t) = w(t, \omega)$  ( $t \in [0, +\infty)$ ,  $\omega \in \Omega$ ) é um processo de Wiener padrão num espaço de probabilidade  $(\Omega, \mathcal{F}, P)$ ,  $f(t, x)$  e  $g(t, x)$  são funções reais de variável real,  $X(t) = X(t, \omega)$  tem valores reais e  $X_0$  é uma constante ou uma v.a. independente do processo de Wiener. Tudo isto pode ser generalizado a várias dimensões. Aqui  $\omega$  representa o "acaso". O processo de Wiener  $w(t)$  é um processo de Markov com incrementos independentes em que os incrementos  $w(t) - w(s)$  ( $s \leq t$ ) têm distribuição normal com média nula e variância  $t - s$ . Este processo é quase certamente (q.c.) de variação ilimitada e por isso a sua derivada  $\varepsilon(t) = dw(t)/dt$ , a que se chama ruído branco padrão (em tempo contínuo), não existe no sentido corrente mas apenas como processo estocástico generalizado. O ruído branco tem a propriedade de haver independência entre instantes distintos. Embora a forma (1) seja geralmente preferida pelos matemáticos, talvez seja mais intuitivo escrever a EDE (1) na forma

$$\frac{dX(t)}{dt} = f(t, X(t)) + g(t, X(t))\varepsilon(t), \quad X(0) = X_0, \quad (2)$$

em que o primeiro termo representa a tendência da dinâmica de  $X(t)$  e o segundo termo representa as flutuações aleatórias em torno dessa tendência. Na ausência do segundo termo temos uma equação diferencial ordinária que poderá descrever a dinâmica do fenómeno em estudo na ausência de perturbações aleatórias. O efeito destas é descrito pelo segundo termo, onde  $g$  é como que um desvio-padrão que mede a intensidade do efeito do ruído sobre a dinâmica. Também se podem considerar EDE com ruídos coloridos (isto é, autocorrelacionados), a que alguns autores preferem chamar equações diferenciais aleatórias, mas o seu tratamento matemático, salvo no caso linear, é difícil. Além disso, um ruído branco é uma boa aproximação para ruídos coloridos cuja autocorrelação decaia rapidamente para zero. Naturalmente, entende-se por solução da EDE (1) (ou (2)), a solução  $X(t) = X(t, \omega)$  da equação integral

$$X(t) = X_0 + \int_0^t f(s, X(s))ds + \int_0^t g(s, X(s))dw(s). \quad (3)$$

Sob condições de regularidade adequadas, para cada trajectória  $\omega$  fixa, o primeiro integral pode ser interpretado como um vulgar integral de Riemann. Porém, o segundo integral não pode ser interpretado como integral de Riemann-Stieltjes porque o processo integrador  $w(t)$  tem variação ilimitada para quase todas as trajectórias, pelo que o limite das somas de Riemann-Stieltjes para uma sucessão de decomposições (com diâmetro convergente para zero) do intervalo de integração depende da escolha dos pontos intermédios onde é calculada a função integranda  $g(s, X(s))$ . Se fizermos a escolha (não-antecipativa) dos pontos iniciais de cada subintervalo (a escolha em que a dinâmica presente não é afectada pelas perturbações aleatórias futuras), temos o integral de Itô, que tem excelentes propriedades probabilísticas mas não segue as regras usuais de cálculo. Tal obriga a um novo cálculo, o cálculo estocástico de Itô. Aqui usa-se a convergência em média quadrática (convergência  $L^2$  com respeito a  $\omega$ ) para a obtenção do limite das somas de Riemann-Stieltjes, embora, ao generalizar a uma classe mais vasta de funções integrandas, se possa substituir essa convergência por uma convergência em probabilidade. Há outros integrais correspondentes a outras escolhas ou combinações de escolhas dos pontos intermédios, sendo a mais popular deles o integral de Stratonovich que, não tendo tão boas propriedades probabilísticas, segue as regras usuais de cálculo. Como é natural, as EDE de Itô e de Stratonovich com idênticas funções  $f$  e  $g$ , têm em geral soluções diferentes. Aliás, a EDE de Itô  $dX=fdt+gdw$  é equivalente à (tem a mesma solução da) EDE de Stratonovich (S)  $dX=f^*dt+gdw$  (o “(S)” é para assinalar que se usa o cálculo de Stratonovich) em que  $f^*=f-(1/4)\partial g^2/\partial x$ . Em qualquer dos cálculos, as funções  $f$  e  $g$  devem satisfazer certas condições de regularidade para se poder garantir a existência e unicidade de solução da EDE; felizmente, com condições de regularidade adequadas, a solução é um processo de difusão que satisfaz as equações de Kolmogorov e, portanto, é um processo de Markov. Se a EDE for autónoma ( $f(t,x)=f(x)$  e  $g(t,x)=g(x)$ ), a solução é mesmo um processo de difusão homogéneo. Mais pormenores podem ver-se nos livros sobre EDE acima referidos ou em Øksendal (1998)<sup>12</sup> ou ainda, para uma versão resumida, em Braumann (1998)<sup>13</sup>.

## 5. Modelos de crescimento populacional em ambiente aleatório

No caso das populações, vamos pensar em  $X(t)$  como o tamanho (número de indivíduos, biomassa ou densidade) da população no instante  $t$ . Se representarmos por  $r$  a taxa de crescimento *per capita*, o modelo determinístico seria  $dX/dt=rX$ , nada impedindo que  $r$  possa depender do tamanho da população  $x$  e do instante  $t$  [ $r=r(t,x)$ ]. Mas, se

<sup>12</sup>Øksendal, B. (2003). *Stochastic Differential Equations. An Introduction with Applications*. (6<sup>th</sup> edition). Springer, Berlin.

<sup>13</sup>Braumann, C. A. (1998). O Acaso, a bolsa e a vida. Em *Estatística: A Diversidade na Unidade. Actas do V Congresso Anual da SPE*, M. Souto de Miranda e I. Pereira (eds.), p. 29-55, Sociedade Portuguesa de Estatística e Edições Salamandra, Lisboa.

admitirmos condições ambientais constantes (salvo no efeito que sobre elas tem o próprio tamanho da população), podemos usar um modelo autónomo em que  $r=r(x)$ , vindo  $dX(t)/dt=r(X(t))X(t)$ . Neste caso, além do equilíbrio  $X=0$ , os valores de  $x$  para os quais  $r(x)=0$  são tamanhos de equilíbrio para a população. Vamos admitir, por exemplo, que são válidas as hipóteses biologicamente razoáveis de  $r(x):(0,+\infty)\rightarrow(-\infty,+\infty)$  ser de classe  $C^1$  decrescente (o que traduz o efeito deletério que o aumento da população e a consequente competição pelos recursos tem sobre a sobrevivência e reprodução individuais) com  $r(0^+)>0$  e  $r(+\infty)<0$ . Neste caso, há um único valor de equilíbrio positivo  $K$  (o único para o qual  $r(K)=0$ ), chamado capacidade de sustento do meio, e ele é um equilíbrio globalmente estável do qual o tamanho da população se aproxima quando  $t\rightarrow+\infty$  (salvo se a população inicial for nula).

Se o ambiente tiver flutuações aleatórias em torno de um valor médio constante, então elas irão afectar a taxa de crescimento *per capita* e podemos ter o seguinte modelo:  $dX/dt=(r+\sigma\varepsilon(t))X$ . Aqui  $r$  representa uma taxa "média" de crescimento *per capita* e  $\sigma\geq 0$  mede a intensidade das perturbações em torno dessa média provocadas pelas variações aleatórias do ambiente (usamos o ruído branco como aproximação dos ruídos naturais, que poderão ser coloridos). Note-se que  $\varepsilon(t)=\varepsilon(t,\omega)$  depende do "acaso"  $\omega$  e, portanto, a solução da EDE também depende de  $\omega$ . Aqui  $\omega$  pode interpretar-se como um cenário ambiental escolhido ao "acaso" pela natureza (de acordo com a lei de probabilidade  $P$ ) de entre todos os cenários ambientais possíveis (que formam o conjunto  $\Omega$ ); por cenário ambiental referimo-nos à descrição de como evolui ao longo do tempo o estado do ambiente em todos os aspectos que influenciem o crescimento da população. Admitindo um ambiente aleatoriamente variável mas estacionário (isto é, com as distribuições de probabilidade de dimensão finita insensíveis a translações no tempo), poderemos usar um modelo autónomo em que  $\sigma$  poderá ser constante ou depender do tamanho da população mas não dependerá directamente do tempo  $t$ . Teremos então modelos que são EDE da forma

$$\frac{dX(t)}{dt} = r(X(t))X(t) + \sigma(X(t))X(t)\varepsilon(t) , \quad X(0)=X_0>0, \quad (4)$$

ou, se preferirmos a notação equivalente, da forma

$$dX(t) = r(X(t))X(t)dt + \sigma(X(t))X(t)dw(t) , \quad X(0)=X_0>0.$$

Foram alguns casos particulares deste modelo geral, correspondentes a expressões específicas para as funções  $r$  e  $\sigma$ , que foram tratados na literatura nos artigos acima referidos e noutros que surgiram posteriormente. Um exemplo típico é o modelo logístico com ruído aditivo em que  $r(x)=ax(1-x/K)$  (com  $a>0$  e  $K>0$  constantes) e  $\sigma(x)\equiv\sigma$  (intensidade do ruído constante).

Põem-se várias questões interessantes como:

- a) Verificar se estão satisfeitas as condições de existência e unicidade.
- b) Obter, se possível, uma solução explícita  $X(t)=X(t,\omega)$ , dependente naturalmente dos parâmetros envolvidos nas expressões concretas de  $r(x)$  e  $\sigma(x)$ , da condição inicial  $X_0$ , do instante  $t$  e do cenário ambiental  $\omega$ . Claro que a dependência de  $\omega$  vem expressa através dos valores da trajectória correspondente do ruído branco  $\varepsilon(s,\omega)$  (ou, equivalentemente, do seu integral, o processo de Wiener  $w(s,\omega)$ ), pelo que não temos felizmente que lidar com o complicado espaço dos cenários do ambiente mas apenas com as distribuições de probabilidade dos seus efeitos sobre o crescimento da população. Se não conseguirmos uma expressão explícita para a solução, podemos obter soluções simuladas pelo método de Monte Carlo.
- c) Estudar, quando possível, propriedades importantes da solução, como as densidades de probabilidade de transição, a distribuição de probabilidade em cada instante, a sua esperança matemática ou o desvio-padrão.
- d) Verificar se a solução converge para um valor de equilíbrio positivo quando  $t \rightarrow +\infty$ . Ao contrário dos modelos determinísticos razoáveis que considerámos há pouco, isso só sucede em modelos estocásticos muito especiais (geralmente pouco realistas). Nos modelos mais realistas em que a intensidade do ruído  $\sigma(x)$  é limitada inferiormente por uma quantidade positiva, a existência de perturbações aleatórias desfaz qualquer candidato  $x > 0$  a valor de equilíbrio. Podemos no entanto almejar a um “equilíbrio estocástico” estável (ou “regime sustentado estocástico”) em que a distribuição de probabilidade do tamanho da população no instante  $t$  converge, quando  $t \rightarrow +\infty$ , para uma distribuição de equilíbrio (dita distribuição estacionária) com densidade de probabilidade (dita densidade estacionária). Claro que, neste caso, o tamanho da população nunca estabiliza, só a sua distribuição de probabilidade é que o faz.
- e) Verificar se é possível a extinção da população (no sentido “matemático” de  $X(t)=0$  para algum  $t$  ou de  $X(t) \rightarrow 0$  quando  $t \rightarrow +\infty$ ) e qual a probabilidade de ela ocorrer.
- f) A extinção “matemática” não é muito interessante. Como sucede para modelos baseados em equações diferenciais, o tamanho da população é uma variável contínua que pode tomar valores muito próximos de zero sem no entanto se extinguir “matematicamente”, ainda que na realidade tais valores sejam biologicamente impossíveis. Deve pois considerar-se um limiar mínimo  $x_e > 0$  abaixo do qual a população é suposta estar extinta<sup>14</sup>. Então a probabilidade de

<sup>14</sup> Ou porque não são biologicamente possíveis valores inferiores ou porque, para valores inferiores, há outros factores não contemplados nos modelos, como efeitos de Allee, que garantem a extinção da população.

extinção “realista” num horizonte temporal finito ou infinito será a probabilidade de a população alguma vez descer abaixo desse limiar nesse horizonte. O tempo de extinção será a variável aleatória que define o tempo de primeira passagem por esse limiar e é interessante estudar a sua distribuição de probabilidade.

- g) Como estimar os parâmetros envolvidos nas expressões consideradas para as funções  $r$  e  $\sigma$  a partir de observações do tamanho da população  $X(t)$  para vários instantes  $t_1, t_2, \dots, t_n$ ? Devemos considerar o caso de várias trajectórias (por exemplo, se se trata de populações laboratoriais em que podemos ter várias réplicas da experiência, isto é, vários valores de  $\omega$ ). Mas também devemos considerar o caso mais frequente de termos apenas uma trajectória pois só observamos aquela população com aquele cenário do ambiente que a natureza por “acaso” seleccionou.
- h) Como prever o tamanho da população num instante futuro?
- i) Como testar hipóteses sobre os parâmetros ou fazer testes de comparação de parâmetros entre populações diferentes?
- j) Como escolher entre expressões alternativas para as funções  $r$  e  $\sigma$  (selecção de modelos)?
- k) Será que se obteriam resultados semelhantes caso se usasse um ruído colorido em vez de usar o ruído branco como sua aproximação?

Trabalhei em quase todas estas alíneas para diversos modelos concretos com formas funcionais específicas para  $r$  e  $\sigma$ .

O modelo mais simples é o modelo malthusiano de crescimento populacional, em que  $r(x) \equiv r$  (coeficiente de tendência) e  $\sigma(x) \equiv \sigma$  (coeficiente de difusão). Este modelo é também o modelo tradicional de Black-Scholes para a cotação de uma acção na bolsa de valores e também tem sido aplicado ao crescimento inicial de organismos (incluindo fetos), pelo que o seu estudo é útil em diferentes áreas de aplicação. A solução da EDE neste caso é o conhecido movimento browniano geométrico, cujas principais propriedades estão bem estudadas. Vários estudantes meus fizeram trabalhos de seminário ou de fim de curso em que sistematizaram, generalizaram ou aplicaram (geralmente usando dados financeiros, mais fáceis de obter) este modelo ou desenvolveram métodos de estimação e previsão para situações amostrais mais complexas<sup>15</sup>. No caso de várias trajectórias do movimento browniano, desenvolvi<sup>16</sup> uma técnica tipo ANOVA de comparação de tendências, que foi

<sup>15</sup> Podem obter-se os trabalhos de fim de curso na Universidade de Évora de Clara Carlos, Cláudia Patrício, Patrícia Filipe, Sónia Soares.

<sup>16</sup> Pode ver-se a versão mais completa em: -Braumann, C. A. (1999). Estimação de parâmetros para uma ou várias trajectórias do movimento browniano geométrico. Em *Afirmar a Estatística: Um Desafio para o Século XXI. Actas do VI Congresso Anual da SPE*, C. D. Paulino, A. Pacheco, A. Pires e Ferreira da Cunha (eds.), p. 149-156, Sociedade Portuguesa de Estatística, Lisboa.

depois generalizada<sup>17</sup> para detectar a presença de sazonalidade na tendência<sup>18</sup>. Este último teste é útil para populações vivendo em ambientes com componente sazonal. Também é útil para detectar sazonalidades na bolsa, tendo sido aplicado<sup>19</sup> ao estudo do efeito do dia da semana.

Os problemas estatísticos de estimação, previsão, testes de hipóteses e selecção de modelos são geralmente esquecidos na literatura internacional nesta área, pelo que esta foi uma matéria a que dediquei sempre (inclusivamente na tese de doutoramento) particular atenção, a fim de que os modelos pudessem ser efectivamente aplicados. Debrucei-me mais<sup>20</sup> sobre modelos mais realistas do que o modelo malthusiano<sup>21</sup>, como o modelo logístico  $r(x)=ax(1-x/K)$ , o modelo de Gompertz  $r(x)=ax \ln(K/x)$  ou outros.

Os problemas de extinção têm sido tratados no sentido “matemático” do termo, mas raramente no sentido “realista”. Por isso dei particular atenção à determinação da probabilidade de extinção (incluindo a sua estimação) e ao estudo do tempo de extinção interpretada no sentido “realista”<sup>22</sup>. Mais recentemente foram, a esse respeito, estudados

<sup>17</sup> Braumann, C. A. (2004). Inferência estatística para o movimento browniano geométrico com coeficiente de tendência sazonal de período múltiplo do período de observação. Em *A Estatística com Acaso e Necessidade. Actas do XI Congresso Anual da SPE*, P. M. M. Rodrigues, E. L. Rebelo e F. Rosado (eds.), p. 101-111, Edições SPE, Lisboa.

<sup>18</sup> Uma generalização do modelo permite a introdução da sazonalidade.

<sup>19</sup> Patrício, C. e Braumann, C. A. (2004). Modelo de Black-Scholes com taxas de rendimento variáveis com o dia da semana. Em *A Estatística com Acaso e Necessidade. Actas do XI Congresso Anual da SPE*, P. M. M. Rodrigues, E. L. Rebelo e F. Rosado (eds.), p. 583-594, Edições SPE, Lisboa.

<sup>20</sup> A lista de trabalhos é extensa, mas limito-me a indicar, a título exemplificativo:

-Braumann, C. A. (1999). Population growth in random environments: some modelling and statistical issues. Em *Applied Stochastic Models and Data Analysis. Quantitative Methods in Business and Industrial Society. Proc. IX Intl. Symp. Appl. Stoch. Models and Data Analysis*, H. Bacelar-Nicolau, F. Costa Nicolau e J. Janssen (eds.), p.36-42, Instituto Nacional de Estatística, Lisboa.

-Braumann, C. A. (1997). Parameter estimation in population growth and fishing in random environments. *Bull. Intl. Statistical Institute* LVII, CPI: 21-32.

-Braumann, C. A. (1996). Estimação de parâmetros em modelos de crescimento e pesca em ambientes aleatórios. Em *Bom Senso e Sensibilidade. Traves Mestras da Estatística. Actas do III Congresso da SPE*, J. Branco, P. Gomes e J. Prata (eds.), p. 103-117, Sociedade Portuguesa de Estatística e Edições Salamandra, Lisboa.

-Braumann, C. A. (1994). Escolha de modelos de crescimento populacional em ambiente aleatório e predição. Em *A Estatística e o Futuro e o Futuro da Estatística. Actas do I Congresso Anual da SPE*, D. Pestana, A. Turkman, J. Branco, L. Duarte e A. Pires (eds.), p. 39-46, Sociedade Portuguesa de Estatística e Edições Salamandra, Lisboa.

<sup>21</sup> O modelo malthusiano só poderá ser útil em fases iniciais de crescimento em que não se faça sentir o efeito da competição dos indivíduos pelos recursos disponíveis.

<sup>22</sup> Indicam-se os trabalhos mais relevantes:

-Braumann, C. A. (1995). Threshold crossing probabilities for population growth models in random environments. *J. Biological Systems* 3: 505-517.

-Braumann, C. A. (1994). Crescimento populacional em ambiente aleatório e probabilidades de cruzamento de limiares críticos. *Actas do II Congresso Anual da SPE*, p. 91-105, Sociedade Portuguesa de Estatística e Departamento de Matemática da Faculdade de Ciências da Universidade de Coimbra.

- Braumann, C. A. (1991). Estimação de probabilidades de extinção em ambiente aleatório. *Actas I<sup>as</sup> Jornadas de Estatística e Aplicações*, p. 393-404, Universidade do Minho, Braga.

-Braumann, C. A. (1983). Population extinction probabilities and methods of estimation for population stochastic differential equation models. Em *Nonlinear Stochastic Problems*, R.S. Bucy e J.M.F. Moura (eds.), p. 553-559, D. Reidel Publ. Comp., NATO ASI Series, Dordrecht.

os modelos logístico e de Gompertz com ruído aditivo ( $\sigma(x) \equiv \sigma$ ) na tese de mestrado de Clara Carlos<sup>23</sup>.

A questão k) foi abordada na minha tese de doutoramento e nalguns outros trabalhos.

Nos trabalhos anteriores (e noutros não mencionados) foram, naturalmente abordadas questões mais gerais como as referidas em a), b), c) e d). Um trabalho de síntese sobre estas questões gerais é a tese de mestrado de Maria Manuela Maia<sup>24</sup>, orientada por Teresa Arede e Francisco Calheiros. Aliás Teresa Arede orientou recentemente a tese de mestrado<sup>25</sup> sobre tempos de primeira passagem de Catarina Santos.

A questão j) põe o dedo numa ferida. É muito difícil saber se as formas funcionais escolhidas para  $r$  e  $\sigma$  são ou não “correctas”. Ficamos pois na dúvida se importantes propriedades que se possam deduzir para o modelo, como sejam as relativas a haver ou não extinção ou a existir ou não uma densidade estacionária, são propriedades da população ou da forma específica que escolhemos para as funções  $r$  e  $\sigma$ . Há pois vantagem em deduzir tais propriedades, não para modelos concretos correspondentes a formas funcionais específicas das funções  $r$  e  $\sigma$ , mas para o modelo geral em que não se especificam essas formas funcionais, apenas se requer que  $r$  e  $\sigma$  satisfaçam certos pressupostos qualitativos ditadas pelo comportamento biológico conhecido da população. Estamos a falar de pressupostos como as hipóteses biologicamente razoáveis há pouco consideradas no modelo determinístico em que, por exemplo, pressupomos que  $r(x)$  decresce com  $x$  mas não especificamos a forma concreta desse decrescimento (se é, por exemplo, do tipo logístico ou outro proposto ou não na literatura), forma concreta que não temos meio seguro de saber se é a que a natureza segue.

Nesta situação em que estabelecemos apenas pressupostos biológicos qualitativos gerais, põem-se questões muito mais difíceis como:

- l) Verificar se é possível a extinção matemática?
- m) Tentar obter expressões para a probabilidade de extinção realista em certo horizonte temporal. Tentar obter a expressão da distribuição do tempo de extinção.
- n) Verificar se a população converge para um valor de equilíbrio positivo. Verificar se existe uma densidade estacionária (“equilíbrio estocástico”).
- o) Estimar  $r(x)$  e/ou  $\sigma(x)$  por métodos não-paramétricos ou semiparamétricos.

<sup>23</sup> Carlos, C. (2004). *Tempos de Extinção para Populações em Ambiente Aleatório*. Tese de mestrado em Matemática Aplicada, Universidade de Évora.

Poderá ser mais fácil consultar (espera-se que brevemente) o artigo (que condensa o essencial da tese):

-Carlos, C. e Braumann, C. A. Tempos de extinção para populações em ambiente aleatório. *Actas do XII Congresso Anual da SPE* (submetido).

<sup>24</sup> Maia, M. M. F. (2003). *Equações Diferenciais Estocásticas e Dinâmica de Populações*. Tese de mestrado em Estatística Aplicada e Modelação, Faculdade de Engenharia da Universidade do Porto.

<sup>25</sup> Santos, C. P. D. (2004). *Tempos de Primeira Passagem por Fronteiras Variáveis e Aplicações*. Tese do mestrado em Estatística, Faculdade de Ciências da Universidade do Porto.

Trabalhei com particular intensidade nas questões l), m) e n), no sentido de obter resultados para modelos gerais que traduzissem assim propriedades da população e não tanto do modelo específico escolhido.

Um dos resultados<sup>26</sup> de que mais me orgulho foi o de demonstrar a ausência de extinção matemática e a existência de densidade estacionária (“equilíbrio estocástico”) para o modelo geral no pressuposto de que  $r(x)$  satisfaz as hipóteses biologicamente razoáveis que acima formulei para o modelo determinístico e que  $\sigma(x)$  (que mede a intensidade com que as flutuações aleatórias do ambiente afectam a taxa de crescimento *per capita*) é, para  $x > 0$ , limitada inferiormente por uma constante positiva. Aliás, demonstrei o resultado para hipóteses mais fracas mas tecnicamente mais complexas que não vale a pena especificar aqui. Já, porém, alterando o pressuposto  $r(0^+) > 0$  para  $r(0^+) < 0$  (caso de uma população em as condições são desfavoráveis mesmo para manter uma população pequena), haverá extinção com probabilidade um. Assim, poderíamos concluir que o que determina a extinção ou a existência de equilíbrio estocástico é o facto de a taxa “média” de crescimento *per capita* para populações pequenas ( $r(0^+)$ ) ser negativa ou positiva. Este resultado, que pode parecer intuitivamente óbvio mas que é de demonstração difícil, foi obtido usando o cálculo de Stratonovich.

## 6. O problema das diferenças qualitativas entre os cálculos de Itô e Stratonovich

Retomemos o resultado referido no final da secção anterior e reportemo-nos agora, por comodidade, ao caso de  $\sigma(x) \equiv \sigma$ . Se usarmos o cálculo de Itô em vez do de Stratonovich, o que decidiria haver extinção ou haver densidade estacionária seria saber se  $r(0^+) < \sigma^2/2$  ou  $r(0^+) > \sigma^2/2$ . Isto é, poderíamos ter extinção com probabilidade um mesmo para valores positivos (desde que inferiores a  $\sigma^2/2$ ) da taxa “média” de crescimento *per capita* para populações pequenas. Este resultado é qualitativamente diferente do que se obtém com o cálculo de Stratonovich. Em qual confiar? Isto é, qual o cálculo adequado?

Esta mesma questão deu uma controvérsia na literatura, porque May e MacArthur (1972)<sup>27</sup> tinham apresentado uma teoria da limitação da semelhança de nichos que estabelecia um limite para essa semelhança para que uma comunidade de espécies pudesse persistir. Isso baseava-se num modelo de competição de Lotka-Volterra em ambiente aleatório e no uso do cálculo de Itô. Contudo, como mostraram Feldman and Roughgarden (1975)<sup>28</sup> e Turelli (1978)<sup>29</sup>, a teoria falha se se usar o cálculo de Stratonovich. A teoria tem

<sup>26</sup> Embora os primeiros resultados estejam noutras publicações, pode ver o resultado mais geral em: -Braumann, C. A. (2001). Crescimento de populações em ambiente aleatório: generalização a intensidades de ruído dependentes da densidade da população. Em *A Estatística em Movimento. Actas do VIII Congresso Anual da SPE*, M. M. Neves, J. Cadima, M. J. Martins e F. Rosado (eds.), p. 119-128, Sociedade Portuguesa de Estatística, Lisboa.

<sup>27</sup> May, R. M. e MacArthur, R. H. (1972). Niche overlap as a function of environmental variability. *Proc. Natl. Acad. Sci. USA* 69: 1109-1113.

ainda outras falhas<sup>28</sup>, mas o que é relevante é a diferença qualitativa entre os dois cálculos. A controvérsia prosseguiu, incluindo a emissão de recomendações sobre qual o cálculo apropriado, o que *grosso modo* dependia de o crescimento populacional ocorrer intrinsecamente em tempo discreto (recomendava-se o cálculo de Itô) ou em tempo contínuo (recomendava-se o cálculo de Stratonovich)<sup>30</sup>. Isso tinha a ver com o facto de as EDE serem obtidas como limite de equações às diferenças estocásticas no primeiro caso ou de equações diferenciais aleatórias com ruído colorido no segundo caso e de se conhecerem teoremas limite que indicavam (mas apenas em condições apropriadas) a convergência das soluções para a solução da EDE de Itô ou de Stratonovich, respectivamente. O problema é saber qual das situações se verifica em cada caso concreto. De facto, os nascimentos e mortes são acontecimentos discretos mas podem ocorrer em qualquer instante num intervalo de tempo contínuo (mesmo em espécies com estações reprodutivas curtas, os nascimentos espalham-se ao longo da estação). Considerando as diferenças dramáticas nas predições de importantes matérias como a extinção, a questão de saber qual o cálculo em que se devia confiar era um obstáculo ao uso destes modelos estocásticos.

Consegui resolver essa controvérsia Itô-Stratonovich (não é uma controvérsia entre eles mas entre utilizadores dos seus cálculos estocásticos), primeiro para modelos mais simples<sup>31</sup> e, depois, no caso de modelos gerais (4) com intensidade de ruído constante ( $\sigma(x) \equiv \sigma$ )<sup>32</sup>. O facto é que, como é vulgar na literatura e é a causa da controvérsia, se usa a mesma letra “*r*” em ambos os cálculos como se estando a referir à mesma taxa “média” de crescimento *per capita*. Não se clarificava de que tipo de “média” se tratava, pressupondo implicitamente que se falava da mesma média. Contudo, se olharmos para o que “*r*” significa em termos da dinâmica da população, verifica-se que significa coisas diferentes. Ela significa a taxa média aritmética de crescimento *per capita* quando se usa o cálculo de Itô e a taxa média geométrica de crescimento *per capita* quando se usa o cálculo de Stratonovich. Se tomarmos em conta a diferença entre estes dois tipos de média, os dois cálculos dão exactamente a mesma solução e todas as diferenças

<sup>28</sup> Feldman, M. W. e Roughgarden, J. (1975). A population's stationary distribution and chance of extinction with remarks on the theory of species packing. *Theoret. Popul. Biol.* 7: 197-207.

<sup>29</sup> Turelli, M. (1978). A reexamination of stability in randomly varying environments with comments on the stochastic theory of limiting similarity. *Theoret. Popul. Biol.* 13: 244-267.

<sup>30</sup> Ver, por exemplo:

Capocelli, R. M. e Ricciardi, L. M. (1974). A diffusion model for population growth in random environments. *Theoret. Popul. Biol.* 5: 28-41.

-Nobile, A. G. e Ricciardi, L. M. (1979). Growth and extinction in random environments. In *Proc. INFO II*, Patras.

-Ricciardi, L. M. (1979). On a conjecture concerning population growth in random environment. *Biol. Cybernet.* 32: 95-99.

<sup>31</sup> Pode ver-se (embora não seja nem o primeiro nem o mais importante artigo), uma explicação com preocupações mais didácticas em:

-Braumann, C. A. (2003). O uso de diferentes cálculos estocásticos na modelação do crescimento populacional em ambiente aleatório. Em *Literacia e Estatística. Actas do X Congresso Anual da SPE*, P. Brito, A. Figueiredo, F. Sousa, P. Teles e F. Rosado (eds.), p. 141-153, Sociedade Portuguesa de Estatística, Lisboa.

<sup>32</sup> Braumann, C. A. (2005). Ito versus Stratonovich calculus in random population growth. *Mathem. Biosci.* (aceite para publicação).

qualitativas e quantitativas desaparecem. As diferenças eram, portanto, meramente semânticas e devidas à falta de clarificação do significado da função  $r(x)$ . Temos é que usar a média adequada em cada cálculo. Se o fizermos, obteremos sempre resultados idênticos, sem necessidade de nos preocuparmos se o crescimento populacional ocorre intrinsecamente em tempo discreto ou contínuo. Assim, no exemplo acima, ambos os cálculos dão o mesmo resultado: há extinção da população ou existência de densidade estacionária conforme a taxa média **geométrica** de crescimento *per capita* para populações pequenas seja negativa ou positiva (note-se que a diferença entre a média aritmética e geométrica é precisamente  $\sigma^2/2$ ). Na vida real, como a taxa média, aritmética ou geométrica, de crescimento *per capita* tem de ser estimada dos dados populacionais, a resolução da controvérsia está intimamente ligada ao problema da estimação. Conseguimos mais recentemente estender a resolução da controvérsia ao modelo geral (4) sem necessidade de nos restringirmos ao caso de intensidades de ruído constantes; a diferença é apenas de que, no cálculo de Stratonovich,  $r(x)$  representa agora uma média geométrica generalizada (que coincide com a média geométrica no caso de a intensidade de ruído ser constante).

Com a resolução da controvérsia é agora seguro usar estes modelos de EDE e utilizar indiferentemente qualquer dos cálculos. Fica-se também alertado para o verdadeiro significado das taxas “médias” utilizadas e para a necessidade de usar a média adequada ao cálculo a que se pretende recorrer.

## 7. Modelos com capturas

No caso de a população viver em ambiente aleatório e estar sujeita a capturas (caça ou pesca), teremos que subtrair a taxa de capturas  $C(t)$ . Obtemos

$$\frac{dX(t)}{dt} = r(X(t))X(t) - C(t) + \sigma(X(t))X(t) \alpha(t), \quad X(0) = X_0 > 0. \quad (5)$$

Admitindo que, para um esforço fixo de pesca ou caça, se obtém uma taxa de captura  $C(t)$  proporcional ao tamanho da população, faz sentido definir o esforço de captura como  $c^*(t) = C(t)/X(t)$ , quantidade que mede assim a intensidade dos meios utilizados na caça ou pesca quando a população tem o tamanho  $X(t)$ . Podemos definir a política de capturas modulando os meios afectos à pesca ou caça de modo a que  $C(t)$  seja uma função escolhida de forma a atingir certos objectivos. Um objectivo típico é o de maximizar o total de capturas acumuladas (ou o lucro acumulado da exploração), com ajustamento temporal por uma taxa social de desconto, até a um determinado horizonte temporal ou até à extinção da população, conforme o que ocorra primeiro. Trata-se de um problema de controlo óptimo estocástico que recentemente, e para modelos relativamente simples, tem

sido tratado na literatura<sup>33</sup>. As políticas de captura “ótimas” têm, porém, o grave inconveniente de alternarem freneticamente entre períodos de captura ao ritmo mais intenso possível seguidos de períodos de total paragem das capturas. Isto para não falar dos problemas de execução e dos problemas sociais e ecológicos de tais políticas. Nós preferimos escolher políticas que conduzam a uma optimização em regime sustentado (e em que a extinção, ainda que mais lucrativa, não seja uma opção), pelo que preferimos políticas em que o esforço de capturas não varie arbitrariamente com o tempo mas dependa apenas do tamanho da população,  $c^*(t)=c(X(t))$ . Isso garante um modelo autónomo onde são possíveis densidades estacionárias (regimes sustentados). O esforço de capturas pode ser constante  $c(x)\equiv c$  (se utilizarmos uma intensidade constante de pesca ou caça) ou depender do tamanho da população<sup>34</sup>. Obtemos o modelo autónomo

$$\frac{dX(t)}{dt} = r(X(t))X(t) - c(X(t))X(t) + \sigma(X(t))X(t) \xi(t), \quad X(0)=X_0>0. \quad (6)$$

Põem-se para este modelo questões semelhantes às colocadas na secção 5 para o modelo (4).

Na sequência dos trabalhos pioneiros<sup>35</sup>, desenvolvi estudos mais completos sobre os mais importantes modelos paramétricos, dando ênfase [sem descurar o estudo de matérias como as referidas nas alíneas a) a. d) da secção 5] a uma matéria importantíssima mas sempre ignorada na literatura da área: os problemas estatísticos<sup>36</sup> de estimação, previsão e selecção de modelos.

Para as famílias paramétricas mais estudadas (modelos logístico e de Gompertz) de crescimento natural e para a classe de políticas de pesca de quotas constantes  $C(t)\equiv C$  ou de esforço constante  $c(x)\equiv c$ , estudámos<sup>37</sup> a questão da determinação da política óptima (no sentido de maximizar a taxa média de capturas) em regime sustentado (isto é, nos casos,

<sup>33</sup> Ver, a título de exemplo (há outros trabalhos):

-Alvarez, L. H. R. e Shepp, L. A. (1997). Optimal harvesting of stochastically fluctuating populations. *J. Math. Biol.* 37: 155-177.

-Lungu, E. e Øksendal, B. (1997). Optimal harvesting from a population in a stochastic crowded environment. *Math. Biosci.* 145: 47-75.

<sup>34</sup> Veja-se, apenas a título de exemplo, o caso de políticas de capturas de quota constante  $C(t)\equiv C$ , caso em que  $c(x)=C/x$ .

<sup>35</sup> Com destaque para:

-Beddington, J. R. e May, R. M. (1977). Harvesting natural populations in a randomly fluctuating environment. *Science* 197: 463-465.

-Gleit, A. (1978). Optimal harvesting in continuous time with stochastic growth. *Math. Biosci.* 41: 112-123.

-May, R. M., Beddington, J. R., Horwood, J. H., Shepherd, J. G. (1978). Exploiting natural populations in an uncertain world. *Math. Biosci.* 42: 219-252.

<sup>36</sup> Com destaque para:

-Braumann, C. A. (1996). Estimação de parâmetros em modelos de crescimento e pesca em ambientes aleatórios. Em *Bom Senso e Sensibilidade. Traves Mestras da Estatística. Actas do III Congresso da SPE*, J. Branco, P. Gomes e J. Prata (eds.), p. 103-117, Sociedade Portuguesa de Estatística e Edições Salamandra, Lisboa, 1996.

<sup>37</sup> Veja-se, por exemplo:

-Braumann, C. A. (1985). Stochastic differential equation models of fisheries in an uncertain world: extinction probabilities, optimal fishing effort, and parameter estimation. Em *Mathematics in Biology and Medicine*, V. Capasso, E. Grosso e S. L. Paveri-Fontana (eds.), p. 201-206, Springer, Berlin.

também determinados, em que existe uma densidade estacionária). Estas classes de políticas de pesca são as habitualmente consideradas mas, em condições extremas, são pouco realistas. Por isso, estudámos<sup>38</sup>, para os mesmos modelos de crescimento populacional: i) políticas de quota constante com truncatura (paragem da pesca) quando a população desce abaixo de certo limiar (por não ser realista manter a quota nestas condições e para evitar a extinção certa); ii) políticas de esforço constante com truncatura (fixação da quota provocada por razões logísticas) quando a população excede certo limiar. O seu estudo é particularmente difícil dado que o esforço de pesca  $c(x)$  não é função diferenciável e não são aplicáveis os teoremas clássicos de existência e unicidade de solução para EDE. Conseguimos contudo, determinar condições suficientes de existência de densidade estacionária (“equilíbrio estocástico”), determinando neste caso a taxa média de capturas em regime sustentado e o seu valor óptimo. O ideal, contudo, seria determinar a política ótima para classes mais vastas (não-paramétricas) de políticas de pesca, objectivo difícil que ainda não foi tratado.

À semelhança do que fizemos para os modelos sem capturas (4), demos particular importância ao estudo de modelos gerais (6), apenas com hipóteses biologicamente razoáveis sobre a taxa de crescimento *per capita*  $r(x)$  e a intensidade do ruído  $\sigma(x)$ . Com as mesmas hipóteses que antes, conseguimos mostrar<sup>39</sup> para o modelo geral (6) haver extinção ou haver densidade estacionária (“equilíbrio estocástico”) conforme a taxa média geométrica líquida de crescimento *per capita* para populações pequenas seja negativa ou positiva. A taxa líquida é a diferença entre a taxa de crescimento *per capita* natural e o esforço de capturas  $c(x)$  (que é uma taxa suplementar de mortalidade provocada pela caça ou pesca).

A resolução da controvérsia Itô-Stratonovich para o modelo geral com capturas (6) é semelhante à do modelo sem capturas (4)<sup>40</sup>.

## 8. Breve referência à difusão das equações diferenciais estocásticas em Portugal

Depois de regressar ao País após o doutoramento, foi com muito gosto que fui aceitando os vários convites para, de uma forma ou de outra, divulgar as EDE em geral e

<sup>38</sup> Veja-se:

-Braumann, C. A. (2002). Política de pesca com truncatura em ambiente aleatório. Em “Novos Rumos em Estatística. Em *Novos Rumos em Estatística. Actas do IX Congresso Anual da SPE*, L. Carvalho, F. Brilhante e F. Rosado (eds.), p. 137-145, Sociedade Portuguesa de Estatística, Lisboa.

-Braumann, C. A. (2001). Constant effort and constant quota fishing policies with cut-offs in random environments. *Natural Resource Modelling* 14 (2): 199-232.

-Braumann, C. A. (2001). Políticas de pesca de quota constante com truncatura em ambiente aleatório. Em *Um Olhar sobre a Estatística. Actas do VII Congresso Anual da SPE*, P. Oliveira e E. Athayde (eds.), p. 431-446, Sociedade Portuguesa de Estatística, Lisboa.

<sup>39</sup> O resultado mais geral (saíram primeiro outros menos gerais) pode ver-se em:

-Braumann, C. A. (2002). Variable effort harvesting models in random environments: generalization to density-dependent noise intensities. *Math. Biosci.* 177 & 178: 229-245.

<sup>40</sup> Braumann, C. A. (2005). Comparação de modelos determinísticos e estocásticos de capturas em ambiente aleatório. Em *Actas do XII Congresso Anual da Sociedade Portuguesa de Estatística* (aceite para publicação).

a sua aplicação em dinâmica de populações em particular (neste último caso, não só em Portugal mas em vários Países europeus, mas falarei do nosso). Normalmente eram conferências ou ciclos de conferências em instituições ou em reuniões científicas, mas houve intervenções mais estruturadas como o ter leccionado em dois anos lectivos consecutivos (1985/86 e 1986/87) uma disciplina de Equações Diferenciais Estocásticas no mestrado em Matemática Aplicada do Instituto Superior Técnico, o de ter leccionado disciplina similar numa Escola de Verão da Sociedade Portuguesa de Matemática (1989) e também no Instructional Meeting in Financial Mathematics no CMAF (1998) e, mais recentemente (2004), na Autumn School on Financial Mathematics organizada pelo Centro Internacional de Matemática em Coimbra<sup>41</sup>.

Mas também não esqueço a conferência convidada (“O Acaso, a Bolsa e a Vida”) no V Congresso Anual da SPE na Curia<sup>42</sup> em 1997, logo a seguir à atribuição do Nobel a Scholes e Merton pelo seu trabalho nas aplicações financeiras Julgo que ela atraiu algumas pessoas para as EDE (mas não para as aplicações biológicas), a julgar por trabalhos<sup>43</sup> que apareceram depois em Actas de Congressos da SPE.

Claro que já havia nas Actas dos Congressos da SPE artigos de pessoas como Paula Milheiro de Oliveira<sup>44</sup>, que tem trabalho continuado na área das EDE e dos problemas de filtragem associados a observações com ruído e que muito tem contribuído também para a sua divulgação.

<sup>41</sup> O manual do curso encontra-se em:

-Braumann, C. A. (2004). Introduction to stochastic differential equations. Em *Stochastic Finance*, P. E. Oliveira (ed.), p. 5-41, Centro Internacional de Matemática 26, Coimbra.

<sup>42</sup> Pode ser lida em:

-Braumann, C. A. (1998). O Acaso, a bolsa e a vida. Em *Estatística: A Diversidade na Unidade. Actas do V Congresso Anual da SPE*, M. Souto de Miranda e I. Pereira (eds.), p. 29-55, Sociedade Portuguesa de Estatística e Edições Salamandra, Lisboa.

<sup>43</sup> Por exemplo, em *Um Olhar Sobre a Estatística. Actas do VII Congresso Anual da SPE*, P. Oliveira e E. Athayde (eds.), Sociedade Portuguesa de Estatística, 2001, podem ver-se:

-Figueira, J. e Ferreira, M. A. Aplicação de uma equação diferencial estocástica linear ao estudo de um fundo de pensões, p. 111-122.

-Mexia, J. T. Differential equations for moment generating functions of diffusions, p. 179-185.

-Guerra, H., Oliveira, M., Inácio, S. e Mexia, J. T. Inferência para o processo de Ornstein-Uhlenbeck, p. 306-316.

-Almeida, R. M. Modelos financeiros com saltos, p. 391-404.

Também se encontra:

-Figueira, J. e Ferreira, M. A. (2001). Sobre o tempo de primeira passagem em processos de difusão, a propósito da modelação de um fundo de pensões. Em *A Estatística em Movimento. Actas do VIII Congresso da SPE*, M. M. Neves, J. Cadima, M. J. Martins e F. Rosado (eds.), p. 185-191, Sociedade Portuguesa de Estatística, Lisboa.

<sup>44</sup> Como por exemplo:

-Milheiro de Oliveira, P. e Valente, P. A. (1997). Um modelo estocástico em tempo discreto para a simulação de trajetórias de uma bóia sujeita a correntes de maré e à força do vento. Em *A Estatística a Decifrar o Mundo. Actas do IV Congresso Anual da SPE*, R. Vasconcelos, I. F. Alves, L. Canto e Castro, D. Pestana (eds.), p. 399-408, Sociedade Portuguesa de Estatística e Edições Salamandra, Lisboa.

-Basto, M. J. e Milheiro de Oliveira, P. (1997). Seguimento da trajetória de uma bóia sujeita a uma corrente marítima. *Idem*, p. 409-417.

-Milheiro de Oliveira, P. (1998). Um filtro aproximado para uma difusão não linear bidimensional medida através de observações unidimensionais com ruído fraco. Em *Estatística: A Unidade na Diversidade. Actas do V Congresso Anual da SPE*, M. Souto de Miranda e I. Pereira, p. 523-528, Sociedade Portuguesa de Estatística, Lisboa.

Na minha própria Universidade, para além de divulgações em seminários, foi possível criar nos últimos anos condições para, na disciplina de Processos Estocásticos II (optativa da licenciatura em Matemática e Ciências da Computação), fazer uma introdução às EDE e suas aplicações. A nível do Mestrado em Matemática e Aplicações tem havido nos últimos anos divulgação num capítulo da disciplina de Modelos Matemáticos em Biologia. Este trabalho está já a dar frutos, começando a aparecer estudantes interessados em fazer trabalho de fim de curso<sup>45</sup>, mestrado ou doutoramento na área das EDE, embora só alguns virados para aplicações biológicas. Brevemente deve ser possível pôr a funcionar no Mestrado uma disciplina autónoma de EDE e suas aplicações.

Várias outras instituições universitárias têm também ensino de final de licenciatura ou de mestrado em EDE, principalmente nas áreas de Estatística e de Economia e Gestão, de que falei ou falarei um pouco. Mas também investigadores inicialmente provenientes da área da análise estocástica (certo de cometer omissões recordo Ana Bela Cruzeiro, Isabel Simão, Teresa Arede, Manuel Esquível, Jorge Salazar) e da área das Telecomunicações (como José Fonseca de Moura) têm dado o seu contributo para o ensino e divulgação das EDE, mas como conheço pouco destes sectores, fico-me por esta referência genérica. Provavelmente haverá mais áreas onde tal também sucede.

A minha formação e tema de investigação nunca me tinham levado a entrar nas aplicações financeiras e económicas, que mais tarde vim a achar particularmente interessantes. Lembro-me que em 1981 no Fundão, na exposição-síntese que fui convidado a apresentar no II Colóquio de Estatística e Investigação Operacional, ter falado vagamente do possível interesse das aplicações financeiras. Lembro-me também da prudência, para não dizer cepticismo, que o Professor Bento Murteira manifestou. Mas, entretanto, essas aplicações foram ganhando dimensão e o Professor Bento Murteira não teve mais dúvidas sobre a sua importância e desafiou-me a fazer um ciclo de conferências sobre EDE no Instituto Superior de Economia e Gestão em 1995. Não se comoveu com os meus argumentos de que eu nada sabia sobre aplicações financeiras e económicas (ele pretendia que eu falasse da teoria das EDE) e foi com entusiasmo que aceitei o convite e que até estudei um pouco (mesmo pouco) sobre tais aplicações para pelo menos poder dar um ou outro exemplo mais motivador. Foi a minha intervenção mais profícua, já que, passado poucos anos vários Professores do ISEG começaram a trabalhar na área e a orientar teses de mestrado de grande qualidade, como pude constatar na minha participação nalguns desses júris. Mas o fruto mais produtivo desse trabalho foi o interesse do agora Professor João Nicolau em fazer a sua tese de doutoramento<sup>46</sup> na área, orientado pelo Professor Nuno Cassola e Barata e por mim. Fez um excelente trabalho com teoria nova em problemas de estimação em EDE (especialmente estimação não-paramétrica) e

---

<sup>45</sup> Referi atrás alguns, mas também Gonçalo Jacinto e Telma Bernardo fizeram trabalhos de fim de curso sobre EDE, mas com teorias e aplicação viradas para o problema da obtenção da fórmula de Black-Scholes para as opções europeias.

<sup>46</sup> Nicolau, J. (2001). *Modelação e Estimação de Séries Financeiras através de Equações Diferenciais Estocásticas*. Tese de doutoramento em Matemática Aplicada à Economia e Gestão, Instituto Superior de Economia e Gestão.

na modelação inovadora de diversas séries financeiras. Tem continuado a trabalhar a bom ritmo após o doutoramento e a também ele a divulgar as EDE.

Na sequência dessa intervenção no ISEG e do que, como consequência dela, aprendi sobre as aplicações financeiras, proporcionou-se fazer a conferência no Congresso da Cúria já referida. Introduzindo-lhe pequenas modificações de cada vez, viria a repeti-la quatro vezes nos anos de 1998 e 1999. Primeiro, a convite do Instituto do Mercado de Capitais, apresentei-a em Lisboa (no ISEG) e no Porto (na Bolsa de Derivados), a que se seguiu o Seminário de Estatística e Aplicações do Porto e a Universidade dos Açores. Começava por usar o modelo de Black-Scholes de uma acção (coincidente com o modelo malthusiano de crescimento de uma população em ambiente aleatório) como pretexto para apresentar as bases teóricas das EDE e, em seguida abordava aplicações biológicas e financeiras, com relevo para a aplicação às opções europeias e correspondente fórmula de Black-Scholes. Foram conferências muito frequentadas, o que constituiu uma excelente oportunidade de dar a conhecer esta importante área do saber.

Mas as EDE têm aplicações em praticamente todos os sectores da Ciência e da Tecnologia, pelo menos sempre que os modelos tradicionais prevalecentes sejam equações diferenciais e o fenómeno por elas descrito sofra perturbações aleatórias. Há que dar a conhecer a esses investigadores que dispõem de mais um utensílio de modelação que poderão utilizar, com a ajuda dos Matemáticos e dos Estatísticos quando necessário.

Felizmente há um número razoável de pessoas a trabalhar em EDE em Portugal que podem envolver-se nessa nobre função e no ensino regular destas equações e das suas aplicações que, como já referi, está a espalhar-se. Façamos todos nós um esforço continuado para que esteja disponível nos currículos dos programas de reciclagem dos actuais Estatísticos e dos cursos que formarão os Estatísticos de amanhã. Para que, num futuro não muito longínquo, as EDE sejam mais um poderoso instrumento ao serviço da VIDA profissional desses domadores do ACASO que são os Estatísticos.

## **Carlos Alberto dos Santos Braumann**

Nascido em 1951.09.04, licenciou-se em Matemática Aplicada em 1973 na Universidade de Luanda, onde ingressou como assistente eventual, tendo transitado em 1975 para a Universidade de Évora. Doutorou-se em 1979 na State University of New York at Stony Brook. É actualmente (desde 1989) professor catedrático da Universidade de Évora, tendo aí desempenhado, entre muitas outras funções, os cargos de Presidente do Departamento de Matemática, Director do Centro de Investigação em Matemática e Aplicações, Presidente do Conselho Científico e Vice-Reitor.

A sua investigação e maioria das publicações inserem-se maioritariamente na área das equações diferenciais estocásticas e suas aplicações biológicas (e também financeiras). É, desde 1992, membro eleito do International Statistical Institute.

Foi Presidente da Comissão Organizadora e Editor das Actas das XV Jornadas Luso-Espanholas de Matemática (Évora, 1990). Foi representante do Conselho de Reitores das Universidades Portuguesas no Conselho Superior de Estatística e integrou a Comissão de Avaliação das Universidades Portuguesas na área de Matemática. É, pela segunda vez, Presidente da Mesa da Assembleia Geral da Sociedade Portuguesa de Estatística e preside à Comissão Organizadora do XII Congresso Anual da Sociedade (Évora, 2004).

# Estatística Robusta: contribuição portuguesa

João A. Branco

## 1. Introdução

Como se sabe, e é conhecido desde longa data, se as hipóteses em que assentam os procedimentos da estatística clássica não forem satisfeitas, esses procedimentos podem perder a sua qualidade de procedimentos óptimos ou ter mesmo um comportamento desastroso. Este problema de falta de robustez sempre preocupou os investigadores e utilizadores da estatística em geral, mas só a partir dos anos sessenta do século XX foi atacado de forma sistemática, o que levou ao estabelecimento e desenvolvimento dos fundamentos da teoria da robustez. Neste texto pretende-se realçar a necessidade da utilização de procedimentos robustos na prática e rever algumas das ideias e ferramentas que formam a chave para a avaliação da robustez e para a construção de procedimentos robustos. Faz-se uma breve referência aos estudos de robustez em análise multivariada e à contribuição dada pelos investigadores portugueses no desenvolvimento deste campo de trabalho, contribuição que, por sinal, toda ela ocorreu no período dos 25 anos de existência da Sociedade Portuguesa de Estatística que agora se celebram.

A palavra robustez é usada na linguagem corrente para designar a qualidade daquilo que é robusto (termo derivado da palavra latina *robustus*). Por sua vez o significado do termo robusto é: forte, vigoroso, resistente, saudável, bem constituído. Por exemplo, tanto se fala de um atleta robusto como se fala de um vinho robusto.

Do ponto de vista técnico o termo robusto aplica-se geralmente a um sistema ou processo que é capaz de manter determinado comportamento mesmo quando haja perturbação das condições habituais do seu funcionamento.

Este entendimento genérico do que é um sistema robusto assume interpretações específicas em cada uma das áreas em que o conceito de robustez é relevante. Estudos de robustez são de facto importantes em muitas áreas de trabalho como as ciências naturais, a engenharia, a sociologia, a estatística e outras áreas. É comum considerar-se a robustez de um processo de produção (engenharia), de um *software* (informática), de um sistema social (sociologia), de uma medida económica (economia), de um ecossistema (ecologia), de uma decisão política (política) e de um procedimento estatístico (estatística).

Traduzindo a ideia geral de robustez nos vários contextos particulares leva a que se diga, por exemplo, que: (i) um ecossistema é robusto (Gunderson and Holling, 2001) se tem a capacidade de manter as suas funções e equilíbrio, quando submetido a perturbações, sejam elas ambientais, invasões por espécies diferentes ou outras, (ii) um *software* é robusto (Huhns and Holderfield, 2002) se funciona correctamente dentro das especificações dos seus programas e é ainda capaz de reagir bem em circunstâncias estranhas, fora das especificações definidas no seu delineamento (é capaz, por exemplo, de tolerar um grande volume de erros e *bugs* detectando essas faltas e recuperando dos seus efeitos).

Uma outra área em que o conceito de robustez é muito requerido é a área do processo de apoio à decisão. Aí consideram-se vários tipos de robustez e a expressão “Análise de robustez” tornou-se parte de muitos estudos nesta área. A Análise de robustez pode ter várias interpretações. Segundo Rosenhead (2001) trata-se de um procedimento que consiste em escolher uma acção, de um conjunto de acções possíveis, que seja suficientemente flexível para poder servir muitas das opções que possam surgir no futuro. Por exemplo, sabendo que vou viajar em Janeiro de 2007, precisamente daqui a dois anos, o que devo fazer relativamente a reservar/comprar já o bilhete, e que tipo de bilhete, para me resguardar das surpresas que o meu próprio futuro e o do país para onde vou viajar me reservam? poderei concretizar a viagem e a minha situação económica actual manter-se-á?, poderá acontecer, no país que vou visitar, uma calamidade nacional ou uma mudança política com consequências na alteração do valor do câmbio local?

Em <http://discuss.santafe.edu/robustness> encontra-se uma lista de possíveis definições de robustez.

Como já se percebeu o conceito de robustez não é único, havendo várias interpretações. Aparece muitas vezes ligado às ideias de flexibilidade, insensibilidade, resistência e estabilidade, com as quais por vezes se confunde. Em qualquer dos casos o estudo da robustez tem por objectivo principal a construção de sistemas robustos, proporcionando concomitantemente uma melhor compreensão de todo o sistema e um conhecimento das hipóteses cruciais para o seu bom funcionamento, um suproduto da maior importância e que não deve ser menosprezado. A robustez está associada à ideia de tranquilidade do funcionamento ou até de sobrevivência do próprio sistema.

A necessidade de sistemas e processos robustos é muito evidente em diversas actividades. Pensando, por exemplo, na área da visão, em robótica, um bom (robusto) robot de reconhecimento deve ser capaz de absorver e interpretar novos objectos e cenários que não lhe sejam familiares. Trata-se de uma situação em que o número de *outliers* (objectos novos) pode ser muito maior do que o número de observações boas, um caso que poderá deixar muitos estatísticos surpreendidos. No caso da estatística o uso de métodos robustos, isto é, métodos que se comportam bem mesmo quando as hipóteses ideais de funcionamento não são rigorosamente verificadas, é muito apreciado pelos estatísticos uma vez que a aplicação forçada dos métodos tradicionais, quando se verifica a violação daquelas hipóteses, pode conduzir a resultados bastante insatisfatórios.

Este artigo ocupa-se da robustez em estatística e tem por objectivo dar uma visão resumida de aspectos gerais da estatística robusta e referir, brevemente, o problema da robustez em análise multivariada, destacando a contribuição que os investigadores portugueses têm dado para o desenvolvimento deste tema. Assim: na Secção 2 ilustram-se os perigos da falta de robustez e conta-se um pouco da evolução do interesse pela estatística robusta; na Secção 3 faz-se uma breve análise crítica ao fraco uso da estatística robusta na prática; na Secção 4 resumem-se os principais conceitos das teorias de Huber e Hampel; na Secção 5 indicam-se as abordagens mais comuns usadas na construção de estimadores robustos multivariados; finalmente na Secção 6 destaca-se a contribuição dos investigadores portugueses neste último tópico.

## 2. Estatística robusta

Os ingredientes de uma análise estatística incluem, em geral, um conjunto de dados, um modelo e procedimentos estatísticos vários (testes e métodos de estimação). O bom funcionamento destes procedimentos requer que se respeitem certas hipóteses como, por exemplo, a normalidade das observações, a sua independência e identidade em termos de distribuição (i.i.d.), e ainda homogeneidade de variâncias, linearidade e estacionaridade. Se alguma ou várias destas hipóteses forem violadas os resultados de muitos dos procedimentos estatísticos clássicos podem tornar-se tão aberrantes que deixam de merecer qualquer credibilidade.

Procedimentos com este comportamento são designados de não robustos. Por sua vez procedimentos robustos são aqueles cujos resultados não mostram grandes alterações em presença de pequenos desvios das hipóteses assumidas. Tendo em conta que as hipóteses consideradas não passam de simples idealizações e que realmente não se verificam na prática percebe-se imediatamente a importância que têm os procedimentos robustos em qualquer análise estatística.

Antes de prosseguir vale a pena analisar um exemplo simples que serve para ilustrar os efeitos severos que pequenos desvios da hipótese da normalidade podem acarretar. Considere-se a normal  $X \sim \mathcal{N}(\mu, \sigma^2)$  e a partir dela forme-se a normal simetricamente contaminada cuja f.d. é

$$F(x) = (1 - \varepsilon)\Phi\left(\frac{x - \mu}{\sigma}\right) + \varepsilon\Phi\left(\frac{x - \mu}{k\sigma}\right),$$

onde  $\Phi$  representa a f.d. da normal reduzida,  $\varepsilon$  ( $0 \leq \varepsilon \leq 1$ ) a probabilidade de contaminação e  $k > 0$ .

Como  $\Phi((x - \mu)/\sigma)$  está associada à normal  $\mathcal{N}(\mu, \sigma^2)$  e  $\Phi((x - \mu)/(k\sigma))$  está associada à normal  $\mathcal{N}(\mu, k^2\sigma^2)$ ,  $F(x)$  é a função de distribuição da normal contaminada cujo valor médio é  $\mu$  e cuja variância é  $(1 - \varepsilon + \varepsilon k^2)\sigma^2$ . A nova variável aleatória, embora não sendo normal (apresenta caudas mais pesadas do que a normal), desvia-se pouco

da normal  $X \sim \mathcal{N}(\mu, \sigma^2)$ . De facto a distância de Kolmogorov, usada para comparar duas distribuições,

$$\max_x \left| F(x) - \Phi \left( \frac{x - \mu}{\sigma} \right) \right|,$$

é muito pequena, 0.04 para  $k = 10$  e  $\varepsilon = 0.1$  (para qualquer  $\mu$  e  $\sigma$ ). Apesar disso os efeitos dessa pequena diferença são bem visíveis, se notarmos o que se passa com a variância. Usando ainda  $\varepsilon = 0.1$  e três valores distintos de  $k$ , obtêm-se os seguintes valores para a variância da normal contaminada:

$k$	$(1 - \varepsilon + \varepsilon k^2) \sigma^2$
6	$4.5 \sigma^2$
10	$10.9 \sigma^2$
16	$26.5 \sigma^2$

O que se pode concluir é que a variância é muito sensível a desvios, mesmo que pequenos, efectuados na distribuição, neste caso concreto nas caudas da distribuição. Ou seja, uma pequena proporção da população pode ter efeitos dominantes em certos aspectos da população, como, por exemplo, a variância, o que vai, por sua vez, reflectir-se em certos procedimentos estatísticos como é o caso dos habituais intervalos de confiança. Uma amostra da distribuição contaminada daria intervalos de confiança para o valor médio da população com amplitudes maiores do que as amplitudes produzidas a partir da amostra da distribuição não contaminada, mesmo considerando, como é o caso, que a variância é conhecida (aproximadamente 2 vezes maior se  $k = 6$ , 3 vezes maior se  $k = 10$  e 5 vezes maior se  $k = 16$ ).

Quanto a estimadores sabe-se que o desvio padrão amostral,  $S_n$ , é um estimador assintoticamente óptimo para o desvio padrão da população normal (é cerca de 12% mais eficiente do que o desvio absoluto médio,  $D_n = \sum_{i=1}^n |X_i - \bar{X}|/n$ ), mas se há desvios da normal  $\tilde{S}_n$  perde rapidamente as suas qualidades. Tukey (1960), usando uma distribuição contaminada com  $k = 3$ , mostrou que com  $\varepsilon$  tão pequeno como  $\varepsilon = 0.002$ , isto é para uma contaminação de apenas 2 observações da normal  $\mathcal{N}(\mu, 9\sigma^2)$  entre 1000 observações da normal  $\mathcal{N}(\mu, \sigma^2)$ ,  $S_n$  é logo ultrapassado por  $D_n$  que se torna mais eficiente. Pode então concluir-se que nem o parâmetro  $\sigma$  da população, nem o estimador  $S_n$ , são robustos relativamente à hipótese de normalidade, e o mesmo se pode concluir para  $\mu$  e  $\bar{X}$ .

A questão que agora se coloca é a de saber se desvios em relação à normal são frequentes e o recurso a procedimentos robustos é inevitável ou se a normalidade deve prevalecer. Como já se deixou entender a distribuição normal não se encontra na prática, trata-se de uma distribuição ideal. A ocorrência de distribuições com caudas mais pesadas

do que a normal é também justificada em Huber (1981) e em Hampel *et al.* (1986) afirma-se que mesmo em dados de grande qualidade os erros estão geralmente presentes.

A necessidade de procedimentos robustos foi sentida desde longa data por estatísticos famosos que demonstraram a não robustez de várias estatísticas e usaram métodos mais robustos (segundo Stigler (1986) o desvio absoluto médio era usado em regressão antes do método dos mínimos quadrados). Entre outros destacam-se Newcomb (1886), Student (1927), Pearson (1929, 1931), Box (1953), a quem é atribuída a introdução do termo "robustez", e Tukey (1960).

As vantagens oferecidas pela robustez foram despertando o interesse de outros estatísticos e o conhecimento foi-se acumulando dando lugar a uma teoria própria da estatística robusta. Essa visão moderna da robustez surgiu só na década de sessenta do século XX, pelas mãos de Huber e de Hampel.

Huber (1964) introduz o conceito de vizinhança de um modelo paramétrico e sugere que o comportamento dos procedimentos estatísticos deve ser analisado não só usando o próprio modelo paramétrico mas também os modelos pertencentes a uma vizinhança desse modelo. Partindo do princípio que os modelos estatísticos são apenas aproximações da realidade, conclui que um bom método de estimação é aquele que se comporta bem na vizinhança do modelo paramétrico assumido. Esta ideia é depois explorada de forma a permitir a construção de estimadores robustos com boas propriedades.

Hampel (1968) apresenta uma via diferente e que se baseia na equivalência entre estimador e funcional estatístico. Para clarificar a ideia pode começar-se por associar um parâmetro a um funcional. Por exemplo, pensando no valor médio (medida de localização) pode introduzir-se o funcional  $T$  que transforma cada distribuição  $F$  no número real  $E[X]$  isto é,

$$T(F) = E[X] = \int x dF(x).$$

Esta equivalência pode ser aproveitada para tentar atingir o objectivo pretendido, ou seja, a robustez da medida de localização, se existir. Dizendo de outra forma, pretende-se que pequenas alterações em  $F$  produzam pequenas alterações em  $E[X]$ , o que leva a sugerir que  $T$  deve possuir a propriedade de ser contínuo. Este tipo de robustez designa-se por robustez qualitativa e a sua análise é feita, como se percebe, usando ferramentas do cálculo diferencial. Outros conceitos essenciais ao estudo da robustez segundo esta abordagem, apresentados também em Hampel (1968), são a função de influência e o ponto de rotura.

Claro que  $T$  pode ser avaliado na distribuição empírica  $F_n$ , materializando-se então a equivalência entre estimador e funcional, como já se disse. O uso adequado das ferramentas apresentadas por Hampel conduz também a estimadores robustos com boas propriedades.

O aparecimento dos trabalhos de Huber e Hampel dá início a um novo período na história do estudo da robustez. Este período é caracterizado pela divulgação generalizada

de conhecimentos sobre robustez e por uma intensa produção científica. Os artigos de Stahel (1991) e de Portnoy and He (2000) são elucidativos quanto ao número de publicações produzidas. Surgem mais investigadores na área e começa a haver uma maior comunicação entre eles. Alguns grupos destacam-se pelo interesse continuado nos estudos de robustez como, por exemplo, o grupo suíço do ETH (Swiss Federal Institute of Technology), a escola argentina e certamente o grupo norte-americano e mais recentemente o grupo belga, talvez o grupo mais numeroso e activo na Europa.

Mais recentemente também a comunidade de estatísticos interessados no estudo da robustez sentiu força e dimensão para dar início às conferências anuais ICORS (International Conference on Robust Statistics) iniciadas em 2001 com o objectivo de constituir um fórum para apresentação de novos desenvolvimentos e aplicações da estatística robusta e interacção com outros campos da ciência em geral.

Entretanto estão em marcha outras iniciativas apoiadas pela ESF (European Science Foundation) que permitem a mobilidade dos investigadores, a sua reunião em *workshops*, tudo em prol do intercâmbio científico à volta da estatística robusta.

A par de um volumoso número de artigos a produção de livros tem sido notável: para além dos já referidos Huber (1981) e Hampel *et al.* (1986), podem citar-se, por ordem cronológica, Andrews *et al.* (1972); Rey (1978); Launer and Wilkinson (1979); Bustos e James (1980); Bierens (1981); Box *et al.* (1983); Hoaglin *et al.* (1983); Kadane (1984); Rasch and Tiku (1984); Franke *et al.* (1985); Tiku and Balakrishnan (1986); Rousseeuw and Leroy (1987); Kariya and Sinha (1989); Lawrence and Arthur (1990); Staudte and Sheather (1990); Stahel and Weisberg (1991); Marazzi (1993); Morgenthaler *et al.* (1993); Rieder (1994); Huber (1996); Jurecková and Sen (1996); Rieder (1996); Maddala and Rao (1997); Muller (1997); Hettmansperger and McKean (1998); Atkinson and Riani (2000); Insua and Ruggeri (2000); Shevlyakov and Vilchevski (2002); Dutter *et al.* (2003); Dell'Aquila and Ronchetti (2004); Hubert *et al.* (2004); Wilcox (2004); Lucas *et al.* (2005).

### 3. A estatística robusta na prática

Perante o remédio que a estatística robusta fornece para tratar o deficiente funcionamento de muitos procedimentos estatísticos clássicos, em consequência do irrealismo das hipóteses em que assentam, cabe perguntar se as capacidades da estatística robusta são de facto bem aproveitadas pelos utilizadores na prática. A resposta talvez seja “ainda não”. Huber e Hampel introduziram os fundamentos de uma teoria longa e complexa que muitos investigadores continuam a construir, possivelmente atraídos pelas muitas questões em aberto e pelo aliciante trabalho matemático requerido na procura de soluções. Um interesse e esforço semelhantes não parece terem sido devotados à preparação dos resultados dessas teorias para utilização imediata dos praticantes da estatística. Esta preocupação existe desde longa data como se percebe, por exemplo, lendo o artigo “Do robust estimators work with real data?” de Stigler (1977), o artigo de Yohai

*et al.* (1991) cujo resumo começa com a frase elucidativa “Even if robust regression estimators have been around for nearly 20 years, they have not found widespread application”, ou reflectindo ainda nos objectivos do programa da rede europeia de estatísticos actualmente em actividade, SACD (Statistical Analysis of Complex Data with Robust and Related Statistical Models).

Os exemplos desta preocupação são muitos e as razões para o menor sucesso da utilização da estatística robusta na prática são várias, entre as quais se destacam:

- (i) os métodos robustos disponíveis são em grande número e não há geralmente uma directiva para que o utilizador possa fazer a selecção que lhe convém.
- (ii) o *software* necessário à operacionalidade dos métodos robustos não está suficientemente divulgado, com excepção feita aos programas disponíveis no S-Plus, R e SAS.
- (iii) a divulgação dos métodos robustos não tem sido eficiente, isto é, feita em termos acessíveis e atractivos para o utilizador – basta observar que a maior parte dos livros publicados são de índole teórica e os livros sobre aplicações são menos e não são simples (uma excepção é certamente o livro de Rousseeuw and Leroy, 1987) e ainda que a inclusão de conceitos básicos de estatística robusta e de explicações sobre os seus objectivos em livros de estatística de nível elementar ou intermédio é praticamente inexistente.

A falta de robustez de certos procedimentos clássicos e a razão dessa falta de robustez deve ser denunciada cedo para que não continue a perpetuar-se o mito da normalidade e de outras hipóteses irrealistas em que estes procedimentos assentam. A compreensão dos perigos da falta de robustez é, por isso, muito útil para que se possa conduzir com consciência uma análise estatística e é talvez mais importante, pelo menos no momento actual, do que o uso inseguro de métodos robustos sofisticados.

Ao comum utilizador da estatística convém respostas claras relativamente às dúvidas que lhe possa suscitar o funcionamento dos métodos que utiliza no seu dia a dia. Por exemplo, o teste  $t$  para comparação de duas médias ou o teste  $F$  em que se baseia a análise de variância são robustos?

Pensando no teste  $t$  as hipóteses envolvidas são a independência (quer entre as observações, quer entre as duas amostras), a normalidade e a igualdade das variâncias. Quais as consequências se alguma ou várias destas hipóteses são violadas e o que fazer nesse caso? No caso da falha da independência entre as duas amostras o teste não é válido embora possa ser substituído pelo teste  $t$  para amostras emparelhadas. No caso da normalidade não se verificar sabe-se que pequenos desvios da normal não perturbam significativamente a estatística  $t$ , principalmente se as amostras têm dimensões grandes e aproximadamente iguais, e a hipótese nula é verdadeira. No entanto a potência do teste é muito sensível até a pequenos desvios da normalidade. Pode dizer-se que o teste é robusto em relação a desvios da normalidade (desvios esses que podem assumir a forma de

*outliers*), sob  $H_0$ , mas pode apresentar-se com fraca potência para certas distribuições não normais, isto é, não é robusto sob  $H_1$ . É sabido, no entanto, que o teste é o mais potente no caso das hipóteses ideais se verificarem. A transformação dos dados é por vezes a salvação desta situação indesejável. Outra possibilidade consiste no uso de testes não paramétricos.

Quanto à hipótese da igualdade das variâncias sabe-se que o teste  $t$  é razoavelmente robusto se as dimensões das amostras são iguais. Mas se as dimensões das amostras forem diferentes e a amostra de menor dimensão tiver a maior variância a falha da igualdade das variâncias pode conduzir a grandes alterações nos resultados. Neste caso aconselha-se o teste  $t$  de Welch-Satterthwaite que tem propriedades de robustez semelhantes às do teste  $t$  na situação das variâncias serem iguais.

Perante esta descrição, como reagir à afirmação corrente, encontrada principalmente na literatura das aplicações da estatística: os métodos de análise de variância são robustos. Qual o significado desta afirmação?

E o que fazer quando se lê?

The power (of the t-test) is very sensitive even to small deviations from normality. For not too small samples, there are other tests, such as the Wilcoxon-(Mann-Whitney U)-test, with a much better behavior. (Hampel, 2000)

On the other hand the t-test is so robust against non-normality that there is really no need to use the Wilcoxon test. (Rasch and Guiard, 2004)

#### 4. Ferramentas básicas para avaliar a robustez

O trabalho que se apresenta nesta secção limita-se a aspectos de robustez relativos a estimadores.

A teoria desenvolvida por Huber (1964), conhecida por abordagem minimax de Huber, considera o chamado modelo para erros grosseiros ("gross error model")

$$F(x) = (1 - \varepsilon)\Phi(x) + \varepsilon H(x),$$

onde  $H(x)$ , é uma distribuição arbitrária. Este modelo pode ser visto como uma mistura de dados puramente normais com erros grosseiros (resultantes de anomalias ligadas ao processo de medição e registo, podendo também ser *outliers*) em quantidade  $\varepsilon$ . Partindo da família de estimadores-M, também introduzidos em Huber (1964) e usando o modelo  $F$  é possível construir estimadores com boas propriedades e úteis nas aplicações.

A teoria de Hampel (1968), conhecida por abordagem infinitesimal, é próxima da de Huber mas a sua aplicação na prática é mais eficaz e mais apelativa. Os três principais

conceitos desta teoria, amplamente analisados em Hampel *et al.* (1986), são a robustez qualitativa, a função de influência e o ponto de rotura (*breakdown point*). Fazendo uso destes conceitos pode avaliar-se o grau de robustez de um estimador e construir estimadores robustos, cumprindo o que é um dos principais objectivos da estatística robusta.

A robustez qualitativa, já mencionada, traduz-se essencialmente pela continuidade de um estimador entendido como funcional.

### Função de influência

Considere-se uma estimativa construída a partir de uma amostra recolhida numa qualquer distribuição  $F$ . Se a essa amostra juntarmos uma nova observação num ponto  $x$  o valor da estatística fica naturalmente alterado. Essa alteração na estimativa é a influência a que nos referimos. A função de influência ( $IF$ ) pretende medir, de forma standardizada, essa alteração (em cada ponto  $x$ ). Formalmente tem-se

$$IF(x; T, F) = \lim_{\varepsilon \rightarrow 0} \frac{T((1 - \varepsilon)F + \varepsilon\Delta_x) - T(F)}{\varepsilon}$$

se o limite existe, onde  $\Delta_x$  é a função de distribuição correspondente à situação degenerada de probabilidade unitária no ponto  $x$  e  $T(\cdot)$  representa o funcional equivalente ao estimador em causa.

A função de influência descreve a primeira derivada do funcional  $T(\cdot)$  em  $F$  e é esta a justificação responsável pela designação de abordagem infinitesimal.

Um caso em que a  $IF$  é fácil de calcular é o caso do estimador média amostral,  $T_n = \sum_{i=1}^n X_i/n$ , construído com base na amostra da população  $X$ , com  $E[X] = \mu$ . Mostra-se que  $IF(x; T, F) = x - \mu$ . O facto desta  $IF$  não ser limitada indica que a alteração (influência) provocada por uma só observação no valor da estimativa pode ser tão grande quanto se quiser, o que revela um comportamento não robusto.

Do ponto de vista da robustez interessa, ao contrário do exemplo, que a  $IF$  seja limitada. Diz-se até que um estimador é B-robusto se a chamada sensibilidade a grandes erros, definida por

$$\gamma^* = \sup_x |IF(x; T, F)|,$$

for finita.

### Ponto de rotura

Enquanto que a  $IF$  fornece uma medida do comportamento local do estimador, o ponto de rotura é uma medida global de robustez (quantitativa ou global) que indica a

percentagem máxima de contaminação tolerada pelo estimador, antes que os valores da estimativa se tornem absurdos. Claro que, do ponto de vista da robustez, interessam os estimadores com pontos de rotura que sejam grandes (50% é o máximo valor que faz sentido para o ponto de rotura). No caso do estimador  $T_n = \sum_{i=1}^n X_i/n$  o ponto de rotura é zero, o pior que se pode desejar e que vem confirmar a falta de robustez deste estimador já denunciada pela sua  $IF$ .

Em termos matemáticos a ideia de rotura pode ser interpretada como uma singularidade e o ponto de rotura é a distância (medida em termos de contaminação) entre zero e a próxima singularidade.

Uma analogia que facilita a compreensão destes conceitos é a analogia com a estabilidade, quer seja ela relativa ao estado psicológico de uma pessoa sujeita a *stress* ou ao estado de uma estrutura (por exemplo, uma ponte) sujeita a forças (cargas) diversas. A  $IF$  mede a alteração ou deformação momentânea devida a *stress* (carga) pontual e o ponto de rotura corresponde ao valor do *stress* (carga) que provoca o esgotamento (colapso).

Com base nestes conceitos Hampel (1974) considera que as propriedades básicas que se devem exigir de um estimador são:

- que seja qualitativamente robusto, isto é, que responda a pequenas perturbações com pequenos efeitos;
- que tenha um ponto de rotura elevado, isto é, que produza informação útil mesmo quando o modelo está muito contaminado;
- que tenha sensibilidade a grandes erros finita, isto é, que a influência provocada por uma única observação seja limitada;
- que tenha baixa sensibilidade local a deslocamentos,

$$\lambda^* = \sup_{x \neq y} \frac{|IF(x; T, F) - IF(y; T, F)|}{|x - y|},$$

isto é, que reaja suavemente, por exemplo, a erros de arredondamento e/ou agrupamento de dados – note-se que para a mediana  $\lambda^* = +\infty$  enquanto que o ponto de rotura é 50%, ou seja um estimador pode ser robusto segundo um critério e não o ser segundo um critério diferente;

- que seja capaz de isolar observações muito afastadas da massa dos dados, isto é, que seja capaz de detectar *outliers* – Hampel (1974) define ponto de rejeição de *outliers*

$$\rho^* = \inf \{r > 0 : IF(x; T, F) = 0 \text{ para } |x| > r\},$$

que convém, do ponto de vista da robustez, ser um valor finito não demasiado elevado;

- que seja consistente sob o modelo, isto é, que não tenha um comportamento desastroso sob o modelo;
- que não perca muita eficiência (não mais do que 5% ou 10%) comparativamente com um bom estimador clássico para o modelo assumido.

Com as ferramentas apresentadas acima é possível analisar as diversas propriedades dos muitos estimadores que têm sido propostos como alternativa aos mais variados estimadores clássicos. Desses estimadores, os estimadores-M (Huber, 1964) são possivelmente os mais conhecidos por serem uma extensão lógica dos populares estimadores de máxima verosimilhança. Como se sabe o estimador de máxima verosimilhança de  $\theta$  associado ao modelo paramétrico  $f(x; \theta)$  minimiza  $\sum -\log(f(X_i; \theta))$ , ou nas condições de regularidade habituais, é solução de

$$\sum_{i=1}^n \frac{\partial}{\partial \theta} \log(f(X_i; \theta)) = 0.$$

A um estimador que minimiza  $\sum \rho(X_i; \theta)$  ou, ainda mais geralmente que é solução de  $\sum \psi(X_i; \theta) = 0$ , onde  $\psi$  é uma função arbitrária que deve ser escolhida convenientemente, chama-se estimador-M. Assim todo o estimador de máxima verosimilhança é um estimador-M, mas a inversa só é verdadeira sob determinadas condições para a função  $\psi$ .

Como já se evidenciou os vários critérios podem entrar em conflito. É difícil obter estimadores que sejam ótimos segundo tantos requisitos. O estimador conveniente deverá surgir de uma análise cuidada de toda a informação disponível, que vai desde a formulação inteligente da pergunta inicial até ao bom uso das ferramentas e da experiência do utilizador.

Depois desta descrição fica claro que a estatística robusta vai mais além do que aquilo que a ideia “estatística de modelos paramétricos aproximados com vista a construir métodos estatísticos robustos” pode sugerir.

A estatística robusta serve com eficácia o problema da detecção de *outliers*, a sua prática lança luz na análise estatística de problemas e as suas teorias brigam com os próprios fundamentos da inferência estatística.

Não se pense, contudo, que a estatística robusta é a única saída para os problemas relacionados com os desvios das hipóteses habituais e com a instabilidade dos procedimentos estatísticos. Outras abordagens são possíveis, como os métodos não paramétricos e semiparamétricos, os métodos de diagnóstico e os métodos para identificação e rejeição de *outliers*. Os objectivos destes métodos partilham também os objectivos da estatística robusta.

## 5. Métodos multivariados robustos

A necessidade de métodos robustos em análise multivariada é bem mais premente do que no caso univariado. A razão desta necessidade acrescida prende-se com os seguintes factos: (i) o ambiente multidimensional é menos propício à verificação das hipóteses estatísticas habitualmente assumidas, (ii) a presença de *outliers* em dados multivariados acarreta grandes dificuldades ao analista, já que estes são difíceis de detectar e os métodos multivariados são, em geral, muito sensíveis a estas observações.

O problema da robustez em análise multivariada reduziu-se inicialmente a procurar estimadores robustos para os habituais parâmetros de localização/posição ( $\mu$ ) e de dispersão/escala ( $\Sigma$ ). Os estimadores habituais destes parâmetros que são, respectivamente, a média ( $\bar{x}$ ) e a matriz de covariâncias ( $S$ ) da amostra têm um comportamento que é, como se sabe, óptimo sob o modelo normal multivariado, mas não são robustos.

Uma ideia que surge imediatamente é substituir cada componente do vector de localização e cada elemento da matriz de covariâncias por um estimador univariado robusto adequado, método que colocaria à disposição do utilizador uma extensa lista de estimadores. Outra ideia inicialmente seguida por Gnanadesikan and Kettenring (1972) consiste em considerar a distância de Mahalanobis de cada elemento  $p$ -dimensional  $x_i$ , da amostra ( $x_1, \dots, x_i, \dots, x_n$ ) ao centro dos dados,  $d_i^2 = (x_i - \bar{x})^T S^{-1} (x_i - \bar{x})$ . O uso desta distância permite detectar *outliers* que depois são eliminados. Os dados restantes são usados para recalculiar  $\bar{x}$ ,  $S$  e  $d_i^2$ . O processo é repetido até cumprir um critério de paragem. Porém os dois tipos de estimadores não apresentam boas propriedades.

Maronna (1976) generalizou os estimadores-M de Huber (1964) criando os estimadores-M de localização e dispersão multivariados. Estes, embora sejam, sob certas condições de regularidade, consistentes e assintoticamente normais, sofrem de uma grande desvantagem pois o seu ponto de rotura é baixo e decresce para zero com  $p$ .

A procura de bons estimadores, nomeadamente estimadores com alto ponto de rotura, conduziu à proposta de dois estimadores importantes apresentados em Rousseeuw (1984):

- (i) elipsóide de volume mínimo (MVE, Minimum Volume Ellipsoid) e
- (ii) matriz de covariâncias de determinante mínimo (MCD, Minimum Covariance Determinant).

O estimador MVE selecciona ao acaso  $h$  pontos dos  $n$  pontos disponíveis, calcula o volume do elipsóide contendo esses pontos e repete este processo. O conjunto de pontos a que corresponde o elipsóide com o volume mínimo é o elipsóide de volume mínimo a partir do qual são calculados a localização (centro do elipsóide) e a matriz de dispersão (matriz que define o elipsóide). Na construção do estimador MCD usa-se basicamente o mesmo princípio mas agora é o determinante associado ao elipsóide que é minimizado.

O problema destes estimadores é que requerem grande esforço computacional. Ambos têm ponto de rotura elevado mas o MVE não tem boas propriedades assintóticas e tem

taxa de convergência abaixo da usual. O MCD é mais eficiente do que o MVE, é assintoticamente normal, tem a taxa de convergência usual e na sua versão FAST-MCD (Rousseeuw and van Driessen, 1999) não tem as dificuldades do passado sendo por isso um estimador muito querido.

Muitos dos métodos da análise multivariada actuam essencialmente sobre a matriz de covariâncias da população. Uma estratégia comum para robustificar os métodos multivariados consiste em obter estimativas robustas da matriz de covariâncias e aplicar sobre ela os métodos multivariados. Esta estratégia está hoje em desuso como se explica na secção seguinte.

## 6. Contribuição dos investigadores portugueses

O interesse pelo estudo da robustez de métodos de estatística multivariada parece ter surgido em Portugal em meados da década de 80 do século XX quando dois investigadores e docentes no Instituto Superior Técnico deram início à realização de um projecto com o objectivo de investigar a robustez das componentes principais, na expectativa de vir a construir novos métodos robustos para esta metodologia multivariada (o autor desconhece a existência em Portugal de quaisquer estudos sistemáticos ou isolados na área da estatística robusta antes desse período).

Determinante para o sucesso deste projecto e para a continuidade do trabalho de investigação nesta área foi a participação daqueles investigadores no segundo curso ECAS (European Courses in Advanced Statistics), intitulado “Robustness in Statistics - Theory and Applications”, que teve lugar em Schloss Reisenburg, Alemanha, de 2 a 7 de Outubro de 1989. O curso foi leccionado por vários dos maiores especialistas de robustez na altura e teve como objectivo introduzir os fundamentos teóricos da robustez bem como as aplicações práticas com recurso ao uso de *software* apropriado.

Lembrando que a ideia de robustez de uma estatística, entendida como a propriedade da estatística se comportar bem mesmo quando as hipóteses associadas à estatística não são inteiramente respeitadas, introduzida por Huber (1964), tinha desencadeado uma caudalosa corrente de contribuições científicas, este curso é da maior importância para todos os participantes que assistem cheios de curiosidade ao desfilar dos avanços e resultados fundamentais, muitos dos quais apresentados pelos próprios autores. O momento é mesmo de grande excitação, acabavam de ser publicadas duas obras fundamentais, uma de índole teórica (Hampel *et al.*, 1986) e outra de índole aplicada (Rousseeuw and Leroy, 1987).

O projecto seria concluído em 1990, Pires (1990). O seu resultado principal traduziu-se na construção de novos estimadores-M das componentes principais obtidas a partir da matriz de covariâncias. O novo método de estimação robusta é inspirado no método de Campbell (1980) e a novidade reside no facto do processo de construção de estimadores requerer a função de influência, não só dos valores próprios, mas também dos vectores próprios.

Mas certamente mais proveitosa do que esta nova contribuição foi o resultado do esforço de investigação realizada e os ensinamentos recolhidos numa área (métodos multivariados robustos) onde as preocupações se tinham até então centrado principalmente, em alargar ao caso multidimensional os conceitos e resultados já estabelecidos para o caso unidimensional e relativos essencialmente às questões de estimação robusta conjunta de localização e escala, e em construir estimadores robustos da matriz de covariâncias e de correlações. Aliás, no que toca à construção de métodos multivariados robustos, a directiva vigente até então era simplesmente aplicar os métodos multivariados sobre matrizes de covariância robustas. Assim se obtinham os métodos multivariados robustos. No projecto de que se fala esta tradição é abandonada e em vez do procedimento geral opta-se por uma procura directa de métodos robustos específicos de cada método multivariado, que neste caso concreto era o método das componentes principais. Esta nova abordagem usada na procura de métodos robustos em estatística multivariada passaria a ser adoptada em outros projectos de investigação que se seguiram, Souto de Miranda (1994), Pires (1995), Oliveira (2002), Rodrigues (2003) e Amado (2003). É também esta abordagem que a maioria dos investigadores usa hoje nos seus estudos de pesquisa de métodos multivariados robustos.

Num espaço de cerca de uma dúzia de anos o grupo inicial tinha-se alargado a mais cinco investigadores com os seus cursos de pós-graduação concluídos. Entretanto os conhecimentos tinham-se consolidado e as opções de investigação tinham-se clarificado. Os resultados deste trabalho continuado, apresentados regularmente sob a forma de comunicações em conferências e de publicações em revistas de estatística, têm contribuído para divulgar os métodos estatísticos robustos e alertar para a necessidade da sua aplicação na prática, sobretudo entre a comunidade estatística nacional. Um dos primeiros trabalhos realizados precisamente no espírito de divulgação foi apresentado, curiosamente, no I Congresso Anual da Sociedade Portuguesa de Estatística realizado no Vimeiro em 1993 e mais tarde publicado nas respectivas actas, Pires e Branco (1993). A mensagem terá passado e hoje sente-se que os estatísticos portugueses, com interesse em outras áreas de estudo, contam com as vantagens dos métodos robustos nos seus trabalhos de investigação, como se nota, por exemplo, em Brilhante (2003) e Figueiredo (2002).

Actualmente vários investigadores mantêm colaboração regular com especialistas de outros países e no próximo ano terá lugar em Portugal a realização da conferência ICORS2006, facto que confirma a qualidade do trabalho desenvolvido em Portugal e o reconhecimento da actividade dos seus investigadores pela comunidade internacional. O trilho foi aberto, o caminho está feito e preparado para novas viagens que se desejam robustas, para que sejam longas.

## Referências Bibliográficas

- Amado, C. (2003). *Bootstrap Robusto com base na Função de Influência*. Tese de Doutoramento, Instituto Superior Técnico, UTL, Lisboa.
- Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H. and Tukey, J. W. (1972). *Robust Estimates of Location: Survey and Advances*. Princeton University Press, Princeton, NJ.
- Atkinson, A. and Riani, R. (2000). *Robust Diagnostic Regression Analysis*. Springer-Verlag, New York.
- Bierens, H. J. (1981). *Robust Methods and Asymptotic Theory in Nonlinear Econometrics*. Springer-Verlag, New York.
- Box, G. E. P. (1953). Non-normality and tests on variances. *Biometrika*, 40, 318-335.
- Box, G. E. P., Leonard, T. and Wu, C. F. (Eds.) (1983). *Scientific Inference, Data Analysis and Robustness*. Academic Press, New York.
- Brilhante, M. F. (2003). Exponencialidade versus Pareto Generalizada - um teste rápido e robusto. In P. Brito, A. Figueiredo, F. Sousa, P. Teles e F. Rosado (Eds.), *Literacia e Estatística*, Actas do X Congresso Anual da Sociedade Portuguesa de Estatística, pp. 155-166. Edições SPE, Lisboa.
- Bustos, O. e James, K. L. (1980). *Procedimentos Robustos*, 4º Simpósio Nacional de Probabilidade e Estatística. Rio de Janeiro.
- Campbell, N.A. (1980). Robust procedures in multivariate analysis I: robust covariance estimation. *Applied Statistics*, 29, 231-237.
- Dell'Aquila, R. and Ronchetti, E. M. (2004). *Introduction to Robust Statistics with Economic and Financial Applications*. John Wiley, New York.
- Dutter, R., Filzmoser, P., Gather, U. and Rousseeuw, P. J. (Eds.) (2003). *Developments in Robust Statistics: International Conference on Robust Statistics 2001*. Physica-Verlag, Heidelberg.
- Figueiredo, F. O. (2002). *Controlo Estatístico da Qualidade e Métodos Robustos*. Tese de Doutoramento, Faculdade de Ciências, UL, Lisboa.
- Franke, J., Hardle, W. and Martin, D. (Eds.) (1985). *Robust and Nonlinear Time Series Analysis*. Springer-Verlag, New York.
- Gnanadesikan, R. and Kettenring, J. R. (1972). Robust estimates, residuals and outlier detection with multiresponse data. *Biometrics*, 28, 81-124.
- Gunderson, L. H. and Holling, C. S., (Eds.) (2001). *Panarchy: Understanding Transformations in Human and Natural systems*. Island Press, Washington, D. C..
- Hampel, F. R. (1968). *Contributions to the Theory of Robust Estimation*. Ph. Thesis, University of California, Berkeley.
- Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, 69, 383-393.
- Hampel, F. R. (2000). *Statistical Inference*. Research Report No. 93, Seminar für Statistik, ETH, Zürich.

- Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986). *Robust Statistics: The Approach based on Influence Functions*. John Wiley, New York.
- Hettmansperger, T. P. and McKean, J. W. (1998). *Robust Nonparametric Statistical Methods*. Arnold, London.
- Hoaglin, D. C., Mosteller, F. and Tukey (Eds.) (1983). *Understanding Robust and Exploratory Data Analysis*. John Wiley, New York.
- Huber, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.*, 35, 73-101.
- Huber, P. J. (1981). *Robust Statistics*. John Wiley, New York.
- Huber, P. J. (1996). *Robust Statistical Procedures* (2nd ed.). SIAM, Philadelphia.
- Hubert, M., Pison, G., Struyf, A. and Van Aelst, S (Eds.) (2004). *Theory and Applications of Recent Robust Methods*. Birkhäuser Verlag, Basel.
- Huhns, M. N. and Holderfield, V. T. (2002). Robust Software. *IEEE Internet Computing*, 6, 2, 80-82.
- Insua, D. R. and Ruggeri, F. (Eds.) (2000). *Robust Bayesian Analysis*. Springer-Verlag, New York.
- Jurecková, J. and Sen, P. K. (1996). *Robust Statistical Procedures: Asymptotics and Intercorrelations*. John Wiley, New York.
- Kadane, J. B. (Ed.) (1984). *Robustness of Bayesian Analysis*. North-Holland, Amsterdam.
- Kariya, T. and Sinha, B. K. (1989). *Robustness of Statistical Tests*. Academic Press, New York.
- Launer, E. and Wilkinson G. (Eds.) (1979). *Robustness in Statistics*. Academic Press, New York.
- Lawrence, K. D. and Arthur, J. L. (1990). *Robust Regression: Analysis and Applications*. Marcel Dekker, New York.
- Lucas, A., Franses, P. H. and Van Dijk, D. (2005). *Outlier Robust Analysis of Economic Time Series*. Oxford University Press, Oxford.
- Maddala, G. S. and Rao, C. R. (Eds.) (1997). *Robust Inference*, Handbook of Statistics 15. Elsevier Science B. V., Amsterdam.
- Marazzi, A. (1993). *Algorithms, Routines and S Functions for Robust Statistics*. Wadsworth, Inc., Belmont, CA.
- Maronna, R. A. (1976). Robust M-estimators of multivariate location and scatter. *The Annals of Statistics*, 4, 51-67.
- Morgenthaler, S., Ronchetti, E. and Stahel, W. A. (Eds.) (1993). *New Directions in Statistical Data Analysis and Robustness*. Birkhäuser Verlag, Basel.
- Muller, C. H. (1997). *Robust Planning and Analysis of Experiments*. Springer-Verlag, New York.
- Newcomb, S. (1886). A generalized theory of the combination of observations so as to obtain the best result. *American Journal of Mathematics*, 8, 343-366.
- Oliveira, M. R. (2002). *Métodos Robustos em Análise de Correlações Canónicas*. Tese de Doutoramento, Instituto Superior Técnico, UTL, Lisboa.

- Pearson, E. S. (1929). The distribution of frequency constants in small samples from non-normal symmetrical and skew populations. *Biometrika*, 21, 259-286.
- Pearson, E. S. (1931). The analysis of variance in cases of non-normal variation. *Biometrika*, 23, 114-133.
- Pires, A. M. (1990). *Estimação Robusta e sua Aplicação a Componentes Principais*. Tese de Mestrado, Instituto Superior Técnico, UTL, Lisboa.
- Pires, A. M. (1995). *Análise Discriminante: Novos Métodos Robustos de Estimação*. Tese de Doutoramento, Instituto Superior Técnico, UTL, Lisboa.
- Pires, A. M. e Branco, J. A. (2004). Estatística robusta: passado, presente e futuro. In D. Pestana, A. Turkman, J. Branco, L. Duarte e A. Pires (Eds.), *A Estatística e o Futuro e o Futuro da Estatística*, Actas do I Congresso Anual da Sociedade Portuguesa de Estatística, Novas Tecnologias/Estatística, pp. 531-549. Edições Salamandra, Lisboa.
- Portnoy, S. and He, X. (2000). A robust journey in the New Millennium. *Journal of the American Statistical Association*, 95, 1331-1335.
- Rasch, D. and Gujard, V. (2004). The robustness of parametric statistical methods. *Psychology Science*, 46, 175-208.
- Rasch, D. and Tiku, M. L. (Eds.) (1984). *Robustness of Statistical Methods and Nonparametric Statistics*. D. Reidel Publishing Company, Dordrecht, Netherlands.
- Rey, W. J. (1978). *Robust Statistical Methods*. Springer-Verlag, New York.
- Rieder, H. (1994). *Robust Asymptotic Statistics*. Springer-Verlag, New York.
- Rieder, H. (Ed.) (1996). *Robust Statistics, Data Analysis and Computer Intensive Methods*. Springer-Verlag, New York.
- Rodrigues, I. M. (2003). *Métodos Robustos em Análise de Componentes Principais Comuns*. Tese de Doutoramento, Instituto Superior Técnico, UTL, Lisboa.
- Rosenhead, J. (2001). Robustness analysis: keeping your options open. In J. Rosenhead and J. Mingers (Eds.), *Rational Analysis for a Problematic World Revisited: problem structuring methods for complexity, uncertainty and conflict*, pp. 181-207. John Wiley, Chichester, England.
- Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association*, 79, 871-881.
- Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust Regression and Outlier Detection*. John Wiley, New York.
- Rousseeuw, P. J. and van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41, 212-223.
- Shevlyakov, G. L. and Vilchevski, N. O. (2002). *Robustness in Data Analysis: Criteria and Methods*. Brill Academic Publishers, Leiden, Netherlands.
- Souto de Miranda, M. M. (1994). *Estimação Robusta com Variáveis Instrumentais em Modelos com Erros-nas-Variáveis*. Tese de Doutoramento, Universidade de Aveiro, Aveiro.

- Stahel, W. A. (1991). Research directions in robust statistics. In W. A. Stahel and S. Weisberg (Eds.), *Directions in Robust Statistics and Diagnostics, Part II*, pp. 243-278. Springer-Verlag, New York.
- Stahel, W. A. and Weisberg, S. (Eds.) (1991). *Directions in Robust Statistics and Diagnostics, Vol. 1,2*. Springer-Verlag, New York.
- Staudte, R. G. and Sheather, S. J. (1990). *Robust Estimation and Testing*. John Wiley, New York.
- Stigler, S. M. (1977). Do robust estimators work with real data? (with discussion). *The Annals of Statistics*, 5, 1055-1077.
- Stigler, S. M. (1986). *The History of Statistics*. Belknap Press, Cambridge, MA.
- Student (W. S. Gosset) (1927). Errors of routine analysis. *Biometrika*, 19, 151-164.
- Tiku, M. L. and Balakrishnan, N. (1986). *Robust Inference*. Marcel Dekker, New York.
- Tukey, J. (1960). A survey of sampling from contaminated distributions. In I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow and H. B. Mann (Eds.), *Contributions to Probability and Statistics*, pp. 448-485. Stanford University Press, Stanford, CA.
- Wilcox, R. R. (2004). *Introduction to Robust Estimation and Testing* (2nd ed.). Academic Press, San Diego, CA.
- Yohai, V.J., Stahel, W.A. and Zamar, R.H. (1991). A procedure for robust estimation and inference in linear regression. In W. A. Stahel and S. Weisberg (Eds.), *Directions in Robust Statistics and Diagnostics, Part II*, pp. 365-374. Springer-Verlag, New York.

## **João António Branco**

É licenciado em Ciências Matemáticas pela Faculdade de Ciências da Universidade de Lisboa (1966). Fez os seus estudos de pós-graduação nas universidades de Oxford (MSc. em Estatística Aplicada - 1975), Leeds e Newcastle upon Tyne (Ph. D. em Estatística - 1979). É Professor Associado do Departamento de Matemática do Instituto Superior Técnico tendo sido seu Presidente de 1984 a 1986. É membro eleito do International Statistical Institute desde 1998. Foi Presidente da Sociedade Portuguesa de Estatística durante dois mandatos consecutivos (1994 - 2000). Desde o seu primeiro contacto com a Análise Multivariada que tem interesse por esta área onde tem desenvolvido os seus estudos. Em meados dos anos oitenta, alertado para a grande actividade à volta dos métodos estatísticos robustos passou a interessar-se pelo estudo da robustez em métodos multivariados. Foi professor de seis dos sete investigadores portugueses mencionados nas referências aqui incluídas, tendo sido orientador de quatro das teses sobre robustez que estes produziram. As publicações resultantes do seu trabalho de investigação científica centram-se no domínio da Análise Multivarida e incluem também trabalhos de divulgação na área das Probabilidades e Estatística. É autor de um livro sobre Análise de Clusters publicado pela Sociedade Portuguesa de Estatística.



# Estatística dos doutoramentos em Estatística

*Maria Antónia Amaral Turkman*

Em 28 de Novembro de 1980 um entusiasmado pequeno grupo de jovens (e não tão jovens), entre os quais eu me contava, acompanhou Tiago de Oliveira a um cartório notarial na Avenida Almirante Reis para fazer a escritura da “Sociedade Portuguesa de Estatística e Investigação Operacional” (SPEIO). Nessa altura o número de doutorados nas áreas de Estatística e Investigação Operacional contava-se pelos dedos. Entretanto o rápido desenvolvimento da Investigação Operacional justificou a formação da APDIO e como tal a SPEIO passou a designar-se por SPE. Apesar da mudança de nome ter acontecido em 1991 é perfeitamente lícito celebrar durante o ano de 2005, o 25º aniversário da SPE.

Quando a Direcção da SPE, por intermédio do seu Presidente Fernando Rosado, me lançou o desafio (assim como a muitos outros) de escrever um artigo para comemorar esta efeméride, eu lembrei-me, recordando aquele dia e recordando quão poucos doutorados éramos na altura, que seria interessante fazer um pequeno estudo estatístico da evolução dos doutoramentos na área de Estatística ao longo destes 25 anos. Da ideia ao acto levou um ano... e por essa razão o trabalho que aqui apresento está bastante aquém do que eu almejava fazer.

Para coligir os dados que necessitava fiz um apelo por e-mail a um grupo inicial de cerca de 50 colegas que prontamente me responderam fornecendo não só dados sobre os seus doutoramentos como também de estudantes que orientaram. Posteriormente consegui pesquisar outros dados que me faltavam usando a base de dados dos doutoramentos do Observatório da Ciência e Ensino Superior (<http://www.oces.mcies.pt>). Nesta base de dados podem procurar-se elementos sobre doutoramentos realizados em Portugal e no Estrangeiro, com reconhecimento por Universidades Portuguesas, por área científica, ramo de conhecimento, especialidade e instituição onde o grau foi obtido, de 1970 a 2003. Embora tenha conhecimento de muitos doutoramentos realizados no Estrangeiro (nomeadamente de ex-alunos do meu departamento), optei por não incluir os daqueles que não regressaram a Portugal para evitar o enviesamento que daí adviria. Seria muito difícil não cometer graves omissões. Outro problema que se me deparou foi “que doutoramentos posso legitimamente incluir na área que genericamente designo por Estatística”? Por

exemplo, se só incluísse na lista doutoramentos que estivessem registados num ramo de conhecimento ou especialidade que contivesse a palavra “estatística” então ficariam injustamente excluídos, por exemplo, os doutoramentos realizados no Departamento de Matemática do IST, já que estes estão registados em Matemática. Assim considereei como candidatas a figurar na lista todas as teses de doutoramento da área científica da matemática, ou outra área científica, cujo tema seja notoriamente sobre Probabilidades, Processos Estocásticos ou Estatística. Apesar de me ter esforçado não fui com certeza exaustiva e peço, desde já, desculpa por qualquer omissão. Surgiu assim a longa lista que aparece em anexo (anexo II) a este texto. Os dados nessa tabela estão ordenados por ano de conclusão do grau e por apelido dentro do ano. Ao “nome” encontra-se anexada a Instituição onde a pessoa se encontrava em Janeiro de 2005. Na entrada relativa ao “local” é indicada a Instituição que conferiu o grau. Não considereei qual a Instituição Portuguesa que reconheceu o grau. Aparece também, em anexo (anexo III), uma outra lista de teses de doutoramentos que estão em curso e que, em princípio, serão discutidas em 2005. Num anexo I aparece a descrição das siglas dos locais de obtenção dos doutoramentos e de outras instituições referidas nos gráficos e no anexo II.

É interessante referir que o primeiro doutoramento realizado em Portugal sob um tema de Estatística foi o de Bento Murteira em 1953 com “Algumas Propriedades dos Processos Auto-regressivos” sob a orientação do famoso M.G. Kendall. Após Bento Murteira só passados 17 anos é que em 1970 Fátima Fontes de Sousa realizou o seu doutoramento com uma tese na área de Planeamento de Experiências. Em 1977 Fernanda Ramalhoto regressou de Inglaterra com o seu doutoramento concluído. Ela foi a primeira de uma série de doutoramentos realizados no estrangeiro graças a uma política de abertura verificada após o 25 de Abril, que permitiu a muitos assistentes das Universidades Portuguesas realizarem os seus estudos fora do País. Esse primeiro ciclo de doutoramentos terminou em 1980 e veio contribuir para o aparecimento de um pequeno grupo de doutorados na FCUL, no IST e na Universidade de Évora, que deu o ímpeto necessário ao desenvolvimento da estatística que hoje se observa em Portugal. Outra nota de interesse é que o primeiro doutoramento realizado na Universidade de Aveiro foi precisamente o de Beatriz Matias, em 1980, na especialidade de Estatística. Após 1980 começou um ciclo duplo de doutoramentos no DEIO, uns ainda fruto de estudos no estrangeiro (Helena Nicolau, Fernando Nicolau, Lucília Carvalho, Daniel Muller) e outros já genuinamente portugueses (Eugénia Graça Martins, Fernando Rosado, Helena Iglésias, Salomé Cabral). Ao mesmo tempo regressavam do estrangeiro com o seu doutoramento João Amaral para o IST, Corália Vicente, Margarida Brito e Paulo Gomes para a Universidade do Porto, Lourdes Centeno para o ISEG, Manuela Magalhães Hill e Elisabeth Reis para o ISCTE. Também Álvaro Nunes fez nessa altura o seu doutoramento em Inglaterra, estando desde há muitos anos radicado no Departamento de Matemática da Universidade de Macau. Já no fim da década de 80 regressou, de França, Esmeralda Gonçalves para Coimbra e Daniel Paulino, do Brasil, para o IST. Durante esta década poucos foram os doutoramentos realizados em Portugal fora do DEIO. Nazaré Mendes Lopes realizou o seu doutoramento

na Universidade de Coimbra sob a orientação de Jean Geffroy (depois de ter estudado em França e lá ter feito uma tese de “troisième cycle”), Francisco Calheiros realizou o seu doutoramento na Universidade do Porto, José Rodrigues Dias na Universidade de Évora (embora orientado por um docente do DEIO) e João Tiago Mexia na Universidade Nova de Lisboa, orientado por Tiago de Oliveira que entretanto tinha abandonado a FCUL e entrado para o Departamento de Matemática da FCT/UNL. Assim, no fim da década de 80, estavam lançadas as sementes para a Estatística poder germinar por todas as Instituições do País. E é esse facto que está bem patente no gráfico 1. Embora se note um crescimento suave do número de doutoramentos no País durante os anos 90, foi já no ano 2000 que se deu a verdadeira explosão. Nos anos 80 a maioria dos doutoramentos foram realizados no Estrangeiro (14) e no DEIO/FCUL (9), onde se tinha iniciado a primeira licenciatura e o primeiro Mestrado nas áreas de Probabilidades e Estatística em Portugal. Nos anos 90 o DEIO/FCUL continuou a formação dos seus assistentes e simultaneamente começou a contribuir para a formação de docentes de outras Instituições. As Universidades da Madeira e dos Açores, por exemplo, devem o seu impulso à contribuição de Dinis Pestana na formação dos seus docentes.

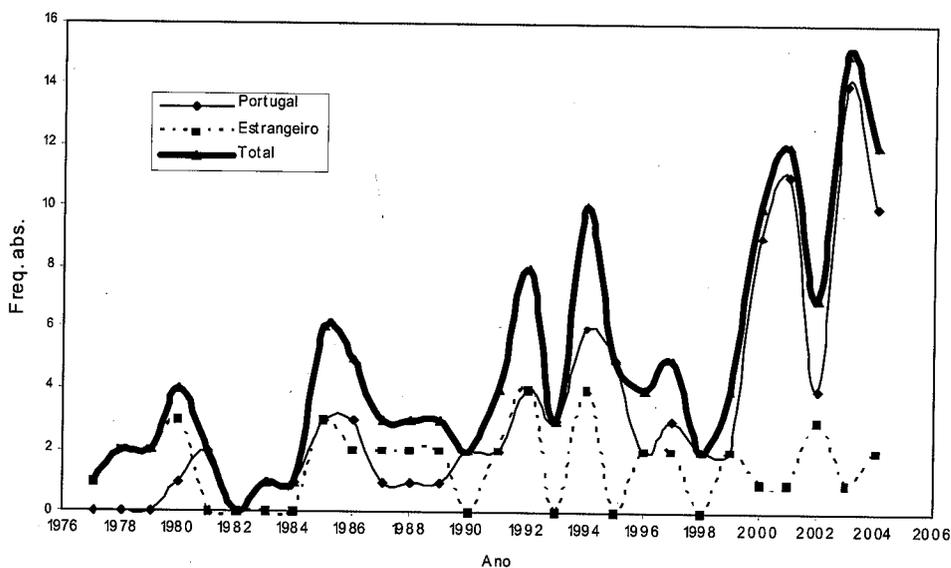


Figura 1 – Evolução do número de doutoramentos na área de Estatística e afins realizados em Portugal e no Estrangeiro de 1977 a 2004

O número total de doutoramentos desde 1977 até Fevereiro de 2005 foi de 139, dos quais 96 foram realizados em Portugal, em 18 diferentes Instituições, e 43 no Estrangeiro em 6 Países.

O gráfico da figura 2 mostra a distribuição do número de doutoramentos por local e na tabela 1 encontra-se o número de doutoramentos por local e ano. Dentro das Instituições Portuguesas verifica-se que o DEIO/FCUL detém o recorde com 34 doutoramentos, regularmente distribuídos ao longo dos anos. Claro que esta circunstância se deve ao facto de este ser o único departamento totalmente vocacionado para a Estatística (e Investigação Operacional), não tendo que partilhar os seus programas de doutoramento com outras áreas, quer da Matemática, quer de outro ramo do conhecimento ou área científica, como acontece com quaisquer outros departamentos das Instituições que estão em análise. Em 2º lugar aparece o IST/UTL com 11 doutoramentos, praticamente todos já obtidos no sec. XXI, e em 3º a FCT/UNL com 9 regularmente distribuídos desde 89, data do 1º doutoramento em estatística nessa Instituição.

Na década de 80, o ano mais produtivo foi o de 1985 (com 6 doutoramentos) e na década de 90 o de 1994 com um total de 10 doutoramentos distribuídos por 6 locais. O recorde do número de doutoramentos no sec. XXI vai em 16, conseguido no ano 2003, mas em termos do número de locais onde foram realizados, o ano 2001 detém esse recorde com 9 (incluindo um no estrangeiro). A última linha da tabela 1 leva-nos a crer que o crescimento observado não deve esmorecer tão cedo.

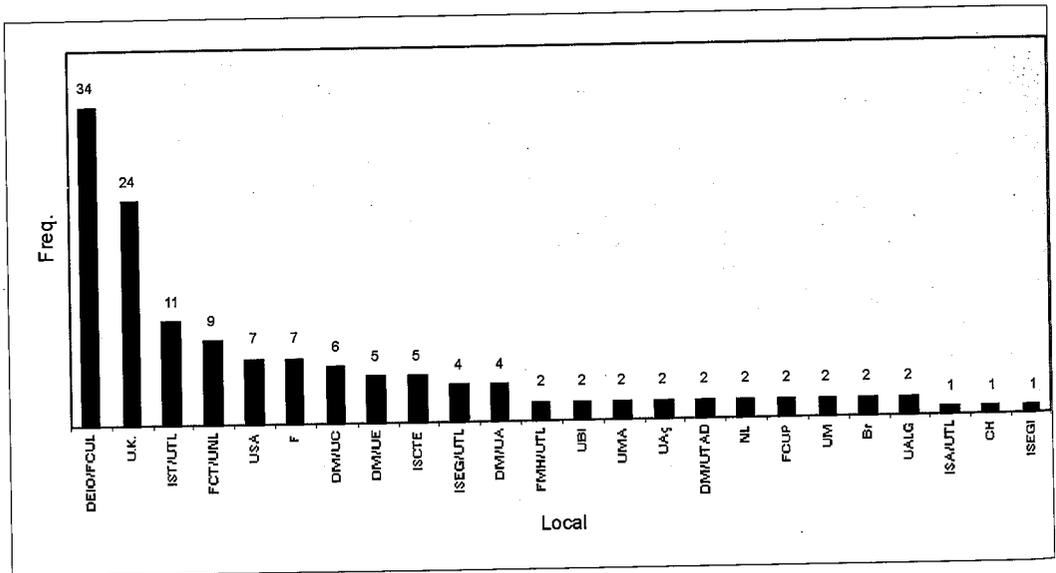


Figura 2: Distribuição do número de doutoramentos, em Estatística e afins, realizados por local

**Tabela 1**  
Distribuição do número de doutoramentos, em Estatística e afins,  
realizados por local e por ano

Ano	Estrangeiro	DEIO/FCUL	IST/UTL	FCT/UNL	DM/JC	DM/UE	ISCTE	ISEG/UTL	DM/JA	UM	DM/UTAD	FCUP	FMH/UTL	UAç	UBI	UMA	UALG	ISA/UTL	ISEGI	FEUP	TOTAL
1977	1																				1
1978	2																				2
1979	2																				2
1980	3								1												4
1981		2																			2
1982																					0
1983		1																			1
1984		1																			1
1985	3	2			1																6
1986	2	2										1									5
1987	2					1															3
1988	2			1																	3
1989	2	1																			3
1990				1							1										2
1991	2			1				1													4
1992	4	4																			8
1993		2			1																3
1994	4	2		1	1	1										1					10
1995		1			1		1		1					1							5
1996	2	1	1																		4
1997	2			1					1								1				5
1998		1							1												2
1999	2										1			1							4
2000	1	2	1	1	2			1								1			1		10
2001	1	2	2	2		1	1	1		1		1									12
2002	3		2							1					1						7
2003	1	6	2			2	3											1			15
2004	2	2	3	1				1					2				1				12
2005		2													1						3
<b>Total</b>	<b>43</b>	<b>34</b>	<b>11</b>	<b>9</b>	<b>6</b>	<b>5</b>	<b>5</b>	<b>4</b>	<b>4</b>	<b>2</b>	<b>1</b>	<b>1</b>	<b>0</b>	<b>139</b>							
2005*		6	1			1	3		1			2	1	1	3	1				1	21

No início do texto referi a minha dificuldade em decidir que teses de doutoramento deveria incluir na lista. Uma pesquisa posterior na base de dados do Observatório da Ciência e Ensino Superior permitiu-me fazer uma classificação da grande maioria das teses de doutoramento (até 2003 inclusive) por especialização, ou na falta dela, por ramo de conhecimento. Essa classificação permitiu-me construir (embora incompleta) a tabela que se segue.

**Tabela 2**

Distribuição por ramo de conhecimento ou área de especialização

Bioestatística	2
Economia	2
Engenharia de Sistemas	1
Eng. Electrotécnica e Informática	1
Estatística ou Probabilidades e Estatística	8
Matemática	20
Matemática Aplicada à Economia e Gestão	5
Métodos Quantitativos de Gestão	8

É de notar, aliás como também já referi, que há Universidades que não conferem especialização em Estatística, sendo as teses dessa área classificadas dentro da Matemática.

Há ainda um aspecto a referir relativamente à contribuição das diferentes instituições, quer para a formação do seu pessoal docente (que eu designo por formação interna), quer para a formação de docentes de outras instituições (formação externa). Para poder ser exaustiva nesta classificação teria de ter um conhecimento mais profundo de todas as instituições intervenientes. Irei pois apenas referir, dentre as instituições com maior número de doutoramentos, aquilo que é do meu conhecimento, esperando não cometer erros ou injustiças. Assim, começando pelo que melhor conheço, dos 34 doutoramentos registados no DEIO/FCUL, 20 correspondem a formação interna. Dos restantes 14, 3 correspondem a doutorados que abandonaram o DEIO, durante a década de 80, para irem para outras instituições, sendo os restantes 11 doutoramentos correspondentes a formação externa. De referir que um dos doutoramentos realizados no DEIO foi orientado por um docente exterior ao DEIO (Paulo Gomes) e correspondeu a uma formação externa. Relativamente ao IST/UTL, 8 dos doutoramentos realizados corresponderam a formação interna e 3 a formação

externa. A FCT/UNL tem um percurso diferente, tendo contribuído numa maior percentagem do que as instituições anteriores para a formação externa (44%). Relativamente à Universidade de Coimbra apenas um dos doutoramentos correspondeu a uma formação que passou a ser externa (“migração” para a UBI). A Universidade de Aveiro beneficiou, para a formação dos seus docentes, da colaboração, primeiramente do IST e posteriormente do DEIO. A colaboração a este nível, isto é docentes de uma instituição orientarem doutoramentos fora da sua própria instituição, é bastante relevante. Tiago Mexia e Ivette Gomes, por exemplo, têm dado um contributo importante nesse aspecto.

Finalmente seria interessante fazer um estudo estatístico do ponto de vista temático. Esse foi um trabalho que ainda tentei fazer à custa das palavras-chave que algumas pessoas me forneceram. Contudo, a certa altura, perdi-me na “floresta” e desisti. Com tempo poderia ter solicitado a todos que aqui estão referidos a indicação de uma área temática para a sua tese e fazer a correspondente classificação. Espero que alguém (que não eu) pegue na ideia e apresente um trabalho num próximo Congresso da SPE.

Acho que é também justo aqui fazer referência a tantos docentes, que não sendo da área da Estatística, ou não tendo formação inicial em Estatística a nível de doutoramento, tiveram um papel de relevo e incentivaram o desenvolvimento da Estatística nas suas Instituições. Muitos são já nossos conhecidos e já são figuras históricas. Outros, de certo, serão referidos, algures neste livro, por quem melhor do que eu lhes sabe prestar o verdadeiro tributo. Eu, por mim, gostava de prestar o meu apreço ao Matemático Luís Loura da Faculdade de Motricidade Humana, por ao reconhecer a importância da Estatística nas Ciências do Desporto, se ter preocupado em desenvolver, na sua escola, um pequeno, mas entusiástico, grupo na área.

Para terminar, quero apresentar uma palavra de agradecimento a todos os que prontamente colaboraram comigo para a elaboração deste texto. Espero que ele, para além de uma simples curiosidade estatística, seja útil como referência para, quanto mais não seja, a escolha de elementos jovens para possíveis Júris de Mestrado, Doutoramentos, etc. Também espero que seja útil para nos conhecermos melhor e melhor sabermos “quem somos”.

Ainda um agradecimento muito especial ao Fernando Rosado, por toda a ajuda que me deu em todas as fases deste trabalho. Sem o seu apoio, sugestões e constante encorajamento rapidamente teria desistido!

## ANEXO I

SIGLAS usadas para as Instituições e Países referidos nos gráficos e anexo II

SIGLA	LOCAL
Br	Brasil
CH	Suíça
DEIO/FCUL	Departamento de Estatística e Investigação Operacional/Faculdade de Ciências de Lisboa
DM/UA	Departamento de Matemática/Universidade de Aveiro
DM/UC	Departamento de Matemática/Universidade de Coimbra
DM/UE	Departamento de Matemática/Universidade de Évora
DM/UTAD	Departamento de Matemática/Universidade de Trás-os-Montes e Alto Douro
ESTGL	Escola Superior de Tecnologia e Gestão de Leiria
F	França
FCT/UNL	Faculdade de Ciências e Tecnologia/Universidade Nova de Lisboa
FCUP	Faculdade de Ciências da Universidade do Porto
FEP	Faculdade de Economia da Universidade do Porto
FEUP	Faculdade de Engenharia da Universidade do Porto
FMH/UTL	Faculdade de Motricidade Humana/Universidade Técnica de Lisboa
FPCE/UL	Faculdade de Psicologia e Ciências da Educação/Universidade de Lisboa
ICBAS/UP	Instituto de Ciências Biomédicas Abel Salazar/Universidade do Porto
ISA/UTL	Instituto Superior de Agronomia/Universidade Técnica de Lisboa
ISCAA	Instituto Superior de Contabilidade e Administração de Aveiro
ISCAP	Instituto Superior de Contabilidade e Administração do Porto
ISCTE	Instituto Superior de Ciências do Trabalho e da Empresa
ISEG/UTL	Instituto Superior de Economia e Gestão/Universidade Técnica de Lisboa
ISEGI	Instituto Superior de Estatística e Gestão de Informação
IST/UTL	Instituto Superior Técnico/Universidade Técnica de Lisboa
NL	Holanda
U.K.	Reino Unido
UAB	Universidade Aberta
UAç	Universidade dos Açores
UALG	Universidade do Algarve
UBI	Universidade da Beira Interior
UM	Universidade do Minho
UMA	Universidade da Madeira
UPT	Universidade Portucalense Infante D. Henrique
USA	Estados Unidos da América

**ANEXO II**

Lista de doutoramentos defendidos até Fevereiro 2005

Apelido	Nome	Título	Orientador	Local	Data
Murteira	Bento José Ferreira Murteira	Algumas Propriedades dos Processos Auto-regressivos	M. G. Kendall	London School of Economics e ISCEF/UTL	1953
Fontes de Sousa	Maria de Fátima Fontes de Sousa	Novas Condições de Existência de Alguns Planos de Experiência	Dugué	FCUL	1970
Ramalhoto	Maria Fernanda Neto Ramalhoto (IST/UTL)	Processos Estocásticos Aplicados	Grimes	University College, U.K.	1977
Gomes	Maria Ivette Leal de Carvalho Gomes (DEIO/FCUL)	Some Probabilistics and Statistical Problems in Extreme Value Theory	Clive Anderson	University of Sheffield, U.K.	1978
Pestana	Dinis Duarte Ferreira Pestana (DEIO/FCUL)	Some Further Results in Unimodality, Infinite Divisibility and Related Topics	D. Shanbag	University of Sheffield, U.K.	1978
Branco	João António Branco (IST/UTL)	Factor Analysis Methods with Application to Measurement in Medicine	J. A. Anderson	Newcastle upon Tyne, U.K.	1979
Braumann	Carlos Alberto dos Santos Braumann (DM/EU)	Population Growth in Random Environments	L. R. Ginzburg.	State University of New York at Stony Brook	1979
Amaral Turkman	Maria Antónia C. A. Amaral Turkman (DEIO/FCUL)	Some Theoretical and Practical Applications of Predictive Distributions	Ian Dunsmore	University of Sheffield, U.K.	1980
Matias	Maria Beatriz Fernandes Matias	Filas de Epera Com Serviço Por Acompanhamentos Dependente Do Número De Clientes Presentes No Sistema		DM/UA	1980
Sernadas	Maria Cristina Sales Viana Seródio Sernadas (IST/UTL)	Multivariate Stochastic Models For Epidemics		University of London, U.K.	1980
Turkman	Kamil Feridun Turkman (DEIO/FCUL)	Limiting Distributions of Maxima of Certain Types of Non-Stationary Stochastic Processes	Morris Walker	University of Sheffield, U.K.	1980

Apelido	Nome	Título	Orientador	Local	Data
Nicolau	Maria Helena Bacelar Nicolau (FPCE/UL)	Contribuição em Estudos de Coeficientes de Comparação em Análise Classificatória	José Tiago de Oliveira	DEIO/FCUL	1981
Nicolau	Fernando Augusto da Costa Nicolau	Crítérios de Análise Classificatória e Hierárquica Baseadas na Função de Distribuição	José Tiago de Oliveira	DEIO/FCUL	1981
Graça Martins	Maria Eugénia Graça Martins (DEIO/FCUL)	Modelos Estáveis: Estudo Preliminar com vista às Aplicações	Dinis Duarte Ferreira Pestana	DEIO/FCUL	1983
Rosado	Fernando Manuel Fialho Rosado (DEIO/FCUL)	Existência e Detecção de Outliers-Uma Abordagem Metodológica	José Tiago de Oliveira	DEIO/FCUL	1984
Muller	Daniel de Assunção Muller (ISEG/UTL)	Inferência Estatística nos Processos Estocásticos: suas Implicações no Processo de Galton-Watson	Marie Duflo	DEIO/FCUL	1985
Amaral	João Alexandre Ferreira Pena do Amaral (IST/UTL)	Aspects of Optimal Sequential Resource Allocation	John Gittins	Universidade de Oxford, U.K.	1985
Centeno	Maria de Lourdes Carças Centeno (ISEG/UTL)	Theoretical Aspects of Combining Proportional and Non-Proportional Reinsurance Treaties.	Howard Waters	Heriot-Watt University, Edimburgo, U.K.	1985
Mendes-Lopes	Maria de Nazaré Simões Quadros Mendes Lopes (DM/UC)	Análise Estatística dos Processos Pontuais Cromáticos	Jean Geffroy	DM/UC	1985
Pereira	Helena Maria Iglésias Pereira (DEIO/FCUL)	Escolha Estatística em Modelos Estáveis	Dinis Duarte Ferreira Pestana	DEIO/FCUL	1985
Vicente	Corália M. Fortuna de Brito Vicente (ICBAS/UP)	Prognostic Factors in Acute Myelogenous Leukaemia	H. Boothroyd	University of Warwick, U.K.	1985
Brito	Margarida Maria Araújo Brito (FCUP)	Encadrement Presque Sur des Statistique D'ordre	P. Deheuvels	Université Paris VI, França	1986
Cabral	Maria Salomé Esteves Cabral (DEIO/FCUL)	Inferência Estatística em Mistura de duas Populações	José Tiago de Oliveira	DEIO/FCUL	1986
Calheiros	Francisco José Lage Campelo Calheiros (FEUP)	Análise Combinatória no Estudo das Transições de Fase dos Sistemas de Spin Vectorial	O próprio	FCUP	1986
Carvalho	Maria Lucília Salema e Carvalho (DEIO/FCUL)	Estimação em Processos de Galton-Watson Multitipo	Marie Duflo	DEIO/FCUL	1986
Nunes	Álvaro Manuel Duarte Nunes (UMacau)	Identification of Nonlinear (quadratic) Systems from Operating Records		University of Manchester, U.K.	1986

Apelido	Nome	Título	Orientador	Local	Data
Dias	José Rodrigues Dias (DM/UE)	Políticas de Inspeção de Sistemas	Maria de Fátima Fontes de Sousa	DM/UE	1987
Gomes	Paulo Jorge de Pinho Gomes (ISEGI)	Distribution de Bingham sur la n-sphère: une Nouvelle Approche de l'Analyse Factorielle	Yves Escoufier	Université de Montpellier; França	1987
Magalhães Hill	Maria Manuela Magalhães Hill (ISCTE)l	Time varying Markov Models of School Enrolment	John L. R. Proops	Universidade de Keele, U.K.	1987
Gonçalves	Maria Esmeralda Elvas Gonçalves (DM/UC)	Processus Fractionnaires	Christian Gouriéroux	Université des Sciences et Techniques de Lille, França	1988
Mexia	João Tiago Mexia (FCT/UTL)	Heteroscedasticidade Controlada, Espaços Vectorias Quociente e testes F para Hipóteses Sobre Valores Médios	José Tiago de Oliveira	FCT/UNL	1988
Reis	Elisabeth Azevedo Reis (ISCTE)	The Spatial Demography of Portugal in the Late Nineteenth Century: Evidences From the 1864 and 1878 Population Censuses		University of Southampton, U.K.	1988
Alpuim	Maria Teresa Alpuim (DEIO/FCUL)	Contribuições à Teoria de Extremos em Sucessões Dependentes	Maria Ivette Leal de Carvalho Gomes	DEIO/FCUL	1989
Paulino	Carlos Daniel Paulino (IST/UTL)	Análise de Dados Categorizados Incompletos - Fundamentos, Métodos e Aplicações	Carlos Alberto Bragança Pereira	Universidade de São Paulo, Brasil	1989
Silva	Maria Carolina da Silva Tavares da Costa e Silva (ISBAS/UP)	Health Service Usage in the Ribeira de Pena - Concelho of Portugal		University of Exeter, U.K.	1989
Macedo	Fernando Aníbal Wolfango de Macedo	Analysis of Mixed Linear Models with Applications to Animal Breeding	Daniel Gianola	DM/UTAD	1990
Neves Figueiredo	Maria Manuela Costa Neves Figueiredo (ISA/UTL)	Estimação por Blocos dos Parâmetros da Distribuição de Fréchet . Comparação de Métodos Expeditos	José Tiago de Oliveira	FCT/UNL	1990

Apelido	Nome	Título	Orientador	Local	Data
Andrade e Silva	João Manuel de Sousa Andrade e Silva (ISEG/UTL)	Estruturas Tarifárias nos Ramos Reais da Indústria Seguradora - Uma Aplicação ao Seguro Automóvel em Portugal	Maria de Lourdes Centeno	ISEG/UTL	1991
Brito	Maria Paula de Pinho de Brito Duarte Silva (FEP)	Analyse de Données Symboliques. Pyramides d'Héritage	Edwin Diday	Universidade de Paris-IX Dauphine	1991
Dias	Gerberto Fernandes de Carvalho Dias	Testes F Selectivos: Construção e Propriedades	JoãoTiago Mexia	FCT/UNL	1991
Oliveira	Paulo Eduardo Aragão Aleixo Neves Oliveira (DM/UC)	Infinite Divisibilité, Principes D'Invariance et Estimation de Noyaux de Transition em Théorie des Mesures Aleatoires	Pierre Jacob	Université Catholique De Lille, França	1991
Agra Coelho	Carlos Manuel Agra Coelho (ISA/UTL)	Generalized Canonical Analysis	Anant M. Kshirsagar	University of Michigan, USA	1992
Cadima	Jorge Filipe de Campinos L. Cadima (ISA/UTL)	Topics in Descriptive Principal Component Analysis	Ian Joliffe	University of Kent at Canterbury, U.K.	1992
Canto e Castro	Luisa da Conceição Canto e Castro (DEIO/FCUL)	Velocidade de Convergência em Teoria de Valores Extremos	Maria Ivette Leal de Carvalho Gomes	DEIO/FCUL	1992
Crato	Nuno Paulo de Sousa Arrobas Crato (ISEG/UTL)	Some Misspecification Problems in Long-memory Time Series Models	Howard M. Taylor	University of Delaware, U.S.A.	1992
Faria	João Pedro Ferreira de Faria (DEIO/FCUL)	Compatibilidade Crisscross em Estruturas de Probabilidade Comparativa.	Maria Antónia Amaral Turkman	DEIO/FCUL	1992
Fraga Alves	Maria Isabel Fraga Alves (DEIO/FCUL)	Inferência Estatística em Modelos Extremos	Maria Ivette Gomes	DEIO/FCUL	1992
Oliveira	Maria Fernanda dos Santos Adão de Oliveira (DEIO/FCUL)	Leis Limite em Sucessões Dependentes de uma Cadeia	Kamil Feridun Turkman	DEIO/FCUL	1992
Oliveira	Pedro Nuno Ferreira Pinto de Oliveira (UM)	Optimal Scheduling of Hydro-Thermal Power Generation Systems	Sean McKee	University of Strathclyde, U.K.	1992

Apelido	Nome	Título	Orientador	Local	Data
Barão	Maria Isabel Frade Calisto Barão (DEIO/FCUL)	Comparação de Populações de Gumbel	J. Tiago de Oliveira e M Ivette Gomes	DEIO/FCUL	1993
Mendes Leal	Margarida Maria Teixeira Diniz Mendes Leal (DEIO/FCUL)	Análise de Tabelas Tridimensionais com Restrições Múltiplas	Helena Bacelar Nicolau	DEIO/FCUL	1993
Nogueira	Maria Emília Mesquita Nogueira (DM/UC)	Estimação Funcional em Presença de Observações Ergódicas	Michel Delecroix e Nazaré Mendes-Lopes	DM/UC	1993
Athayde	Maria Emília Feijão Queirós de Athayde (UM)	Estudo de Algumas Sucessões Markovianas com Relevo para a Teoria de Extremos	Maria Ivette Leal Carvalho Gomes	DEIO/FCUL	1994
Braumann	Maria Manuela Abreu Braumann	Testes de Detecção de Outliers em Populações Exponenciais	Fernando Manuel Fialho Rosado	DM/UE	1994
Ferreira	Helena Maria Simões Ferreira (UBI)	Condições de Dependência Local em Teoria de Valores Extremos	Maria Ivette Gomes	DM/UC	1994
Pacheco Pires	António Manuel Pacheco Pires (IST/UTL)	Markov-additive Processes Arising in Storage Models for Communication Systems	N. U. Prabhu	Cornell University, USA	1994
Pereira	Maria Teresa Themido Pereira (DEIO/FCUL)	Contribuições à Teoria de Valores Extremos	J. Tiago de Oliveira e Maria Ivette Gomes	DEIO/FCUL	1994
Pereira	Gilda Maria de Carvalho Fernandes Soromenho Pereira (FPCE/UL)	Avaliação Do Número de Componentes de Uma Mistura: Aplicações em Classificação	Fernando da Costa Nicolau	FCT/UNL	1994
Reis	Alfredo Egídio dos Reis (ISEG/UTL)	On the Duration of Negative Surplus		Heriot-Watt University, U.K.	1994
Rosa	Ana Cristina Martins Rosa (DM/UC)	Prevision Robuste Sous Une Hypothese Ergodique		Université de Toulouse, França	1994
Silva	Maria Eduarda da Rocha Pinto Augusto da Silva (FEUP)	Some Contributions to the Analysis of Bilinear Time Series Models	T. Subba Rao	University of Manchester, U.K.	1994

Apelido	Nome	Título	Orientador	Local	Data
Vasconcelos	Rita Maria César e Sá Fernandes Vasconcelos (UMA)	Contribuição à Análise de Dados Categorizados	Dinis Duarte Ferreira Pestana	UMA	1994
Ferreira	Manuel Alberto Martins Ferreira (ISCTE)	Comportamento Transeunte e Período de Ocupação de Sistemas de Filas de Espera sem Espera	Augusto Afonso de Albuquerque	ISCTE	1995
Miranda	Maria Manuela Souto de Miranda (DM/UA)	Estimação Robusta com Variáveis Instrumentais em Modelos com Erros nas Variáveis	João António Branco	DM/UA	1995
Rocha	Cristina Maria Tristão Simões Rocha (DEIO/FCUL)	Modelos com Fragilidade em Análise de Sobrevivência	Maria Antónia Amaral Turkman	DEIO/FCUL	1995
Rocha	José Andrade Rocha (UAç)	Localização e Escala em Situações não Clássicas	Dinis Duarte Ferreira Pestana	UAç	1995
Tenreiro da Cruz	Carlos Manuel Rebelo Tenreiro da Cruz (DM/UC)	Estimação Funcional: Aplicação aos Testes de Ajustamento e de Parâmetro Constante	Christian Gouriéroux e Esmeralda Gonçalves	DM/UC	1995
Gomes	João José Ferreira Gomes (DEIO/FCUL)	Valores Extremos e alguns Processos Markovianos	Maria Teresa Alpuim	DEIO/FCUL	1996
Parente	Ana Maria Nobre Vilhena Nunes Pires de Melo Parente (IST/UTL)	Análise Discriminante, Novos Métodos Robustos de Estimação	João António Branco	IST/UTL	1996
Pinto da Costa	Joaquim Fernando Pinto da Costa (FCUP)	Coefficients d'Association et Binarisation par la Classification Hiérarchique dans les Arbres de Décision. Application à l'Identification de la Structure Secondaire des Protéines	Israel César Lerman	Université de Rennes I, França	1996
Rodrigues	António José Rodrigues (DEIO/FCUL)	Dynamic Regression and Supervised Learning Methods in Time Series Modelling and Forecasting	Peter Young	Lancaster University, U.K.	1996
Alpizar-Jara	Russell Alpizar-Jara (DM/UE)	Assessing Assumption Violations in Line Transect Sampling	Kenneth H. Pollock	North Carolina State University, USA	1997

Apelido	Nome	Título	Orientador	Local	Data
Esquível	Manuel Leote Tavares Inglês Esquível (FCT/UNL)	Applications of fourier methods to the analysis of some stochastic processes	Jean-Pierre Kahane	FCT/UNL	1997
Magalhães	Fernando José Malheiro de Magalhães (ISCAP)	Prediction in Poisson and other Errors in Variables Models	Ian Dunsmore	University of Sheffield, U.K.	1997
Pereira	Isabel Maria Simões Pereira (DM/UA)	Análise Bayesiana Em Séries Temporais Com Erros Exponenciais	Maria Antónia Amaral Turkman	DM/UA	1997
Rebelo	Efígenio da Luz Rebelo (UALG)	A Gauss-Newton Regression Approach to Tests of Nonnested Hypothesis in Some Linear Econometric Models	Carlos Silva Ribeiro	UALG	1997
Freitas	Maria Adelaide Valente de Freitas (DM/UA)	Nova Classe de Aproximações em Teoria de Valores Extremos	Maria Ivette Leal de Carvalho Gomes	DM/UA	1998
Hall	Andreia Oliveira Hall (DM/UA)	Extremos de Sucessões de Contagens - do Outro Lado do Espelho	Maria Ivette Leal de Carvalho Gomes	DEIO/FCUL	1998
Brilhante	Fátima Almeida Brilhante (UAç)	Modelos Inferência Estatística em Modelos não Gaussianos com Recurso a Spacings e outras funções de Estatísticas Ordinais	Dinis Duarte Ferreira Pestana	UAç	1999
Nascimento	Maria Manuel da Silva Nascimento (UTAD)	Análise de Resultados de uma Simulação de um Processo Não-Ergódico. Sobre um Exemplo da Indústria Extractiva	Acácio Porta Nova e Carlos Novais Madureira	DM/UTAD	1999
Rodrigues	Paulo Manuel Marques Rodrigues (UALG)	Inference in Seasonal Nonstationary Processes	D. R. Osborne e L. Gill	University of Manchester, U.K.	1999
Teles	Paulo João Figueiredo Cabral Teles (FEP)	The Effects of Temporal Aggregation on Time Series Tests	William Wei	Temple University, USA	1999
Cardoso	Maria Margarida dos Santos Cardoso (ISCTE)	Segmentação - uma Aplicação no Mercado do Turismo	Isabel Themido	IST/UTL	2000

Apelido	Nome	Título	Orientador	Local	Data
Coelho	Pedro Miguel Pereira Simões Coelho (ISEGI)	Estimação em Pequenos Domínios sob o Modelo Linear Geral Misto com Informação Cronológica e Espacial	Jean-Jacques Dreesbeke	ISEGI	2000
Ferreira	Ana Portela Nunes de Sousa Ferreira (FPCE/UL)	Combinação de Modelos em Análise Discriminante sobre Variáveis Qualitativas	Helena Bacelar Nicolau	FCT/UNL	2000
Figueiredo	Adelaide Maria de Sousa Figueiredo (FEP)	Classificação de Variáveis no Contexto de um Modelo Probabilístico Definido na n-Esfera	Paulo Gomes	DEIO/FCUL	2000
Lopes	Sofia Mucharreira de Azeredo Lopes (DEIO/FCUL) pós-doc	Statistical Analysis of Particle Distributions in Composite Materials	Nick Fieller	University of Sheffield, U.K.	2000
Martins	Cristina Maria Tavares Martins (DM/UC)	Modelos Bilineares em Séries Temporais: Propriedades Probabilísticas e Decisão Estatística	Nazaré Mendes Lopes e Esmeralda Gonçalves	DM/UC	2000
Mendonça	Sandra Mendonça (UMA)	Tópicos sobre a Convergência Fraca de Variáveis Aleatórias	Dinis Duarte Ferreira Pestana	UMA	2000
Oliveira	Teresa Paula Costa Azinheira Oliveira (UAB)	Planeamento de Experiências - Novas Perspectivas	Maria de Fátima Fontes de Sousa	DEIO/FCUL	2000
Oliveira	Cristina Maria Correia Teles Garcia Oliveira (ISCAP)	Função de autocorrelação Estendida Generalizada Amostral: Contributo para a Identificação dos Modelos de Função Transferência	Daniel Muller	ISEG/UTL	2000
Temido	Maria da Graça Temido (DM/UC)	Classes de Leis Limites em Teoria de Valores Extremos. Estabilidade e Semiestabilidade	Maria Ivette Gomes e Luísa Canto e Castro	DM/UC	2000
Antunes	Nelson Rodrigues Gama Antunes (UALG)	Modelação e Análise de Redes Sem Fios	António Manuel Pacheco Pires	IST/UTL	2001
Braga	Ana Cristina da Silva Braga (UM)	Curvas ROC: Aspectos Funcionais e Aplicações	Pedro Nuno Ferreira Pinto de Oliveira	UM	2001
Corte Real	Pedro Alexandre da Rosa Corte Real (FCT/UNL)	Modelos Lineares Normais com Conexão	João Tiago Mexia	FCT/UNL	2001
Costa	Francisco Camões Costa (ISCTE)	Um Modelo de Notação de Cartões de Crédito Incorporando Rentabilidade Esperada	Maria Manuela Magalhães Hill	ISCTE	2001

Apelido	Nome	Título	Orientador	Local	Data
Ferrão	Maria Eugénia Neto Ferrão (UBI)	Modelo Multinível para Dados Longitudinais com Variável Resposta Discreta	Harvey Goldstein	P: U. C. do Rio de Janeiro, Brasil	2001
Martins	Maria João Teixeira Martins (ISA/UTL)	Estimação de Caudas Pesadas - Variantes ao Estimador de Hil	Maria Ivette Gomes e Manuela Neves	DEIO/FCUL	2001
Nicolau	João Carlos H. C. Nicolau (ISEG)	Modelação e Estimação de Séries Financeiras através de Equações Diferenciais Estocásticas	Carlos Braumann	ISEG/UTL	2001
Oliveira	Maria Manuela Melo Oliveira (DM/UE)	Modelação de Séries de Emparelhadas de Estudos com Estrutura Comum	João Tiago Mexia	DM/UE	2001
Oliveira	Vera Lúcia Oliveira (ISCAP)	Modelos Autorregressivos para Sucessões Cronológicas de Contagem: Caracterização e Modelação	Maria Eduarda da Rocha Pinto Augusto da Silva	FCUP	2001
Scotto	Manuel González Scotto (DM/UA)	On the Extremes of Certain Time Series	Kamil Feridun Turkman	DEIO/FCUL	2001
Silva	Giovani Loiola Silva (IST/UTL)	Análise Bayesiana de Modelos de Sobrevivência com Fragilidade	Maria Antónia Amaral Turkman	IST/UTL	2001
Sousa	Fernanda Maria Campos de Sousa (FEUP)	Novas Metodologias e Validação em Classificação Hierárquica Ascendente	Fernando da Costa Nicolau	FCT/UNL	2001
Azevedo	Cecília Maria Vasconcelos Costa e Castro Azevedo (UM)	Estimador do Operador de Covariância do Limite do Processo Empírico	Paulo Oliveira e Emília Athayde	DM/UM	2002
Henriques	Ana Ferreira Henriques (ISA/UTL)	Statistics of Extremes: Estimation and Optimality	Laurens de Haan e Einmull	University of Tilburg, Holand	2002
Morais	Manuel João Cabral Moraes (IST/UTL)	Ordenação Estocástica na Análise de Desempenho de Esquemas de Controlo de Qualidade	António Manuel Pacheco Pires	IST/UTL	2002
Oliveira Silva	Maria do Rosário de Oliveira Silva (IST/UTL)	Métodos Robustos em Análise de Correlações Canónicas	João António Branco	IST/UTL	2002
Pereira	Lúisa Maria Jota Pereira (UBI)	Valores Extremos Multidimensionais de Variáveis Dependentes	Helena Maria Simões Ferreira	UBI	2002

Apelido	Nome	Título	Orientador	Local	Data
Ramos	Alexandra Patrícia Horta Ramos (FEP)	Multivariate Joint Modelling and Score Tests of Independence	Anthony Ledford e Trevor Sweeting	University of Surrey, U.K.	2002
Sousa	Bruno Cecílio de Sousa (UM)	Contribution to the Estimation of the Tail Index of Heavy-Tailed	Bruce Hill e George Michailidis	University of Michigan, USA	2002
Antunes	Marília Cristina de Sousa Antunes (DEIO/FCUL)	Some Problems in Non-linear Prediction	Kamil Feridun Turkman	DEIO/FCUL	2003
Bermudez	Patrícia de Zea Bermudez (DEIO/FCUL)	Bayesian Approach to Extreme Quantile Estimation	Kamil Feridun Turkman	DEIO/FCUL	2003
Curto	José Joaquim Dias Curto (ISCTE)	Modelação das Taxas de Rendibilidade: Estudo Comparativo dos Mercados de Capitais Português, Alemão e Norte-Americano	Elizabeth Reis e José Paulo Esperança	ISCTE	2003
Figueira	João Pedro da Silva Horta Reis Figueira (ISCTE)	Aplicação dos Processos de Difusão e da Teoria do Renovamento num Estudo de Reservas Aleatórias	Manuel Alberto Ferreira	ISCTE	2003
Figueiredo	Fernanda Otilia de Sousa Figueiredo (FEP)	Controlo Estatístico da Qualidade - Métodos Robustos	Maria Ivette Leal de Carvalho Gomes	DEIO/FCUL	2003
Gonçalves	Maria Helena Neves de Queirós Gonçalves (UALG)	Likelihood Methods for Discrete Longitudinal Data	A.Azzalini e Maria Salomé Cabral	DEIO/FCUL	2003
Nunes Philipart	Cláudia Coelho Nunes Philipart (IST/UTL)	Cadeias Markov Moduladas	António Manuel Pacheco Pires	IST/UTL	2003
Oliveira	Orlando Aníbal Lopes de Oliveira	In Extremis	Maria Ivette Leal de Carvalho Gomes	DEIO/FCUL	2003
Oliveira	Irene Oliveira (UTAD)	Correlated Data in Multivariate Analysis	Ian Joliffe e Jorge Cadima	University of Kent at Canterbury, U.K.	2003
Pereira	Dulce Gamito Santinhos Pereira (DM/UE)	Análise Conjunta Pesada de Regressões em Redes de Ensaios	João Tiago Mexia	DM/UE	2003
Pestana	Maria Helena Aguiar Pereira Pestana (ISCTE)	Turismo e Envelhecimento Demográfico: Uma realidade de futuro. O caso particular de Lisboa	Manuel Alberto Ferreira, Elisabete Reis e João Albino Matos Silva	ISCTE	2003

Apelido	Nome	Título	Orientador	Local	Data
Rodrigues	Isabel Maria Alves Rodrigues (IST/UTL)	Métodos Robustos em Análise de Componentes Principais Comuns	Ana Pires Parente	IST/UTL	2003
Santos	Paulo de Jesus Infante dos Santos (DM/UE)	Métodos de Amostragem em Controlo de Qualidade	José Rodrigues Dias	DM/UE	2003
Seco	Maria Alexandra Abreu Henriques Seco (ESTGL)	Modelação de Dados Agrupados nas Ciências do Ambiente	António St'Aubyn	ISA/UTL	2003
Severino	Joaquim Eduardo Gonçalves Severino (DEIO/FCUL)	Space-Time Linear Prediction. An Application to Area Rainfall Measurement	Maria Teresa Alpuim	DEIO/FCUL	2003
Amado	Maria da Conceição Esperança Amado (IST/UTL)	Bootstrap Robusto com Base na Função de Influência	Ana Pires Parente	IST/UTL	2004
Bruno	Paula Marta Bruno (FMH/UTL)	Modelos de Séries Temporais com Aplicação às Ciências da Motricidade	Luísa Canto e Castro	FMH/UTL	2004
Carita	Ana Isabel Andrade Dinis Carita (FMH/UTL)	Planeamento de Experiências e Análise Estatística de Modelos Associados à Motricidade Humana	Maria Antónia Amaral Turkman e Luís Loura	FMH/UTL	2004
Dias	Alexandra da Costa Dias (FCT/UNL)	Copula Inference for Finance and Insurance	Paul Embrechts	ETH Zurique, Suíça	2004
Dias	José Gonçalves Dias (ISCTE)	Finite Mixture Models. Review, Applications, and Computer-intensive Methods	A.G.M. Steerneman, J.K. Vermunt e F.J. Willekens	Universidade de Groningen, Holanda	2004
Garcia	Jorge Manuel Afonso Garcia (ISEG)	As Transformadas de Fourier e Laplace na Teoria do Risco	Maria de Lourdes Centeno e Alfredo Egídio dos Reis	ISEG/UTL	2004
Gonçalves	Luzia Augusta Pires Gonçalves (IHMT/UNL)	Metodologias Estatísticas em Biologia Molecular: Construção de Mapas Físicos	M. Antónia Amaral Turkman e Hildete P Pinheiro	DEIO/FCUL	2004

Apelido	Nome	Título	Orientador	Local	Data
Miguens	Maria de Fátima Varregoso Miguens (FCT/UNL)	Modelo Bivariado para Caudais Máximos	Maria de Fátima Fontes de Sousa	FCT/UNL	2004
Pereira	Carla Mónica Santos Dias Pereira (UPT)	Reconhecimento de Padrões: Classificação Supervisionada com Rejeição de Observações	Ana Pires Parente	IST/UTL	2004
Soares	Paulo Soares (IST/UTL)	Análise Bayesiana de Dados Deficientemente Categorizados	Carlos Daniel Paulino	IST/UTL	2004
Sousa	Lisete Maria Ribeiro de Sousa (DEIO/FCUL)	Development of Statistical Methods for the Analysis of Molecular Genetic Data.	M. Antónia Amaral Turkman e Wolfgang Urfer	DEIO/FCUL	2004
Valle	Patrícia Susana Pinto Oom do Valle (UALG)	Modeling Consumer Behavior in the Portuguese Recycling Program: The Logistics and Communication Planning Implications	Elisabeth Reis, Efigénio Rebelo e João Menezes	FE/UALG	2004
Durão	Natércia Felgueiras Seabra Durão (UPT)	Bayesiana na Análise de Problemas de Triagem	M. Antónia Amaral Turkman	DEIO/FCUL	2005
Martins	Ana Paula André Martins (UBI)	Coefficientes Extremais	Helena Maria Simões Ferreira	UBI	2005
Miranda	Maria Cristina Miranda (ISCAA)	Estatística de Extremos: Estimção dos Índices Extremal e de Cauda	Maria Ivette Gomes e Andreia Hall	DEIO/FCUL	2005

**ANEXO III**

Lista de teses de doutoramento a terminar em 2005

Apelido	Nome	Título	Orientador	Local
Abreu	Ana Maria Figueira da Silva Abreu	Análise de Sobrevivência para Populações com Indivíduos Imunes	Cristina Maria Rocha	UMA
Couto	Gualter Couto	Opções Reais e Decisão sob Incerteza na escolha da Localização	Claúdia Nunes Philippart (co-orientadora)	IST/UTL
Cruz	Manuel Mendes da Cruz	Fundamentos do Processo de Decisão Estratégica na Indústria Seguradora	Nélson António e José Dias Curto	ISCTE
Dionísio	Andreia Teixeira Marques Dionísio	Medidas de Informação Aplicadas a Mercados Financeiros	Rui Menezes	ISCTE
Ferreira	Sandra Maria Bargão Saraiva Ferreira	Teoria das Componentes de Variância	João Tiago Mexia	UBI
Ferreira	Nuno Rafael B. Ferreira	Modelação não Linear dos Mercados Bolsistas	Rui Menezes	ISCTE
Ferreira	Dário Jorge da Conceição Ferreira	Teoria dos Modelos Mistos	João Tiago Mexia	UBI
Freitas	Ana Cristina Moreira de Freitas	Estimação do Coeficiente de Cauda Exponencial: Aplicação à Teoria do Risco	Margarida Brito	FCUP
Gomes	Dulce Maria de Oliveira Gomes	Processos Auto-regressivos de Coeficientes Aleatórios na Modelação de Dados de Contagem	Luísa Canto e Castro	DM/UE
Gonçalves	Rui Jorge de Almeida Soares Gonçalves	Sistemas Naturais Complexos: Que Tendências? -O Estocástico e o Determinístico	Francisco José Lage Campelo Calheiros	FEUP

Apelido	Nome	Título	Orientador	Local
Gonçalves	Arminda Manuela Andrade Pereira Gonçalves	Modelação Estatística da Qualidade de Águas de Superfície	Maria Teresa Alpuim	DEIO/FCUL
Moniz	Ana Isabel Moniz	A Sustentabilidade do Turismo em Ilhas de Pequena Dimensão - o Caso dos Açores	Maria Manuela Magalhães Hill	UAç
Mendes	Jorge Morais Mendes	Bayesian Hierarchical Models for Spatio-Temporal Processes	Kamil Feridun Turkman	DEIO/FCUL
Natário	Isabel Cristina Maciel Natário	Hierarchical Bayesian Models for Epidemiological Analysis of Rare Events	Maria Lucília Salema e Carvalho	DEIO/FCUL
Nunes	Maria Helena Mouriño Silva Nunes	Discovering the Dynamics that Relate Phytoplankton Abundance to Upwelling Events. A New Approach to the Problem of Missingness in the Gaussian Context	Maria Teresa Alpuim	DEIO/FCUL
Nunes	Célia Maria Pinto Nunes	Testes F e Relacionados em Modelos Mistos com Cross- Nesting Ortogonal	João Tiago Mexia	UBI
Palma	José António da Conceição Palma	Medidas de Performance para Testes em Populações Normais	Fernando Rosado	DEIO/FCUL
Ramos	Maria do Rosário Olaia Ramos	Testes de Tendência não Paramétricos e Sua Comparação com Métodos Paramétricos	Maria Teresa Alpuim	DEIO/FCUL
Silva	Nélia Maria Marques da Silva	Análise Bayesiana de Séries Temporais de Valores Inteiros com Observações Repetidas	Isabel Pereira e Eduarda Silva	DM/UA
Silva	Isabel Maria Marques da Silva	Contributions to the Analysis of INAR Processes	Eduarda Silva	FCUP
Teles	Julia Maria Vitorino Teles	Determinação de Modelos: uma Abordagem Bayesiana	Maria Antónia Amaral Turkman e Luís Loura	FMH/UTL

## **Maria Antónia Amaral Turkman**

É Professora Catedrática da Faculdade de Ciências da Universidade de Lisboa (FCUL) e investigadora do Centro de Estatística e Aplicações dessa mesma Universidade.

Licenciou-se em Matemática Aplicada pela FCUL em 1971 e obteve os graus de Mestre e de Doutor em Probabilidades e Estatística pela Universidade de Sheffield em 1977 e 1980, respectivamente.

Tem exercido desde sempre a sua actividade docente no Departamento de Estatística e Investigação Operacional da FCUL, onde tem leccionado essencialmente disciplinas de Inferência e Decisão Estatística, Fundamentos e Metodologias da Estatística, Estatística Multivariada, Estatística Computacional e Projecto de Estatística.

Publicou vários artigos de investigação no domínio da Estatística Bayesiana e artigos com aplicações à Medicina, Genética e Meio Ambiente.

É co-autora de um livro sobre Modelos Lineares Generalizados, publicado pela Sociedade Portuguesa de Estatística e de outro sobre Estatística Bayesiana publicado pela Fundação Calouste Gulbenkian.

Foi sócia fundadora da Sociedade Portuguesa de Estatística, tendo desempenhado os cargos de Tesoureira entre 1990 e 1994 e Vice-Presidente entre 1994 e 2000.



# Pluridisciplinaridade, Interdisciplinaridade e Convergências

*Helena Bacelar-Nicolau*

A imagem que a memória me traz instantaneamente e em registo nítido, quando começo a escrever este apontamento para o “memorial dos 25 anos da SPE” é a de um grupo de estatísticos jovens, sentados numa sala de aulas escura, já noite, em torno do Professor Tiago de Oliveira, de que não poucos tínhamos sido alunos e éramos actuais assistentes. No fim de um dia de trabalho conversávamos sobre estatutos, discutíamos prioridades nas futuras actividades a desenvolver no âmbito da nova sociedade que acabávamos de fundar. O papel do (jovem) Professor Tiago de Oliveira na criação e na consolidação da Sociedade Portuguesa de Estatística (e Investigação Operacional, na origem) foi, desde o início, fundamental e a sua marca indelével. O Mestre, que ensina e passa o testemunho aos seus Aprendizes, eis a fotografia base do meu álbum.

Outras imagens vêm logo de seguida, configurando alguns dos momentos do quarto de século de vida da Sociedade, que intersectam e marcam a minha. Aqui o álbum que abro é necessariamente curto e peculiar, por urgência no tempo que me é concedido, por estado de alma no tempo que atravesso.

Assim, as impressões mais nítidas sucedem já nas primeiras reuniões científicas organizadas pela SPE e associam-se aos congressos do Fundão e de Tróia. O primeiro teve lugar em 1981, poucos meses depois de eu e o Fernando obtermos o doutoramento. Os métodos, as técnicas e o software específico de Análise de Dados Multivariados entravam pela nossa mão, nos congressos da SPE - “pela nossa mão e pelo nosso pé”, afirmava um colega recente, ex-aluno de há pouco, que acrescentava, enquanto dansávamos à noite no bar “e parece que continuam em estado normal, apesar de agora serem doutores!”.

Entre o Fundão e Tróia a investigação em ADM deu um salto em frente, com a entrada na “Linha” de investigação do CEA e na Sociedade, de um grupo de gente nova, competente e entusiasmada com as recentes abordagens à ADM e suas aplicações. Vários artigos neste domínio e sua vizinhança começaram a ser, desde então, publicados nas actas dos congressos da SPE. Cito na bibliografia alguns trabalhos pioneiros nesta matéria em Portugal, que foram apresentados ou publicados nos congressos da SPE. Parte destes trabalhos foram fruto de programas de investigação com colegas europeus, especialmente franceses, eles próprios sócios da Sociedade Francesa de Estatística e, mais tarde, também membros da Associação Francófona de Classificação. O desenvolvimento das relações científicas internacionais foi uma prioridade da SPE, desde o princípio, e a sua entrada como membro do ISI, o “International Statistical Institute” veio solidificar o processo.

O progresso rápido da análise estatística de dados multivariados, particularmente das metodologias empíricas e probabilísticas de análise classificatória e áreas afins, no seio da SPE e do DEIO, sua sede, esteve correlacionado significativamente desde logo, com três factores: a ocorrência de tal desenvolvimento no âmbito da cooperação internacional, a firme preparação de base dos jovens investigadores e a riqueza das próprias metodologias e seus resultados nas aplicações aos dados, reais ou simulados.

Entretanto, a difusão e a relevância, teórica e prática, de certas áreas específicas e necessárias da Estatística e da ADM levaram a que, paralelamente e em diferentes países, novas associações fossem fundadas, com objectivos fortemente direccionados para o investimento nessas áreas e nas relações científicas inter-associações. Destacamos aqui, naturalmente, as associações onde as metodologias e as técnicas de Classificação ou Análise Classificatória e áreas afins constituem o principal objecto. Assistiu-se neste campo à criação, quase simultânea, das Associações Francófona, Alemã, Norte-Americana e do Reino Unido, dando nascimento pouco tempo depois, à IFCS, a "International Federation of Classification Societies". Seguindo o mesmo movimento, nasceu em 1994 a Associação Portuguesa de Classificação e Análise de Dados, de que tive o gosto de ser primeira presidente, que em 1995 entrava, por sua vez, na IFCS. Muitos de nós passávamos assim a ser sócios da SPE e da CLAD, e através delas, do ISI e da IFCS, respectivamente.

Nas fotos do meu álbum revejo professores, colegas e/ou alunos, muitos daqueles com quem tenho trabalhado na investigação, e mais (de) uma vez, na participação ou na organização dos congressos, os locais privilegiados dos nossos encontros. E no entanto, alguns partiram já "au delà des étoiles", um modo de pensar em que acompanho o meu colega e amigo Erwin Diday. E me recorda a "Contribuição à estatística", de Wislawa Szymborska :

"Em cem pessoas,  
sabendo tudo melhor que as outras: cinquenta e duas,  
incertas em cada passo: quase todas as outras,  
.....  
dignas de compaixão: noventa e nove  
mortais: cem por cento.  
Número que, por enquanto, não pôde ser modificado."

Quando voltei do "Distancia '92", o primeiro congresso associado a uma rede de Laboratórios de Estatística Europeus em que o Fernando e eu participávamos como co-organizadores, o Mestre partira "au delà des étoiles".

A par do fortalecimento interno das associações, assistiu-se à consolidação da cooperação pluri e interdisciplinar. As possibilidades de troca de saberes e criação de mais saber, cresceram naturalmente, para todos. Nos congressos do ISI e das suas secções, como nos da IFCS, têm lugar, regularmente, sessões convidadas "on behalf" de uma das

outras instituições. A IASE, International Association of Statistical Education, que organiza quadrianualmente a International Conference on Teaching Statistics - ICOTS, e a IASC, International Association of Statistical Computing, que organiza bianualmente o bem conhecido COMPSTAT, além de workshops e escolas avançadas, nos anos intervalares, são-me particularmente próximas, pois fui correspondente nacional da primeira e membro do "board" de directores da Seccção Europeia da segunda, duas experiências estimulantes para que incentivo outros colegas de espírito missionário. Cresce a tendência para que as escolas e os workshops sejam apoiados por duas ou mais destas instituições. Investe-se no ensino e na formação da estatística e da análise de dados em domínios ligados à indústria, à gestão, à economia, à medicina, às ciências humanas e sociais, à própria educação estatística, e suas combinações, citando os que nos têm vindo a interessar prioritariamente. A SPE, como membro do ISI, acompanha naturalmente este movimento.

Note-se que a pluridisciplinaridade tem aqui como denominador comum a matemática. Mas lembremo-nos que já Richard P. Feynman, prémio Nobel de Física, afirmava que "Mathematics is only looking for patterns...". Assim voltamos aos modelos de análise classificatória e, com eles, à classificação - de indivíduos ou de variáveis. Fecho o álbum sobre uma última representação que ficara lá atrás.

Corria o ano de 1996, o trabalho pesado mas gratificante da docência ia a meio, acumulava-se o trabalho para os projectos de investigação, para a participação e/ou organização de reuniões científicas, amontoavam-se os relatórios e os pareceres inacabados, ..., enfim, o cenário habitual. Acabava precisamente de regressar a Lisboa, vinda de uma missão breve integrada no âmbito de um dos nossos programas de cooperação científica e tecnológica luso-franceses, quando fui informada de que o Director do ISI estava em Lisboa. Desde que fora admitida como membro do ISI, dois anos antes, tínhamos desenvolvido uma agradável e enriquecedora correspondência, mistura de conteúdos próximos da estatística com conceitos e questões próximos da organização da estatística e dos estatísticos. A minha qualidade de sócia fundadora e activa na Sociedade Portuguesa de Estatística, uma sociedade membro do "seu" ISI, aliada à recém-adquirida qualidade de sócia fundadora e primeira presidente da Associação Portuguesa de Classificação e Análise de Dados, uma "sister society" nascente, membro da jovem Federação Internacional das Sociedades de Classificação, analisava-as ele com particular interesse. Não esquecerei esse nosso primeiro encontro, no CCB, durante uma reunião internacional promovida pelo INE. Era um homem de estatura pouco superior ao meu metro e meio, de humor fácil, tolerância, humanismo e competência à flor da pele. Fomos conversando enquanto avançávamos na fila do café, continuámos enquanto o bebíamos à mesa, com a esposa, e no fim do pequeno intervalo já havia todo um plano de trabalho definido. Dele fazia parte no imediato, a "imprescindível" organização por mim de um "Invited Paper Meeting" para o próximo ISI-1997 em Istambul, cujo tema ficou logo ali decidido: "Desenvolvemos métodos de classificação de indivíduos, mas estou igualmente empenhada na classificação de variáveis, que é fundamental nas áreas afins às

ciências humanas”, “Que tal Classificação de Indivíduos Versus Classificação de Variáveis?” “Não parecerá provocador?” “Não, é simplesmente estimulante.” Apercebi-me depois de que todos os prazos já estavam entretanto ultrapassados, e era “bastante razoável” a responsabilidade que me coubera, mas em Julho de 1997 lá apresentei o meu primeiro IPM num congresso do ISI, IPM que reunia investigadores convidados representativos de diversas “sister societies” do ISI e da IFCS, como previsto. Foi provocante e estimulante q.b.. Dez anos passados, e depois de mais alguns IPM organizados nos congressos do ISI ou das suas associações, e/ou da IFCS, de umas tantas escolas / seminários / workshops nacionais ou internacionais, vejo com satisfação muito especial a SPE e a CLAD prepararem-se para patrocinar, conjuntamente, o próximo ISI-2007, que terá lugar em Lisboa.

### Referências Bibliográficas

- Bacelar-Nicolau, H. (1987), *On the distribution equivalence in cluster analysis*, in NATO ASI Series, vol F30, Patt. Recogn. Theory and Applic., P.A.Devijver/J.Kitler (eds.), Springer-Verlag, pp. 73-9.
- Bacelar-Nicolau, H. (1997), *Classification of Variables VS Classification of Subjects*, Invited Paper Meeting in Proceedings of the 51st Session of the International Statistical Institute (ISI 97), Istanbul, Turkey, Book 2, pp 343-360.
- Bacelar-Nicolau, H. (1997), *Sur le Développement et les Applications de la Classification et l'Analyse des Données au Portugal: Tendances Principales*, in Bulletin de la Société Francophone de Classification, pp.3-5.
- Bacelar-Nicolau, H. (1999), *Teaching and Training Multivariate Data-Analysis*, IPM in Proceedings of the 52nd Session of the International Statistical Institute (ISI 99), Helsinki, Finland, Book 2, pp.209-226.
- Bacelar-Nicolau, H. (2000), *The Affinity Coefficient. Analysis of Symbolic Data Exploratory Methods for Extracting Statistical Information from Complex Data*, H.H.Bock and E.Diday (Eds.), Springer, pp. 160-5.
- Bacelar-Nicolau, H. (2001), *On the paper “Research in Statistical Education: Some Priority Questions” by Carmen Batanero et al*, Invited paper in IASE Statistical Education Research Newsletter, vol 2(2), pp.3-4.
- Bacelar-Nicolau, H. e Mendes Leal, M. (1990), *Análise Multivariada de Dados Discretos. Uma aplicação à distribuição regional do vírus da SIDA em Angola*, in Actas da I Conferência em Estatística e Optimização, Tróia, Portugal, pp 405-417.

- Bacelar-Nicolau, H. e Nicolau, F. C. (1990), *Analyse classificatoire de l'identité nationale chez les portugais. Une étude exploratoire fondée sur le coefficient d'affinité* in *Mathématiques, Informatique et Sciences Humaines*, 28eme année, 109, pp. 55-64.
- Bacelar-Nicolau, H. e Nicolau, F. C. (1981), *Novos Aspectos da Análise Classificatória*, in *Actas do II Colóquio de Estatística e Investigação Operacional*, INIC/UL, Fundação/Covilhã, pp. 322-341.
- Bacelar-Nicolau, H. et al. (1998), *LEASP97: An Improvement in Teaching and Analysing New Methodology on Probabilistic Clustering Models - Topic 7: The Role of Technology in the Teaching of Statistics*, in *Proceedings of the Fifth International Conference on Teaching of Statistics, ICOTS-5, Singapore*, pp. 863-9.
- Bacelar-Nicolau, H. e Figueira, M.L. (1989), *Análise Exploratória da Evolução do Auto-conceito nos Adolescentes através de Modelos Probabilísticos de Classificação hierárquica*, in *Psicologia Clínica*, vol. 10, nº1, pp. 43-8.
- Bacelar-Nicolau, H. e Lobo Antunes, J. (1992), *A entrevista no processo de admissão a Faculdade de Medicina de Lisboa. Correlação com outros parâmetros de avaliação* (with) - *Revista da Sociedade Portuguesa de Educação Médica*, vol 3, no3, 88-97.
- Bacelar-Nicolau, H. e Nicolau, F. C. (2003), *Teaching and Learning Hierarchical Clustering Probabilistic Models for Categorical Data*, in *Proceedings of the 54nd Session of the International Statistical Institute (ISI-54), Berlim, Alemanha*, pp. 333-349.
- Bacelar-Nicolau, H. e Nicolau, F. C. (2003), *Sur des Modèles Probabilistes de Classification pour le Data Mining*, in *Méthodes et Perspectives en Classification*, Y. Dodge and G. Melfi (Eds), PAN-Presses Académiques Neuchâtel, pp. 3-6.
- Bacelar-Nicolau, H., Nicolau, F. C. e Dias, O. (1990), *Modelos de Classificação Hierárquica Ascendente baseados no Coeficiente de Afinidade: Uma Aplicação à Psiquiatria* in: *Actas das XV Jornadas Luso-Espanholas de Matemática*, Universidade de Évora, Portugal, pp.265-270.
- Bacelar-Nicolau, H. e Pinto Dória, I. (1990), *O Coeficiente de Afinidade em Análise de Dados Multivariados: Uma Aplicação à Economia*, in *Actas de XV Conferência em Estatística e Optimização*, Tróia, Portugal, pp. 389-404
- Bacelar-Nicolau, H. e Silva, Danilo (2003), *Um padrão ou vários padrões de características de resposta de Rorschach num grupo de dependentes de heroína?* in *Toxicodependências-Ministério da Saúde-Inst. da Droga e da Toxicodependência*, vol. 9, nº 1, pp.47-61.
- Nicolau, F. C. e Sousa, F. (1995), *Métodos de Classificação Baseados em Estatísticas de Tendência Central*, in *Bom Senso e Sensibilidade. Traves Mestras da Estatística* (J. Branco, P. Gomes e J. Prata, eds.), SPE e Ed. Salamandra pp. 583-595.
- Nicolau, F. C. e Nascimento, J.L (1995), *Avaliação Estatística Multivariada do Enquadramento Funcional de Recursos Humanos de uma Empresa*, *Comportamento Organizacional e Gestão*, vol. 1, nº 1, ISPA (ed.), 126-135.
- Nicolau, F. C. e Bacelar-Nicolau, H. (1998), *Some Trends in the Classification of Variables*, in *Data Science, Classification and Related Methods*, Hayashi, Oshumi, Yajima, Tanaka, Bock, Baba (Edit.), Springer, pp. 89-98.

- Nicolau, F. C. e Bacelar-Nicolau, H. (1999), *Clustering Symbolic Objects Associated to Frequency or Probability Laws by the Weighted Affinity Coefficient. Applied Stochastic Models and Data Analysis*, in Quantitative Methods, in Business and Industry Society. H. Bacelar-Nicolau, F. C. Nicolau and Jacques Janssen (Eds.), INE, Lisboa, Portugal, pp.155-8.
- Nicolau, F. C. e Brito, P. (1989), *Improvements in NHMEAN Method* in E. Diday (ed.), *Data Analysis, Learning Symbolic and Numeric Knowledge*, Nova Science Publishers, New York-Budapest, pp. 109-116.
- Phillips, Brian and Hawkins, Anne (1999), *Statistical Education for Life*, IPM in Proceedings of the 52nd Session of the International Statistical Institute, Helsinki, Finland, pp.241-254
- Silva, A. L., Bacelar-Nicolau, H. e Saporta, G. (2002), *Missing Data in Hierarchical Classification of Variables - a Simulation Study* in Classification Clustering and Data Analysis, in Data Analysis, Classification and Related Methods, Springer, pp.121-128.
- Silva, A. L., Saporta, G. e Bacelar-Nicolau, H. (2004). Missing data and imputation methods in partition of variables, in: *Classification, Clustering and Data Mining Applications*, Springer, pp.631-637.
- Sousa, F. e Nicolau, F. (2001), *Uma Abordagem ao Problema da Comparação de Estruturas Classificatórias*, in A Estatística em Movimento (M. M. Neves, J. Cadima, M. J. Martins e F. Rosado, eds.), SPE, pp. 409-418.
- Sousa Ferreira, A., Bacelar-Nicolau, H., Pinto Dória, I., Dias, O. (2005), *Classificação e Discriminação Num Estudo Sobre Corrosão*, in Actas do XII Congresso da SPE (a publicar)
- Sousa Ferreira, A., Pinto Dória, I., Dória, M., Bacelar-Nicolau, H., (2002), *Análise Multivariada do Risco e da Confiança em Sistemas de Tratamento de Resíduos Sólidos* in Novos Rumos Em Estatística, Actas do IX Congresso da SPE, SPE Ed., 215-226.
- Szymborska, Wislawa (1997), *Je ne Sais Quelles Gens*, Poésie Fayard

## **Helena Bacelar-Nicolau**

É Professora catedrática na Universidade de Lisboa, é coordenadora do LEAD na FPCEUL, convidada na FFUL e na FML e coordenadora da Linha de Investigação de Análise de Dados Multivariados (ADM) no CEAUL. É doutorada pela Universidade de Lisboa, Faculdade de Ciências, no domínio da ADM.

O seu e o de Fernando da Costa Nicolau, foram os primeiros doutoramentos portugueses nesta área, introduzindo aqui, em particular, uma abordagem probabilística à metodologia de análise classificatória. Ambos foram orientados pelo Professor Tiago de Oliveira, no desenvolvimento de estudos efectuados na Universidade de Paris VI e na Maison des Sciences de l'Homme. Tais estudos prévios - que obtiveram doutoramento francês do terceiro ciclo em Paris VI - levaram, regressada a Portugal, ao início da colaboração com a FPCE, então como membro do grupo de matemática e estatística da respectiva comissão instaladora, grupo presidido pelo Professor Tiago de Oliveira.

Responsável ou co-responsável, também com Fernando da Costa Nicolau, pela criação das primeiras disciplinas de ADM e áreas afins, no DEIOC, e em várias faculdades e universidades portuguesas, tem orientado ou co-orientado mestrados e doutoramentos neste campo, alguns no âmbito de programas de cooperação. Editora ou co-editora, autora ou co-autora e "reviewer" de publicações nacionais e internacionais, tem sido também organizadora ou co-organizadora e/ou membro do comité científico de congressos, IPM e workshops, bem como coordenadora ou colaboradora de diversos projectos científico-tecnológicos ou pedagógicos, nacionais e internacionais.

Sócia fundadora da SPE e da CLAD, foi primeira presidente da CLAD em 1994-2000.

Coordenadora da Secção de Formação e Ensino da CLAD desde 1998, é Presidente do Comité Educacional da IFCS desde 2002. Foi membro do Comité Financeiro da IFCS em 1995-2000, Correspondente Nacional Portuguesa da IASE em 2000-2003, membro do "IASC BoD-ERS", em 2000-2004. É membro "Eleito" do ISI, International Statistical Institute desde 1994.



# O Sistema Estatístico Nacional

Uma Componente Indispensável das Sociedades Democráticas e Desenvolvidas

*Manuel José Vilares*

## 1 - Introdução

Este artigo não se debruça sobre os métodos estatísticos, mas sim sobre a organização e o funcionamento do Sistema Estatístico Nacional, centrando-se numa reestruturação muito profunda deste Sistema ocorrida em 1989 a que o autor esteve estreitamente associado. São historiadas as tentativas anteriores, desde a fundação da Sociedade Portuguesa de Estatística (SPE), de reestruturação e identificados os principais estrangulamentos que o Sistema enfrentava. As principais alterações introduzidas com a reestruturação de 1989 para combater tais estrangulamentos e portanto melhorar o funcionamento do SEN são também explicadas. O artigo não cobre o funcionamento do SEN na última década, dada a existência de outro artigo nesta publicação e que se debruça sobre este período.

Apesar da actividade estatística remontar a vários milénios (os primeiros recenseamentos da população e da agricultura parecem ter sido feitos na China, no ano 2238 a.c. promovidos pelo Imperador Yao.), a criação de organismos oficiais encarregados especificamente da produção de informação estatística começa na Europa apenas no século XVIII, seguindo-se muito frequentemente a criação de tais organismos ao nascimento dos estados modernos. Aliás o termo estatística, intimamente associado ao próprio conceito de Estado, é pela primeira vez utilizado pelo filósofo alemão Gottfried Achenwall apenas nos meados do século XVIII.<sup>1</sup>

---

<sup>1</sup>Servi-nos-emos neste artigo da publicação do autor : Vilares, M.J. *Sistemas de Informação Estatística* – Instituto Superior de estatística e gestão de Informação da Universidade Nova de Lisboa, ISBN 972-8093-00-4, 1993. O leitor poderá encontrar nesta referência, bibliografia sobre os temas abordados neste artigo.

O desenvolvimento verificado nos últimos séculos tornou a informação cada vez mais um elemento indispensável à tomada de decisões, quer sejam da responsabilidade do sector público, quer do sector privado. As empresas modernas passaram a tratar a informação como um recurso essencial para a sua gestão, existindo indiscutivelmente uma elevada correlação positiva entre o desenvolvimento económico e social de uma sociedade e o desenvolvimento do seu sistema estatístico.

A adesão de Portugal, em 1986, às então Comunidades Europeias veio tornar ainda mais premente a necessidade de dispor de um sistema estatístico eficiente, na medida em que grande parte das políticas comunitárias já então eram baseadas em indicadores estatísticos que, ou eram fornecidos com rigor pelos países membros, ou então eram estimados pelos serviços da Comissão com todos os riscos que daí podiam recorrer<sup>2</sup>.

Existiam então à data condições particularmente oportunas para a uma reestruturação do Sistema Estatístico Nacional (SEN)<sup>3</sup> a qual, a ser adiada, poderia acarretar elevados custos para o desenvolvimento do nosso país.

Este artigo centrar-se-á sobre a reestruturação do SEN operada neste período, que para simplificar, será designada por Reestruturação de 1989. Começar-se-á (na secção 2) por apresentar uma breve súmula dos projectos de reestruturação do SEN anteriores, mas posteriores a 1980. Na secção 3, serão identificados os estrangulamentos mais importantes do SEN e, na secção 4, serão apresentadas as principais componentes da reestruturação destinadas a combater tais estrangulamentos. Este artigo não se ocupará da evolução do SEN na última década, dado que a análise de um tal período constitui o objectivo de outro artigo desta mesma publicação da Sociedade Portuguesa de Estatística.

## 2 - Projectos anteriores de reestruturação do SEN

O Sistema Estatístico Nacional orientou-se, até 1989, por determinados princípios básicos os quais se mantiveram praticamente inalterados desde 1935, data em que foram pela primeira vez estabelecidos através da lei n.º 1911 de 25 de Maio que criou o Instituto Nacional de Estatística (INE). Culminou-se, deste modo, o período mais fecundo da história das estatísticas portuguesas durante o qual foi construído o ainda edifício sede do INE. Esta situação manteve-se, não obstante o Sistema ter sido sujeito a duas importantes

---

<sup>2</sup> O papel da informação estatística aparece pouco depois da sua adesão reforçado em dois documentos de particular importância para Portugal visto que regulavam quer os fundos recebidos do orçamento comunitário, quer as contribuições para esse mesmo orçamento : o Regulamento Quadro das Intervenções Comunitárias (de Julho de 1988) e a Directiva PNB (de Fevereiro de 1989). O primeiro definia as regiões abrangidas pelos objectivos 1 e 2 com base num conjunto de indicadores estatísticos e a Directiva do PNB estabelecia como recurso comunitário adicional (4.º Recurso) uma percentagem do Produto Nacional Bruto.

<sup>3</sup> O Sistema estatístico Nacional (SEN) é definido como o conjunto de organismos públicos com competência legal na produção e difusão de informação estatística. Por seu lado, a informação estatística é definida como a informação que satisfazendo um dado conjunto de conceitos, definições e classificações se integra harmoniosamente num dado sistema coerente. Já a informação estatística oficial corresponde à informação estatística produzida e difundida pelo SEN.

alterações – a primeira em 1966 (Decretos-Lei nº 46925 e 46926 de 29 de Março) e a segunda em 1973 (Decretos-Lei nº 427/73 e 428/73, ambos de 25 de Agosto).

Dadas as insuficiências reveladas pelo Sistema para cumprir as suas obrigações foram feitas com uma certa periodicidade, várias tentativas para o reestruturar. Entre estas tentativas, são de referir, a partir de 1980, as seguintes:

Em 18 de Dezembro de 1980 – Comissão Permanente do Desenvolvimento e Planeamento Estatísticos. O CNE decidiu, nessa data, que esta comissão procedesse à “Reflexão sobre o SEN e os seus princípios de funcionamento”. Esta Comissão funcionou na forma de grupo de trabalho, reunindo pontualmente entre Janeiro de 1981 e Janeiro de 1982 e tendo produzido um documento demasiado genérico, atendendo aos objectivos em vista e à complexidade do problema.

Em 3 de Maio de 1982 – Grupo de Trabalho “Ad-hoc” para a reflexão sobre os princípios do SEN. Tendo realizado algumas reuniões, dele não resultou qualquer decisão digna de registo.

Em 28 de Dezembro de 1982 – Grupo de Trabalho para análise dos princípios do SEN. O grupo de trabalho, aparecido em 3 de Maio de 1982, foi agora recriado por despacho do Ministro das Finanças e do Plano. Reuniu, como tal, apenas uma vez.

Em 4 de Março de 1983 – Subcomissão especializada para a reflexão sobre os princípios do SEN. Esta é uma mera mudança de designação do grupo de trabalho criado em 28/12/1982. Reuniu 7 vezes entre 22/3/1983 e 30/6/1983. A reflexão foi consubstanciada num relatório de progresso que não chegou a dar origem a qualquer decisão.

É ainda de referir uma tentativa de reestruturação dos serviços do INE, levada a efeito em Fevereiro de 1984, através de um projecto de diploma que viria a ser apreciado em estudo realizado por uma equipa de trabalho, criada no âmbito da Secretaria de Estado da Administração Pública, e integrando representantes da Direcção Geral da Organização Administrativa (DGOA) e da Direcção geral da Administração e Função Pública (DGAFP). Esse projecto e diploma não viria, no entanto, a ser aprovado.

Face aos insucessos e reduzido alcance das acções anteriores, o X Governo Constitucional criou a Comissão de Reestruturação do Sistema Estatístico Nacional, por resolução do Conselho de Ministros nº 48-B/86, de 25 de Junho, a qual foi mandatada para proceder (até Fevereiro de 1987) a um rigoroso levantamento da situação do sistema estatístico nacional e preparar a programação de todo um conjunto de acções e decisões em ordem a atingir um quadro de objectivos para aquele sistema. Esta Comissão, para além de ser nomeada por uma Resolução do Conselho de Ministros, apresentou a particularidade, que veio a revelar-se fundamental para o bom andamento dos trabalhos, do seu Presidente ter sido simultaneamente nomeado Presidente do Conselho de Direcção do INE.

Nos pontos seguintes apresenta-se, numa forma necessariamente resumida, os principais estrangulamentos do Sistema Estatístico Nacional e as propostas para os ultrapassar, constantes no Relatório desta Comissão (entregue em Fevereiro de 1987)<sup>4</sup>.

<sup>4</sup> Este Relatório foi objecto de publicação: Vilares, M. J.; Marques, M.O.; Santos, A.D. – *Sistema Estatístico Nacional - Situação Actual e Propostas de Reestruturação* - Ministério do Planeamento e da administração do Território, 1989

### 3 - Principais estrangulamentos do SEN

Entre os principais estrangulamentos do Sistema Estatístico Nacional detectados, em 1986, podem apontar-se para além de deficiências na organização interna, no funcionamento, na estratégia e nas instalações do INE, diversos desajustamentos de ordem legal e estrutural, designadamente:

- *Um inadequado ordenamento jurídico.* A legislação que regulava o Sistema Estatístico Nacional estabelecia, com excessivo detalhe, o modo de funcionamento de cada um dos elementos do Sistema, constituindo um factor de grande rigidez, bloqueador da sua actividade. Numa área de actuação em constante mutação como é a informação, os elementos do Sistema e em particular o INE têm de ter a possibilidade de se adaptar a tais alterações.
- *Deficiências na composição e no funcionamento do Conselho Nacional de Estatística.* Por ser constituído essencialmente por representantes da Administração Pública, este Conselho não reflectia os interesses dos utilizadores, dos produtores e dos informadores do sistema estatístico, afectando-se assim a sua capacidade de intervenção, facto que, aliado às suas limitadas competências e às deficiências do seu funcionamento interno, acarretaram a inoperância do CNE, impedindo-o de coordenar e orientar superiormente o Sistema;<sup>5</sup>
- *A excessiva rigidez da aplicação das actuais normas do segredo estatístico e a forma incoerente como elas tinham vindo a ser interpretadas.* A legislação na garantia o respeito pela confidencialidade da informação e, por outro lado, dificultava o cumprimento da missão fundamental que compete aos produtores do sistema estatístico nacional que é informar;<sup>6</sup>
- *A excessiva centralização geográfica e funcional da actividade estatística, aliada à incapacidade de os órgãos do sistema estatístico nacional de corresponderem às exigências de uma tal centralização*<sup>7</sup>. Verificou-se assim, a proliferação dos serviços estatísticos ministeriais à revelia dos princípios que legalmente norteavam o sistema estatístico e em desprezo de todas as conveniências de ordem funcional e técnica. Abandonou-se, por completo, a tarefa de coordenação do sistema. Poder-se-á afirmar que, nestas condições, o Sistema Estatístico Nacional tinha deixado de existir, daí resultando vários e graves inconvenientes, tais como:
  - i) insatisfação dos utilizadores devido à existência de uma produção estatística

<sup>5</sup> A última reunião plenária do CNE verificou-se em 21-03-1985, portanto cerca de ano e meio antes da Comissão de Reestruturação do Sistema Estatístico Nacional iniciar as suas funções.

<sup>6</sup> Exemplo: De acordo com estas normas, não era claro que o INE pudesse fornecer informação sobre o nº de alunos e os professores de *um* qualquer liceu ou sobre o número de camas de *um* dado hospital visto tratar-se de informação sobre *uma* unidade estatística e não ser feita *qualquer* distinção sobre a *proveniência da informação*.

<sup>7</sup> Um sistema estatístico é *centralizado funcionalmente* se as actividades de produção, difusão e gestão de informação estatística oficial estiverem confiadas a *um único serviço* de estatística o qual pode, no entanto, possuir uma descentralização regional. Um sistema é *centralizado geograficamente* se tais actividades estiverem concentradas num *único local* (regra geral, a capital) do país.

repleta por um lado de lacunas e, por outro lado, de duplicações; ii) contestação crescente por parte dos informadores já que, eram sobrecarregados por pedidos da mesma informação de base, e não menos importante,(iii) o desperdício na utilização dos recursos públicos.

- *O inadequado estatuto de que gozava então o INE* – Este estatuto de organismo público simples, sem qualquer autonomia, impunha-lhe uma rigidez de funcionamento administrativo que se traduzia em tratar um vasto organismo entregue à actividade de produzir informação estatística como um serviço administrativo típico. De facto este estatuto não era compatível nem com as suas tarefas de organismo central de produção estatística e de órgão coordenador do sistema estatístico nacional, nem com a dimensão e características dessas tarefas em tudo semelhantes às de uma empresa produtora de informação. Por outro lado, o permanente esforço de ajustamento exigido ao INE, dada a área da sua actuação, por demais dinâmica e determinantemente afectada pelo actual desenvolvimento das modernas tecnologias de informação, não se compadecia com uma estrutura rígida.
- *A escassez de quadros superiores especializados no domínio da concepção, tratamento e análise da informação estatística.* Esta escassez resultava não só do número reduzido de quadros superiores empregues no Sistema, mas também da inexistência de uma estrutura de formação ligada ao Sistema que uniformizasse e complementasse a formação prestada pelas universidades nesta matéria, à qual faltava a orientação para as exigências concretas da prática profissional.

#### **4 - As alterações introduzidas com a reestruturação de 1989**

O objectivo central da reestruturação do SEN, como não podia deixar de ser, era produzir, atempadamente e ao mais baixo custo, informação fiável e adequada às necessidades dos utilizadores públicos e privados de modo a que o nosso país pudesse enfrentar os múltiplos desafios que se avizinhavam designadamente os decorrentes da instauração da então União Económica e Monetária.

Para permitir alcançar este objectivo central tornava-se necessário proceder simultaneamente:

- a uma alteração no ordenamento jurídico;
- a uma redefinição dos princípios básicos do SEN;
- a uma reorganização da estrutura institucional do SEN e à concessão de um novo estatuto ao INE;
- à criação, no seio da Universidade, de uma estrutura de formação para quadros superiores do SEN;

#### **4.1 - Um Novo Ordenamento Jurídico**

O ordenamento jurídico do SEN passou a assentar nos dois seguintes diplomas (todos publicados em 1989):

- Lei de Bases do Sistema Estatístico Nacional (Lei nº 6/89 de 15 de Abril)<sup>8</sup>.
- Decreto-Lei para os estatutos do INE (Decreto - Lei 280/89 de 23 de Agosto).

Por outro lado, o regime, carreira, categorias e remuneração de pessoal do INE passam a ser aprovados por Despacho do Ministro da Tutela do INE. Já a estrutura organizacional do INE, bem como o funcionamento das suas Direcções Regionais passam a constar de simples Regulamento Interno, aprovado pela Direcção do INE.

Trata-se de um ordenamento lógico e simples que permitiu revogar algumas dezenas de diplomas legais que tornavam a legislação pouco clara e transparente.

#### **4.2 - Redefinição dos Princípios Básicos do SEN**

Os sistemas estatísticos dos países desenvolvidos regem-se por um conjunto de princípios básicos: segredo estatístico, grau de centralização, coordenação, autoridade estatística e autonomia técnica. A lei de bases do SEN de 1989 introduziu alterações em todos eles, aproximando-os dos adoptados nos sistemas estatísticos dos países mais desenvolvidos e dos recomendados pelas organizações internacionais na área da estatística oficial.

##### **4.2.1 - Segredo Estatístico**

É um dos problemas mais delicados com que os sistemas estatísticos se vêm confrontados. Por um lado, é indispensável para salvaguardar a privacidade dos cidadãos, preservar a concorrência e, por conseguinte, garantir a confiança dos prestadores de informação no Sistema. Por outro lado, o segredo estatístico pode impedir os produtores de informação de cumprir cabalmente a sua missão fundamental que é a de informar. A nova legislação foi no sentido do equilíbrio e do compromisso entre estes dois aspectos.

Sem pôr em causa a privacidade individual e a defesa da concorrência, optou-se por uma orientação transparente e flexível. Assim, contrariamente à situação então vigente onde nenhuma distinção era feita, passou a adoptar-se comportamento diverso consoante se tratava: a) informação sobre as famílias; b) informação sobre a administração pública; c) informação proveniente de outras fontes (empresas públicas e privadas, cooperativas, etc.)

---

<sup>8</sup> Esta lei foi aprovada no parlamento sem um único voto contra.

No que respeita ao primeiro grupo, o princípio foi o do segredo absoluto pelo que as informações sobre os indivíduos ou as famílias nunca poderão ser divulgadas. Quanto ao segundo grupo, e salvo disposição legal em contrário, as informações sobre a Administração Pública deixaram de estar abrangidas pelo segredo estatístico. No que respeita ao terceiro grupo, a informação continua sob alçada do segredo estatístico, admitindo-se, contudo, derrogações a conceder pontualmente. Alterou-se, no entanto, a entidade que se deve pronunciar quanto à derrogação: deixou de ser o ministro que tutela o INE para passar a ser o Conselho Superior de Estatística, no qual têm assento representantes dos próprios prestadores de informação.

Naturalmente que esta legislação é só por si claramente insuficiente para impedir, de facto, a identificação não autorizada de uma empresa, duma família ou mesmo de um indivíduo, a partir da informação que estas entidades prestaram ao SEN em geral e ao INE, em particular. Afigura-se neste (como aliás na generalidade dos casos) indispensável regulamentar os aspectos práticos. Foi com este objectivo que o Conselho Superior de Estatística criou uma secção permanente, designada por Secção do Segredo Estatístico, presidida pelo Ministério da Justiça que criou um Regulamento onde são reguladas as condições para a libertação do segredo estatístico. Foi também para reforçar o cumprimento da confidencialidade que o Regulamento de Pessoal do INE integrou a violação não autorizada do segredo estatístico como uma falta profissional grave.

#### **4.2.2 - Centralização estatística**

Ao analisar o grau de centralização dos sistemas estatísticos dos países desenvolvidos, constatam-se dois pontos. Por um lado, não existem sistemas puros, ou seja inteiramente centralizados ou inteiramente descentralizados, quer funcionalmente quer geograficamente. Por outro lado, existem países com sistemas centralizados e outros com sistemas descentralizados, existindo argumentos a favor e contra os dois modelos de centralização. A opção por um destes modelos está estreitamente associada à história e à realidade económica e social de cada país. No caso de Portugal, o pendor demasiado centralizado do Sistema Estatístico Nacional foi abandonado quer avançando inequivocamente na descentralização geográfica, quer flexibilizando a descentralização funcional.

A descentralização geográfica foi implementada através da criação de Direcções Regionais do INE, com âmbito coincidente com o nível II das NUTS – Nomenclatura das Unidades Territoriais para Fins Estatísticos. Foram assim criadas, de imediato, as Direcções Regionais do Norte (com sede no Porto), do Centro (com sede em Coimbra), no Alentejo (com sede em Évora) e mais tarde as Direcções Regionais do Algarve (com sede em Faro) e de Lisboa e Vale do Tejo (com sede em Lisboa). Foram atribuídas a estas Direcções Regionais áreas de responsabilidade e funções relevantes na produção e difusão da informação estatística e, em particular, na de carácter regional, mantendo-se igualmente, dentro desta óptica, as competências dos já existentes Serviços Regionais de Estatística dos Açores e da Madeira.

Esta descentralização não se justificava por si própria mas em função de um conjunto de objectivos destinados a melhorar a eficiência do SEN. Entre estes objectivos destacavam-se: i) diminuir os custos e melhorar a qualidade da informação estatística; ii) aproximar a informação dos utilizadores e dos prestadores de informação; iii) fomentar a utilização de estatísticas de âmbito regional, aspecto tanto mais importante, quanto maior for a atenção acordada ao desenvolvimento equilibrado do conjunto do país.

A delegação funcional foi flexibilizada na medida em que, por um lado, o Conselho Superior de Estatística passou a poder delegar competências do INE noutras entidades, mesmo contra a vontade do próprio INE e, por outro lado, apenas se passou a exigir a satisfação de uma restrição a esta delegação: a necessidade absoluta da informação delegada ser devidamente coordenada.

#### 4.2.3 - Coordenação técnica

O reforço dos meios de coordenação foi um dos objectivos prioritários da reestruturação do SEN, necessidade acrescida pela maior descentralização permitida. A experiência tinha demonstrado que a ausência de coordenação induzia a destruição do próprio Sistema Estatístico Nacional.

A legislação aprovada em 1989 veio criar condições para reforçar a coordenação técnica do Sistema<sup>9</sup>. De facto:

- A aprovação de projectos de diplomas que criem *serviços de estatística ou contenham quaisquer normas com incidência na estrutura ou funcionamento do Sistema Estatístico Nacional* passa a ser obrigatoriamente precedida de audição do Conselho superior de Estatística. (Artigo 24º da Lei 6/89). Procura-se deste modo evitar repetir a proliferação de serviço de estatística pelos Ministérios, sem qualquer integração no SEN.
- A aprovação de conceitos, definições e nomenclaturas estatísticas passa a constituir competência exclusiva do Conselho Superior de Estatística. Pretende-se que, não só o INE, mas todos os serviços de estatística com competência delegada, passassem a utilizar os mesmos conceitos, definições e nomenclaturas.
- A delegação de competências do INE noutros serviços públicos passa a poder cessar quando estes não respeitassem as exigências de coordenação estatística. A certificação da capacidade destes serviços constituía um objectivo a prosseguir.
- A criação de uma instituição universitária especialmente vocacionada para a formação de quadros superiores de estatística. Esta Instituição (identificada na secção 44) deveria tornar-se num verdadeiro veículo de difusão da “cultura estatística” devendo em consequência constituir um instrumento de coordenação por excelência.

<sup>9</sup> Este reforço diz sobretudo respeito à coordenação do conjunto do SEN (também conhecida como *coordenação externa*). No entanto, estes mesmos factores contribuem para a melhoria da coordenação dentro do próprio INE (também conhecida como *coordenação interna*).

A alteração do estatuto do INE e o reforço dos seus meios de intervenção deveriam permitir a este Instituto exercer *de facto* as funções de coordenação, tarefa delicada e para a qual se exigem técnicos altamente qualificados.

#### 4.2.4 - Autoridade estatística

O princípio de autoridade estatística foi reforçado em dois domínios:

- *o domínio das transgressões estatísticas*: Procedeu-se a um agravamento dos limites das multas aplicáveis e foi instituído um mecanismo de actualização. Este reforço justificava-se dada a profunda ineficácia dos processos de transgressão em geral e a total desactualização das penalizações aplicadas (a última actualização tinha sido estabelecida pelo Decreto-Lei nº 131/82 de 23 de Abril). Realce-se, contudo, que anteriormente a qualquer processo de transgressão, deve o INE fomentar um bom relacionamento com os fornecedores de informação.
- *no domínio do acesso do INE às fontes administrativas*. Nos termos do novos estatutos (artigo 4º -3 do DL 280/89 de 23 de Agosto, O INE passa a *poder aceder à informação individualizada relativa às cooperativas, empresas pública e privadas, instituições de crédito e outras pessoas singulares, recolhidas no quadro da sua missão, pelas administrações, autarquias locais, ou instituições de direito privado que tenham como atribuição a gestão de um serviço público*. Este acesso deve permitir aproveitar as vantagens inquestionáveis desta fonte de informação, designadamente o baixo custo (frequentemente nulo) e a diminuição da sobrecarga sobre os prestadores de informação (que é praticamente reduzida a zero), constituindo um importante factor de desburocratização e consequente modernização da Administração Pública. De facto, deste modo, os prestadores de informação (e muito particularmente as empresas) passam a responder *simultaneamente* às exigências administrativas e às exigências do sistema estatístico.

#### 4.2.5 - Autonomia Técnica

A reestruturação do SEN criou condições de reforço do exercício da autonomia técnica. Com efeito, enquanto na legislação anterior, era apenas salientado que *no desempenho das suas atribuições, o INE goza de autonomia técnica (cf art 11º do DL 427/73)*, com a reestruturação, verificam-se as seguintes alterações:

- a) A lei de bases do SEN passa a incluir um artigo (o 4º) exclusivamente dedicado a este princípio:

1. *No exercício da sua actividade os órgãos do INE gozam de autonomia técnica.*

2. *A autonomia técnica consiste no poder conferido aos órgãos de definir livremente os meios tecnicamente mais ajustados à prossecução das atribuições do INE, agindo no âmbito da sua competência técnica, com inteira independência.*
3. *O INE tem competência para tornar disponíveis, divulgar e difundir os resultados da actividade desenvolvida no quadro das suas atribuições, sem prejuízo do respeito pelas regras de segredo estatístico.*

- b) É acordado ao INE um estatuto de grande autonomia o que permite reforçar as condições de exercício da autonomia técnica.
- c) As competências e a composição do Conselho Superior de Estatística (CSE) são reforçadas em detrimento dos poderes de tutela governamental do INE, sendo este Conselho que se torna *de facto* a verdadeira tutela do INE;

#### **4.3 - Uma Nova Estrutura Organizacional do Sistema**

A nova estrutura do Sistema Estatístico Nacional passa assentar em dois pilares fundamentais:

- a) Conselho Superior de Estatística
- b) O Instituto Nacional de Estatística

*O Conselho Superior de Estatística passa a ser definido como o órgão de Estado que superiormente coordena e orienta o Sistema Estatístico Nacional. Este Conselho substituiu o Conselho Nacional de Estatística por reforço de competências e atribuições deste, bem como pela profunda alteração da sua composição. Assim, passaram a ser novas competências do CSE: a proposta de delegação de competências do INE noutros serviços públicos; a libertação do segredo estatístico e a apreciação do plano de actividades do INE e do correspondente relatório final.*

Por outro lado, a composição do CSE foi profundamente alterada de modo a que este passasse a constituir um fórum representativo dos interesses dos prestadores e dos utilizadores da informação estatística. Assim, contrariamente à situação anterior onde existiam apenas representantes da Administração Pública, o novo CSE passou a incluir representantes das centrais sindicais, das associações empresariais, das associações de consumidores e dois professores universitários nas áreas dos métodos estatísticos e econométricos.

Paralelamente, eliminaram-se as Comissões Consultivas de Estatística, nas quais apenas tinham assento representantes da Administração Pública, passando as suas atribuições a serem exercidas por secções criadas no âmbito do Conselho Superior de Estatística.

*O Instituto Nacional de Estatística é definido como o organismo central do Sistema Estatístico Nacional, que passou a ser profundamente reestruturado.*

Ao alterar o estatuto de organismo simples sem qualquer tipo de autonomia para o de instituto público, com autonomia administrativa, financeira e patrimonial, visavam-se entre outros, os seguintes objectivos: i) alterar a filosofia de gestão do INE de modo que a componente económica e financeira passasse a intervir claramente e directamente nas decisões; ii) conferir mobilidade aos meios e flexibilidade ao funcionamento interno e às ligações ao exterior, possibilitando a adequação da gestão às características do processo de obtenção de produtos estatísticos muito semelhante a um processo emoesarial típico; iii) reforçar a capacidade institucional necessária às exigências acrescidas de coordenação estatística; iv) incentivar a produção da informação estatística na perspectiva dos utilizadores facilitando a repercussão dos custos nos mesmos, aliviando deste modo os encargos a suportar pelo Orçamento de Estado que deveriam tendencialmente limitar-se à função social da estatística a exemplo, aliás, do que vinha a verificar-se com os modernos institutos de estatística dos países desenvolvidos.

Conforme se pode constar pela análise então efectuada dos sistemas estatísticos países da União Europeia, do Canadá e dos Estados Unidos da América, Portugal passou a dispor dos sistemas estatísticos com maior autonomia (designadamente em relação à influência governamental), quer em termos dos estatutos do INE, quer em termos da composição e das competências do Conselho Superior de Estatística.<sup>10</sup>

#### 4.4 - Formação de Quadros Superiores

A lei de bases incumbiu o INE (no artigo 14º-4) de *“promover em conjunto com instituições de ensino superior universitário a criação de cursos nos domínios da concepção e da aplicação estatística aos quadros da Administração Pública, empresas públicas e privadas e, em particular, aos quadros do sistema Estatístico Nacional, bem como acções de cooperação, nomeadamente com os países africanos de língua oficial portuguesa, no âmbito da formação de quadros superiores de estatística”*

Neste contexto, foram tomadas duas iniciativas complementares: a criação do Instituto Superior de Estatística e Gestão de Informação da Universidade Nova de Lisboa (ISEGI) e a criação do Centro de Estudos para os Países em Vias de Desenvolvimento (CESD-Lisboa)

O ISEGI (criado através do Decreto-Lei nº417/89 de 30 de Novembro), tem na origem dois tipos de atribuições. (Artigo 2º) a) organizar e orientar actividades de ensino e e investigação nos domínios da concepção, tratamento, análise e difusão da informação estatística, e b) promover acções de cooperação nos domínios referidos, nomeadamente com os Países Africanos de Língua Oficial Portuguesa Países Africanos de Língua Oficial Portuguesa (PALOP).

Um objectivo fundamental deste Instituto é assim tornar-se num veículo de difusão da “cultura estatística” no âmbito do Sistema Estatístico Nacional e paralelamente um

<sup>10</sup> Naturalmente que esta autonomia ainda podia teoricamente ser reforçada, sobretudo em relação à forma de nomeação e exoneração do Presidente do INE e à composição (incluindo a presidência) do Conselho Superior de Estatística.

instrumento de ligação deste Sistema à Universidade e aos PALOP. Este Instituto deveria portanto contribuir para que fosse possível estender a delegação funcional, alargar a cobertura estatística e, simultaneamente, reforçar a capacidade de coordenação do Sistema. Deveria ainda possibilitar a produção de informação estatística de qualidade ao mais baixo custo já que, num tal contexto, o sistema estatístico passava a estar em condições de beneficiar do extraordinário impacto que as novas tecnologias de informação e comunicação tiveram nos domínios da recolha, produção, análise e difusão de informação estatística. O ISEGI deveria, em síntese, contribuir para eliminar um dos principais estrangulamentos com que se debatia a Sistema Estatístico Nacional e o INE em particular e paralelamente responder, de modo adequado, às necessidades de reforço de base institucional para a cooperação com os PALOP.

O CESD-Lisboa - Centro Europeu de Estatística para os Países em Vias de Desenvolvimento -, criado por iniciativa conjunta do INE, da Universidade Nova de Lisboa, do Instituto para a Cooperação Portuguesa e da Comissão das Comunidades Europeias, visava precisamente beneficiar do facto de Portugal ser um país membro das então Comunidades Europeias e de, simultaneamente, ter uma relação histórica profunda e uma língua comum com os PALOP.

O CESD-Lisboa tinha assim como objectivo central a formação de quadros estatísticos de nível superior para os Países em Vias de Desenvolvimento e, em particular, para os PALOP. O CESD-Lisboa deveria igualmente fornecer cursos de curta duração e estar em condições de realizar estudos, investigação e desenvolvimento. As actividades do CESD-Lisboa funcionavam no seio e em articulação com o ISEGI.

O CESD-Lisboa passou a integrar uma rede que incluía o CESD-Paris (o primeiro), o CESD-Madrid, o CES-Roma e o CESD-Comunitário que coordenava as actividades do conjunto da rede.

## 5. Conclusão

Passaram dezasseis anos sobre a publicação dos diplomas legais relativos á profunda reestruturação do SEN e do INE em particular, a qual na opinião de Fernando de Sousa, autor da História da Estatística em Portugal<sup>11</sup> *pode ser considerada um marco na história do Sistema Estatístico Nacional* (p. 236) pois *só em 1989 se conseguiu introduzir a urgente reforma de fundo que concedeu ao INE personalidade jurídica, autonomia administrativa e financeira e património próprio, no sentido de lhe permitir dar a resposta adequada aos novos desafios das tecnologias de informação e comunicação e às exigências decorrentes da adesão de Portugal às Comunidades Europeias* (p.247). Impunha-se agora analisar se os princípios que nortearam a reestruturação foram respeitados e se os objectivos então pretendidos foram alcançados. Tais tarefas caem fora do âmbito deste artigo.

---

<sup>11</sup> Sousa, F. *História da Estatística em Portugal*, INE, Lisboa, 1995

## **Manuel José Vilares**

É Professor catedrático e Presidente do Conselho Científico do Instituto Superior de Estatística e Gestão de Informação da Universidade Nova de Lisboa (ISEGI/UNL) e consultor do Banco de Portugal. É também Peer Reviewer de projectos financiados pela União Europeia.

Licenciatura em Economia pela Universidade do Porto (1975), Doutoramento de Estado em Economia pela Universidade de Dijon (1983) e Agregado em Econometria pela Universidade Nova de Lisboa (1987).

É autor de livros e artigos publicados em Portugal e no estrangeiro nas áreas dos sistemas de informação estatística, métodos econométricos e modelos macroeconómicos.

Foi Presidente da Comissão de Reestruturação do Sistema Estatístico Nacional (1986/87) e Presidente do Instituto Nacional de Estatística (1986/92). Foi Presidente da Comissão Instaladora (1990-1994) e primeiro Director (1994-1998) do ISEGI/UNL. Foi igualmente consultor de alto nível da União Europeia para os assuntos de estatística e economia (1995-1999).



# “Outliers” em Português

*Fernando Rosado*

## 1. Na década de 70 (I)

A Estatística tornou-se conhecida no século XX como um instrumento matemático para analisar os dados e afirmou-se como ciência na sequência da contribuição inovadora desenvolvida por grandes cientistas do século XIX que foram introduzindo o “pensamento estatístico” nas diversas áreas do saber. A (criação da) Estatística é, portanto, o culminar de uma “tradição de pensamento científico”. Quando e como, despertou Portugal para essa realidade (estatística) mundial?

Situemo-nos na década de 70 do século passado. Portugal vive “momentos novos” nos mais diversos caminhos e também no campo da ciência. Com especial incidência a partir dos anos 60, muitos portugueses migravam no seu país, procurando melhores condições de vida ou, (pelo menos,) trabalho. Alguns, não poucos, partiam para países terceiros onde “a certeza” de “melhor vida” ajudava a enfrentar esse desafio. Este estado de espírito português também arrastava e desafiava os jovens para novos caminhos. Alguns descobriam assim o percurso para dizer “não à guerra colonial”. Muitos jovens de Portugal, terão aproveitado esta “motivação” que lhes trouxe a feliz consequência de progressão (também) no saber. Na sua grande maioria eram oriundos das grandes cidades e das zonas “mais evoluídas”. Ser jovem no interior do país e com ambições de “estudos avançados” implicava a migração - eventualmente de toda a família - para uma das três cidades com universidade. Para essa minoria, na maioria dos casos, os percursos eram bastante sinuosos (e aleatórios?) tornando difícil atingir o “grande objectivo familiar” - completar “um curso superior”.

Nesta época, na língua portuguesa, as palavras mestrado e doutoramento tinham um significado não muito bem definido e, sempre se relacionavam com graus científicos “do estrangeiro”. No início dos anos 80 começaram a ser criados os Mestrados em Probabilidades e Estatística. Doutoramento em Portugal era um acontecimento raro e a especialidade de Probabilidades e Estatística não existia. Muito se avançou nos últimos 20 anos... Os estudos pós-graduados estão a dar os primeiros passos...

Nestes anos - de pós-guerra colonial - Portugal iniciava-se pois, para um lugar na Ciência. É uma excelente referência temporal para se iniciar um memorial que – como “aquele outro” de enorme sucesso - também envolve “homens e formigas” que levam “isto daqui para ali porque as forças não dão para mais, e depois vem outro homem que transportará a carga até à próxima formiga, até que, como de costume, tudo termina num buraco...” (citando *Memorial do Convento* de José Saramago). Na década em referência, alguns obreiros, despertavam, tentavam transformar... Uns, com a ajuda de bolsas de estudo peregrinavam lá para fora, às vezes para bem longe, por outras universidades dando passos fundamentais... Outros, poucos e em muito menor número, avançaram “dentro de portas”... com diferentes dificuldades!

Referindo uns e outros, de pioneiros estamos a falar!

Nos “idos anos oitenta” surgiu pois o despertar português para a investigação científica – nessa época incipiente (também) em estatística. Alguns históricos estatísticos portugueses - na sua maior parte incentivados por Tiago de Oliveira - juntavam-se à diáspora lusitana. Foram, com sucesso, até outras universidades aprender; para até nós trazer ciência estatística. Outros - e neste grupo me integro - pelas razões mais diversas, decidiram ficar e, acumulando ensino e investigação, também ajudaram a sedimentar novos cursos universitários ajudando a Matemática Aplicada a “dar à luz a Estatística” nas Universidades Portuguesas e aqui, é justo referir o protagonismo da Faculdade de Ciências da Universidade de Lisboa. Uns e outros, alguns anos mais tarde, congregavam esforços para alcançar “novas perspectivas” científicas em Portugal, também na ciência estatística que entretanto vinha sendo implantada a partir de pioneiros como Tiago de Oliveira, Bento Murteira e outros.

Estava portanto “acontecendo o acaso” que seria a génese da moderna Estatística em Portugal.

Um primeiro grande fruto, em 1980, foi a fundação da Sociedade Portuguesa de Estatística - SPE, (durante alguns anos incluindo também a Investigação Operacional) com a designação de Sociedade Portuguesa de Estatística e Investigação Operacional - SPEIO.

No início da década de 90, Ivette Gomes liderou o grupo que sedimentaria a SPE – associação de onde a Investigação Operacional se tinha separado pois entretanto tinha sido criada a APDIO – Associação Portuguesa para o Desenvolvimento da Investigação Operacional, onde os investigadores dessa área se congregaram.

## 2. Na década de 70 (II)

Em 1978, Barnett e Lewis publicaram a primeira edição de *Outliers in Statistical Data* – livro de base para o estudo de outliers em dados estatísticos tanto do ponto de vista teórico como prático. Nesta obra fundamental foi, pela primeira vez, agregada e sistematicamente organizada toda a vasta literatura sobre outliers.

Na segunda edição, em 1984, os autores incluíram novos temas do estudo estatístico de outliers e outros que sofreram grande evolução metodológica desde a publicação da edição anterior. Em 1994 foi publicada a terceira edição e nela foram incluídas novas abordagens para dados univariados e multivariados, apresentando ainda tópicos especiais nos métodos bayesianos e em sucessões cronológicas com os aditivos e os inovadores.

As "observações difíceis" de uma amostra sempre desafiaram os estatísticos. O conceito de outlier tem fascinado (em especial) os cientistas que numa primeira abordagem querem interpretar os dados. Os mais diversos nomes têm sido aplicados a uma observação (ou a um grupo de observações) que se apresenta diferente; desde "não representativa" até "espúria" ou "discordante", numa terminologia tão vaga quanto as outras. De facto, para uma observação ser discordante, é fundamental que se indique o modelo do qual discorda... relevando portanto o modelo de discordância.

Na época em que estamos, o registo da informação, ainda com mais ênfase permitia admitir como erros todas as observações que ao experimentador parecessem mal vindas. E as reacções foram desde os seguidores da "incondicional inclusão" - como admitem Barnett e Lewis na primeira edição da obra acima referenciada - porque "nunca devemos violar a santidade dos dados" atrevendo-nos a julgar as suas propriedades até aqueles que sempre usam "na dúvida deita-se fora" como regra prática.

Em 1976, Barnett publicou "The Ordering of Multivariate Data", um estudo fundamental cujo lema é "order properties... exist only in one dimension" e com discussão pelos melhores especialistas. É um artigo de referência que desperta para a importância da ordenação na detecção de observações discordantes. Conjugado com a dimensão dos dados estatísticos esse artigo "atravessa" muitos domínios, novos à época, como o estudo de dados multivariados e a sua relação com as sub-ordens. O "termo outlier" surge "no contexto" onde vai adquirindo cada vez mais importância à medida que se avança no estudo desse texto. Este pode ser um sinal, a palavra-chave, para o despertar de um novo campo de investigação (nesta década ainda) sem história em Portugal (e muito novo no mundo científico de então!). E assim pode acontecer (mais) um acaso científico! Este, (verificado em 1982) levaria à elaboração da tese *Existência e Detecção de Outliers - Uma Abordagem Metodológica* - Rosado (1984) - para obtenção de doutoramento na área dos outliers - o primeiro em Portugal.

Numa perspectiva actual os pontos de vista são mais sofisticados. A teoria estatística dos outliers já possui diversas metodologias de tratamento de observações discordantes ou contaminantes; têm sido propostos modelos de discordância que permitem explicar a geração dos dados; os procedimentos robustos têm tido bastante avanço (cf. Barnett and Lewis (1994)).

### 3. Um relato "na primeira pessoa"

No contexto deste Memorial, à minha modesta contribuição não pode deixar de ficar associado o nome do Professor Tiago de Oliveira que foi o meu orientador de

doutoramento e que me fez descobrir o caminho, também para ele novo, da teoria dos outliers. Recordando o Professor Tiago desloco-me no tempo e revivo bons momentos que têm início nas aulas da, na altura recém criada, Licenciatura em Matemática Aplicada na Faculdade de Ciências da Universidade de Lisboa – na “velha” Escola Politécnica e que se prolongam até às sessões de acompanhamento do meu trabalho de investigação conducente ao doutoramento que regular e semanalmente mantínhamos como agenda onde, na maior parte das vezes, eu era apenas um ouvinte atento da sua vasta cultura e eloquência que me deram a oportunidade de muito aprender e de muito crescer. O Prof. Tiago foi o meu Mestre desde os tempos da Faculdade, onde me iniciei como estudante universitário e de onde, até hoje, apenas me “afastei” para cumprir o serviço militar obrigatório nos anos “de referência” - 1973/75.

O Prof. Tiago tinha grande capacidade para o cálculo científico e era enorme a rapidez como manobrava as mais intrincadas expressões matemáticas. Quando lhe apresentei aquele que viria a ser um dos meus primeiros resultados, também para ele inesperados e inovadores, o seu grau de surpresa foi tal que replicou: “Os cálculos estarão certos?” Felizmente estavam e tive a oportunidade de ver e viver a (primeira) alegria da descoberta na presença de um grande cientista. Estes resultados iniciais conduziram à inovação científica no estudo de observações discordantes “no meio da amostra” a que (com alguma lógica) chegámos a admitir chamar inliers. Mas, o modelo de discordância é o instrumento estatístico fundamental que discrimina (e condiciona!) a condição outlier (de uma ou várias observações). Tal como não devemos distinguir entre outliers superiores e outliers inferiores pois ambos são discordantes em relação a um modelo e essa condição em nada os distingue, também abandonámos a designação inlier. É o modelo de discordância - que apelidei de generativo com alternativa natural - que “condiciona” e “permite definir” uma observação que deve ser declarada outlier; depois de esta ser descoberta através de um teste de homogeneidade à amostra.

Outliers são observações que “estatisticamente” nos surgem diferentes. No entanto, a condição outlier é fortemente condicionada pelo modelo de discordância que admitimos para os dados. E cada vez mais é uma noção usada nos computadores. Todo o pacote estatístico invoca a sua actualidade com variadas aplicações na detecção de outliers nos diferentes ramos. Sabemos como o Prof. Tiago pouco simpatizava com os computadores. Talvez as suas críticas fossem bem mais mordazes com o avanço que nos levou até à Internet, onde aparentemente qualquer leigo se pode “cultivar”, incluindo na teoria dos outliers. Basta saber “navegar” e escolher um bom “site”. Pois bem! Desde grupos artísticos e musicais (<http://www.fuzzyco.com/outliers/>) até outliers arqueológicos (<http://ecolan.sbs.ohio-state.edu/jhm/arch/outliers.html>) podemos encontrar nos momentos seguintes ao “toque do rato” no sítio certo da respectiva “home-page”. São os sinais dos tempos que nos fazem reflectir sobre caminhos percorridos e percursos vindouros. Como se deturpará uma noção pelo seu mau uso, não rigoroso e completamente vago!? Não podemos confundir a divulgação científica com o marketing. Mas, de facto, já nesta nossa época, procurando outliers na “rede global de informação”

chegamos primeiro ao acessório e só os especialistas conseguem (não o necessitando) analisar onde estão os verdadeiros outliers (o fundamental!). Com alguma dificuldade conseguimos encontrar as referências à obra base de Barnett e Lewis sobre o estudo de outliers (<http://www.amazon.com/exec/obidos/tg/detail/-/0471930946/002-4576153-0685610?v=glance>).

Numa linguagem para todos compreensível, o Prof. Tiago foi um outlier. Tal como na estatística, que tanto amou e tão apaixonadamente fez crescer e criar escola em Portugal, qualquer observação discordante só é confirmada na sua condição outlier desde que assumido algum modelo de discordância. Cientificamente perfeito, o modelo da vida não nos permite construir o respectivo teste de discordância. Com o seu desaparecimento prematuro, a mãe natureza (que costumava invocar) não lhe permitiu ver reconhecida muita da sua obra. Nesta época jubilar para a Sociedade Portuguesa de Estatística que ele tanto quis e da qual foi o principal dinamizador, em breves palavras registo, a "mais sincera homenagem"!

#### 4. "Outliers" em Português!

O glossário de termos estatísticos disponível na página web do International Statistical Institute - ISI, associação prestigiada de congregação mundial de estatísticos e onde a SPE é associação filiada desde 1988, pode ser um ponto de partida. Numa consulta àquele documento – e cada vez mais este gesto se tornará trivial – é proposta a seguinte correspondência para a palavra outliers:

- valeurs aberrantes, observations aberrantes para a língua francesa,
- valori anomali para a língua italiana,
- valores extremos, valores atípicos para a língua espanhola,
- valores de exceção (sic!) para a língua portuguesa.

Destes exemplos, que, em termos linguísticos, nos são "mais próximos" podemos concluir que é pouco eficaz o efeito prático da existência deste glossário com a agravante de existirem várias sugestões de tradução que, naturalmente, obrigam a um esclarecimento pormenorizado do sentido que se pretende dar – valores extremos não serão sempre valores atípicos. Nalguns casos, entenda-se nalgumas línguas, não há tradução para outliers – por exemplo, em dinamarquês, norueguês ou sueco.

Devemos registar ainda que a "navegação" no glossário apenas permite a correspondência num sentido. Por exemplo, não conseguiremos facilmente concluir que "valeurs aberrantes" da língua francesa deveria dar "valores de exceção" de um texto em português e que de outliers se tratava nesse trabalho. Para resolver essa questão teremos sempre de passar pela palavra outlier - "intermédia e de ligação".

Assim, a consulta do referido glossário consolida a opção pela não tradução, em palavras como outliers, embora seja muito útil na correspondência para outros termos estatísticos com tradução na língua de Camões (e aqui os exemplos serão muitos).

Desenvolvendo esta questão podemos questionar sobre a vantagem de traduzir Bootstrap ou p-value ou outros “termos estatísticos” internacionalmente esclarecidos e por todos usados como pertencendo a uma “linguagem comum” dos estatísticos; com vantagens se for universal?

Sempre que, em português, lemos “valores de exceção” ou “valores atípicos” ou “valores discordantes” – considerando apenas 3 alternativas – aparece (e é exigida?) a (necessária?) correspondente explicação de que aquela tradução corresponde a outliers. Assim, cada uma dessas expressões, em cada texto científico, não é mais do que um código de palavras que faz corresponder “valores aberrantes” a outliers para (apenas) um determinado texto e não para o mesmo autor que, noutro artigo, usa (ou pode usar) outra terminologia.

Esta questão passa perto (ou será que não?) da polémica surgida em Portugal há quase 20 anos, entre os que defendem a “... obrigatoriedade do uso do português nas ... dissertações...” como a proposta ao Governo e ao Conselho de Reitores das Universidades Portuguesas em 1988 pela Comissão Nacional da Língua Portuguesa e os que, no campo oposto, asseguram desde logo que, como primeira consequência, muito nefasta, essa será uma grave intromissão na autonomia universitária.

A questão está em aberto e enquanto assim estiver será decerto um alento para a ciência.

O fundamental, de facto, é o dinamismo da investigação e a publicação científica em português, sem prejuízo e com o maior incentivo à sua internacionalização.

A SPE, internamente, também já aflorou esta polémica e, em especial, registam-se os artigos publicados nos Boletins Informativos 2 e 3/99 e Jan/Abr 2000. Uma tão interessante quanto importante tarefa (que lhe cabe?) é a criação de um Dicionário de Estatística. É um grande desafio, dada a vastidão de assuntos e a variedade de termos e temas mas que tem garantido a divulgação enciclopédica da estatística, o que assegura o sucesso de uma obra com esse objectivo.

Para já, com a certeza de que se aumenta a sistematização e, para que mais facilmente se possa concluir em que área se inclui um determinado artigo científico através de alguma das suas “palavras - chave”, aceitemos que, em português, se escreva “outliers”. Todos saberemos de que se trata e muito mais facilmente “buscaremos” artigos do nosso interesse científico.

## 5. E o futuro?

Apesar da sua longa história, “o problema outlier” continua a despertar o maior interesse tanto do ponto de vista teórico como prático. Nos mais diversos campos e aplicações, sendo uma eventual explicação para a proliferação na terminologia da teoria dos outliers, as revistas científicas internacionais contêm cada vez mais contribuições nessa área de estudo. Vejam-se os mais importantes, por exemplo, em Applied Statistics,

Technometrics, Biometrika ou Journal of the American Statistical Association. A investigação mais recente desenvolve ainda alguns métodos informais para pesquisa de observações discordantes em modelos estruturados e apresenta questões do maior relevo para amostras multivariadas.

Da etimologia da palavra estatística resulta que o seu uso (mais ou menos) sempre se associa à colheita e ao uso de dados de modo a apoiar a administração de um estado. O sistema de justiça é, na realidade, um dos pilares fundamentais de um moderno estado e é basilar na política da maior parte dos países.

Metodologias probabilísticas já são usadas desde o século XIV para modelar e apoiar a decisão na aplicação da justiça.

Os mais recentes avanços da teoria dos "outliers" têm surgido baseados na inferência estatística para interpretar dados de um ponto de vista legal. Os tribunais estão introduzindo novos desafios para os estatísticos que assim são solicitados a pronunciar-se em domínios de trabalho não tradicionais – por exemplo a correcta aplicação da legislação envolvendo os direitos de autor ou, com muito maior impacto, as evidências bioestatísticas ou genéticas em determinada prova.

Toda a prova admissível, e não apenas a prova científica, pode desempenhar um papel fundamental em tribunal. Torna-se aqui fulcral o termo "admissível". O "julgamento" feito por um estatístico poderá ser o apoio (também científico) na decisão do tribunal. São novos temas para a estatística e, por consequência, para a teoria dos "outliers". Este é decerto o mais recente desafio para os "estatísticos dos outliers" e que se vem juntar a alguns outros objectivos científicos ainda por atingir tais como os que envolvem as metodologias multivariadas e todos os que mais directamente se relacionam com questões de modelação estatística e inferência robusta. Esse desafio envolve a própria designação e terminologia pois se poderá seguir para a "nomo-estatística" (se optarmos pela etimologia do latim) ou "dicometria" (se usarmos as origens gregas), dando pois a possibilidade de, em breve, se começar a usar "nomo-outliers" ou dico-outliers". E esse futuro dos "outliers em tribunal" já começou. É bastante a referência histórica dos exemplos enunciados em Barnett and Lewis (1994, p. 4-7)

No futuro, cada vez mais, os "outliers" continuarão a ocupar um lugar do centro na ciência estatística e nos métodos estatísticos, pois sempre uma observação discordante será um desafio para o analista e da qual poderá depender o seu relatório final para a mais importante tomada de decisão. Mas, quando tudo está dito e feito o principal problema no estudo de observações eventualmente suspeitas, continua a ser aquele que desafiou os primeiros investigadores - O que é um "outlier" e como se deve trabalhar com essa observação?

## Referências Bibliográficas

- Barnett, V. (1976) – The Ordering of Multivariate Data (with discussion). *Journal of Royal Statistical Society A*, p. 318-354.
- Barnett, V. and Lewis, T. (1994) – *Outliers in Statistical Data*. 3<sup>rd</sup> edition. Wiley.
- Rosado, F. (1984) - *Existência e Detecção de Outliers - Uma Abordagem Metodológica*. Tese de Doutoramento. Universidade de Lisboa.

## Anexo

Uma Lista Bibliográfica, (obviamente incompleta) com “Outliers em português”

- Almeida, C. (2001) – *Máxima Verosimilhança e Detecção de Outliers*. Dissertação de Mestrado. Universidade de Lisboa / Faculdade de Ciências.
- Alpiarça, I. (1995) – *Outliers em Localização e Escala para Populações Exponenciais e Gama*. Dissertação de Mestrado. Universidade de Lisboa. Faculdade de Ciências.
- Braumann, M.M. (1989) – *Testes de Discordância para “Outliers” em Populações Normais e Gama*. Provas de Aptidão Pedagógica e Capacidade Científica. Universidade de Évora.
- Braumann, M. M. (1994) – *Sobre Testes de Detecção de “Outliers” em Populações Exponenciais*. Dissertação de Doutoramento. Universidade de Évora.
- Costa, S. (2005) – *Análise Estatística Multivariada na Segmentação de uma Companhia de Seguros*. Dissertação de Mestrado. Universidade de Lisboa. Faculdade de Ciências.
- Jorge, A. (1999) – *Sobre a Definição de Outlier no Domínio Específico dos Modelos Lineares e Séries Temporais*. Dissertação de Mestrado. Universidade Técnica de Lisboa. Instituto Superior de Economia e Gestão.
- Martins, S. (2000) – *Medidas de “Performance” em Modelos de Discordância Exponenciais*. Dissertação de Mestrado. Universidade de Lisboa. Faculdade de Ciências.
- Mendes, Z. (1993) – *Modelos e Testes de Discordância para Outliers em Populações Normais*. Dissertação de Mestrado. Universidade de Lisboa. Faculdade de Ciências.
- Oliveira, P. (1988) – *Tratamento Estatístico de “Outliers”*. Provas de Aptidão Pedagógica e Capacidade Científica. Universidade do Minho.
- Palma, J. (1998) – *Outliers em Séries Temporais. Uma abordagem no domínio dos modelos ARMA*. Dissertação de Mestrado. Universidade de Lisboa. Faculdade de Ciências.
- Passos, J. (1992) – *Influência das Observações nos Coeficientes Estimados no Modelo de Regressão Múltipla*. Dissertação de Mestrado. Universidade Técnica de Lisboa/Instituto Superior de Economia e Gestão.

- Figueira, M.M. (1995) – *Identificação de Outliers: uma aplicação ao conjunto das maiores empresas com actividade em Portugal*. Dissertação de Mestrado. Universidade Técnica de Lisboa/Instituto Superior de Economia e Gestão.
- Rosado, F. (1982) – Análise Qualitativa de Densidades de Gram - Charlier e de Edgeworth como Modelos de Alternativas Inerentes para Outliers. Nota n° 27 do Centro de Estatística e Aplicações da Universidade de Lisboa.
- Rosado, F. (1982) – Distribuições Assintóticas sobre Testes de Discordância para Outliers em Populações Exponenciais. Nota n° 31 do Centro de Estatística e Aplicações da Universidade de Lisboa.
- Rosado, F. (1982) – Testes de Discordância para Outliers em Distribuições Exponenciais - Resultados Assintóticos. *Actas ds IX Jornadas Hispano - Lusas de Matemática, vol. II, pp.623-26.*
- Rosado, F. (1984) – The Null Distribution Function of Discordancy Tests for Outlier in Exponential Populations. *METRON, Rivista Internazioanale di Statistica*, vol.XLII, n° 1-2, pp.51-7.
- Rosado, F. (1984) – *Existência e Detecção de Outliers - Uma Abordagem Metodológica*. Tese de Doutoramento. Universidade de Lisboa.
- Rosado, F. (1984) – Outliers a posteriori. *Actas do III Colóquio de Estatística e Investigação Operacional*, pp.273-9.
- Rosado, F. (1986) – Identificação de Outliers. Conferência apresentada no VII Simpósio Nacional de Probabilidades e Estatística na Universidade Estadual de Campinas (Brasil); pré-print na nota n° 31/86 do Centro de Estatística e Aplicações da Universidade de Lisboa.
- Rosado, F. (1987) – Outliers in Exponential Populations. *METRON, Rivista Internazionale di Statistica*, vol.XLV, n.1-2,pp.85-91.
- Rosado, F. (1987) – Algumas Reflexões sobre a Condição Outlier. *Actas das XII Jornadas Luso-Espanholas de Matemática*, volume III, pp. 175-80.
- Rosado, F. (1990) – Outliers, Inliers e Observações Influentes. *Actas das XV Jornadas Luso-Espanholas de Matemática vol. IV p.227-229.*
- Rosado, F. (1996) – Detecção de Outliers e Análise em Componentes Principais. Centro de Estatística e Aplicações da Universidade de Lisboa, nota 2/96.
- Rosado, F. (1996) – Testes de Discordância para Outliers e sua Dependência da Dimensão da Amostra in *A Estatística a Decifrar o Mundo* (R. Vasconcelos et al editores). *Actas do IV Congresso Anual da Sociedade Portuguesa de Estatística p. 173-181.* Edições Salamandra.
- Rosado, F. (1997) – Sobre a influência de observações discordantes na modelação estatística. Nota CEAUL 2/97 do Centro de Estatística e Aplicações da Universidade de Lisboa.
- Rosado, F. (1997) – Sobre a Detecção de Outliers Utilizando Meios Computacionais. in *Estatística: a diversidade na unidade* (M. Souto de Miranda e I. Pereira editores). *Actas do V Congresso Anual da Sociedade Portuguesa de Estatística p. 199-206.* Edições Salamandra

- Rosado, F. (1998) – Efeitos de uma Única Observação na Estimação de Parâmetros em Modelos Lineares. Nota CEAUL 1/98 do Centro de Estatística e Aplicações da Universidade de Lisboa.
- Rosado, F. (1998) – Outlier(s). *Boletim Informativo da Sociedade Portuguesa de Estatística*. Número Especial de Homenagem a Tiago de Oliveira, p. 39-40.
- Rosado, F. (2000) – A Note on Detection of Discordant Observations. Nota CEAUL 18/2000 do Centro de Estatística e Aplicações da Universidade de Lisboa.
- Rosado, F. (2000) – O que é a Estatística. Dossier Especial “Ano Mundial da Matemática”. *Jornal Primeiro de Janeiro* de 2 de Outubro.
- Rosado, F. (2001) – Outliers em Dados Estatísticos – o passado e o presente. E o futuro? *Actas do VII Congresso Anual da Sociedade Portuguesa de Estatística – Um Olhar sobre a Estatística*, p. 90-110.
- Rosado, F. (2001) – Using Maximum Likelihood in the Detection of Outliers. Comunicação apresentada na 53ª Sessão do ISI - International Statistical Institute realizada em Seoul (Coreia).
- Rosado, F. (2005) – On the Statistical Interpretation of Outlier on Forensic Statistics. Comunicação a apresentar na 55ª Sessão do ISI.
- Rosado, F. e Almeida, C. (2001) – Máxima Verosimilhança e a Detecção de Outliers. Nota CEAUL 13/2001.
- Rosado, F. e Alpiarça, I. (1994) – Outliers Múltiplos em Modelos de Discordância para Populações Exponenciais. Nota 7/94 do CEAUL.
- Rosado, F. e Alpiarça, I. (1994) – Sobre Modelos de Discordância para Outliers em populações Exponenciais e Gama. *Actas do II Congresso Anual da Sociedade Portuguesa de Estatística* p. 67-71.
- Rosado, F. e Braumann, M.M. (1990) - Critical Values for a Lower Outlier in a Gamma Sample. *METRON, Rivista Internazionale di Statistica*, vol.XLVIII, n.1-4, pp.19-25.
- Rosado, F. e Mendes, Z. (1993) – Sobre a Detecção de Observações Discordantes em Populações Normais. *Actas do I Congresso Anual da Sociedade Portuguesa de Estatística - A Estatística e o Futuro e o Futuro da Estatística* (p 271-286). Edições Salamandra.
- Rosado, F. e Oliveira, I.(1996) – Selecção Multivariada de Outliers - Uma Aplicação às Variedades de Castanheiro em Trás-os-Montes. in *A Estatística a Decifrar o Mundo* (R. Vasconcelos et al editores). *Actas do IV Congresso Anual da Sociedade Portuguesa de Estatística* p. 163-173. Edições Salamandra
- Rosado, F. e Palma, J. (2001) – Measures of Performance for Discordancy Tests in Normal Populations. *Revista de Estatística. Contributed Papers*, vol.II, 2º Quad. 357-358.
- Rosado, F. e Palma, J. (2001) – Sobre a Qualidade de Testes de Discordância em Populações Normais. Nota CEAUL 12/2001.
- Rosado, F. e Palma, J. (2003) – Problemas e Limitações da Detecção de Outliers. Nota CEAUL 2/2003.
- Rosado, F. e Palma, J. (2005) – Outliers em Dados Circulares. Nota CEAUL 1/2005.

- Rosado, F. e Santos, E. (2000) – Outliers em Regressão Linear Simples – O Efeito de Alargamento. Nota CEAUL 19/2000 do Centro de Estatística e Aplicações da Universidade de Lisboa.
- Rosado, F., Turkman, A., Medeiros, L., Santos, C. e Silva, M. (2000) – Quem somos, no Vigésimo Aniversário. *Boletim Especial Aniversário da Sociedade Portuguesa de Estatística*.

## **Fernando Manuel Fialho Rosado**

Licenciado em Matemática Aplicada, em 1971, pela Faculdade de Ciências da Universidade de Lisboa, é doutorado em Ciências, na especialidade de Probabilidades e Estatística, pela Universidade de Lisboa. É membro do Centro de Estatística e Aplicações desta Universidade e a sua investigação científica desenvolve-se, em especial, na Teoria dos Outliers.

Professor Catedrático da Faculdade de Ciências da Universidade de Lisboa, desde 1993.

Membro eleito do International Statistical Institute, desde 1999.

Presidente da Sociedade Portuguesa de Estatística, em segundo mandato.

# Ciências Actuarias

*Maria de Lourdes Centeno*

## 1. Resenha histórica

Por Ciências Actuarias entende-se um corpo de disciplinas que estudam e modelam os fenómenos aleatórios comuns à generalidade das companhias seguradoras. Os actuários são como que arquitectos de sistemas matemáticos que descrevem os fenómenos típicos da actividade seguradora, de modo a assegurarem uma gestão sem grandes embaraços financeiros, embaraços esses que podem pôr em causa a solvabilidade das companhias, com os inerentes prejuízos para os segurados. As ideias próprias destes sistemas têm origens em muitas disciplinas da área da matemática, da economia e da gestão, mas a pedra basilar é sem dúvida a Teoria da Probabilidade.

Nos países mediterrânicos há evidência da existência de formas de seguros anteriores à era Cristã. É de referir que um dos primeiros trabalhos sobre seguros, senão o primeiro, se deve ao português Pedro de Santarém, que em 1522 publica em Veneza um tratado de seguros. O desenvolvimento das Ciências Actuarias remonta ao fim do século XVII, princípios do século XVIII, muitos séculos após a existência de seguros.

O primeiro ministro holandês John De Witt apresentou em 1671 um relatório sobre rendas vitalícias onde as anuidades eram calculadas como valores esperados, utilizando para o efeito tabelas de juro composto existentes na altura. Edmund Haley embora mais conhecido pelos seus trabalhos na área da astronomia, publica em 1693, no seguimento dos trabalhos de John Graunt, uma das primeiras tabelas de mortalidade, conhecida por tabela de *Breslau* e apresenta, à semelhança de John de Witt, um trabalho sobre rendas vitalícias (veja-se [41]). Abraham de Moivre escreveu um texto sobre o assunto, sintetizando os trabalhos anteriores.

O primeiro plano de pensões com benefícios definidos, parece ser o “*Scotish Ministers Widows Fund*”, iniciado em 1743. Em 1762 surge em Londres a primeira companhia de seguros assente em princípios científicos – a *Equitable*, princípios elaborados por James Dobson, pupilo de de Moivre. Um dos consultores da *Equitable* é Richard Price, que é autor de inúmeros contributos na área dos seguros e autor de [53], manual com imensa repercussão.

O desenvolvimento das Ciências Actuarias no século XIX dá-se simultaneamente ao da Demografia. Desde de Moivre [33] que se sugeria a lei de mortalidade como uma fórmula matemática. Várias fórmulas foram sendo propostas, sendo as mais conhecidas as de Gompertz [43] e de Makeham [47]. Elas mais não são do que uma tentativa de ajustamento de uma curva a observações empíricas e têm estado ligadas às técnicas estatísticas de análise de dados sobre mortalidade.

Outro cientista ligado aos desenvolvimentos das Ciências Actuarias no século XIX, princípio do século XX foi Thiele, com a famosa equação diferencial para cálculo da reserva dos prémios, trabalho nunca publicado, mas tornado conhecido por outros matemáticos.

É também no século XIX que se inicia na Europa o estudo da Teoria do Risco, que veio dar outra dimensão às Ciências Actuarias. O modelo então preconizado é o modelo individual, onde as indemnizações provenientes das diferentes apólices de uma carteira são somadas. É utilizada uma forma do teorema do limite central para justificar a utilização da distribuição normal.

No início do século XX, Philip Lundberg, com o seu modelo de risco no colectivo, dá um incremento significativo no desenvolvimento da Teoria do Risco. O seu primeiro artigo data de 1903. Segundo Cramér (veja-se, por exemplo, o texto de Carl Philipson [51]), a teoria de Philip Lundberg foi pioneira, tendo antecipado ideias utilizadas num contexto mais geral, passados alguns anos, na Teoria dos Processos Estocásticos. Note-se que foi Lundberg que iniciou o estudo do Processo de Poisson Misto. O desenvolvimento do modelo clássico da teoria da ruína, apresentado por Lundberg, ficou a dever-se a Cramér, Ammeter, Ove Lundberg – filho de Philip Lundberg – e muitos outros.

A teoria da decisão é incorporada no modelo por Karl Borch, que num conjunto de artigos (veja-se [8]), muitos deles apresentados nos Congressos Internacionais de Actuários, divulga junto dos actuários, as ideias apresentadas dois séculos antes por Daniel Bernoulli. De facto, o conceito de utilidade marginal decrescente, que estabelece que o decisor está disposto a pagar mais do que o valor esperado de uma determinada perda, para se livrar do risco associado, tinha sido apresentado em 1738 por Daniel Bernoulli [7]. É sobre esta ideia que floresce a indústria seguradora. Surpreendentemente, é apenas na segunda metade do século XX que ela é interiorizada pelos actuários.

Os actuários não ficaram imunes às ideias defendidas pelos Bayesianos. Embora havendo referências anteriores, nomeadamente de Ove Lundberg em 1940 (veja-se [46]), o desenvolvimento da teoria da credibilidade utilizando técnicas Bayesianas, deve-se a Bailey [3], [4] e [5]. Refira-se que o pioneiro de teoria da credibilidade é Whitney [56] que, em 1918, propõe o cálculo do prémio como uma média ponderada de duas médias, uma calculada com base na experiência individual e uma outra baseada no colectivo de apólices semelhantes. Os trabalhos de Bailey não colheram a simpatia dos actuários da altura. É Bühlmann, que nos anos sessenta, com dois artigos, [10-11], publicados no então recém criado *ASTIN Bulletin*, convence os actuários sobre a utilidade daquelas ideias. Seguem-se vários artigos sobre o assunto, nomeadamente o modelo de credibilidade multidimensional de Bill Jewell.

A utilização das cadeias de Markov na modelação de sistemas de *bonus malus* iniciou-se nos anos sessenta. O primeiro *ASTIN Colloquium*, realizado em França em 1959, é inteiramente dedicados a estes sistemas.

Muitos outros tópicos da Estatística são usados nas Ciências Actuarias. Refiram-se, a título de exemplo, os modelos lineares generalizados na tarificação de vários ramos e os filtros de Kalman na provisão de sinistros.

Para uma análise mais detalhada sobre a história das Ciências Actuarias veja-se o artigo de James Hickman [44].

## 2. Os últimos vinte e cinco anos

O desenvolvimento recente das Ciências Actuarias está intimamente ligado ao desenvolvimento dos sistemas informáticos.

Muitos dos modelos da teoria do risco, anteriormente desenvolvidos, são de difícil solução. A simulação estocástica veio introduzir um meio excelente de resolução de problemas actuariais. É possível construir hoje um modelo para toda a companhia, que tenha em consideração os vários factores que afectam cada ramo de negócios: o processo relativo ao número de sinistros, a função de distribuição das indemnizações particulares, o tipo e nível de resseguro, prémios, rendimento financeiro, inflação, distribuição de dividendos, etc. Estes modelos que simulam a actividade seguradora são algumas vezes designados por Dynamic Financial Analysis (DFA).

Os pioneiros no desenvolvimento destes métodos foram os Finlandeses. O modelo por eles desenvolvido e cujos autores são designados por “The Finnish Working Party” está apresentado em [50]. A estes seguiram-se os britânicos com o “The British Solvency Working Party”, cujo modelo está, por exemplo, explicado em [32].

O desenvolvimento das fórmulas recursivas para o cálculo da função de distribuição das indemnizações agregadas - distribuição composta - foi sem dúvida uma das inovações das últimas décadas. Estes métodos foram introduzidos nas Ciências Actuarias, em 1981, por Harry Panjer [48] que resolve o problema quando a distribuição do número de sinistros pertence à família Katz e o montante dos sinistros individuais tem distribuição aritmética. Desenvolvimentos às fórmulas originais devem-se a autores como Sundt e Wilmot.

As Ciências Actuarias não ficaram imunes à revolução ocorrida, na última metade do século XX, na área da Economia Financeira – veja-se o número de Prémios Nobel atribuídos a economistas nesta área. O actuariado financeiro sofreu um tal crescimento que a “International Actuarial Association” considerou como necessária a criação em 1986 da secção *AFIR* (Actuarial Approach for Financial Risks), autonomizando-a da secção *ASTIN* (Actuarial Studies In Non-life insurance) e passando a organizar anualmente, à semelhança do *ASTIN*, um colóquio internacional. Para além destes, os Congressos Internacionais associados à revista *Insurance: Mathematics and Economics*, têm contado com um número significativo de artigos na área financeira, focando assuntos como os da imunização, “capital asset pricing model” ou do preço de opções. Para um resumo sobre o actuariado financeiro veja-se Panjer [49].

### 3. O ensino do actuariado em Portugal

Com o forte crescimento da indústria seguradora no século XIX surgem as organizações profissionais de actuários. O “Institute of Actuaries” foi criado em Londres em 1848, seguido da “Faculty of Actuaries” em Edimburgo em 1856. Em 1897 foi criada “The American Society of Actuaries” e, em 1913, “The American Institute of Actuaries” que se fundiram, em 1949, originando a “Society of Actuaries”. Em 1914, é criada a “Casualty Actuarial and Statistical Society”, tendo visto o seu nome alterado para “Casualty Actuarial Society” em 1921. Nestes institutos a admissão é, desde o seu início, feita por exames.

Na Europa, a seguir aos ingleses, seguiram-se, ainda no fim do século XIX, os franceses, os belgas e os italianos. No início do século XX surge o suíço e o alemão. O Instituto dos Actuários Portugueses foi criado apenas em 1945, dois anos após o espanhol.

O modelo britânico é seguido nos países anglo-saxónicos, sendo o caminho seguido na Europa continental diferente. Aqui, a prática actuarial tende a ser mais regulada pelas autoridades de controle, e influenciada pelas universidades, do que pelas associações profissionais.

Assim, o ensino do actuariado em Portugal, tem sido feito essencialmente pelas Universidades com alguns impulsos recentes do Instituto de Seguros, do Instituto dos Actuários e da Associação Portuguesa de Seguradores.

No XIV Congresso Internacional de Actuários, que teve lugar em Madrid, em 1954, Caetano Beirão da Veiga e Bastos Martins resumem o estado do ensino actuarial existente à época em Portugal (veja-se [6]). Segundo esse documento, a maioria dos actuários era detentor de uma licenciatura em Ciências Matemáticas da Faculdade de Ciências da Universidade de Lisboa ou de uma licenciatura em Ciências Económicas e Financeiras do Instituto Superior de Ciências Económicas e Financeiras (actual Instituto Superior de Economia e Gestão) da Universidade Técnica de Lisboa. Actualmente, é ainda o ISEG o principal responsável pela formação actuarial em Portugal, tendo cabido à Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa e à Universidade Católica Portuguesa um papel importante, como se verá nas secções seguintes.

#### 3.1 O ensino do actuariado no ISEG/UTL

Teria sido no Instituto Superior do Comércio - que originou o Instituto Superior de Ciências Económicas e Financeira (ISCEF), aquando da criação em 1930 da Universidade Técnica de Lisboa, em 1972 denominado Instituto Superior de Economia e desde 1989 designado por Instituto Superior de Economia e Gestão - que se teriam dado os primeiros passos no ensino do actuariado. De facto, e segundo o decreto nº1909 de 23 de Setembro de 1915, a 21ª cadeira do Curso Superior do Comércio, curso iniciado em 1913, era a cadeira de “Seguros. Instituições de previdência. Contabilidade de Seguros.”, de cujo programa, constavam matérias próprias do cálculo actuarial como: Tábua de mortalidade e

seu ajustamento; Equações de mortalidade; Tábua de comutação; Seguros em caso de vida; Seguros em caso de morte; Contrasseguro; Combinações especiais; Seguro Complementar; Tarifas; Reservas; Resgates.

A disciplina de Cálculo Actuarial surge em 1931, na Secção de Finanças do Curso Superior de Ciências Económicas e Financeiras, como disciplina anual - 4ª cadeira, a ser leccionada no 4º ano do curso, após o ensino de duas disciplinas anuais de Matemáticas Superiores, que incluíam o Cálculo das Probabilidades e suas Aplicações e uma outra de Estatística Geral e Aplicada.

O grande dinamizador do ensino do Cálculo Actuarial terá sido Caetano Beirão da Veiga, que para além de ter leccionado e regido a 21ª cadeira do Curso Superior do Comércio, rege a 4ª cadeira Curso Superior de Ciências Económicas e Financeiras. Da sua Biografia, veja-se [55], constam as “Lições de Cálculo Actuarial – Lições no ISCEF).

Em 1949, houve uma profunda alteração do plano de estudos, tendo os Cursos Aduaneiro, Consular, de Comércio e de Finanças sido substituídos pelos cursos de licenciatura em Economia e em Finanças. Nesta altura passam a existir duas disciplinas anuais no curso de Finanças – Cálculo Actuarial I e Cálculo Actuarial II - sendo as cadeiras precedentes: Matemáticas Gerais, Análise Matemática e Estatística.

Outros professores do ISCEF, como Leite Pinto, Bento Murteira, Rómulo Rodrigues e Luís Santos Fernandes, leccionaram as disciplinas de Cálculo Actuarial, podendo no entanto atribuir-se a Santos Fernandes, a responsabilidade pelo Actuariado a partir do início dos anos cinquenta até início dos anos oitenta, publicando vários artigos nos Anais do ISCEF e as “Lições de Cálculo Actuarial”.

Assistiu-se na reforma de 1972, altura em que o Instituto se passou a designar Instituto Superior de Economia e em que o curso de Organização e Gestão de Empresas substituiu o curso de Finanças, a um declínio do ensino do Cálculo Actuarial, declínio esse que se acentua com as alterações ao plano de estudos de 1976 e de 1986.

Em 1990 o ISEG, lança o Curso de Pós-Graduação em Actuariado e Gestão de Riscos Financeiros, que em 1993/94 é promovido a Curso de Mestrado com a mesma designação. A necessidade, por parte do mercado, de formação de actuários não-vida, foi incrementada com a publicação de legislação que obriga as Seguradoras dos Ramos Reais a possuir desde 30/6/98, à semelhança das Seguradoras do Ramo Vida, um actuário responsável (aprovado pelo Instituto de Seguros de Portugal). Foi para corresponder a esta necessidade das empresas seguradoras, que o ISEG reformulou em 1997 o referido mestrado, passando a ser designado por Mestrado em Ciências Actuarias.

O mestrado do ISEG foi e ainda é o único curso de mestrado do país nestas matérias. Tem periodicidade bienal e o elenco das disciplinas e respectivos docentes na última edição foi: Cálculo Actuarial (Onofre Simões), Complementos de Estatística (João Andrade e Silva), Gestão de Empresas Seguradoras (Carlos Pereira da Silva), Provisões para Sinistros (Walther Neuhaus), Técnicas de Construção de Tarifas (João Andrade e Silva), Teoria dos Mercados de Capitais (João Duque), Teoria do Risco (Maria de Lourdes Centeno), Complementos Actuarias (Gregório Luís), Complementos de Teoria do Risco

(Alfredo Egídio dos Reis), Fundos de Pensões (Jorge Garcia), Gestão de Activos-Passivos (João Duque) e Sistemas de *Bonus Malus* (Maria de Lourdes Centeno). Existem já 49 mestres em Actuariado formados pelo ISEG.

A licenciatura em Matemática Aplicada à Economia e à Gestão, criada em 1990, tem desde o seu início algumas disciplinas da área do actuarido, existindo mesmo uma área de concentração em Matemática Financeira e dos Seguros, desde a reformulação do seu plano de estudo, ocorrida em 2002.

### **3.2 O ensino do actuariado na FC/UL**

Em 1968/69 e por três anos consecutivos, Jorge Garcia lecciona na Faculdade de Ciências da Universidade de Lisboa, a disciplina de Cálculo Actuarial, que segundo o próprio, já em tempos idos havia sido leccionada naquela Faculdade. Entre 1974/75 e 1979/80 o ensino da disciplina é da responsabilidade de José Tiago de Oliveira, tendo tido como assistente Maria Manuela Corte Real. Em 1998/99 é retomado o ensino da disciplina nas licenciaturas de Probabilidades e Estatística e de Estatística e Investigação Operacional, que tem estado a cargo de Maria Fernanda Fernandes de Oliveira.

A disciplina de Teoria do Risco funciona para os mesmos cursos desde 1996/97 e tem estado a cargo de Maria Isabel Fraga Alves.

### **3.3 O ensino do actuariado na FCH/UCP**

Em 1985 a Faculdade de Ciências Humanas da Universidade Católica Portuguesa lança a primeira experiência de pós-graduação em Ciências Actuarias em Portugal. O Curso de Pós-Graduação em Actuariado, teve o patrocínio do Instituto de Seguros de Portugal, e formou vários dos actuais actuários, com destaque para os do ramo vida. A última realização foi a de 1989.

### **3.4 O ensino do actuariado na FCT/UNL**

Em 1988 a Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa cria, na licenciatura em Matemática, o ramo de Matemática Aplicada em Ciências Actuarias, tendo sido César de Freitas o seu impulsionador. Os professores das disciplinas de actuariado foram Jorge Garcia, José Tiago de Oliveira e João Tiago Mexia. Entre 1989/90 e 2003/2004 licenciaram-se neste ramo 242 alunos. É de referir que muitos dos alunos dos mestrados do ISEG eram oriundos desta licenciatura.

Em 2002 assiste-se a uma reformulação dos planos de estudo, que acaba com o ramo em Ciências Actuarias, continuando algumas das disciplinas a serem leccionadas no ramo de Matemática Aplicada que integra também disciplinas de Estatística e Investigação Operacional.

### 3.5 O ensino do actuariado noutras instituições

Para além das quatro escolas já citadas, onde o ensino do actuariado tem ou teve um papel relevante, é de referir que a licenciatura em Matemática da Universidade de Trás-os-Montes e Alto Douro tem um ramo de Matemáticas Financeiras. O curso de Matemática da Universidade do Minho tem uma disciplina de Cálculo Actuarial na área de especialização em Aplicações à Economia. O mestrado em Matemática Aplicada da Universidade de Coimbra conta também, a partir do corrente ano, com uma disciplina de Teoria do Risco.

## 4. Investigação em Ciências Actuariais em Portugal

Sendo o ISEG a instituição universitária com um maior historial no ensino do actuariado, não é de estranhar que seja também esta a escola com uma maior produção científica na área. O Departamento de Matemática e o CEMAPRE – Centro de Matemática Aplicada à Previsão e Decisão Económica – contam com o maior grupo de professores/investigadores do país em Ciências Actuariais. O Departamento tem cinco professores com teses de doutoramento em Ciências Actuariais, todos membros do CEMAPRE, centro que conta ainda com Isabel Cordeiro da Universidade do Minho. Existem ainda dois professores no departamento de gestão doutorados na área dos seguros.

Refira-se, no entanto, que os primeiros artigos publicados em revistas científicas internacionais emanaram da Faculdade de Ciências da Universidade de Lisboa, por Maria Amélia Cabral e Jorge Garcia, [12-13] que, em meados dos anos setenta, publicam no *ASTIN Bulletin* dois artigos sobre teoria da credibilidade. Na mesma altura Tiago de Oliveira, [54], publica na mesma revista um artigo sobre grandes indemnizações. Sobre a mesma problemática, mas já em 1986, Ivette Gomes e Dinis Pestana publicam [42].

A investigação no CEMAPRE tem sido diversificada. Os trabalhos iniciais [16-18] foram sobre resseguro, área onde me doutorei, a que se seguiram [19], [21-24] e [27-28]. João Andrade e Silva doutorou-se com uma tese sobre tarifação no ramo automóvel, quer *a priori* quer *a posteriori*, área que continuámos a investigar e onde publicámos [25-26] e [2]. O princípio de cálculo de prémio do ajustamento do risco, foi também motivo do artigo [1]. Alfredo Egídio dos Reis doutorou-se com uma tese sobre teoria da ruína, área que continuou a investigar, tendo publicado, parte em co-autoria com o seu orientador de doutoramento, [34-40], [14] e [45]. A teoria da ruína foi também o tema escolhido por Jorge Garcia para a sua recente tese de doutoramento. Isabel Cordeiro concluiu a sua tese de doutoramento sobre seguros de saúde de longo prazo e publicou [29-30]. Onofre Simões fez a sua tese de doutoramento sobre seguros de vida. Note-se que alguns dos trabalhos referidos, nomeadamente, [21], [14], [45] e [52] foram fruto de teses de mestrado. É de referir que metade dos membros deste grupo fizeram o seu doutoramento na Heriot-Watt University em Edimburgo, escola com grande tradição na área.

Rui Cardoso, docente da FCT da UNL, na sequência da sua tese de mestrado no ISEG, doutorou-se, também na Heriot-Watt, em teoria da ruína e publicou com o seu orientador, [15].

Ainda no ISEG, nomeadamente no Centro de Investigação sobre Economia Financeira, tem-se privilegiado a investigação sobre a sustentabilidade do sistema de segurança social português, tendo publicado, entre outros, [31].

Sendo difícil de delimitar o campo das Ciências Actuariais, pois existem áreas de fronteira, como a da Teoria de Valores Extremos, onde as aplicações aos seguros são óbvios, optou-se por apenas se referir os artigos publicados nas revistas internacionais da área dos seguros. Assim na área da Teoria dos Valores Extremos, e para além dos artigos [54] e [42], atrás citados, apenas se refere [9], trabalho realizado por Margarida Brito e Ana Cristina Freitas, mas poder-se-ia citar, sobre o mesmo assunto, a obra extensa da linha de investigação do Centro de Estatística e Aplicações da Faculdade de Ciências da Universidade de Lisboa, liderada por Ivette Gomes.

Tiveram lugar em Portugal dois congressos internacionais na área do actuariado. O primeiro, o *XII ASTIN Colloquium*, ocorreu em pleno verão quente de 1975, em Portimão, organizado pelo Instituto dos Actuários Portugueses. O segunda, o “VI International Congress on Insurance: Mathematics and Economics” foi organizado pelo CEMAPRE e teve lugar no ISEG em 2002.

## Referências Bibliográficas

- [1] Andrade e Silva, J. M and Centeno, M. L. (1998). Comparing risk adjusted premiums from the reinsurance point of view, *ASTIN Bulletin* **28**, 221-239.
- [2] Andrade e Silva, J. M. and Centeno, M. L., A note on bonus scales, *The Journal of Risk and Insurance*, aceite para publicação.
- [3] Bailey, A.L. (1942). Sampling theory in casualty insurance, *Proceedings of the Casualty Actuarial Society* **29**, 50-93.
- [4] Bailey, A.L. (1945). A generalized theory of credibility, *Proceedings of the Casualty Actuarial Society* **32**, 13-20.
- [5] Bailey, A.L. (1945). Credibility procedures: Laplace’s generalization of Bayes’ rule and the combination of collateral knowledge with observation data, *Proceedings of the Casualty Actuarial Society* **37**, 7-23.
- [6] Beirão da Veiga, C. e Bastos Martins E. (1954). O ensino actuarial em Portugal, *Transactions of the XIV International Congress of Actuaries*.
- [7] Bernoulli, D. (1738). Exposition of a new theory on the measurement of risk, *Econometrica* **22**, 23-26, 1954 (translation of a paper in Latin published in St. Petersburg in 1738).
- [8] Borch, K. (1974). *The mathematical theory of insurance: an annotated selection of papers on insurance published 1960-1972*. Lexington Books.
- [9] Brito, M. and Freitas, A. (2003). Limiting behaviour of a geometric-type estimator for tail indices. *Insurance: Mathematics and Economics* **33** (2), 211-226.

- [10] Bühlmann, H. (1967). Experience rating and credibility, *ASTIN Bulletin* **4**, 199-207.
- [11] Bühlmann, H. (1969). Experience rating and credibility, *ASTIN Bulletin* **5**, 157-165.
- [12] Cabral, M.A. and Garcia, J. A. (1974). Calculation of provisions using credibility theory, *ASTIN Bulletin* **7(3)**, 277-292.
- [13] Cabral, M.A. and Garcia, J. A. (1977). Study of factors influencing the risk and their relation to Credibility Theory, *ASTIN Bulletin* **9(C)**, 84-104.
- [14] Cardoso, R. and Egídio dos Reis, A. (2002). Recursive calculation of time to ruin distributions. *Insurance: Mathematics and Economics* **33**, 659-676.
- [15] Cardoso, R. and Waters, H. (2003). Recursive calculation of finite time ruin probabilities under interest force. *Insurance: Mathematics and Economics* **30**, 219-230.
- [16] Centeno, L. (1985). On combining quota-share and excess of loss, *ASTIN Bulletin* **15**, 49-63.
- [17] Centeno, L. (1986). Some mathematical aspects of combining proportional and non-proportional reinsurance, in *Insurance and Risk Theory*, 247-266, edited by M. Goovaerts, F. de Vylder e J. Haezendonck. D. Reidel Publishing Company, Holland.
- [18] Centeno, L. (1986). Measuring the effects of reinsurance by the adjustment coefficient, *Insurance: Mathematics and Economics* **5**, 169-182.
- [19] Centeno, M. L. (1988). The expected utility applied to reinsurance, in *Risk, Decision and Rationality*, 679-689, edited by Bertrand R. Munier. D. Reidel Publishing Company, Holland.
- [20] Centeno, L. (1989) The Buhlmann-Straub model with the premium calculated according to the variance principle, *Insurance: Mathematics and Economics* **8**, 3-10.
- [21] Centeno, L. and Simões, O. (1991) Combining quota-share and excess of loss treaties on the reinsurance of  $n$  independent risks, *ASTIN Bulletin*, **21**, 41-55.
- [22] Centeno, L. (1991) An insight into the excess of loss retention limit, *Scandinavian Actuarial Journal*, 97-102.
- [23] Centeno, M. L. (1995) The effect of the retention limit on the risk reserve, *ASTIN Bulletin* **25**, 67-74.
- [24] Centeno, M. L. (1997) Excess of loss reinsurance and the probability of ruin in finite horizon, *ASTIN Bulletin* **27**, 59-70.
- [25] Centeno M. L. and Andrade e Silva J. M. (2001). Bonus systems in an open portfolio, *Insurance: Mathematics and Economics*, **28(3)**, 341-350.
- [26] Centeno M. L. and Andrade e Silva J. M., (2002). Optimal bonus scales under path-dependent bonus rules, *Scandinavian Actuarial Journal*, 129-136.
- [27] Centeno, M. L. (2002). Measuring the effect of reinsurance by the adjustment coefficient in the Sparre Andersen model, *Insurance: Mathematics and Economics* **30** (1), 37-50.
- [28] Centeno, M. L. (2002). Excess of loss reinsurance and Gerber's inequality in the Sparre Andersen model, *Insurance: Mathematics and Economics* **31**, 415-427.

- [29] Cordeiro, I.M.F. (2002). A multiple state model for the analysis of permanent health insurance claims by cause of disability, *Insurance: Mathematics and Economic* **30**, 167-186.
- [30] Cordeiro, I.M.F. (2002). Transition intensities for a model for permanent health insurance. *ASTIN Bulletin*, **32** (2), 319-346.
- [31] da Silva C.M.P., Calado J.P.T. and Garcia M.T.M. (2004). The financial sustainability of the Portuguese social security system, *Geneva Papers on Risk and Insurance - Issues and Practice* **29** (3), 417-439.
- [32] Daykin, C., Bernstein, G., Coutts, S., Devitt, E., Hey, G., Reynolds, D. and P. D. Smith, (1989) "The solvency of a general insurance company in terms of emerging Costs", in *Financial Models of Insurance Solvency*, ed. by J. David Cummins and Richard Derrig. Boston: Kluwer Academic Publishers.
- [33] de Moivre, A. (1925). In *Annuities Upon Lives*, W. Pearson, London.
- [34] Dickson, D. and Egídio dos Reis, A. (1994). Ruin problems and dual events. *Insurance: Mathematics and Economics* **14**, 51-60.
- [35] Dickson, D., Egídio dos Reis, A. and Waters, H. (1995). Some stable algorithms in ruin theory and their applications. *ASTIN Bulletin* **25**, 153-175.
- [36] Dickson, D. and Egídio dos Reis, A. (1996). On the distribution of the duration of negative surplus. *Scandinavian Actuarial Journal*, n° 2, 148-164.
- [37] Dickson, D. and Egídio dos Reis, A. (1997). The effect of interest on negative surplus. *Insurance: Mathematics and Economics* **21**, 1-16.
- [38] Egídio dos Reis, A. (1993). How long is the surplus below zero? *Insurance: Mathematics and Economics* **12**, 23-38.
- [39] Egídio dos Reis, A. (2000). On the moments of ruin and recovery times. *Insurance: Mathematics and Economics* **27**, 331-343.
- [40] Egídio dos Reis, A. (2002). How many claims does it take to get ruined and recovered? *Insurance: Mathematics and Economics* **31** (2), 235-248.
- [41] Halley, E. (1693). An estimate of the mortality of mankind drawn from curious tables of the births and funerals at the city of Breslau; with an attempt to ascertain the price of annuities upon lives, *Philosophic Transactions of the Royal Society of London* **17**, 596-610.
- [42] Gomes, M. I. and Pestana D. (1986) Large claims – extreme value models, in *Insurance and Risk Theory*, 301-323, edited by M. Goovaerts, F. de Vylder e J. Haezendonck. D. Reidel Publishing Company, Holland.
- [43] Gompertz, B. (1825) On the nature of the function expressive of the law of human mortality, *Philosophic Transactions of the Royal Statistical Society of London* **115**, 513-583.
- [44] Hickman, J. (2004) History of actuarial science, *Encyclopedia of Actuarial Science*, vol. **2**, 838-848.
- [45] Lima, F.D.P., Garcia, J.A. and Egídio dos Reis, A.D. (2002). Fourier/Laplace transforms and ruin probabilities. *ASTIN Bulletin* **32** (1), 91-104.

- [46] Lundberg, O. (1940) *On Random processes and their applications to sickness and accident statistics*, Inaugural dissertation, Uppsalla; 2<sup>nd</sup> Edition, Almqvist & Wiksells, Upsala, 1964.
- [47] Makeham, W. M. (1867) On the law of mortality. *Journal of the Institute of Actuaries* **8**, 301-310; *Journal of the Institute of Actuaries* **13**, 325-358.
- [48] Panjer, H. (1981) Recursive evaluation of a family of compound distributions, *ASTIN Bulletin* **12**, 22-26.
- [49] Panjer, H. H., ed., (1998) *Financial economics with applications to investments, insurance and pensions*. The Actuarial Foundation, Schaumburg.
- [50] Pentikäinen, T. and Rantalla, J. (1982) *Solvency of insurers*. Two Volumes. Helsinki: Insurance Publishing Co.
- [51] Philipson, C. (1968) A Review of the collective risk theory, *Skandinavisk Aktuarietidskrift*.
- [52] Pinheiro, P., Andrade e Silva, J. M. and Centeno, M. L. (2003) Bootstrap methodology in claim reserving, *The Journal of Risk and Insurance* **70** (4), 701-714.
- [53] Price, R. (1771) *Observations on reversionary payments: on schemes for providing annuities for persons of old age; on the method of calculating the values of assurances; and on the national debt*, Cadell & Davies, London.
- [54] Tiago de Oliveira, J. (1977) Statistical methodology for large claims. *ASTIN Bulletin*, **9**, 1-9.
- [55] Universidade Técnica de Lisboa. (1956) *A Universidade Técnica de Lisboa e os seus mestres. Notas biobibliográficas*.
- [56] Whitney, A. W. (1918) The theory of experience rating, *Proceedings of the Casualty Actuarial Society*, **4**, 274-292.

### **Maria de Lourdes Centeno**

É professora catedrática do departamento de Matemática do Instituto Superior de Economia e Gestão (ISEG) da Universidade Técnica de Lisboa e investigadora do Centro de Matemática Aplicada à Previsão e Decisão Económica (CEMAPRE).

Licenciou-se em 1975 em Matemática Aplicada pela Faculdade de Ciências da Universidade de Lisboa e doutorou-se em 1985 na Heriot-Watt University.

É editora associada da revista "Insurance: Mathematics and Economics", membro do corpo editorial do "ASTIN Bulletin" e editora da secção de Resseguro da "Encyclopedia of Actuarial Science".

Publicou vários artigos de investigação em diferentes tópicos das Ciências Actuariais, nomeadamente em resseguro, teoria da credibilidade, sistemas de bonus e provisões para sinistros.

# A Evolução do Sistema Estatístico Nacional após a Reestruturação de 1989 Pontos Fortes e Pontos Fracos

*Paulo Gomes*

## **1. Introdução**

O funcionamento das modernas sociedades democráticas determinou uma crescente importância da informação estatística oficial, decorrente, não já exclusivamente do seu carácter de instrumento essencial à tomada de decisões de política geral ou sectorial, mas também da sua relevância como elemento essencial da avaliação pelos cidadãos do desempenho dos governos aos diferentes níveis.

As questões relacionadas com a transparência na produção e disponibilização das estatísticas oficiais, directamente ligadas à confiança nas mesmas, emergiram igualmente como elementos merecedores de análise e debate, visando assegurar a consolidação de uma crescente percepção pública da independência das estatísticas oficiais.

A garantia dos atributos da qualidade das estatísticas oficiais, isto é, da pertinência, precisão, actualidade, pontualidade, acessibilidade, clareza, comparabilidade e abrangência, passou a surgir intimamente ligada à transparência da respectiva produção e disponibilização.

Em Portugal, como na totalidade dos países da União Europeia, a responsabilidade pela produção das estatísticas oficiais é partilhada entre o INE e o Banco de Portugal – no contexto europeu entre o Eurostat e o Banco Central Europeu – entidades às quais a lei outorga um estatuto de independência e de autonomia técnica indispensáveis à disponibilização de estatísticas de qualidade.

Os dados estatísticos oficiais têm também visto a sua importância reforçada, ao longo dos últimos anos, no contexto da União Europeia, sendo, hoje, clara a existência de um Sistema Estatístico Europeu cujas linhas de desenvolvimento condicionam fortemente as opções ao nível de cada Sistema Estatístico Nacional.

A informação estatística exige, pela sua própria natureza de sistema de informação baseado num conjunto de normas específicas, definidas, em regra, de acordo com padrões

acordados ao nível das principais organizações internacionais, que a sua análise se processe com base no conhecimento daquelas regras e logo das potencialidades e limites das conclusões que dela é legítimo retirar.

As circunstâncias internas quanto à avaliação das estatísticas oficiais produzidas, aliadas à cada vez maior exigência com que, no quadro da União Europeia, as entidades responsáveis pela produção daquelas estatísticas se confrontam, colocam a estas e em particular aos institutos nacionais de estatística e aos bancos centrais desafios complexos. Estes desafios obrigam estas instituições a levar a cabo, periodicamente, uma avaliação do estado dos sistemas estatísticos nacionais, revisitando o caminho percorrido e identificando os pontos fortes e os pontos fracos no contexto das actuais disposições e obrigações comunitárias, sem prejuízo da consideração das especificidades nacionais, nomeadamente o estado de desenvolvimento do país e das suas instituições.

Propomo-nos neste artigo visitar sumariamente os traços mais significativos da evolução das estatísticas oficiais portuguesas no período 1989-2003, na sequência do processo de reestruturação do Sistema Estatístico Nacional iniciado em 1989, e que teve como suportes jurídicos, a lei de bases do Sistema Estatístico Nacional e o decreto lei nº 280/89 que conferiu novos estatutos ao INE.

## **2. O Sistema Estatístico Nacional e o meio envolvente**

O Sistema Estatístico Nacional tem vindo a ser desafiado pelas transformações da economia e da sociedade em geral, nomeadamente pelos fenómenos da globalização e do desenvolvimento tecnológico. Estes dois fenómenos relacionam-se com outros factores relevantes, como a emergência de novos produtos e serviços, mudanças qualitativas nos produtos e serviços tradicionais, mudanças estruturais nas empresas, nos mercados de trabalho e financeiros e a crescente importância da economia do conhecimento.

Tais desafios criaram novas necessidades em várias áreas das estatística económicas e sociais, com repercussões, por exemplo, no cálculo do índice de preços, na existência de lacunas de informação do sector dos serviços, na necessidade de novas metodologias capazes de avaliar o papel do capital social e humano no crescimento económico e na revisão das classificações e definições usadas na produção estatística.

A produção de informação estatística sobre a organização e actividades empresariais apresenta novas exigências: os contratos de emprego atípicos generalizaram-se, os sistemas clássicos de retribuição alteraram-se, as práticas contabilísticas tradicionais são insuficientes, num contexto em que a produção imaterial ganha maior relevo e os movimentos verificados nos mercados de capital implicam inovadoras abordagens metodológicas.

Nos últimos anos do período em análise o desenvolvimento das estatísticas económicas concentrou-se sobretudo na resposta às necessidades das políticas de União Europeia, nomeadamente no que respeita ao plano de acção da União Económica e Monetária e à política monetária comum do Banco Central Europeu. Também na área da política fiscal, como consequência do Pacto de Estabilidade e Crescimento e do processo

dos défices excessivos, se tem realçado a importância da monitorização estatística do desenvolvimento económico; a política de emprego da União Europeia e a discussão sobre a coordenação da política económica criaram igualmente novas necessidades de informação estatística. O interesse político centrou-se nas questões estruturais de longo prazo, na qualidade das políticas públicas e no desenvolvimento sustentável.

Neste contexto, os desafios resultantes das novas necessidades dos agentes públicos ou privados promotores do desenvolvimento têm implicado que a produção estatística assuma uma flexibilidade capaz de responder rápida e eficazmente a tais mudanças.

### **3. A evolução do Sistema Estatístico Nacional ao longo do período 1989-2003**

A análise que sucintamente apresentaremos neste texto procurará reflectir com a maior isenção e imparcialidade aquilo que foram os pontos fortes e os pontos fracos do processo de evolução do Sistema Estatístico Nacional, no período em apreço, face não só aos constrangimentos diagnosticados no período de preparação da reestruturação de 1989, mas também considerando aquilo que julgamos que deveria ter sido o percurso de um Instituto Nacional de Estatística inserido numa sociedade de informação e do conhecimento e de economia global com a consequente transposição de novos e complexos desafios para a esfera de intervenção do Sistema Estatístico Nacional.

Nesta missão de balanço do caminho percorrido não teremos naturalmente a pretensão de referir exhaustivamente os avanços conseguidos e os relativos insucessos face às expectativas dos agentes promotores do desenvolvimento ou simplesmente face às expectativas criadas nos planos anuais e plurianuais de actividade do INE e dos seus organismos delegados. Procuraremos tão somente trazer à colação sinais reveladores do significativo progresso alcançado nesse período, da estagnação não superada nalguns domínios emergentes e mesmo da impossibilidade de se terem eliminado certos constrangimentos à adequada produção e difusão de estatísticas oficiais portuguesas.

#### ***3.1. A coordenação do Sistema Estatístico Nacional***

Nos termos da Lei de Bases do Sistema Estatístico Nacional (SEN), o exercício das funções de notação, apuramento, coordenação e difusão de estatísticas oficiais cabe exclusivamente ao INE. Contudo este princípio definidor de um sistema estatístico centralizado é, de alguma forma, mitigado pela possibilidade de serem delegadas competências exclusivas do INE noutros serviços públicos, mediante parecer favorável do Conselho Superior de Estatística.

Note-se que a lei do SEN não prevê que a difusão de estatísticas oficiais seja delegada, acautelando o princípio da centralização e intencionalmente salvaguardando o efectivo controlo do INE sobre as estatísticas oficiais divulgadas. Não obstante, é hoje inquestionável o desajustamento existente entre os estatutos dos organismos simples da administração pública detidos pelos serviços nos quais o INE tem delegado competências

e a verificação dos princípios de independência e autonomia técnica, consagrados na lei do Sistema Estatístico Nacional.

O balanço do processo de delegação de competências, decorrente da reestruturação de 89, evidenciou assinaláveis fragilidades no plano prático da sua concretização, da sua eficiência e da sua eficácia, designadamente no que se refere ao acompanhamento e colaboração técnica do INE junto dos organismos delegados na prossecução dos objectivos definidos nos planos anuais de actividade e na sua própria definição prévia, assistindo-se até 2001 a uma contínua degradação do cumprimento do calendário de execução de projectos estatísticos inscritos em várias áreas delegadas, muito especialmente na área das estatísticas da educação e nas áreas das estatísticas do trabalho e da formação profissional.

Esta situação responsabilizou essencialmente o INE e decorreu da impossibilidade do Instituto ter alocado verbas do seu orçamento para o exercício do acompanhamento permanente da actividade delegada com a insuficiente afectação de recursos técnicos adequados a tal nível de intervenção. A título de exemplo, os indicadores globais relativos às estatísticas de educação, apresentavam em 2001 atrasos superiores a 4 anos, bem explicitados nos sucessivos relatórios anuais da actividade estatística discutidos e apreciados no âmbito do Conselho Superior de Estatística, mas sem nunca terem originado, até 2001, o equacionamento previsto na lei, da cessação da delegação de competências nos serviços de estatística do Ministério da Educação ou o estabelecimento de um rigoroso plano de acção com vista à rápida superação de tão graves insuficiências.

Estas constatações não podem, contudo, pôr em causa os amplos avanços que pautaram o funcionamento do Conselho Superior de Estatística em áreas muito diversas da intervenção estatística bem consubstanciados na dinâmica de inúmeras secções especializadas do conselho e grupos de trabalho criados ao longo do tempo.

Com base nesta apreciação, incluída no *Relatório de Avaliação do Estado do Sistema Estatístico Nacional 1999 – 2001*, o Conselho Superior de Estatística deliberou no final de 2001 que a execução dos planos de actividade do INE e das entidades com delegação de competências seriam objecto de acompanhamento trimestral, atribuindo tal missão a uma secção especializada do conselho.

Em matéria de coordenação técnica, importa agora relevar o baixo nível de observância do dispositivo legal, previsto na lei de bases, e que consagra o dever de audição prévia do Conselho Superior de Estatística dos projectos de diploma que criem serviços de estatística ou que contenham normas com incidência na estrutura e funcionamento do Sistema Estatístico Nacional.

De facto, e a título de exemplo, o relatório de avaliação 1999-2001 explicita que dos serviços com funções e competências no âmbito da estatística, criados entre 1999 e 2001, num total de 22 instrumentos jurídicos, só 5 foram submetidos à audição prévia do Conselho Superior de Estatística, não obstante 3 envolvessem entidades nas quais o INE tinha delegado competências.

Adicionalmente, o Conselho Superior de Estatística realizou, em 2001, um levantamento preliminar dos observatórios com funções de natureza estatística onde foi

possível evidenciar a sobreposição de entidades a produzirem a informação estatística nas mesmas áreas com conseqüente duplicação de esforços, carga sobre os inquiridos e eventual delapidação de recursos.

Em terceiro lugar, fazemos um balanço relativamente positivo do trabalho desenvolvido pelo INE e pelas secções especializadas do conselho, em matéria de aprovação de conceitos, definições e nomenclaturas estatísticas com a respectiva disponibilização na Internet em 1998. A este respeito, o documento estratégico apresentado ao Conselho Superior de Estatística em 2002 e relativo ao período 2003–2007”, incluiu um objectivo de primeira prioridade que consistia na implementação de um sistema integrado de meta – informação estatística, isto é, a conceptualização e implementação de um sistema integrado de gestão de nomenclaturas, dum sistema integrado de gestão de conceitos e de um sistema integrado de gestão de documentos metodológicos.

No que diz respeito à formação especializada de quadros superiores de estatística, enquanto instrumento de coordenação técnica, nomeadamente através do concurso de instituições universitárias vocacionadas para o efeito, releva-se o importante papel desempenhado pelo Instituto Superior de Estatística e Gestão de Informação da Universidade Nova de Lisboa na formação de quadros do INE e das entidades delegadas, ao nível de licenciaturas ou de cursos pós-graduados, nas áreas da estatística e sistemas de informação, pese embora o facto de nunca ter sido executado um plano integrado de formação estatística que abrangesse a equipa técnica do INE e das entidades co-produtoras de estatística oficiais. Por outro lado, o INE não soube aproveitar em toda a sua extensão a capacidade técnica montada nesse Instituto ao longo da década de 90, designadamente em áreas determinantes da actividade estatística tais como as dos desenhos amostrais e metodologias em geral e os métodos de tratamento e análise da informação estatística. Exceptua-se, pela positiva, a importante área dos sistemas de informação geográfica, onde o INE evoluiu muito positivamente na segunda metade da década de 90, assim como o estabelecimento de uma articulação de base contratual entre o INE e esse Instituto nos últimos anos do período em análise.

Por último, a alteração dos estatutos do INE ocorrida em 1989, visava, entre outro objectivos, a migração para um modelo de gestão do tipo empresarial conferindo mobilidade aos meios e flexibilidade ao funcionamento interno e às ligações ao exterior num quadro de autonomia financeira, permitindo uma boa expectativa ao exercício, por parte do INE, das funções de coordenação.

A avaliação que fazemos a este propósito inclui referências muito positivas ao nível do planeamento e gestão financeira, na base da implementação de um sistema de instrumentos de contabilidade analítica com a inerente co-responsabilidade das unidades departamentais na elaboração e execução orçamental, mas também inclui referências negativas no que diz respeito à ausência de uma verdadeira estratégia em matéria de gestão de recursos humanos a médio prazo que tomasse em consideração o “*core business*” do Instituto e de cada um dos seus departamentos de produção. Enfim, não se consolidou, de facto, a reestruturação iniciada em 89 que previa o estabelecimento, a

prazo, de uma nova matriz de recursos humanos adaptada às previsíveis pressões que recairiam crescentemente sobre o Instituto, decorrentes do processo europeu, e acentuadas em finais da última década no quadro da União Económica e Monetária e mais tarde no âmbito da Agenda 2000.

A este respeito os propósitos enunciados no quadro da macro-estrutura orgânica do INE, aprovados em início de 2002, e que visavam um reforço da coordenação nas vertentes interna e externas, nomeadamente as condições de coordenação da actividade delegada, foram, no período subsequente, prejudicados devido à impossibilidade do Instituto renovar contratos com as equipas qualificadas recrutadas em 2001 assim como tais propósitos foram prejudicados pela decisão governamental relativa à perda de autonomia financeira do Instituto, ocorrida no ano de 2002.

### 3.2 *Autoridade Estatística*

Abordaremos agora a problemática da autoridade estatística circunscrita ao importante domínio do acesso do INE às fontes administrativas. A legislação do Sistema Estatístico Nacional atribui ao Conselho Superior de Estatística competências para “*fomentar o aproveitamento de dados administrativos para fins estatísticos formulando recomendações, com vista, designadamente, à utilização nos documentos administrativos das definições, conceitos e nomenclaturas estatísticas*”. Ora, apesar das orientações implícitas naqueles dispositivos legais e dos inúmeros esforços desenvolvidos desde a publicação da lei de bases em 89, verificou-se que a posterior publicação de outros instrumentos legais, visando proteger domínios específicos tais como o segredo fiscal, o segredo bancário e os dados individuais (lei de protecção de dados pessoais de 1998), ao não terem em conta os princípios consagrados na lei do Sistema Estatístico Nacional, vieram introduzir dificuldades no aproveitamento de dados estatísticos não inteiramente superadas no período em análise. Tal situação persistiu apesar da significativa atenção dada a este assunto pelo Instituto Nacional de Estatística e pelo Conselho Superior de Estatística. Em consequência, o Sistema Estatístico Nacional é dos sistemas europeus o que menos recorre ao aproveitamento de actos administrativos para fins estatísticos – o que tem retirado flexibilização e racionalização à produção de estatísticas oficiais.

Consideramos este facto como um dos maiores constrangimentos da produção de estatísticas oficiais em Portugal e que nos distancia claramente das melhores práticas atingidas no âmbito da União Europeia.

Em casos pontuais foi, apesar de tudo, possível evoluir positivamente, diminuindo a carga sobre os respondentes e os custos globais da produção estatística - a título de exemplo podemos referir:

- a utilização da fonte administrativa *Quadros de Pessoal* para a obtenção de informação sobre os trabalhadores estrangeiros e para a actualização dos dados relativos à evolução do emprego e do emprego diplomado no sector das tecnologias de informação e comunicação;

- o recurso aos ficheiros centrais das universidades e da Direcção Geral do Ensino Superior para a obtenção de dados relativos a doutoramentos concluídos nas universidades portuguesas e estrangeiras;
- o recurso a informação administrativa gerida pela Fundação para a Ciência e Tecnologia para produção de dados relativos a formação avançada;
- ou ainda o aproveitamento de dados financeiros de suporte à Conta Geral do Estado para a elaboração das estatísticas relativas às despesas da administração pública em actividades de gestão e protecção ambiental.

Por último, regista-se pela positiva, o regulamento sobre o acesso, recolha e tratamento pelo INE de dados pessoais de carácter administrativo explicitado no decreto lei nº: 294/2001 – no essencial, tal regulamento conciliou os dispositivos previstos na lei de protecção de dados individuais com as necessidades informacionais consideradas relevantes pelo Conselho Superior de Estatística ao eficiente funcionamento do Sistema Estatístico Nacional:

- preconizando a celebração de protocolos entre o INE e os organismos públicos detentores de tais dados, mediante parecer prévio da Comissão Nacional de Protecção de Dados;
- esclarecendo que o diploma prevalece sobre eventuais limitações ou deveres de sigilo ao abrigo dos quais as informações foram recolhidas, salvo casos de espécie que não afectam, de facto, as necessidades do Sistema Estatístico Nacional;

### 3.3 A centralização Estatística

Já referimos, a propósito da coordenação estatística, a descentralização funcional do INE através do estabelecimento do processo de delegação de competências previsto na Lei de Bases do Sistema Estatístico Nacional.

Abordaremos agora uma das medidas mais emblemáticas da reestruturação de 1989 e que se refere à descentralização geográfica do INE consubstanciada na criação das Direcções Regionais, cujo funcionamento foi devidamente regulamentado, mediante ordem de serviço, em 1990: *“As Direcções Regionais constituíram-se serviços desconcentrados ao nível de departamento tendo por finalidades assegurar, a nível regional, a execução das operações estatísticas de âmbito nacional, regional e local, as funções de centros regionais de informação e documentação estatística nacional e a gestão dos respectivos recursos humanos, financeiros e materiais, de acordo com a política e os objectivos de âmbito nacional definidos pelo Instituto para o referido subsistema”*.

Nesse contexto, as Direcções Regionais desenvolveram a sua actividade em articulação funcional com os diferentes departamentos do instituto, tendo-se ainda articulado com as entidades públicas ou privadas das regiões, designadamente com as Comissões de Coordenação de Desenvolvimento Regional, e outros serviços regionais

desconcentrados da administração central, com as autarquias locais, associações empresariais, associações sindicais e associações de consumidores, com implantação regional. Essa articulação ganha mais relevo com as Secções Regionais do Conselho Superior de Estatística, secções permanentes deste Conselho cuja composição incluía representantes das instituições mais relevantes da região, que acompanhavam sistematicamente a execução dos planos anuais de actividade promovidos pelas Direcções Regionais do INE, constituindo-se também um organismo de escuta de novas demandas integradas de estatísticas oficiais portuguesas.

No quadro das orientações e directrizes emanadas dos competentes serviços centrais do instituto, competia às Direcções Regionais do INE, dentro da respectiva área de competência, e segundo parâmetros definidos pela Direcção do Instituto:

- colaborar na concepção de operações estatísticas de âmbito nacional;
- executar, dentro da sua área geográfica de competência, as operações estatísticas de âmbito nacional;
- propor a concessão e execução de operações estatísticas de âmbito regional no âmbito da produção de informação reconhecidamente relevante para o desenvolvimento regional e mediante captação de financiamento à escala regional;
- coordenar e difundir a informação estatística de âmbito regional, bem como exercer as funções de centro regional de informação e documentação estatística nacional;
- Propor a realização de estudos económicos e sociais de âmbito regional.

As Direcções Regionais do Norte, Centro e Alentejo foram as primeiras a serem criadas e só mais tarde, na segunda metade da década de 90 o Instituto criou as Direcções Regionais do Algarve e da região de Lisboa e Vale do Tejo.

É incontornável referirmos aqui, com algum detalhe, uma das manifestações mais positivas da evolução do sistema estatístico nacional ao longo da década de 90, consubstanciada numa crescente implantação do INE no seio das regiões através do estabelecimento de um diálogo profícuo do Instituto com os actores da região, nomeadamente por via de uma progressiva articulação do INE com o tecido institucional das regiões, mormente o universitário. Por outro lado, a nível interno ao INE, essa evolução positiva traduziu-se na contínua superação de inúmeros constrangimentos detectados na década de 80 e que se relacionavam com a qualidade e o custo da recolha directa da informação estatística. A título de exemplo, constatava-se no passado uma insuficiente taxa de resposta em várias operações estatísticas de grande importância.

Para o efeito, constituíram-se em cada Direcção Regional competências especializadas em várias áreas da produção estatística e simultaneamente equipas técnicas polivalentes, com crescentes ganhos de produtividade, e mediante a introdução de novas racionalidades nos processos de produção estatística à custa de soluções tecnológicas inovadoras e sempre que possível, de economias de escala.

Em meados da década de 90 era já muito evidente o contributo qualificado das jovens equipas das Direcções Regionais na prossecução de objectivos cada vez mais exigentes definidos pelas coordenações centrais das operações estatísticas, nomeadamente os decorrentes das prioridades das políticas da União Europeia e da respectiva informação de suporte: União Económica e Monetária, Alargamento da União Europeia, Competitividade, Desenvolvimento Sustentável e Agenda Social, com particular incidência nos domínios seguintes:

- *União Económica e Monetária*: a informação estatística necessária as sucessivas fases da União Económica e Monetária e ao Pacto de Estabilidade e Crescimento;
- *Competitividade, Desenvolvimento Sustentável e Agenda Social*: realce para as estatísticas relativas ao mercado de trabalho, indicadores de apoio ao desenvolvimento sustentável, serviços, condições de vida das famílias, migrações, indicadores de exclusão social e pobreza e *e-Europe*;
- *Indicadores Estruturais*: consolidação do sistema de informação definido na Cimeira de Lisboa, organizado pelos Institutos Nacionais de Estatística dos países membros e divulgados anualmente no chamado Relatório da Primavera da Comissão Europeia;
- *Apoio às estatísticas europeias em curso*: actividades estatísticas nas áreas da agricultura, política regional, comércio externo, sociedade da informação, inovação, investigação e desenvolvimento.

Nos últimos anos do período objecto de análise foi, de facto, crucial o apoio das Direcções Regionais do INE na concretização de novas metas ao nível dos calendários de divulgação de informação, particularmente aqueles que se relacionavam com informação de curto prazo. Sublinhamos, pela sua importância a divulgação trimestral dos agregados macro-económicos, as chamadas contas trimestrais, cujo prazo de divulgação passou em 2003 de 120 dias para 70 dias, em linha com as recomendações do Eurostat. Regista-se também a evolução muito positiva no prazo de divulgação dos indicadores mensais de conjuntura, onde nos situávamos em 2003 entre os mais cumpridores à escala Europeia. Isto é, ao dinamismo e qualificação das equipas dos departamentos centrais coordenadores de tão importantes projectos estatísticos da área económica, somaram-se as valências regionais que intervieram com determinação na prossecução de objectivos prioritários da actividade estatística do SEN.

Paralelamente, as Direcções Regionais do INE desenvolveram operações estatísticas inovadoras em Portugal na sequência da caracterização da procura expressa pelos actores do desenvolvimento regional e local. Sumariamente citaremos, a título ilustrativo, as iniciativas de maior relevo:

- o Inventário Municipal, implementado pela primeira vez na Região Norte em 1992, e progressivamente levado a cabo nas restantes regiões do país;

- o Índice de Poder de Compra per Capita dos concelhos do país;
- o Sub-sistema de Estatísticas Urbanas, com base na delimitação das cidades para fins estatísticos e apoiada em informação cartográfica de detalhe;
- o Sistema de Informação da Construção e Habitação, que se encontrava profundamente deficitário em meados da década de 90, e que consistiu essencialmente na disponibilização de indicadores sobre a pressão construtiva nas áreas metropolitanas de Lisboa e Porto e nos concelho com cidades médias, na produção das primeiras estimativas inter-censitárias do parque habitacional, na informação sobre os custos de construção nova e ainda a relativa aos preços de avaliação bancária da habitação.

Refira-se ainda o papel das Direcções Regionais na dinamização das estatísticas do turismo no período 2001-2003, sob a coordenação da Direcção Regional do Algarve. De facto, no final da década de 90, as estatísticas do turismo encontravam-se numa total estagnação apesar da reconhecida importância do sector do turismo para a economia nacional. O Instituto Nacional de Estatística, em parceria com o Banco de Portugal e com a Direcção Geral do Turismo, realizou em 2002 uma inventariação detalhada das operações estatísticas necessárias à viabilização da elaboração da Conta Satélite do turismo a partir do ano 2005.

Sublinha-se também, pela sua relevância, o papel das Direcções Regionais na preparação e execução das grandes operações censitárias da agricultura (1999) e da População e Habitação (em 1991 e 2001). Registou-se uma eficaz complementaridade de saberes e competências entre as coordenações centrais deste projecto censitário e as unidades regionais do INE constituídas para o efeito, e articuladas com as autarquias locais, o que muito contribuiu para o indiscutível sucesso dessas operações censitárias em sede de qualidade intrínseca do produto estatístico e no que diz respeito ao calendário de divulgação dos resultados, em linha com as melhores práticas à escala europeia.

Em segundo lugar, o Instituto Nacional de Estatística tem na difusão activa da informação estatística um dos principais pilares da sua missão: a informação estatística só é útil se for devidamente conhecida e utilizada. A este respeito as Direcções Regionais do INE tornaram-se observatórios privilegiados para a identificação das necessidades dos utilizadores quer de informação geograficamente agregada, de informação inter-regional ou de informação de base territorial e de geometria variável.

O insuficiente nível de literacia estatística do nosso país aconselha que nos principais centros populacionais haja efectivas competências na área da produção e interpretação do produto estatístico, capazes de promover a comunicação entre os diferentes actores no terreno, fazendo das estatísticas oficiais um efectivo instrumento ao serviço das políticas do desenvolvimento. Sem prejuízo de outro tipo de acções à distância e desde logo nas escolas, como tem sido o caso do projecto *Alea* (Acção Local de Estatística Aplicada), implementado na década de 90, mediante uma frutuosa articulação entre as escolas e o INE.

Produzir e difundir informação estatística oficial **orientada para o utilizador** exige que se aprofunde o círculo virtuoso utilizador / produtor de informação, mediante intervenções qualificadas e em regime de proximidade e por isso protagonizadas pelo INE através das suas Direcções Regionais.

Em todo o caso há ainda um complexo e longo caminho a percorrer para o desenvolvimento de sistemas de informação de base territorial assentes na sistematização das problemáticas do território nas suas diferentes dimensões, estruturando-as enquanto subsistemas de informação estatística – o aprofundamento dos sistemas de informação de base regional, dos sistemas de informação do espaço urbano ou dos sistemas de informação do espaço rural estarão em breve na primeira linha da agenda europeia e tudo indica que merecerão apoios comunitários no âmbito do próximo quadro comunitário de apoio. Por isso, poderá caber ao INE um papel relevante no que diz respeito ao contributo das estatísticas oficiais portuguesas para tais sistemas.

### 3.4. *A análise e a compreensão da informação*

No período 1999-2003, verificou-se uma interessante evolução ao nível da análise e compreensão da informação para efeitos de difusão das estatísticas oficiais – quer no que diz respeito ao conteúdo das notas à comunicação social quer no que se refere à multiplicação de iniciativas, no INE e em alguns organismos delegados, na elaboração e difusão de estudos interpretativos das estatísticas oficiais produzidas, em áreas muito diversas do conhecimento, com o concurso dos departamentos do INE particularmente vocacionados para a realização de estudos sócio-económicos e com a intervenção crescente das Direcções Regionais em trabalhos de abrangência nacional ou especificamente regionais.

Destacamos algumas das principais iniciativas que neste domínio e pela sua qualidade mais visibilidade e prestígio conferiram ao INE ao longo desses anos:

- a *Síntese Económica da Conjuntura*, mensal e trimestral;
- a *Revista de Estudos Demográficos do INE*, interrompida em 1993 e retomada em 2002;
- a *Revista do INE*, mais tarde transformada numa revista científica, a *Revstat*, em particular parceria com a Sociedade Portuguesa de Estatística e com a participação de investigadores nacionais ou estrangeiros.
- As dezenas de estudos realizados pelas Direcções Regionais do INE, muitos deles em regime de parceria com o tecido institucional da região ou do espaço trans-fronteiriço;
- a *Revista Portuguesa de Estudos Regionais*, criada em 2002, iniciativa conjunto do INE e da Associação Portuguesa de Desenvolvimento Regional.

### 3.5. *A gestão da qualidade no âmbito do Sistema Estatístico Nacional*

Em linha com as melhores práticas europeias o INE criou em 1996 uma Comissão Permanente para a qualidade tendo apresentado mais tarde, em 1999, o sistema de qualidade do INE onde foram relatadas algumas experiências de controlo de qualidade abrangendo os ficheiros gerais de unidades estatísticas, o Inquérito Harmonizado às Empresas bem como uma abordagem sobre problemas de qualidade na recolha directa de informação estatística. Em 2001, o INE recorreu mesmo a uma entidade universitária, o Instituto Superior de Estatística e Gestão de Informação da Universidade Nova de Lisboa, para a realização de um inquérito à qualidade em total independência face ao Instituto e ainda nesse ano o INE comprometeu-se na criação de um conjunto de instrumentos relativos a métodos, procedimentos e documentação de referência na prossecução dos objectivos de qualidade e mediante a realização de auditorias de qualidade aos principais processos de produção estatística e a outros processos-chave, prosseguindo a implementação do sistema de gestão da qualidade com base no modelo de excelência da *European Foundation for Quality Management* e nas normas ISO 9000 que recolhe ampla adesão ao nível do Sistema Estatístico Europeu.

No entanto o inquestionável dinamismo gerado no INE em torno da implementação das auditorias de qualidade, coexistiu com o adiamento, ao longo da década de 90, na definição de uma política consequente em matéria de gestão e recrutamento de recursos humanos que projectasse o INE no futuro como um instituto de referência reconhecido como:

- produtor e fornecedor de informação oficial de qualidade;
- organização independente e credível;
- promotor da literacia estatística;
- interveniente activo e eficaz na cooperação internacional.

Esta situação gerou insuficiências técnicas ao nível dos departamentos centrais de produção designadamente no que se refere às competências metodológicas (mais geralmente na esfera da coordenação estatística).

Tais deficiências impossibilitaram em certos casos que o INE colmatasse as áreas estatísticas mais carenciadas e inviabilizaram adequada intervenção do instituto nas áreas estatísticas emergentes no quadro do Sistema Estatístico Europeu. Esta situação foi parcialmente superada através do redireccionamento de algumas áreas da produção estatística para a esfera de intervenção das Direcções Regionais.

A remodelação levada a cabo no INE em 2001, permitiu a criação do Departamento de Metodologia do INE, até então inexistente como tal, conferindo a tal Departamento a missão de coordenar a actividade do desenvolvimento técnico-científico no domínio das metodologias estatísticas, apoiando horizontalmente as unidades orgânicas do INE e complementarmente os organismos integrantes do Sistema Estatístico Nacional – tratou-

se de colmatar uma das maiores deficiências que o Sistema de Estatística Nacional evidenciou na década de 90, pela inexistência de uma unidade departamental com capacidade técnica para integrar a harmonizar determinados processos que condicionam particularmente a qualidade, o tempo de elaboração e os custos da actividade de produção de estatísticas oficiais.

Compete a este Departamento obter ganhos de eficiência e eficácia nos diferentes procedimentos metodológicos, de amostragem, de estimação, de controlo de qualidade, de validação, de tratamento de não respostas, de protecção de confidencialidade, e de apuramento de dados, construindo padrões aplicativos de uso comum e evitando o recurso sistemático ao desenvolvimento de aplicação específicas a cada operação estatística.

### *3.6 A reformulação dos dispositivos legais do sistema estatístico nacional*

A legislação do Sistema Estatístico Nacional – lei do sistema e estatutos do INE – carece de uma reformulação que tome em consideração os pontos fracos do sistema e as actuais exigências da sociedade.

No que diz respeito à lei de bases, destacamos os seguintes pontos, relativamente aos quais enunciamos as seguintes propostas:

- em matéria de noção de Sistema Estatístico Nacional propomos que o dispositivo legal assegure que os intervenientes na actividade de produção e difusão de estatísticas oficiais sejam exclusivamente entidades públicas; por outro lado a noção de SEN deve prever um preceito que consagre que o Sistema Estatístico Nacional é parte integrante do Sistema Estatístico Europeu;
- no que se refere aos Órgãos do Sistema Estatístico Nacional, propomos a inserção do Banco de Portugal na estrutura do SEN, pelas competências que detém nas áreas estatísticas monetárias e financeiras e da balança de pagamentos e pela necessária articulação que Portugal tem que assegurar no cumprimento das obrigações estatísticas comunitárias, particularmente as relacionadas com o Banco Central Europeu;
- relativamente aos Princípios do Sistema Estatístico Nacional importa:
  - consagrar a absorção dos Princípios Constantes do Regulamento Comunitário sobre o Sistema Estatístico Europeu: Autonomia Técnica; Fiabilidade; Imparcialidade; Autoridade Estatística; Pertinência; Segredo Estatístico; Coordenação Estatística;
  - incorporar também o princípio da acessibilidade estatística;
  - flexibilizar a aplicação do princípio do Segredo Estatístico para fins de investigação científica;

- a nova lei de bases do SEN deverá incluir um preceito sobre Princípios de Ética Profissional do Estatísticos Oficiais;
- no quadro da delegação de competências do INE:
  - o SEN deve ter um pendor mais centralizante no plano funcional, e o INE só deve delegar competências em Áreas Sociais cujos Ministérios da tutela sejam depositários de informação administrativa susceptível de efectivo aproveitamento para fins estatísticos oficiais;
  - não delegar competências em áreas estatísticas contributivas do cálculo das contas nacionais e outros indicadores Macro-económicos;
  - assegurar por parte do INE um efectivo acompanhamento técnico e funcional dos Órgãos Delegados.
- sobre o acesso a dados de natureza administrativa consideramos fundamental que a futura lei de bases garanta uma solução jurídica que assegure o efectivo acesso a dados administrativos para fins estatísticos oficiais, bem como a participação do INE na concepção dos correspondentes formulários e registos de suporte;
- por último, em sede de Audição Prévia do Conselho Superior de Estatística a futura lei deverá consagrar um preceito que estimule a obrigatoriedade de tal audição na aprovação de projectos de diplomas que criem Serviços de Estatística ou com outra incidência no SEN, consagrando a possibilidade do Presidente do INE ser chamado a participar nas reuniões do Conselho de Secretários de Estado que apreciem tais projectos;

Relativamente aos Estatutos do INE destacamos os dois pontos seguintes:

- sobre a natureza jurídica é essencial, promover acções adequadas no sentido de incluir o INE nos Regimes Especiais previstos no artigo 48, nº 1 da lei nº 3 / 04, de 15 de Janeiro, que aprovou a Lei Quadro dos institutos públicos; e recuperar a natureza de pessoa colectiva de direito público, dotada de personalidade jurídica, património próprio e autonomia administrativa e financeira;
- sobre a desconcentração geográfica, os novos estatutos do INE deverão consagrar a desconcentração geográfica do INE com a revisão das atribuições dos correspondentes serviços numa perspectiva de ganhos de eficiência e eficácia, e visando reforçar a resposta do INE à escala regional, designadamente a aproximação do INE aos utilizadores e aos inquiridos.

#### 4. Conclusões

O desenvolvimento do mercado introduz novas realidades que exigem estatísticas que permitam monitorar os aspectos determinantes da actual dinâmica dos agentes

económicos. Por isso a evolução dos sistemas estatísticos nacionais deverá incorporar novos desafios relativos à produção de indicadores que retratem fenómenos emergentes tais como a produtividade e a competitividade sectoriais, o grau de penetração de capital estrangeiro nas empresas com actividade no território nacional, a identificação de transferência nos centros de decisão e avaliação do grau de internacionalização e inovação das empresas portuguesas.

Complementarmente, caberá aos institutos nacionais de estatística uma responsabilidade crescente na consolidação do sistema de indicadores estatísticos oficiais que monitorizem uma multiplicidade de fenómenos de exclusão social e de pobreza, numa base territorial, de geometria variável.

As profundas exigências colocadas à produção e oferta de informação estatística oficial pertinente, pressupõem uma forte coesão da rede institucional do SEN, que permita dar uma resposta adequada a tais desafios. O reforço da articulação entre o INE e as restantes entidades surge como fundamental e deve assentar em princípios de cooperação e co-responsabilização, a favor da eficiência e prestígio do SEN mas sobretudo a favor da governança nacional e regional.

O imperativo de intensificar a prática da coordenação do Sistema Estatístico Nacional tanto no domínio dos procedimentos e práticas de gestão, como dos instrumentos técnico-científicos de normalização e da gestão da qualidade, deverá constituir uma bandeira do contínuo processo de modernização do SEN.

As Linhas Gerais da Actividade Estatística Nacional e Respectivas Prioridades para 2003-2007, consubstanciadas em quatro eixos de desenvolvimento estratégico:

- melhorar a qualidade da informação estatística oficial;
- melhorar a eficiência dos processos associados à produção e difusão das estatísticas oficiais;
- potenciar o desenvolvimento dos Recursos Humanos do Sistema Estatístico Nacional;
- rever o quadro jurídico e institucional do Sistema Estatístico Nacional,

deverão constituir o suporte essencial para a definição de um plano de acção em linha com os objectivos estratégicos enunciados e aprovados pelo Conselho Superior de Estatística em 2002 e que abrangem a produção estatística, a difusão, a coordenação, a cooperação internacional e a organização, recursos humanos e meios.

**Paulo Gomes**

Consultor no Instituto Nacional de Estatística

Presidente do Instituto Nacional de Estatística (2001-2003)

Vice-Presidente do Conselho Superior de Estatística (2001-2003)

Director Regional do Instituto Nacional de Estatística (1989-2001)

Vice-Chairman da 56<sup>a</sup> Sessão do Instituto Internacional de Estatística, a realizar em Lisboa em 2007.

Autor ou co-autor de vários trabalhos científicos na área da análise estatística multivariada

Autor ou co-autor de vários trabalhos de aplicação dos métodos da estatística multivariada à temática da caracterização de espaços territoriais

# Modelos Normais Ortogonais: dos Anos 70 ao Presente

*João Tiago Mexia*

## 1. Introdução

No que se segue procuraremos apresentar os aspectos mais relevantes da teoria dos modelos normais ortogonais. Este tema enquadra-se bem com os 25 anos da SPE, já que se pode começar a história moderna destes modelos com os trabalhos seminais de Seely na década de 70.

Iniciaremos a nossa exposição pela estrutura algébrica destes modelos, a qual, como veremos, conduz a uma formulação canónica. Esta formulação, além de ser menos restritiva do que a formulação usual, conduz directamente à obtenção de estatísticas minimais suficientes e permite ainda discutir dessa perspectiva adequada certos problemas de inferência. Seguidamente consideraremos a formulação usual destes modelos. Esta formulação assenta na consideração de factores de efeitos fixos ou aleatórios, pelo que a designaremos como formulação factorial. Mostraremos como a formulação factorial resulta duma particularização da formulação canónica. Uma vez apresentadas as duas formulações dos modelos, passamos à inferência começando por obter UMVUE para os vários parâmetros. Em particular, discutiremos o bem conhecido problema de, para componentes de variância se poderem obter UMVUE negativos. Pode admitir-se que todo o tratamento dos modelos até à estimação pontual é feito duma forma contínua a partir da respectiva formulação canónica. No entanto, quando se pretende passar à construção de regiões de confiança ou de testes de hipóteses a presença de parâmetros perturbadores levanta problemas. Dado dispor-se para esses parâmetros de estimadores funções das estatísticas normais suficientes, podia-se utilizar uma abordagem baseada em variáveis ancilares (ver, por exemplo, [28]). No entanto optamos por uma abordagem alternativa baseada em variáveis pivot e de teste generalizadas (ver [31]). Dado estas noções terem sido introduzidas cerca de vinte anos depois do avanço permitido pelos trabalhos de Seely, a história que apresentamos enquadra-se, como referimos, com a da SPE.

## 2. Estrutura Canónica

Se o vector  $\mathbf{Y}^n$  tiver distribuição normal com vector médio  $\boldsymbol{\mu}^n$  e matriz de covariância  $\mathbf{V}$ , escreveremos  $\mathbf{Y}^n \sim N(\boldsymbol{\mu}^n; \mathbf{V})$ . Dada uma partição ortogonal

$$\mathbb{R} = \boxplus_{i=1}^w \nabla_i, \quad (1)$$

o modelo é ortogonal associado a esta partição se

1.  $\boldsymbol{\mu}^n \in \boxplus_{i=1}^w \nabla_i$ ;
2.  $\mathbf{V} = \sum_{i=1}^w \gamma_i \mathbf{Q}_i$ ;

com  $\mathbf{Q}_1, \dots, \mathbf{Q}_w$  as matrizes de projecção ortogonal sobre os  $\nabla_1, \dots, \nabla_w$ .

Sendo

$$A = \left\{ \sum_{i=1}^w a_i \mathbf{Q}_i, a_1, \dots, a_w \in \mathbb{R} \right\}, \quad (2)$$

é fácil verificar que  $A$  é um espaço vectorial constituído por matrizes simétricas que comutam e, se  $\mathbf{M} \in A$ ,  $\mathbf{M}^2 \in A$ . Logo,  $A$  será uma álgebra de Jordan comutativa. Estas álgebras foram introduzidas por [7] para estudar questões relativas à Mecânica Quântica. Posteriormente, estas estruturas foram redescobertas por Seely (ver [23], [24]) nos seus trabalhos seminais atrás referidos. Neste caso estas estruturas eram utilizadas para resolver questões de inferência estatística, orientação esta que seguiremos. Deveremos também referir como importantes os trabalhos de Seely e Zyskind ([25]), bem com [27]. Em particular, neste último é estabelecido que para toda a álgebra de Jordan comutativa existe uma única base constituída por matrizes de projecção ortogonal mutuamente ortogonais. Para  $A$  essa base, a base principal, é constituída pelas matrizes  $\mathbf{Q}_1, \dots, \mathbf{Q}_w$ . Referimos ainda que Seely designava estas estruturas como espaços vectoriais quadráticos. De facto, trata-se de espaços vectoriais que contêm os quadrados das matrizes que os compõem. No entanto, por razões de prioridade, adaptaremos a designação de álgebras de Jordan.

Se os vectores linha de  $\mathbf{A}_i$  constituem uma base ortonormada para  $\nabla_i$ , ter-se-á  $\mathbf{Q}_i = \mathbf{A}_i' \mathbf{A}_i$ ,  $i = 1, \dots, w$ . Tem-se então

$$\boldsymbol{\mu}_n = \sum_{i=1}^w \mathbf{Q}_i \boldsymbol{\mu}^n = \sum_{i=1}^w \mathbf{A}_i' \boldsymbol{\eta}_i^{g_i}, \quad (3)$$

com  $\boldsymbol{\eta}_i^{g_i} = \mathbf{A}_i \boldsymbol{\mu}^n$ ,  $g_i = \dim(\mathbf{Q}_i) = \text{car}(\mathbf{Q}_i)$   $i = 1, \dots, w$ .

É fácil verificar, multiplicando à esquerda por  $A_j$ ,  $j = 1, \dots, m$ , que  $\sum_{i=1}^m A_i a_i^{g_i} = \sum_{i=1}^m A_i b_i^{g_i}$  se verifica se e só se  $a_j^{g_j} = b_j^{g_j}$ ,  $j = 1, \dots, m$ . Analogamente, multiplicando à esquerda por  $Q_j$ ,  $j = 1, \dots, w$ , mostra-se que  $\sum_{i=1}^w a_i Q_i = \sum_{i=1}^w b_i Q_i$  equivale a  $a_j = b_j$ ,  $j = 1, \dots, w$ . Assim, os parâmetros canônicos constituídos por

1. as componentes de variância  $\gamma_i$ ,  $i = 1, \dots, w$ ;
2. os vectores  $\eta_i^{g_i}$ ,  $i = 1, \dots, m$ ;

identificam  $N(\mu^n, V)$ .

No que se segue, para referir as partes de efeitos fixos e de efeitos aleatórios do modelo admitiremos que

$$\gamma_i = \sum_{j=1}^m c_{i,j} \gamma_j, \quad (i = 1, \dots, w) \quad (4)$$

Para garantir a existência de densidade para  $Y^n$  temos de garantir que  $V$  é definida positiva, o que é equivalente a ter-se  $\gamma_i > 0$ ,  $i = 1, \dots, w$ . Tem-se então, atendendo a que as matrizes da base principal são mutuamente ortogonais,

$$V^{-1} = \sum_{i=1}^w \gamma_i^{-1} Q_i. \quad (5)$$

Por outro lado, os  $\gamma_1, \dots, \gamma_w$  serão os valores próprios de  $V$  com multiplicidade  $g_1, \dots, g_w$ , pelo que

$$\det(V) = \prod_{i=1}^w \gamma_i^{g_i}. \quad (6)$$

É agora fácil, tomando

$$\begin{cases} Z_i^{g_i} = A_i Y^n; i = 1, \dots, m \\ S_i = \|Z_i^{g_i}\|^2; i = 1, \dots, w \end{cases} \quad (7)$$

mostrar que

$$(\mathbf{Y}^n - \boldsymbol{\mu}^n)\mathbf{V}^{-1}(\mathbf{Y}^n - \boldsymbol{\mu}^n) = \sum_{i=1}^m \frac{\|\mathbf{Z}_i^{g_i} - \boldsymbol{\eta}_i^{g_i}\|^2}{\gamma_i} + \sum_{i=m+1}^w \frac{S_i}{\gamma_i}. \quad (8)$$

e, conseqüentemente, que se têm as estatísticas suficientes:  $\mathbf{Z}_1^{g_1}, \dots, \mathbf{Z}_m^{g_m}, S_{m+1}, \dots, S_m$ .

Como o espaço paramétrico da densidade de  $\mathbf{Y}^n$  contém o produto cartesiano de intervalos não degenerados, estas estatísticas são (ver [29]) completas. Assim, atendendo ao teorema de Blackwell-Lehmann-Scheffé:

1. os  $\mathbf{Z}_i^{g_i}, i=1, \dots, m$ , são UMVUE dos  $\boldsymbol{\eta}_i^{g_i}, i=1, \dots, m$ , já que  $\mathbf{Z}_i^{g_i} \sim N(\boldsymbol{\eta}_i^{g_i}, \gamma_i \mathbf{I}_{g_i}), i=1, \dots, m$ ;

2. os  $\tilde{\gamma}_i = \frac{S_i}{g_i}, i=m+1, \dots, w$ , são UMVUE dos  $\gamma_i, i=m+1, \dots, w$ , já que estas somas de quadrados são os produtos pelas componentes de variância canónicas de qui-quadrados centrais com  $g_{m+1}, \dots, g_w$  graus de liberdade. Poremos  $S_i \sim \gamma_i \chi_{(g_i)}^2, i=m+1, \dots, w$ ;

3. os  $\tilde{\gamma}_i = \sum_{j=m+1}^w c_{i,j} \tilde{\gamma}_j, i=1, \dots, m$  são UMVUE dos  $\gamma_i, i=1, \dots, m$ .

Estes estimadores generalizam-se directamente, tendo-se o UMVUE  $\tilde{\theta} = \sum_{j=m+1}^w c_j \tilde{\gamma}_j$  para  $\theta = \sum_{j=m+1}^w c_j \gamma_j$ .

Convém-nos considerar em mais detalhe a estimação de parâmetros da forma

$$\boldsymbol{\lambda}^s = \mathbf{B}\boldsymbol{\mu}^s. \quad (9)$$

Assim,

$$\tilde{\boldsymbol{\mu}}^n = \sum_{i=1}^m \mathbf{A}_i' \mathbf{Z}_i^{g_i} \quad (10)$$

será o UMVUE de  $\boldsymbol{\mu}^n$  e, conseqüentemente, o UMVUE de  $\boldsymbol{\lambda}^s$  será

$$\tilde{\boldsymbol{\lambda}}^s = \mathbf{B} \tilde{\boldsymbol{\mu}}^n = \sum_{i=1}^m \tilde{\boldsymbol{\varphi}}_i^s, \quad (11)$$

com  $\tilde{\boldsymbol{\varphi}}_i^s = \mathbf{B}\mathbf{A}_i' \mathbf{Z}_i^{g_i}$  o UMVUE de  $\boldsymbol{\varphi}_i^s = \mathbf{B}\mathbf{A}_i' \boldsymbol{\eta}_i^{g_i}, i=1, \dots, m$ . Mais geralmente o UMVUE de  $\boldsymbol{\varphi}_i^s = \mathbf{B}\boldsymbol{\eta}_i^{g_i}$  será  $\tilde{\boldsymbol{\varphi}}_i^s = \mathbf{B}\tilde{\boldsymbol{\eta}}_i^{g_i}, i=1, \dots, m$ .

Observe-se ainda que a distribuição conjunta dos  $\mathbf{Z}_i^{g_i}$ ,  $i = 1, \dots, w$ , é normal e que as matrizes de covariância cruzadas destes vectores são nulas, já que

$$\Sigma \left[ \left( \mathbf{Z}_1^{g_1'} \quad \dots \quad \mathbf{Z}_w^{g_w'} \right)' \right] = \mathbf{D} \left( \gamma_1 \mathbf{I}_{g_1}, \dots, \gamma_w \mathbf{I}_{g_w} \right), \quad (12)$$

com  $\mathbf{D} (\cdot, \dots, \cdot)$  uma matriz diagonal por blocos. Assim, os  $\mathbf{Z}_i^{g_i}$ ,  $i = 1, \dots, w$  são independentes.

Vê-se assim que o problema de estimação pontual no quadro da formulação canónica está resolvido. Observe-se que  $\tilde{\gamma}_i \geq 0$ ,  $i = m + 1, \dots, w$ . Ao considerarmos a construção de testes e obtenção de regiões de confiança voltaremos a considerar esta formulação dos modelos.

### 3. Formulação Factorial

Esta formulação é a usual, sendo da forma

$$\mathbf{Y}^n = \sum_{i=1}^u \mathbf{X}_i \beta_i^{c_i} + \mathbf{e}^n, \quad (13)$$

com  $\mathbf{X}_1, \dots, \mathbf{X}_u$  matrizes conhecidas. Os  $\beta_i^{c_i}$ ,  $i = 1, \dots, v$ , com  $v \leq u$ , serão vectores fixos cujas componentes são os efeitos dos níveis ou as interacções entre níveis de factores de efeitos fixos. Os efeitos de factores aleatórios e as interacções entre níveis de factores, pelo menos em parte de efeitos aleatórios são as componentes dos  $\beta_i^{c_i} \sim N(\mathbf{0}^{c_i}, \sigma_i^2 \mathbf{I}_{c_i})$ .

Estes últimos vectores são independentes entre si e de  $\mathbf{e}^n \sim N(\mathbf{0}^n, \sigma^2 \mathbf{I}_n)$ . Assim, o vector médio e a matriz de covariância de  $\mathbf{Y}^n$  serão dadas por

$$\begin{cases} \boldsymbol{\mu}^n = \sum_{i=1}^v \mathbf{X}_i \beta_i^{c_i} \\ \mathbf{V} = \sum_{i=v+1}^u \sigma_i^2 \mathbf{M}_i + \sigma^2 \mathbf{I}_n \end{cases}, \quad (14)$$

com  $\mathbf{M}_i = \mathbf{X}_i \mathbf{X}_i'$ ,  $i = 1, \dots, u$ . Os  $\beta_1^{c_{v+1}}, \dots, \beta_u^{c_u}, (\sigma_{v+1}^2, \dots, \sigma_u^2, \sigma^2)$  são os parâmetros usuais do modelo. Em particular, os  $\sigma_{v+1}^2, \dots, \sigma_u^2$  as componentes de variância usuais.

Pondo-se

$$\mathbf{X} = (\mathbf{X}_1 \quad \dots \quad \mathbf{X}_v) \quad (15)$$

e  $\mathbf{M} = \mathbf{X}\mathbf{X}'$ ,  $\mu^n$  pertence ao espaço imagem  $\Omega = R(\mathbf{X}) = R(\mathbf{M})$  das matrizes  $\mathbf{X}$  e  $\mathbf{M}$ . Ora (ver [20]), as matrizes  $\mathbf{M}$ ,  $\mathbf{M}_{\nu+1}, \dots, \mathbf{M}_u$  são diagonalizadas por uma mesma matriz ortogonal se e só se comutam. Admitamos que isso se verifica, vindo então que  $\mathbf{M}$ ,  $\mathbf{M}_{\nu+1}, \dots, \mathbf{M}_u \in A(\mathbf{P})$ , sendo  $A(\mathbf{P})$  o conjunto das matrizes diagonalizadas por  $\mathbf{P}$ . Ora, é fácil verificar que  $A(\mathbf{P})$  é uma álgebra de Jordan comutativa. É fácil ver que intersectando álgebras de Jordan comutativas se obtém uma álgebra de Jordan comutativa. Existirá então uma álgebra de Jordan comutativa minimal  $A$  que contém  $\mathbf{M}$ ,  $\mathbf{M}_{\nu+1}, \dots, \mathbf{M}_u$ . Sejam  $\mathbf{Q}_1, \dots, \mathbf{Q}_w$  as matrizes da base principal de  $A$ . Podemos ordenar estas matrizes de forma a ter-se

$$\mathbf{M} = \sum_{j=1}^m a_j \mathbf{Q}_j, \quad (16)$$

tendo-se então

$$\mu^n \in \Omega = \bigoplus_{i=1}^m \nabla_i. \quad (17)$$

Por outro lado, com

$$\mathbf{M}_i = \sum_{j=1}^w a_{i,j} \mathbf{Q}_j, \quad (18)$$

tem-se

$$\mathbf{V} = \sum_{i=\nu+1}^u \sigma_i^2 \left( \sum_{j=1}^w a_{i,j} \mathbf{Q}_j \right) + \sigma^2 \sum_{j=1}^w \mathbf{Q}_j = \sum_{j=1}^w \gamma_j \mathbf{Q}_j. \quad (19)$$

As componentes canónicas de variância serão agora dadas por

$$\gamma_j = \sigma^2 + \sum_{i=\nu+1}^u a_{i,j} \sigma_i^2; j = 1, \dots, w, \quad (20)$$

sendo os restantes parâmetros canónicos os

$$\eta_j^{g_j} = \mathbf{A}_j \mu_j^n \quad j = 1, \dots, w. \quad (21)$$

Observe-se que, sendo  $A_j$  a matriz cujos vectores linha são os vectores próprios iguais de  $Q_j$  correspondentes a valores próprios iguais a 1, se tem  $Q_j = A_j' A_j, j = 1, \dots, w$ .

Atendendo às expressões (18) e (19) vê-se que, quando as matrizes  $M, M_{v+1}, \dots, M_u$  comutam, o modelo é ortogonal.

As expressões (20) podem ser interpretadas como um sistema de equações nas  $\sigma_{v+1}^2, \dots, \sigma_u^2$ . Admitamos que esse sistema é resolúvel tendo uma solução única

$$\sigma_i^2 = \sum_{j=1}^{w+1} b_{i,j} \gamma_j; i = v + 1, \dots, u, \quad (22)$$

onde  $\gamma_{w+1} = \sigma^2$ . Têm-se então os UMVUE

$$\sigma_i^2 = \tilde{\sum}_{j=1}^{w+1} b_{i,j} \tilde{\gamma}_j; i = v + 1, \dots, u. \quad (23)$$

Surge agora o problema das probabilidades dum qualquer destes estimadores ser negativo não serem nulas. Classicamente (ver, por exemplo, [11]) existem duas posições:

1. toma-se  $\tilde{\sigma}_i^2 < 0$  como uma forte indicação de que  $\sigma_i^2 \approx 0$ ;
2. opta-se por outros métodos de estimação que garantam estimadores não negativos.

No entanto, estes últimos estimadores não têm justificação teórica quando se trabalha em pequenas amostras.

Ora, como vimos, estes modelos ao serem ortogonais podem receber formulação canónica. Dentro dessa formulação as expressões (23) podem ser interpretadas como definindo  $\sigma_i^2, i = v + 1, \dots, u$ , como combinações lineares das componentes de variância canónicas. Ao exigirmos que  $\sigma_i^2 > 0, i = v + 1, \dots, u$ , estamos a impor restrições à formulação canónica. Assim, a formulação factorial corresponde a uma particularização da formulação canónica. Resulta daqui que a inferência baseada na formulação canónica, menos restritiva, é mais robusta, o que nos leva a propor que se utilizem os  $\tilde{\sigma}_i^2, i = v + 1, \dots, u$ , e que, quando  $\tilde{\sigma}_i^2 \geq 0$ , se admita ter-se  $\sigma_i^2 \approx 0$ .

#### 4. Primeiro Exemplo de Formulação Factorial

Este modelo é uma variante do modelo em [34]. Um primeiro factor aninha um segundo e um terceiro factores que cruzam. Cada um dos  $a$  níveis do primeiro factor contem  $bc$

pares de níveis do segundo e terceiro factores. Haverá pois  $abc$  combinações de níveis para cada uma das quais se tomam  $r$  observações. Assim, ter-se-á  $n = abcr$  bem como

$$\mathbf{Y}^n = \mathbf{1}^n \mu + \mathbf{X}(1) \beta^a(1) + \mathbf{X}(2) \beta^{ab}(2) + \mathbf{X}(3) \beta^{ac}(3) + \mathbf{X}(2,3) \beta^{abc}(2,3) + \mathbf{e}^n, \quad (24)$$

onde os índices se referem a factores e o par (2, 3) à interacção entre o segundo e o terceiro factores. Como o primeiro factor aninha os outros dois, não participa em qualquer interacção. As matrizes do modelo são, representando por o produto de Kronecker de matrizes, dadas por

$$\begin{cases} \mathbf{X}(1) = \mathbf{I}_a \otimes \mathbf{1}^b \otimes \mathbf{1}^c \otimes \mathbf{1}^r \\ \mathbf{X}(2) = \mathbf{I}_a \otimes \mathbf{I}_b \otimes \mathbf{1}^c \otimes \mathbf{1}^r \\ \mathbf{X}(3) = \mathbf{I}_a \otimes \mathbf{1}^b \otimes \mathbf{I}_c \otimes \mathbf{1}^r \\ \mathbf{X}(2,3) = \mathbf{I}_a \otimes \mathbf{I}_b \otimes \mathbf{I}_c \otimes \mathbf{1}^r \end{cases} \quad (25)$$

Admitamos que  $\mu$  é fixo e que  $\beta^a \sim N(\mathbf{0}^a, \sigma^2(1)\mathbf{I}_a)$ ,  $\beta^{ab} \sim N(\mathbf{0}^{ab}, \sigma^2(2)\mathbf{I}_{ab})$ ,  $\beta^{ac} \sim N(\mathbf{0}^{ac}, \sigma^2(3)\mathbf{I}_{ac})$ ,  $\beta^{abc} \sim N(\mathbf{0}^{abc}, \sigma^2(2,3)\mathbf{I}_{abc})$  e  $\mathbf{e}^n \sim N(\mathbf{0}^n, \sigma^2\mathbf{I}_n)$ , sendo estes vectores independentes. Com  $\mathbf{J}_s = \mathbf{1}^s \mathbf{1}^{s'}$ , tem-se

$$\begin{cases} \mathbf{M}(1) = \mathbf{X}(1)\mathbf{X}'(1) = \mathbf{I}_a \otimes \mathbf{J}_b \otimes \mathbf{J}_c \otimes \mathbf{J}_r \\ \mathbf{M}(2) = \mathbf{X}(2)\mathbf{X}'(2) = \mathbf{I}_a \otimes \mathbf{I}_b \otimes \mathbf{J}_c \otimes \mathbf{J}_r \\ \mathbf{M}(3) = \mathbf{X}(3)\mathbf{X}'(3) = \mathbf{I}_a \otimes \mathbf{J}_b \otimes \mathbf{I}_c \otimes \mathbf{J}_r \\ \mathbf{M}(2,3) = \mathbf{X}(2,3)\mathbf{X}'(2,3) = \mathbf{I}_a \otimes \mathbf{I}_b \otimes \mathbf{I}_c \otimes \mathbf{J}_r \end{cases} \quad (26)$$

vendo-se que

$$\begin{cases} \mu^n = \mathbf{1}^n \mu \\ \mathbf{V} = \sigma^2(1)\mathbf{M}(1) + \sigma^2(2)\mathbf{M}(2) + \sigma^2(3)\mathbf{M}(3) + \sigma^2(2,3)\mathbf{M}(2,3) + \sigma^2\mathbf{I}_n \end{cases} \quad (27)$$

Como  $\mathbf{M} = \mathbf{J}_n$  permuta com as  $\mathbf{M}(1)$ ,  $\mathbf{M}(2)$ ,  $\mathbf{M}(3)$ ,  $\mathbf{M}(2,3)$ , o modelo é ortogonal.

As matrizes  $\frac{1}{s}\mathbf{J}_s$  e  $\bar{\mathbf{J}}_s = \mathbf{I} - \frac{1}{s}\mathbf{J}_s$  são matrizes simétricas e idempotentes e, portanto, matrizes de projecção ortogonal. É ainda fácil verificar que o produto de Kronecker de matrizes de projecção ortogonal dá matrizes de projecção ortogonal.

Vê-se ainda que as matrizes

$$\left\{ \begin{array}{l} \mathbf{Q}(0) = \frac{1}{n} \mathbf{J}_a \otimes \mathbf{J}_b \otimes \mathbf{J}_c \otimes \mathbf{J}_r = \frac{1}{n} \mathbf{J}_n \\ \mathbf{Q}(1) = \frac{1}{bcr} \bar{\mathbf{J}}_a \otimes \mathbf{J}_b \otimes \mathbf{J}_c \otimes \mathbf{J}_r \\ \mathbf{Q}(2) = \frac{1}{cr} \mathbf{I}_a \otimes \bar{\mathbf{J}}_b \otimes \mathbf{J}_c \otimes \mathbf{J}_r \\ \mathbf{Q}(3) = \frac{1}{br} \mathbf{I}_a \otimes \mathbf{J}_b \otimes \bar{\mathbf{J}}_c \otimes \mathbf{J}_r \\ \mathbf{Q}(2,3) = \frac{1}{r} \mathbf{I}_a \otimes \bar{\mathbf{J}}_b \otimes \bar{\mathbf{J}}_c \otimes \mathbf{J}_r \\ \mathbf{Q}^\perp = \mathbf{I}_n - \mathbf{Q}(0) - \mathbf{Q}(1) - \mathbf{Q}(2) - \mathbf{Q}(3) - \mathbf{Q}(2,3) \\ = \mathbf{I}_a \otimes \mathbf{I}_b \otimes \mathbf{I}_c \otimes \bar{\mathbf{J}}_r \end{array} \right. \quad (28)$$

constituem a base principal de uma álgebra de Jordan comutativa. Aliás,  $\mu^n$  pertence ao espaço imagem de  $\mathbf{Q}(0)$  e

$$\mathbf{V} = \gamma(0)\mathbf{Q}(0) + \gamma(1)\mathbf{Q}(1) + \gamma(2)\mathbf{Q}(2) + \gamma(3)\mathbf{Q}(3) + \gamma(2,3)\mathbf{Q}(2,3) + \sigma^2\mathbf{Q}^\perp, \quad (29)$$

com

$$\left\{ \begin{array}{l} \gamma(0) = \sigma^2 + r(bc\sigma^2(1) + c\sigma^2(2) + b\sigma^2(3) + \sigma^2(2,3)) \\ \gamma(1) = \sigma^2 + r(bc\sigma^2(1) + c\sigma^2(2) + b\sigma^2(3) + \sigma^2(2,3)) \\ \gamma(2) = \sigma^2 + r(c\sigma^2(2) + \sigma^2(2,3)) \\ \gamma(3) = \sigma^2 + r(b\sigma^2(3) + \sigma^2(2,3)) \\ \gamma(2,3) = \sigma^2 + r\sigma^2(2,3) \end{array} \right. \quad (30)$$

Obtém-se facilmente

$$\left\{ \begin{array}{l} \sigma^2(1) = \frac{1}{rbc} (\gamma(1) - \gamma(2) - \gamma(3) + \gamma(2,3)) \\ \sigma^2(2) = \frac{1}{rc} (\gamma(2) - \gamma(2,3)) \\ \sigma^2(3) = \frac{1}{rb} (\gamma(3) - \gamma(2,3)) \\ \sigma^2(2,3) = \frac{1}{r} (\gamma(2,3) - \sigma^2) \end{array} \right. \quad (31)$$

Uma matriz ortogonal de ordem  $s$  é estandardizada se o seu primeiro vector linha for  $\frac{1}{\sqrt{s}} \mathbf{1}^{s'}$ . Seja  $\mathbf{T}_s$  a matriz obtida retirando a primeira linha a uma matriz ortogonal estandardizada de ordem  $s$ .

Como

$$\frac{1}{s} \mathbf{J}_s + \bar{\mathbf{J}}_s = \mathbf{I}_s = \mathbf{P}'_s \mathbf{P}_s = \frac{1}{s} \mathbf{J}_s + \mathbf{T}'_s \mathbf{T}_s, \quad (32)$$

vê-se que

$$\bar{\mathbf{J}}_s = \mathbf{T}'_s \mathbf{T}_s. \quad (33)$$

Podemos pois tomar

$$\left\{ \begin{array}{l} \mathbf{A}(0) = \frac{1}{\sqrt{n}} \mathbf{1}^{n'} \\ \mathbf{A}(1) = \frac{1}{\sqrt{bcr}} \mathbf{T}_a \otimes \mathbf{1}^{b'} \otimes \mathbf{1}^{c'} \otimes \mathbf{1}^{r'} \\ \mathbf{A}(2) = \frac{1}{\sqrt{cr}} \mathbf{I}_a \otimes \mathbf{T}_b \otimes \mathbf{1}^{c'} \otimes \mathbf{1}^{r'} \\ \mathbf{A}(3) = \frac{1}{\sqrt{br}} \mathbf{I}_a \otimes \mathbf{1}^{b'} \otimes \mathbf{T}_c \otimes \mathbf{1}^{r'} \\ \mathbf{A}(2,3) = \frac{1}{\sqrt{r}} \mathbf{I}_a \otimes \mathbf{T}_b \otimes \mathbf{T}_c \otimes \mathbf{1}^{r'} \\ \mathbf{A}^\perp = \mathbf{I}_a \otimes \mathbf{I}_b \otimes \mathbf{I}_c \otimes \mathbf{T}_r \end{array} \right. \quad (34)$$

para obter as somas de quadrados

$$\left\{ \begin{array}{l} S(1) = \|\mathbf{A}(1)\mathbf{Y}^n\|^2 \sim \gamma(1)\chi_{(g(1))}^2 \\ S(2) = \|\mathbf{A}(2)\mathbf{Y}^n\|^2 \sim \gamma(2)\chi_{(g(2))}^2 \\ S(3) = \|\mathbf{A}(3)\mathbf{Y}^n\|^2 \sim \gamma(3)\chi_{(g(3))}^2 \\ S(2,3) = \|\mathbf{A}(2,3)\mathbf{Y}^n\|^2 \sim \gamma(2,3)\chi_{(g(2,3))}^2 \\ S^\perp = \|\mathbf{A}^\perp\mathbf{Y}^n\|^2 \sim \sigma^2 \chi_{(g^\perp)}^2 \end{array} \right. \quad (35)$$

Como a característica de produto de Kronecker é o produto das características de Kronecker, vem  $g(1) = a - 1$ ,  $g(2) = a(b - 1)$ ,  $g(3) = a(c - 1)$ ,  $g(2, 3) = a(b - 1)(c - 1)$  e  $g^\perp = abc(r - 1)$ . Assim, para as componentes de variância canônicas têm-se os UMVUE

$$\tilde{\gamma}(1) = \frac{S(1)}{g(1)}, \tilde{\gamma}(2) = \frac{S(2)}{g(2)}, \tilde{\gamma}(3) = \frac{S(3)}{g(3)}, \tilde{\gamma}(2,3) = \frac{S(2,3)}{g(2,3)} \text{ e } \tilde{\sigma}^2 = \frac{S^\perp}{g^\perp}. \text{ A partir destes obtêm-se os}$$

$$\left\{ \begin{array}{l} \tilde{\sigma}^2(1) = \frac{1}{rbc} (\tilde{\gamma}(1) - \tilde{\gamma}(2) - \tilde{\gamma}(3) + \tilde{\gamma}(2,3)) \\ \tilde{\sigma}^2(2) = \frac{1}{rc} (\tilde{\gamma}(2) - \tilde{\gamma}(2,3)) \\ \tilde{\sigma}^2(3) = \frac{1}{rb} (\tilde{\gamma}(3) - \tilde{\gamma}(2,3)) \\ \tilde{\sigma}^2(2,3) = \frac{1}{r} (\tilde{\gamma}(2,3) - \tilde{\sigma}^2) \end{array} \right. \quad (36)$$

Sendo  $\bar{F}(z | s, l)$  a distribuição do quociente  $\frac{\chi^2(s)}{\chi^2(l)}$  de qui-quadrados centrais independentes tem-se, dado  $\tilde{\gamma}(2) = \frac{S(2)}{g(2)}$  e  $\tilde{\gamma}(2,3) = \frac{S(2,3)}{g(2,3)}$ ,

$$P[\tilde{\sigma}^2(2) < 0] = P\left[\frac{\chi^2_{(g(2))}}{\chi^2_{(g(2,3))}} < \frac{\gamma(2,3)}{\gamma(2)}\right] = \bar{F}\left(\frac{\gamma(2,3)}{\gamma(2)} |_{g(2), g(2,3)}\right), \quad (37)$$

obtendo-se analogamente

$$\begin{cases} P[\tilde{\sigma}^2(3) < 0] = \bar{F}\left(\frac{\gamma(2,3)}{\gamma(3)} |_{g(3), g(2,3)}\right) \\ P[\tilde{\sigma}^2(2,3) < 0] = \bar{F}\left(\frac{\sigma^2}{\gamma(2,3)} |_{g(2,3), g^\perp}\right) \end{cases} \quad (38)$$

Vê-se que estas probabilidades dependem dos quocientes  $\frac{\gamma(2,3)}{\gamma(2,3)}$ ,  $\frac{\gamma(2,3)}{\gamma(3)}$  e  $\frac{\gamma(2,3)}{\sigma^2}$ , tornando-se estas desprezáveis quando os mesmos são aproximadamente nulos. Aliás, no âmbito da formulação canónica, nada impede que uma ou mais das componentes de variância (factoriais) sejam negativas. Adiante desenvolveremos este ponto de vista através da obtenção de intervalos de confiança para estas componentes.

### 5. Segundo Exemplo de Formulação Factorial

Neste segundo exemplo, um primeiro factor de efeitos aleatórios cruza com um segundo factor de efeitos fixos que aninha um terceiro factor de efeitos aleatórios. Haverá  $a$  e  $b$  níveis para os dois primeiros factores e, para cada nível do segundo factor,  $c$  níveis do terceiro factor. Para cada uma das  $abc$  combinações de níveis tomam-se  $r$  observações. Assim, ter-se-á um total de  $n = abcr$  observações, podendo escrever-se o modelo como:

$$\mathbf{Y}^n = \mathbf{1}^n \mu + \mathbf{X}(1) \beta^a(1) + \mathbf{X}(2) \beta^b(2) + \mathbf{X}(1, 2) \beta^{ab}(1, 2) + \mathbf{X}(3) + \beta^{bc}(3) + \mathbf{X}(1, 3) \beta^{abc}(1, 3) + \mathbf{e}^n. \quad (39)$$

Agora  $\mu$  e  $\beta^a(1)$  são fixos, sendo as componentes de  $\beta^a(1)$  os efeitos do segundo factor. Além disso, para os efeitos do primeiro e terceiro factores, para as interacções entre o primeiro e os outros dois factores e para o erro temos os vectores independentes  $\beta^a(1) \sim N(\mathbf{0}^a, \sigma^2(1)\mathbf{I}_a)$ ,  $\beta^{ab}(1, 2) \sim N(\mathbf{0}^{ab}, \sigma^2(1, 2)\mathbf{I}_{ab})$ ,  $\beta^{bc}(3) \sim N(\mathbf{0}^{bc}, \sigma^2(3)\mathbf{I}_{bc})$ ,  $\beta^{abc}(1, 3) \sim N(\mathbf{0}^{abc}, \sigma^2(1, 3)\mathbf{I}_{bc})$  e  $\mathbf{e}^n \sim N(\mathbf{0}^n, \sigma^2\mathbf{I}_n)$ .

As matrizes que figuram no modelo são dadas por:

$$\left\{ \begin{array}{l} \mathbf{X}(1) = \mathbf{I}_a \otimes \mathbf{1}^b \otimes \mathbf{1}^c \otimes \mathbf{1}^r \\ \mathbf{X}(2) = \mathbf{1}_a \otimes \mathbf{I}_b \otimes \mathbf{1}^c \otimes \mathbf{1}^r \\ \mathbf{X}(1,2) = \mathbf{I}_a \otimes \mathbf{I}_b \otimes \mathbf{1}^c \otimes \mathbf{1}^r . \\ \mathbf{X}(3) = \mathbf{1}_a \otimes \mathbf{1}^b \otimes \mathbf{I}_c \otimes \mathbf{1}^r \\ \mathbf{X}(1,3) = \mathbf{I}_a \otimes \mathbf{I}_b \otimes \mathbf{I}_c \otimes \mathbf{1}^r \end{array} \right. \quad (40)$$

Assim, o vector médio e a matriz de covariância de  $\mathbf{Y}^n$  são

$$\left\{ \begin{array}{l} \boldsymbol{\mu}^n = \mathbf{1}^n \boldsymbol{\mu} + \mathbf{X}(2) \boldsymbol{\beta}^b(2) \\ \mathbf{V} = \sigma^2(1) \mathbf{M}(1) + \sigma^2(1,2) \mathbf{M}(1,2) + \sigma^2(3) \mathbf{M}(3) + \sigma^2(1,3) \mathbf{M}(1,3) + \sigma^2 \mathbf{I}_n \end{array} \right. , \quad (41)$$

tendo-se

$$\left\{ \begin{array}{l} \mathbf{M}(1) = \mathbf{X}(1) \mathbf{X}'(1) = \mathbf{I}_a \otimes \mathbf{J}_b \otimes \mathbf{J}_c \otimes \mathbf{J}_r \\ \mathbf{M}(2) = \mathbf{X}(2) \mathbf{X}'(2) = \mathbf{J}_a \otimes \mathbf{I}_b \otimes \mathbf{J}_c \otimes \mathbf{J}_r \\ \mathbf{M}(1,2) = \mathbf{X}(1,2) \mathbf{X}'(1,2) = \mathbf{I}_a \otimes \mathbf{I}_b \otimes \mathbf{J}_c \otimes \mathbf{J}_r . \\ \mathbf{M}(3) = \mathbf{X}(3) \mathbf{X}'(3) = \mathbf{J}_a \otimes \mathbf{I}_b \otimes \mathbf{I}_c \otimes \mathbf{J}_r \\ \mathbf{M}(1,3) = \mathbf{X}(1,3) \mathbf{X}'(1,3) = \mathbf{I}_a \otimes \mathbf{I}_b \otimes \mathbf{I}_c \otimes \mathbf{J}_r \end{array} \right. \quad (42)$$

Como  $\boldsymbol{\mu}^n$  pertence ao espaço imagem de  $\mathbf{X}(2)$  estas matrizes comutam, o modelo é ortogonal. As matrizes da base principal da álgebra de Jordan são

$$\left\{ \begin{array}{l} \mathbf{Q}(0) = \frac{1}{abc} \mathbf{J}_a \otimes \mathbf{J}_b \otimes \mathbf{J}_c \otimes \mathbf{J}_r = \mathbf{A}'(0) \mathbf{A}(0) \\ \mathbf{Q}(1) = \frac{1}{bcr} \mathbf{J}_a \otimes \mathbf{J}_b \otimes \mathbf{J}_c \otimes \mathbf{J}_r = \mathbf{A}'(1) \mathbf{A}(1) \\ \mathbf{Q}(2) = \frac{1}{acr} \mathbf{J}_a \otimes \bar{\mathbf{J}}_b \otimes \mathbf{J}_c \otimes \mathbf{J}_r = \mathbf{A}'(2) \mathbf{A}(2) \\ \mathbf{Q}(1,2) = \frac{1}{cr} \bar{\mathbf{J}}_a \otimes \bar{\mathbf{J}}_b \otimes \mathbf{J}_c \otimes \mathbf{J}_r = \mathbf{A}'(1,2) \mathbf{A}(1,2) \\ \mathbf{Q}(3) = \frac{1}{ar} \mathbf{J}_a \otimes \mathbf{I}_b \otimes \bar{\mathbf{J}}_c \otimes \mathbf{J}_r = \mathbf{A}'(3) \mathbf{A}(3) \\ \mathbf{Q}(1,3) = \frac{1}{r} \mathbf{J}_a \otimes \mathbf{I}_b \otimes \bar{\mathbf{J}}_c \otimes \mathbf{J}_r = \mathbf{A}'(1,3) \mathbf{A}(1,3) \\ \mathbf{Q}^\perp = \mathbf{I}_n - \mathbf{Q}(0) - \mathbf{Q}(1) - \mathbf{Q}(2) - \mathbf{Q}(1,2) - \mathbf{Q}(3) - \mathbf{Q}(1,3) \\ = \mathbf{I}_a \otimes \mathbf{I}_b \otimes \mathbf{I}_c \otimes \bar{\mathbf{J}}_r = \mathbf{A}'^\perp \mathbf{A}^\perp \end{array} \right. , \quad (43)$$

com

$$\left\{ \begin{array}{l} \mathbf{A}(0) = \frac{1}{\sqrt{abc}} \mathbf{1}^{a'} \otimes \mathbf{1}^{b'} \otimes \mathbf{1}^{c'} \otimes \mathbf{1}^{r'} \\ \mathbf{A}(1) = \frac{1}{\sqrt{bcr}} \mathbf{T}_a \otimes \mathbf{1}^{b'} \otimes \mathbf{1}^{c'} \otimes \mathbf{1}^{r'} \\ \mathbf{A}(2) = \frac{1}{\sqrt{acr}} \mathbf{1}^{a'} \otimes \mathbf{T}_b \otimes \mathbf{1}^{c'} \otimes \mathbf{1}^{r'} \\ \mathbf{A}(1,2) = \frac{1}{\sqrt{cr}} \mathbf{T}_a \otimes \mathbf{T}_b \otimes \mathbf{1}^c \otimes \mathbf{1}^{r'} \\ \mathbf{A}(3) = \frac{1}{\sqrt{ar}} \mathbf{1}^{a'} \otimes \mathbf{1}^b \otimes \mathbf{T}_c \otimes \mathbf{1}^{r'} \\ \mathbf{A}(1,3) = \frac{1}{\sqrt{r}} \mathbf{T}_a \otimes \mathbf{I}_b \otimes \mathbf{T}_c \otimes \mathbf{1}^{r'} \\ \mathbf{A}^\perp = \mathbf{I}_a \otimes \mathbf{I}_b \otimes \mathbf{I}_c \otimes \mathbf{T}_r \end{array} \right. \quad (44)$$

As características destas últimas matrizes são, respectivamente,  $g(0) = 1$ ,  $g(1) = a - 1$ ,  $g(2) = b - 1$ ,  $g(1, 2) = (a - 1)(b - 1)$ ,  $g(3) = b(c - 1)$ ,  $g(1, 3) = (a - 1) b(c - 1)$  e  $g^\perp = abc(r - 1)$ .

Como

$$\left\{ \begin{array}{l} \mathbf{M}(1) = bcr(\mathbf{Q}(0) + \mathbf{Q}(1)) \\ \mathbf{M}(1,2) = cr(\mathbf{Q}(0) + \mathbf{Q}(1) + \mathbf{Q}(1,2) + \mathbf{Q}(2)) \\ \mathbf{M}(3) = ar(\mathbf{Q}(0) + \mathbf{Q}(2) + \mathbf{Q}(3)) \\ \mathbf{M}(1,3) = r(\mathbf{Q}(0) + \mathbf{Q}(1) + \mathbf{Q}(1,2) + \mathbf{Q}(2) + \mathbf{Q}(3) + \mathbf{Q}(1,3)) \end{array} \right. \quad (45)$$

tem-se

$$\begin{aligned} \mathbf{V} = & bcr\sigma^2(1) (\mathbf{Q}(0) + \mathbf{Q}(1)) + cr\sigma^2(1, 2) (\mathbf{Q}(0) + \mathbf{Q}(1) + \mathbf{Q}(1, 2) + \mathbf{Q}(2)) \\ & + ar\sigma^2(3) (\mathbf{Q}(0) + \mathbf{Q}(2) + \mathbf{Q}(3)) \\ & + r\sigma^2(1, 3) (\mathbf{Q}(0) + \mathbf{Q}(1) + \mathbf{Q}(1, 2) + \mathbf{Q}(2) + \mathbf{Q}(3) + \mathbf{Q}(1, 3)) \\ & + \sigma^2 (\mathbf{Q}(0) + \mathbf{Q}(1) + \mathbf{Q}(1, 2) + \mathbf{Q}(2) + \mathbf{Q}(3) + \mathbf{Q}(1, 3) + \mathbf{Q}^\perp). \end{aligned} \quad (46)$$

Esta matriz pode ser reescrita na forma

$$\begin{aligned} \mathbf{V} = & \gamma(0)\mathbf{Q}(0) + \gamma(1)\mathbf{Q}(1) + \gamma(1, 2)\mathbf{Q}(1, 2) + \gamma(2)\mathbf{Q}(2) + \gamma(3)\mathbf{Q}(3) \\ & + \gamma(1, 3)\mathbf{Q}(1, 3) + \sigma^2\mathbf{Q}^\perp, \end{aligned} \quad (47)$$

com

$$\begin{cases} \gamma(0) = \sigma^2 + bcr\sigma^2(1) + cr\sigma^2(1,2) + ar\sigma^2(3) + r\sigma^2(1,3) \\ \gamma(1) = \sigma^2 + bcr\sigma^2(1) + cr\sigma^2(1,2) + r\sigma^2(1,3) \\ \gamma(2) = \sigma^2 + cr\sigma^2(1,2) + ar\sigma^2(3) + r\sigma^2(1,3) \\ \gamma(1,2) = \sigma^2 + cr\sigma^2(1,2) + r\sigma^2(1,3) \\ \gamma(3) = \sigma^2 + ar\sigma^2(3) + r\sigma^2(1,3) \\ \gamma(1,3) = \sigma^2 + r\sigma^2(1,3) \end{cases} \quad (48)$$

Logo,

$$\begin{cases} \sigma^2(1) = \frac{1}{bcr} (\gamma(1) - \gamma(1,2)) \\ \sigma^2(1,2) = \frac{1}{cr} (\gamma(1,2) - \gamma(1,3)) \\ \sigma^2(3) = \frac{1}{ar} (\gamma(3) - \gamma(1,3)) \\ \sigma^2(1,3) = \frac{1}{r} (\gamma(1,3) - \sigma^2) \end{cases} \quad (49)$$

Ora,

$$\begin{cases} S(1) = \|\mathbf{A}(1)\mathbf{Y}^n\|^2 \sim \gamma(1)\chi_{(g(1))}^2 \\ S(2) = \|\mathbf{A}(2)\mathbf{Y}^n\|^2 \sim \gamma(2)\chi_{(g(2))}^2 \\ S(1,2) = \|\mathbf{A}(1,2)\mathbf{Y}^n\|^2 \sim \gamma(1,2)\chi_{(g(1,2))}^2 \\ S(3) = \|\mathbf{A}(3)\mathbf{Y}^n\|^2 \sim \gamma(3)\chi_{(g(3))}^2 \\ S(1,3) = \|\mathbf{A}(1,3)\mathbf{Y}^n\|^2 \sim \gamma(1,3)\chi_{(g(1,3))}^2 \\ S^\perp = \|\mathbf{A}^\perp\mathbf{Y}^n\|^2 \sim \sigma^2\chi_{(g^\perp)}^2 \end{cases} \quad (50)$$

obtendo-se os UMVUE

$$\begin{cases} \tilde{\sigma}^2(1) = \frac{1}{bcr} \left( \frac{S(1)}{g(1)} - \frac{S(1,2)}{g(1,2)} \right) \\ \tilde{\sigma}^2(1,2) = \frac{1}{cr} \left( \frac{S(1,2)}{g(1,2)} - \frac{S(1,3)}{g(1,3)} \right) \\ \tilde{\sigma}^2(3) = \frac{1}{ar} \left( \frac{S(3)}{g(3)} - \frac{S(1,3)}{g(1,3)} \right) \\ \tilde{\sigma}^2(1,3) = \frac{1}{r} \left( \frac{S(1,3)}{g(1,3)} - \frac{S^\perp}{g^\perp} \right) \end{cases} \quad (51)$$

pelo que

$$\begin{cases} P[\hat{\sigma}^2(1) < 0] = \bar{F}\left(\frac{\gamma(1)}{\gamma(1,2)} \mid g(1), g(1,2)\right) \\ P[\hat{\sigma}^2(1,2) < 0] = \bar{F}\left(\frac{\gamma(1,2)}{\gamma(1,3)} \mid g(1,2), g(1,3)\right) \\ P[\hat{\sigma}^2(3) < 0] = \bar{F}\left(\frac{\gamma(2,3)}{\gamma(3)} \mid g(3), g(2,3)\right) \\ P[\hat{\sigma}^2(1,3) < 0] = \bar{F}\left(\frac{\sigma^2}{\gamma(1,3)} \mid g(1,3), g^\perp\right) \end{cases} \quad (52)$$

Existe um tratamento (ver [3]) para um modelo muito semelhante a este, em que o segundo factor é tratado como tendo efeitos aleatórios. O presente tratamento deve-se a [5]. O tratamento dado por [4] foi posteriormente generalizado em [5]) para o caso de  $L$  grupos de factores aninhados que cruzam. Dentro de cada grupo os factores são de efeitos fixos ou aleatórios.

## 6. Variáveis de Teste Generalizadas

Até este ponto a exposição apresentada assenta nos trabalhos de Seely já referidos. Entramos agora na segunda fase do nosso trabalho.

Dado:

1. o vector  $\mathbf{V}^k$  das estatísticas suficientes;
2. o vector  $\mathbf{v}^k$  dos valores tomados pelas estatísticas suficientes;
3. o vector  $\lambda^l$  dos parâmetros perturbadores;
4. o parâmetro  $\theta$  que figura nas hipóteses;

tendo-se, em geral,

$$H_0 : \theta = \theta_0 \quad (53)$$

contra

$$H_1 : \theta > \theta_0, \quad (54)$$

uma estatística de teste será (ver [31]) uma função

$$T(\mathbf{V}^k, \mathbf{v}^k, \lambda^l, \theta) \quad (55)$$

tal que:

1.  $t(\mathbf{v}^k; \mathbf{v}^k; \lambda^l; \theta)$  não depende de valores desconhecidos;
2. a distribuição de  $T$ , dado  $\theta$ , não depende de  $\lambda^l$ ;
3. a distribuição de  $T$ , dado  $\mathbf{v}^k$  e  $\lambda^l$ , é função monótona de  $\theta$ .

Sendo a distribuição de  $T$  uma função não decrescente de  $\theta$ , tomemos

$$C = \{\mathbf{V}^k : T(\mathbf{V}^k, \mathbf{v}^k, \lambda^l, \theta) > t(\mathbf{v}^k, \mathbf{v}^k, \lambda^l, \theta)\}, \quad (56)$$

invertendo-se o sinal da desigualdade se a distribuição for não crescente em  $\theta$ .

Então o teste com estatística generalizada  $T$  tem  $p$ -value generalizado

$$p = \sup_{\theta \leq \theta_0} P[\mathbf{V}^k \in C]. \quad (57)$$

Fixado o nível  $q$  de teste, rejeita-se  $H_0$  quando

$$P_{\theta_0}[\mathbf{V}^k \in C] < q. \quad (58)$$

Suponhamos que após uma reordenação conveniente das componentes de variância canónicas  $(\gamma_{m+1}, \dots, \gamma_w)$  se tem

$$\theta = \sum_{j=1}^r b_j \gamma_j - \sum_{j=r+1}^{2r} b_j \gamma_j. \quad (59)$$

É agora fácil de mostrar que

$$T = \sum_{j=1}^r b_j \gamma_j \frac{s_j}{s_j} - \sum_{j=r+1}^{2r} b_j \gamma_j \frac{s_j}{s_j} - \theta \quad (60)$$

é variável de teste generalizada. Como  $S_j \sim \gamma_j X_{(g_j)}^2$ ,  $j = 1, \dots, 2r$ , ter-se-á

$$p = P \left[ \sum_{j=1}^r b_j \frac{S_j}{\chi_{(g_j)}^2} > \sum_{j=r+1}^{2r} b_j \frac{S_j}{\chi_{(g_j)}^2} \right], \quad (61)$$

visto

$$t(s^{2r}, s^{2r}, \gamma^{2r}, \theta) = 0. \quad (62)$$

Observe-se que nos exemplos apresentados as componentes de variância factoriais se enquadram no caso agora estudado.

## 7. Medidas de Probabilidade Induzidas

Continuando a utilizar a mesma ordenação das componentes de variância canónicas, temos  $S_j \sim \gamma_j X_{(g_j)}^2$ ,  $j = 1, \dots, 2r$ , bem como a probabilidade de qualquer intervalo fechado conter qualquer uma das  $\gamma_j$ ,  $j = 1, \dots, 2r$ .

Como a família dos intervalos fechados (completada pelo conjunto vazio) e fechada para a intersecção finita, a mesma constitui um sistema- $\Pi$ . Assim (ver [32]), as probabilidades de cobertura atrás referidas definem univocamente medidas de probabilidade  $P_j, j = 1, \dots, 2r$ . Segundo [1], diremos que estas medidas de probabilidade foram induzidas para os parâmetros  $\gamma_j, j = 1, \dots, 2r$ . Sendo  $F_j$  a distribuição associada a  $P_j$  e  $x_{g_j, 1-q}$ , o quantil, para a probabilidade  $1 - q$ , de  $\chi_{(g_j)}^2, j = 1, \dots, 2r$ , tem-se

$$F_j \left( \frac{s_j}{x_{g_j, 1-q}} \right) = q; j = 1, \dots, 2r. \tag{63}$$

Podem-se gerar amostras de dimensão arbitrariamente grande com estas distribuições. Assim, sendo  $W_{j,l}$  o quociente de  $s_j$  por um qui-quadrado com  $g_j$  graus de liberdade,  $F_j$  será a distribuição  $\{W_{j,l}, l = 1, \dots, n\}, j = 1, \dots, 2r$ .

Como os  $S_1, \dots, S_{2r}$  são independentes e  $\theta$  é dado por uma função mensurável dos  $\gamma_1, \dots, \gamma_{2r}$ , a mesma induz, a partir das medidas de probabilidade  $\mathbb{P}_1, \dots, \mathbb{P}_{2r}$ , uma medida de probabilidade  $\mathbb{P}^*$  para o parâmetro  $\theta$ . A distribuição  $F^*$  associada a  $\mathbb{P}^*$  será ainda a distribuição dos

$$U_l = \sum_{j=1}^r b_j w_{j,l} - \sum_{j=r+1}^{2r} b_j w_{j,l}; l = 1, \dots, n, \tag{64}$$

visto estas variáveis serem dadas por uma função mensurável das anteriores, que são independentes. Reencontram-se assim os resultados da alínea anterior. A vantagem desta abordagem está na fundamentação obtida em termos da teoria da medida. Consegue-se ainda utilizar métodos de Monte Carlo para calcular os  $p$ -values.

Seja agora  $\varphi_i^s = \mathbf{B}_i \eta_i^{s_i}$  um dos vectores estimáveis atrás introduzidos. Tendo-se para o mesmo o UMVUE  $\varphi_i^s \sim N(\varphi_i^s, \gamma_i \mathbf{B}_i \mathbf{B}_i')$  bem como (ver [15])

$$(\varphi_i^s - \tilde{\varphi}_i^s)' (\mathbf{B}_i \mathbf{B}_i') (\varphi_i^s - \tilde{\varphi}_i^s) \sim \gamma_i \chi_{(g_i)}^2; i = 1, \dots, m. \tag{65}$$

Continuando a admitir que  $\gamma_i = \sum_{j=1}^m a_{i,j} \gamma_j, i = 1, \dots, m$ , podemos proceder como anteriormente para obter amostras  $\{V_l, l = 1, \dots, n\}$  de dimensão arbitrariamente grande com distribuição associada a uma medida de probabilidade induzida por  $\gamma_i, i = 1, \dots, m$ . Sendo  $d_{n, 1-q}$  o quantil empírico, para a probabilidade  $1 - q$ , obtido a partir dessa amostra de dimensão  $n$  ter-se-á, devido ao teorema de Glivenko-Cantelli reverso (ver [5])

$$P \left[ (\varphi_i^s - \tilde{\varphi}_i^s)' (\mathbf{B}_i \mathbf{B}_i') (\varphi_i^s - \tilde{\varphi}_i^s) \leq d_{n, 1-q} \right] \xrightarrow[n \rightarrow \infty]{} 1 - q; i = 1, \dots, m, \tag{66}$$

obtendo-se assim elipsóides de confiança de nível  $1 - q$  para  $\varphi_i^s, i = 1, \dots, m$ .

Através da dualidade obtém-se testes de nível limite  $q$  para as hipóteses

$$H_{0,i}(\varphi_{0,i}^s) : \varphi_i^s = \varphi_{0,i}^s; i = 1, \dots, m. \quad (67)$$

Estes testes rejeitam  $H_{0,i}(\varphi_{0,i}^s)$  se e só se  $\varphi_{0,i}^s$  não pertencer ao correspondente elipsóide de confiança de nível  $1 - q$ . Além disso, tem-se (ver [19])

$$P \left[ \bigcap_{c^s \in R^s} (c^s \varphi_i^s - c^s \tilde{\varphi}_i^s) \leq \sqrt{d_{n,1-q} c^s \mathbf{B}_i \mathbf{B}_i' c^s} \right] \xrightarrow{n \rightarrow \infty} 1 - q, \quad (68)$$

obtendo-se assim intervalos de confiança simultâneos, com nível limite  $1 - q$ , para todas as combinações lineares das componentes de  $\varphi_i^s$ ,  $i = 1, \dots, m$ .

## Referências

- [1] Covas R. (2003). *Inferência Semi-Bayesiana e Modelos de Componentes de Variância*. Master's Degree Thesis in Statistics and Optimization. FCT/UNL.
- [2] Fonseca M., Mexia J.T., Zmyslony R. (2002). Exact Distributions for the Generalized  $F$  Statistic. *Discussiones Mathematicae – Probability and Statistics* **22**, 37–51.
- [3] Fonseca M., Mexia J.T., Zmyslony R. (2003). Estimators and Tests for Variance Components in Cross Nested Orthogonal Models. *Discussiones Mathematicae – Probability and Statistics* **23**, 173–201.
- [4] Fonseca M., Mexia J.T., Zmyslony R. (2003). Estimating and Testing of Variance Components: an Application to a Grapevine Experiment. *Biometrical Letters* **40**, 1–7.
- [5] Fonseca M., Mexia J.T., Zmyslony R. (2003). Jordan Algebras, Generating Pivot Variables and Orthogonal Normal Models. *Unpublished*.
- [6] Garcia J. (2002). *Laplace and Fourier Transforms in Risk Theory*. Ph.D. Thesis. Lisbon Technical Univ.
- [7] Jordan P., von Neumann J., Wigner E.P. (1934). On an algebraic generalization of the quantum mechanical formalism. *Ann. Math. II. Ser.* **35**, 29–64.
- [8] Kallenberg O. (1997). *Foundations of Modern Probability*. Springer.
- [9] Khuri A.I., Mathew T., Sinha B.K. (1997). *Statistical Tests for Mixed Linear Models*. Wiley.
- [10] Krishnamoorthy K., Mathew T. (2004). One-Sided Tolerance Limits in Balanced and Unbalanced One-Way Random Models based in Generalized Confidence Intervals. To be published in *JASA*.
- [11] Lehmann E.L., Casella G. (1998). *Theory of point Estimation. 2nd. Ed.*. Springer.
- [12] Loève M. (1960). *Probability Theory. 2nd. Ed.*. D. Van Nostrand.

- [13] Malley J.D. (1994). *Statistical applications of Jordan algebras*. Springer-Verlag.
- [14] Mexia J.T. (1989). Controlled Heteroscedasticity, Quotient Vector Spaces and F Tests for Hypothesis on Mean Vectors. *Trabalhos de Investigação* **1**. FCT/UNL.
- [15] Mexia J.T. (1995). *Introdução à Inferência Estatística Linear*. Ed. Lusófonas.
- [16] Michalski A., Zmyslony R. (1996). Testing Hypothesis for Variance Components in Mixed Linear Models. *Statistics* **27**, 297–310.
- [17] Michalski A., Zmyslony R. (1999). Testing Hypothesis for Linear Functions of Parameters in Mixed Linear Models. *Tatra Mt. Math. Publ.* **17**, 103–110.
- [18] Nunes C. (2004). *F Tests and Selective F Tests in Orthogonal Mixed Linear Models*. Ph.D. Thesis. Beira Interior Univ.
- [19] Scheffé H. (1959). *The Analysis of Variance*. John Wiley & Sons.
- [20] Schott J.R. (1997). *Matrix analysis for statistics*. Wiley.
- [21] Searle R.S., Casella G., McCulloch C.E. (1992). *Variance Components*. John Wiley & Sons.
- [22] Seber G.A.F. (1980). *The Linear Hypothesis: a General Theory*. 2nd. Ed.. Charles Griffith & Co.
- [23] Seely J. (1970). Linear Spaces and Unbiased Estimation. *Ann. Math. Stat.* **41**, 1725–1734.
- [24] Seely J. (1970). Linear Spaces and Unbiased Estimation. An Application to the Mixed Linear Model. *Ann. Math. Stat.* **41**, 1735–1748.
- [25] Seely J., Zyskind G. (1971). Quadratic Subspaces and Completeness. *Ann. Math. Stat.* **42**, 691–703.
- [26] Seely J., Zyskind G. (1971). Linear Spaces and Minimum Variance Unbiased Estimation. *Ann. Math. Stat.* **42**, 691–703.
- [27] Seely J. (1972). Completeness for a family of multivariate normal distributions. *Ann. Math. Stat.* **43**, 1644–1647.
- [28] Severini T.A. (2000). *Likelihood Methods in Statistics*. Oxford University Press.
- [29] Silvey S.D. (1975). *Statistical Inference*. Reprinted. Chapman & Hall.
- [30] Weerahandi S. (1993). Generalized Confidence Intervals. *J. Am. Stat. Assoc.* **88**, 899–905.
- [31] Weerahandi S. (1996). *Exact Statistical Methods for Data Analysis*. 2nd. Print. Springer-Verlag.
- [32] Williams D. (1991). *Probability with Martingales*. Cambridge Univ. Press.
- [33] Witkovsky V. (2001). Computing the distribution of a linear combination of inverted gamma variables. *Kybernetika* **37**, 79–90.
- [34] Zhou, L., Mathew T. (1994). Some tests for variance components using generalized  $p$  values. *Technometrics* **36**, 394–402.

## **João Tiago Mexia**

Teve duas fases na sua carreira, a primeira das quais no Instituto de Investigação Científica Tropical em que, fundamentalmente, delineava e analisava experiências agrícolas. Esta fase terminou em 1982 quando se auto-propôs a Doutoramento na Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa na especialidade de Estatística. Transferiu-se então para essa faculdade, tendo obtido a Agregação em 1992 e a Cátedra em 1994. A sua actividade exerce-se sobretudo no âmbito da Inferência Estatística Linear, em que procura soluções teóricas para os problemas que encontrou na primeira fase. Privilegia o contacto com os estudantes de Mestrado e Doutoramento, realizando grande parte da sua actividade em colaboração com ex e actuais alunos.

Desde 1999 é director do Centro de Matemática e Aplicações, o que lhe tem permitido apoiar a formação dos Assistentes do Departamento de Matemática da Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa.

# Estabilidade e comportamento limite forte das estatísticas de ordem

Margarida Brito

## 1. Introdução

Vou considerar neste trabalho de revisão o tema em que iniciei a investigação, o estudo dos limites superiores e inferiores das  $k_n$ -ésimas estatísticas de ordem normalizadas e a consequente procura de critérios de majoração e de minoração assintóticas com probabilidade um. Este tipo de critérios conduziu-me ao estudo da estabilidade quase certa, questão de particular importância no estudo do comportamento limite forte das estatísticas de ordem e à qual será dada aqui especial atenção.

Lembra-se que uma sucessão de variáveis aleatórias (v.a.)  $W_1, W_2, \dots$  se diz (absolutamente) **estável** ou **relativamente estável** se existem sucessões  $(a_n)$  e  $(b_n)$ ,  $b_n \neq 0$ , tais que  $W_n - a_n \rightarrow 0$  ou  $W_n / b_n \rightarrow 1$  em probabilidade, respectivamente. Se a convergência é quase certa a sucessão diz-se **estável quase certamente** (q.c.) ou **relativamente estável q.c.**, respectivamente. Por analogia com o estudo das somas de variáveis aleatórias, também se diz que  $(W_n, a_n)$  verifica a lei fraca aditiva dos grandes números (AWL) no primeiro caso e que  $(W_n, b_n)$  verifica a lei fraca multiplicativa dos grandes números (MWL) no segundo caso. Como é usual, as leis fortes, designadas respectivamente por ASL e MSL, correspondem à convergência quase certa.

Seja  $X_1, X_2, \dots$  uma sucessão de variáveis aleatórias independentes e identicamente distribuídas (i.i.d.) com função de distribuição (f.d.)  $F$  e representem-se por  $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$  as estatísticas de ordem correspondentes à amostra  $X_1, X_2, \dots, X_n$ . Como é tradicional neste tema, vamos considerar aqui a estabilidade das  $k$ -ésimas estatísticas de ordem superior. Obviamente, os resultados apresentados, após modificações triviais continuam válidos para as sucessões correspondentes de estatísticas de ordem inferior.

O conceito de estabilidade para as estatísticas de ordem não é recente, os primeiros resultados completos foram publicados em 1949 por Gnedenko para a convergência em probabilidade. Embora nesta apresentação seja visada a convergência quase certa, tendo em

conta o desenvolvimento do estudo, a apresentação ficaria incompleta se a convergência em probabilidade fosse completamente ignorada. Serão então lembrados os resultados pioneiros, apontando-se algumas referências a trabalhos posteriores sobre a estabilidade em probabilidade.

No estudo dos limites superiores e inferiores das estatísticas de ordem será focada essencialmente a caracterização de probabilidades do tipo

$$P\{X_n - k_{n+1,n} > c_n \text{ i.o.}\} \text{ e } P\{X_n - k_{n+1,n} \leq c_n \text{ i.o.}\},$$

onde, como usualmente, i.o. é a abreviatura de "infinitely often" (ver p.e. Galambos (1978/1987)).

É claro que a importância deste estudo não se resume à aplicação à estabilidade. Por exemplo, a questão da caracterização do comportamento limite forte das estatísticas de ordem superior surge naturalmente na estimação de coeficientes de cauda ou quantis elevados.

Começaremos por abordar a estabilidade dos valores extremos –  $k$ -ésimas estatísticas de ordem – e dos valores centrais da amostra na Secção 2. No segundo parágrafo desta Secção, lembram-se algumas caracterizações dos limites superiores e inferiores obtidas até à década de 80.

As caracterizações completas do comportamento limite forte das  $k_n$ -estatísticas de ordem para outras sucessões  $k_n$  foram obtidas a partir da década de 80 e são resumidas na Secção 3, bem como alguns resultados mais recentes sobre a estabilidade quase certa.

Finalmente, alguns comentários e perspectivas de desenvolvimento são apresentados na Secção 4.

## 2. Comportamento limite forte das $k$ -ésimas estatísticas de ordem superior: $k \geq 1$ ou $k = ]np[, 0 < p < 1$

### 2.1 Estabilidade

Vamos considerar em primeiro lugar as propriedades de estabilidade da sucessão das  $k$ -ésimas estatísticas de ordem superior, sendo  $k$  um inteiro positivo fixado. Supõe-se então aqui que  $F(x) < 1$ , para todo o  $x$ . Sem esta condição o estudo da estabilidade dos valores extremos (superiores) é trivial.

No fim dos anos 70 já tinham sido obtidos resultados bastante completos sobre a estabilidade da sucessão  $(X_{n-k+1,n})$ , sendo  $k \geq 1$ . Note-se que, no entanto, em muitos dos trabalhos de investigação consagrados ao estudo de propriedades de estabilidade das estatísticas de ordem, alguns destes resultados não são referidos, ou só parcialmente, tendo, por vezes, sido "redescobertos" posteriormente.

O estudo, na sua generalidade, foi iniciado por Gnedenko (1943) e continuado por Smirnov (1952), Geffroy (1958, 1959), Barndorff-Nielsen (1963) e Geffroy (1965), artigo raramente citado, embora contenha resultados muito importantes e até inesperados sobre a estabilidade quase certa.

As primeiras leis caracterizadas foram naturalmente as leis fracas podendo já ser considerado como um resultado clássico na altura, o critério de Gnedenko, segundo o qual, a sucessão de máximos é estável se, e só se,

$$\lim_{x \rightarrow \infty} \frac{1 - F(x + \epsilon)}{1 - F(x)} = 0,$$

para todo o  $\epsilon > 0$ . A sucessão normalizadora pode ser escolhida como  $a_n = Q(1 - 1/n)$ , onde  $Q$  denota a função quantil, definida como é usual por  $Q(u) = \inf\{x: F(x) \geq u\}$ ,  $0 < u \leq 1$ .

Se forem somente consideradas sucessões normalizadoras crescentes para infinito, um critério para um tipo de estabilidade pode ser convertido num critério para o outro tipo. Assim, a sucessão de máximos é relativamente estável se, e só se,

$$\lim_{x \rightarrow \infty} \frac{1 - F(\delta x)}{1 - F(x)} = 0,$$

para todo o  $\delta > 1$ , ou seja, se, e só se,  $1 - F$  é de variação rápida. Entre os trabalhos pioneiros sobre a relação entre a estabilidade e a propriedade de variação regular destaca-se de Haan (1970).

Sabia-se também que, sendo  $k$  um inteiro positivo fixado, a estabilidade da sucessão  $(X_{n-k+1,n})$  é equivalente à estabilidade da sucessão de máximos  $(X_{n,n})$ . As condições acima referidas são então condições necessárias e suficientes de estabilidade da sucessão das  $k$ -ésimas estatísticas de ordem superior. Note-se que, como seria de esperar, estas condições dizem respeito à cauda superior de  $F$ .

Conheciam-se ainda outras condições necessárias e suficientes ou só suficientes de estabilidade. Em particular, são apresentadas no livro Galambos (1978), condições suficientes de estabilidade quando a sucessão (convenientemente normalizada) converge fracamente para uma distribuição não degenerada, não sendo, por vezes, exigida a independência ou a igualdade em distribuição das variáveis aleatórias.

**Exemplo.** Suponha-se  $X_1, X_2, \dots$  uma sucessão Gaussiana estacionária com  $E(X_n) = 0$  e  $V(X_n) = 1$  e seja  $r_m = E(X_1 X_{m+1})$ .

Então, se  $r_m = o(1/\log m)$ ,  $(X_{n,n}, a_n)$  satisfaz tanto a lei aditiva AWL com a multiplicativa MWL, com

$$a_n = (2 \log n)^{1/2} - \frac{\log \log n + \log 4\pi}{2(2 \log n)^{1/2}}.$$

Relativamente à estabilidade quase certa, segundo Geffroy (1965) e Barndorff-Nielsen (1963), a sucessão de máximos é estável se, e só se

$$\lim_{x \rightarrow \infty} \frac{(1 - F(x)) \log(1 - F(x))}{1 - F(x - \epsilon)} = 0,$$

ou, de modo equivalente, se, e só se

$$\sum_n (1 - F(a_n + \epsilon)) < \infty,$$

para todo o  $\epsilon > 0$ , com  $a_n = Q(1 - 1/n)$ . Assim, se  $F$  verifica as condições acima,  $X_{n,n} - Q(1 - 1/n) \rightarrow 0$  q.c., ou seja

$$P\{X_{n,n} > Q(1 - 1/n) + \epsilon \text{ i.o.}\} = P\{X_{n,n} \leq Q(1 - 1/n) - \epsilon \text{ i.o.}\} = 0,$$

para todo o  $\epsilon > 0$ .

Tem-se também uma propriedade análoga à da estabilidade em probabilidade, nomeadamente, sendo  $k$  um inteiro positivo fixado, a sucessão  $(X_{n-k+1,n})$  é estável q.c. se, e só se,  $(X_{n,n})$  o é, tendo-se neste caso

$$X_{n,n} - X_{n-k+1,n} \rightarrow 0 \text{ q.c.}$$

Lembra-se ainda que, no final dos anos 70, já se conheciam algumas propriedades de estabilidade para classes particulares de variáveis aleatórias dependentes, com relevo para sucessões gaussianas estacionárias (ver p.e. Galambos (1978)).

Considere-se agora o caso em que  $k$  pode variar com  $n$  e suponha-se que  $k_n$  é uma sucessão tal que  $k_n \uparrow \infty$  e  $k_n/n \rightarrow c$ ,  $c \in [0, 1)$ . Naturalmente, o estudo da estabilidade difere substancialmente consoante  $c$  seja igual ou diferente de zero, podendo-se notar que já se conhecem propriedades de estabilidade dos valores centrais ( $c \neq 0$ ) segundo os modos de convergência usuais desde 1949 (Smirnov). Em particular, se existe uma solução única,  $Q(1-c)$ , da equação  $1 - F(x) = c$ , então sabe-se que

$$X_{n-k_n+1,n} \rightarrow Q(1-c) \text{ q.c..}$$

Para  $c = 0$ , estudo das propriedades de estabilidade quase certa da sucessão  $(X_{n-k_n+1,n})$  é mais complexo, o comportamento limite forte da sucessão depende da ordem de crescimento de  $k_n$ . Estuda-se então esta questão num contexto mais geral, da existência de sucessões  $(a_n)$  e  $(b_n)$ ,  $b_n \neq 0$  tais que  $\limsup_{n \rightarrow \infty} (X_{n-k_n+1,n} - a_n) / b_n$  ou  $\liminf_{n \rightarrow \infty} (X_{n-k_n+1,n} - a_n) / b_n$  seja finito e diferente de zero q.c., que, pela sua especificidade, será abordada na secção seguinte.

## 2.2 Limites superior e inferior

A caracterização do comportamento limite superior e inferior das estatísticas de ordem normalizadas foi considerada em vários trabalhos e, tal como para a estabilidade, até final

dos anos 70, foram obtidos resultados bastante completos no caso em que  $k_n = k$ ,  $k$  um inteiro positivo fixado. Este estudo envolve a caracterização de probabilidades da forma

$$P\{X_{n-k_n+1,n} > c_n \text{ i.o.}\} \text{ e } P\{X_{n-k_n+1,n} \leq c_n \text{ i.o.}\}, \quad (1)$$

sendo fundamentais neste estudo os lemas do tipo Borel Cantelli.

Considere-se em primeiro lugar a majoração assintótica. O resultado apresentado a seguir é um melhoramento obtido por Mori (1976) de resultados anteriormente obtidos (ver Kiefer (1972), Barndorff-Nielsen (1963) e Geffroy (1958/59) para  $k=1$ ). Sendo  $c_n \uparrow \infty$ ,

$$P\{X_{n-k+1,n} > c_n \text{ i.o.}\} = 0 \text{ ou } 1,$$

consoante

$$\sum_n n^{k-1} (1 - F(c_n))^k < \infty \text{ ou } = \infty.$$

A determinação de  $P\{X_{n-k+1,n} \leq c_n \text{ i.o.}\}$  é mais laboriosa. Barndorff-Nielsen (1961), através de uma generalização da parte de convergência do Lema de Borel Cantelli e da adaptação de técnicas usadas por Erdős (1949), mostra que, se  $c_n \uparrow$  e  $(F(c_n))^n \downarrow$ , então

$$P\{X_{n,n} \leq c_n \text{ i.o.}\} = 0 \text{ ou } 1,$$

consoante

$$\sum_n (F(c_n))^n \frac{\log \log n}{n} < \infty \text{ ou } = \infty.$$

Vários resultados se seguiram sobre a minoração de  $X_{n-k+1,n}$ , dos quais destacamos a caracterização obtida por Frankel (1976) e melhorada por Shorack e Wellner (1978), que é a extensão de um resultado de Robbins e Siegmund (1972) ao caso  $k > 1$ .

Se  $c_n \uparrow$  e  $n(1 - F(c_n)) \uparrow$  ou  $\liminf(n(1 - F(c_n)) / \log \log n) \geq 1$ , então

$$P\{X_{n-k+1,n} \leq c_n \text{ i.o.}\} = 0 \text{ ou } 1,$$

consoante

$$\sum_n n^{k-1} (1 - F(c_n))^k \exp(-n(1 - F(c_n))) < \infty \text{ ou } = \infty.$$

Como já foi referido, pode-se caracterizar o comportamento limite superior e inferior da sucessão de valores extremos através dos resultados acima. Considere-se, por exemplo, a sucessão de máximos e seja  $c_n(\epsilon) = a_n + (1 + \epsilon)b_n$ ,  $|\epsilon| < \epsilon_0$ , uma sucessão crescente.

Então

$$\limsup_{n \rightarrow \infty} (X_{n,n} - a_n)/b_n = 1 \text{ q.c.},$$

se

$$\sum_n (1 - F(c_n(\epsilon))) < \infty \text{ ou } = \infty,$$

consoante  $\epsilon > 0$  ou  $\epsilon \leq 0$ .

Por outro lado, se  $c_n(\epsilon)$  é ainda tal que  $n(1 - F(c_n(\epsilon))) \uparrow$  ou  $\liminf(n(1 - F(c_n(\epsilon))) / \log \log n) \geq 1$  então

$$\liminf_{n \rightarrow \infty} (X_{n,n} - a_n)/b_n = 1 \text{ q.c.},$$

se

$$\sum_n (1 - F(c_n(\epsilon))) \exp(-n(1 - F(c_n(\epsilon)))) < \infty \text{ ou } = \infty,$$

consoante  $\epsilon > 0$  ou  $\epsilon \leq 0$ .

A partir desta caracterização podem-se obter enquadramentos óptimos da sucessão de valores extremos para distribuições específicas. Limitamo-nos a apresentar dois exemplos bem conhecidos no final dos anos 70.

**Exemplo.** Distribuição normal

Se  $F$  é a f.d. de uma lei  $N(0, 1)$ , então tem-se:

$$\limsup_{n \rightarrow \infty} \frac{\sqrt{2 \log n}}{\log \log n} \left( X_{n,n} - \sqrt{2 \log \left( \frac{n}{\sqrt{\log n}} \right)} \right) = 1 \text{ q.c.}$$

$$\liminf_{n \rightarrow \infty} \frac{X_{n,n}}{\sqrt{2 \log n}} = 1 \text{ q.c.}$$

Conhecidos estes limites, obtém-se de modo trivial a estabilidade da sucessão de máximos, bem como a forma da sucessão normalizadora, em particular,

$$\lim_{n \rightarrow \infty} \frac{X_{n,n}}{\sqrt{2 \log n}} = 1 \text{ q.c..}$$

Considere-se agora o caso da distribuição uniforme, que é de especial importância no estudo do comportamento limite forte das estatísticas de ordem. Para distinguir este caso do caso geral, representam-se por  $U_{1,n} \leq U_{2,n} \leq \dots \leq U_{n,n}$  as estatísticas de ordem correspondentes às  $n$  primeiras observações duma sucessão  $(U_n)$  de v.a. i.i.d. com distribuição  $U(0, 1)$ .

**Exemplo.** Distribuição uniforme

Tendo em conta as aplicações, explicitamos aqui as expressões gerais das sucessões que enquadram  $U_{n-k+1,n}$ . Para  $j \geq 4$  e designando por  $\log_j x$  o logaritmo iterado de  $x$ ,

$$P \left\{ U_{n-k+1,n} \leq 1 - \frac{\log_2 n + (k+1)\log_3 n + \log_4 n + \dots + \log_{j-1} n + (1+\epsilon)\log_j n}{n} \text{ i.o.} \right\} =$$

$$P \left\{ U_{n-k+1,n} > 1 - \frac{(\log n \dots \log_{j-1} n (\log_j n)^{1+\epsilon})^{-1/k}}{n} \text{ i.o.} \right\} = 0 \text{ ou } 1,$$

consoante  $\epsilon > 0$  ou  $\epsilon \leq 0$ . Este enquadramento foi obtido por Deheuvels (1974) por um método diferente, nomeadamente, através do estudo de sucessões de passagem convenientemente definidas.

Em Deheuvels (1974) são também determinados enquadramentos para outras distribuições específicas, em particular para a normal, a partir do enquadramento acima. De facto,  $(Q(U_n))$  define uma sucessão de v.a. i.i.d. com f.d.  $F$  e, tendo em conta a monotonia de  $Q$ , podem-se obter enquadramentos para a sucessão geral  $(X_{k,n})$  a partir dos resultados correspondentes para  $(U_{k,n})$ . Em particular, as caracterizações do comportamento limite forte de  $(X_{k,n})$  podem ser deduzidas directamente. Deste modo e, para simplificar a apresentação, consideramos de seguida o comportamento de  $(U_{n-k+1,n})$ .

Tal como para a convergência em probabilidade, quando  $k$  pode variar com  $n$ , os primeiros resultados completos obtidos dizem respeito às estatísticas de ordem central. Considere-se então  $k_n = \lfloor np \rfloor$ , onde  $\lfloor x \rfloor$  designa o menor inteiro superior ou igual a  $x$  e  $p$  uma constante fixada tal que  $p \in (0, 1)$ . Para sucessões  $c_n \uparrow$ ,

$$P \left\{ \pm \frac{n^{1/2}(U_{n-k_n+1,n} - (1-p))}{(p(1-p))^{1/2}} > c_n \text{ i.o.} \right\} = 0 \text{ or } 1,$$

consoante

$$\sum_n n^{-1} c_n \exp(-c_n^2/2) < \infty \text{ or } = \infty.$$

Esta caracterização foi obtida por Kiefer (1967) através de um método introduzido por Bahadur (1966), que estabeleceu uma representação forte para os processos empíricos e dos quantis, no espírito do princípio de invariância forte de Strassen (1964,1966) para somas de

v.a.. No fim dos anos 70 registou-se um grande interesse por este tipo de representações, destacando-se os trabalhos de Komlos, Major e Tusnady (1975) e Csörgö e Révész (1978).

Para outro tipo de sucessões  $k_n$  a situação é mais complexa, as primeiras caracterizações foram obtidas a partir dos anos 80 e serão portanto abordadas na Secção 3 deste trabalho. Note-se que, no fim dos anos 70, só se conheciam os termos de primeira ordem quer das sucessões majorantes quer das minorantes (Kiefer (1972)).

### 3. Leis fortes para as $k$ -ésimas estatísticas de ordem superior: resultados obtidos a partir da década de 80

Na década de 80 registou-se um grande desenvolvimento no estudo do comportamento limite forte das estatísticas de ordem, visando-se não só a extensão dos resultados anteriormente obtidos a outras situações mais gerais, como também, o enfraquecimento das condições de regularidade assumidas. Inserem-se neste último contexto os trabalhos de Klass (1984, 1985) e, mais recentemente, Tomkins (1996a) e Wang (1997). Sabemos também que algumas das condições assumidas em Kiefer (1972) podem ser enfraquecidas (ver p.e. Deheuvels e Mason (1988, 1990)). Não é nosso objectivo explicitar aqui os sucessivos melhoramentos destas condições de regularidade, o que conduziria a um estudo demasiado técnico (ver p.e. Tomkins e Wang (1998)) mas, tendo em conta que o resultado de Klass é usualmente considerado como uma versão final, notamos que nesse trabalho são relaxadas as hipóteses de monotonia. Nomeadamente, o resultado de Robbins e Siegmund (1972) sobre a minoração é estabelecido supondo somente que  $1 - F(c_n) \rightarrow 0$  e  $n(1 - F(c_n)) \rightarrow \infty$ .

A interessante questão da extensão dos diversos critérios conhecidos para a sucessão de máximos ao caso em que  $k > 1$ , fixado, é abordada em Wang e Tomkins (1992).

Considere-se agora ao caso em que  $k$  pode depender de  $n$ . Os resultados de Kiefer (1972) já aqui referidos, não só confirmaram a grande dependência do comportamento limite forte na ordem de crescimento de  $k_n$  mas também evidenciaram o papel fronteiro desempenhado pela sucessão  $\log_2 n$ . De facto, os limites explicitados diferem consoante  $k_n / \log_2 n \rightarrow 0$ ,  $k_n / \log_2 n \rightarrow c \in (0, \infty)$  ou  $k_n / \log_2 n \rightarrow \infty$ .

O trabalho de Kiefer foi inicialmente desenvolvido e aprofundado em Brito (1986, 1987), Deheuvels (1986, 1989), conduzindo não só a um melhor conhecimento das sucessões majorantes e minorantes, mas também à caracterização das classes superiores e inferiores para alguns casos. Em particular, em Brito (1986) é obtido um enquadramento optimal de  $U_{n-k_n+1, n}$  no caso em que  $k_n = [\tau \log_2 n]$ , onde  $\tau > 0$ . Enquadramentos correspondentes para outras distribuições específicas são determinados em Brito (1987), contemplando distribuições contínuas usuais, como a normal, gama, Cauchy, e também distribuições discretas, binomial negativa, utilizando um prolongamento contínuo da distribuição proposto por Anderson (1970).

As caracterizações obtidas em Deheuvels (1986, 1989) envolvem o estudo da série

$$\sum_n \frac{k_n^{1/2}}{n} \left( \frac{e}{k_n} n(1 - c_n) \right)^{k_n} \exp(-n(1 - c_n)). \quad (2)$$

Começemos por considerar a majoração assintótica. Suponha-se que  $k_n \rightarrow \infty$  tal que  $\limsup k_n / n < 1$  e seja  $c_n$  uma sucessão crescente. Deheuvels (1986) mostrou que, se  $c_n$  é tal que  $k_n^{-1/2} (n(1 - c_n) - k_n) \rightarrow -\infty$ , então a convergência da série (2) implica

$$P\{U_{n-k_{n+1},n} > c_n \text{ i.o.}\} = 0.$$

Segundo o resultado correspondente para a minoração, se  $c_n$  é tal que  $k_n^{-1/2} (n(1 - c_n) - k_n) \rightarrow \infty$ , então a convergência da série (2) implica

$$P\{U_{n-k_{n+1},n} \leq c_n \text{ i.o.}\} = 0.$$

O estudo da optimalidade dos enquadramentos é mais complexo e, em certas situações, só dispomos de resultados parciais envolvendo, por vezes, condições de regularidade intrincadas sobre o crescimento das sucessões  $k_n$  e  $c_n$ .

O caso em que  $k_n = O(\log_2 n)$  foi abordado em Deheuvels (1986). Apresenta-se a seguir um dos resultados obtidos. Seja  $k_n \uparrow \infty$  tal que  $k_n / (np_n) \rightarrow 1$ , para uma sucessão decrescente  $p_n$ . Seja ainda  $c_n$  tal que  $n(1 - c_n) \leq (n + 1)(1 - c_{n+1})$  e  $c_n \leq c_{n+1}$  sempre que  $k_n = k_{n+1}$ . Suponha-se também que se verifica uma das seguintes condições (i) ou (ii):

(i)  $k_n = O(\log_2 n)$ ,  $n(1 - c_n) / \log_2 n \rightarrow A$ , para um  $A > 0$  e  $\liminf_{n \rightarrow \infty} k_n / \log_2 n > 0$ ;

(ii)  $k_n = o(\log_2 n)$  e

$$0 < \liminf_{n \rightarrow \infty} \frac{n(1 - c_n)}{\log_2 n} \leq \limsup_{n \rightarrow \infty} \frac{n(1 - c_n)}{\log_2 n} < \infty.$$

Então, a divergência da série (2) implica

$$P\{U_{n-k_{n+1},n} \leq c_n \text{ i.o.}\} = 1.$$

Segundo Tomkins e Wang (1998), é estabelecido um resultado análogo em Wang (1991) para certas sucessões  $k_n$ , não sendo imposta a condição  $k_n = O(\log_2 n)$ . O problema da determinação de condições suficientes para que

$$P\{U_{n-k_{n+1},n} > c_n \text{ i.o.}\} = 1, \tag{3}$$

embora seja referido em Tomkins e Wang (1998) como um problema em aberto, já tinha sido abordado, conhecendo-se nomeadamente, condições suficientes para certas sucessões  $k_n$ . Em particular, um resultado deste tipo foi estabelecido por Deheuvels (1989). Seja  $k_n \uparrow$

tal que  $k_n / n \rightarrow 0$  e  $k_n / \log k_{2n} \rightarrow \infty$ . Suponha-se também que  $c_n > 0$  é uma sucessão verificando  $n(1 - c_n) / k_n \rightarrow 0$  e

$$\limsup_{n \rightarrow \infty} \left( k_n \left( \frac{(n+1)(1 - c_{n+1}) - n(1 - c_n)}{1 - c_n} \right) 1_{\{k_n = k_{n+1}\}} \right) < 1.$$

Então, a divergência da série (2) implica (3). Note-se que, pelas condições assumidas, as sucessões  $k_n$  de interesse são tais que  $k_n = o(\log_2 n)$ .

Como já referimos, as representações fortes dos processos empíricos tiveram um grande impacto no estabelecimento de leis fortes gerais, conduzindo na década de 80 a um desenvolvimento deste tipo de métodos, designados usualmente por métodos de aproximação forte (ver p.e. Csörgö e Révész (1981)).

Usando este tipo de métodos, em Brito (1987) foi dada uma caracterização completa das probabilidades do tipo (1) para certas sucessões  $k_n$ , através do estudo do comportamento limite forte do processo dos quantis ao longo duma sucessão adequada  $p_n$ . Foi estabelecida uma aproximação forte deste processo por um processo de Kiefer ao longo de  $p_n$ , que permite reduzir o problema ao estudo do problema correspondente para o processo de Kiefer. As técnicas utilizadas conduzem a considerar sucessões  $p_n$  com  $0 < p_n < 1$ , tais que

$$p_n / (\log_2 n (\log n)^4) \rightarrow \infty, \quad (4)$$

satisfazendo as condições de regularidade

$$0 < p_n < 1, \quad p_n \downarrow 0 \quad e \quad (np_n) \uparrow. \quad (5)$$

Para simplificar a notação escreva-se agora  $c_n = \psi(n)$  e seja  $t_0$  uma constante, com  $t_0 \geq \exp(e)$ . Sendo  $\psi$  uma função positiva crescente, então para  $k_n = ]np_n[$ ,

$$P\{(k_n(1 - k_n/n))^{-1/2} n | U_{n-k_n+1,n} - (1 - k_n/n) | > \psi(n) \text{ i.o.}\} = 0 \text{ ou } 1, \quad (6)$$

consoante

$$I(\psi) = \int_{t_0}^{\infty} t^{-1} \psi(t) \exp(-(\psi(t))^2/2) dt < \infty \text{ ou } = \infty,$$

ou, de modo equivalente, consoante

obtendo-se então o seguinte enquadramento optimal

$$P\{(k_n(1 - k_n/n))^{-1/2}n |U_{n-k_n+1,n} - (1 - k_n/n)| > (2(\log_2 n + (3/2)\log_3 n + \log_4 n + \dots + (1+\epsilon)\log_p n))^{1/2} \text{ i.o.}\} = 0 \text{ ou } 1,$$

consoante  $\epsilon > 0$  ou  $\epsilon \leq 0$ .

É interessante notar que, por este resultado, sob as condições de regularidade assumidas, o critério da caracterização de Kiefer (1967) para as estatísticas de ordem central permanece válido para as  $k_n$ -ésimas estatísticas de ordem, quando  $k_n/n \rightarrow 0$  e  $k_n/(\log_2 n(\log n)^4) \rightarrow \infty$ .

Os resultados obtidos por Kiefer (1972), ou do mesmo tipo, podem também ser obtidos a partir de leis fortes funcionais para processos dos quantis (cf. Deheuvels e Mason (1990)). Através destas leis, por vezes designadas por leis funcionais do logaritmo iterado, podem estabelecer-se leis fortes para variadas estatísticas. Por exemplo, em Bacro e Brito (1992) é feita uma aplicação ao estudo do comportamento limite forte da diferença entre duas estatísticas de ordem. No fim dos anos 80 e, fundamentalmente, na década de 90, foram publicados vários trabalhos consagrados à determinação deste tipo de leis para processos empíricos e dos quantis, tendo também sido consideradas diversas aplicações. Entre as possíveis aplicações, destacamos o estudo do comportamento limite forte de estatísticas baseadas nos valores extremos da amostra, como, estimadores de coeficientes de cauda ou de quantis elevados, ou somas aparadas de estatísticas de ordem (ver, p.e. Bacro e Brito (1993) e Deheuvels e Mason (1991)).

Voltando agora à questão da estabilidade, é de referir que este tema teve um grande desenvolvimento na década de 90. O estudo da estabilidade quase certa das  $k$ -ésimas estatísticas de ordem,  $k$  fixado, é retomado por diversos autores, incluindo Tomkins (1999), Tomkins (1996b) onde é estudada a velocidade de convergência e Goldie e Maller (1996) usando técnicas do domínio dos processos pontuais. As propriedades de estabilidade da sucessão  $(X_n - k_{n+1,n})$  foram extensivamente investigadas, mas no que diz respeito à convergência em probabilidade (ver p.e. Gather e Tomkins (1995), Qi (1997) e referências citadas). Em Tomkins e Wang (1992) é também considerada a convergência quase certa, mas para a sucessão particular  $a_n = Q(1 - 1/n)$ . Note-se que esta é a sucessão normalizadora das  $k$ -ésimas estatísticas de ordem superior,  $k$  fixado, sendo  $Q(1 - k_n/n)$  a sucessão normalizadora natural neste caso. O estudo da estabilidade quase certa de  $(X_n - k_{n+1,n})$  tem estado associado ao problema da determinação dos limites fortes (lim sup; lim inf) da sucessão, obtendo-se, em particular, uma condição suficiente de estabilidade a partir da caracterização de (6) para as sucessões  $k_n$  aí consideradas (cf. Brito (1990)).

Este problema é retomado em Barne-Delcroix e Brito (2001) num outro contexto, onde é dado o seguinte critério de estabilidade. Seja  $k_n = \lceil np_n \rceil$  e suponha-se que  $p_n$  verifica (4) e (5). Então a sucessão  $(X_n - k_{n+1,n})$  é estável q.c. se, e só se, existem duas funções crescentes  $\psi_1$  e  $\psi_2$ , com  $\lim_{n \rightarrow \infty} \psi_1(n) = \lim_{n \rightarrow \infty} \psi_2(n) = \infty$  e tais que

$$I(\psi_1) < \infty, \quad I(\psi_2) < \infty$$

$$e \quad \lim_{n \rightarrow \infty} (Q(1 - b_n) - Q(1 - a_n)) = 0,$$

onde  $a_n = (k_n + \psi_1(n)(k_n)^{1/2})/n$  e  $b_n = (k_n - \psi_2(n)(k_n)^{1/2})/n$ .

#### 4. Alguns comentários e perspectivas

Neste trabalho foram consideradas essencialmente sucessões de v.a. i.i.d.. Embora, como foi aqui referido, se conheçam algumas propriedades de estabilidade quase certa em situações não i.i.d., a questão clássica da extensão de resultados sobre v.a. independentes ao caso de variáveis dependentes, continua actual no estudo da estabilidade das estatísticas de ordem. Sabe-se já, por exemplo, que a condição suficiente de estabilidade do máximo de v.a. i.i.d. não basta em geral, para assegurar a estabilidade do máximo em sucessões estacionárias. Surge assim, como particularmente importante, a pesquisa de critérios de estabilidade quase certa para sucessões de v.a. com diferentes tipos de dependência. Ainda recentemente, foi estabelecida em Naveau (2003) uma condição suficiente de estabilidade relativa quase certa da sucessão de máximos de sucessões estacionárias satisfazendo certas condições de dependência. No mesmo trabalho também foi dada uma condição necessária, sob hipóteses mais restritivas.

Como se constatou também na Secção anterior, as caracterizações dos limites fortes das  $k_n$ -ésimas estatísticas de ordem envolvem, por vezes, condições de regularidade complexas. Uma questão interessante é, naturalmente, a de investigar se estas condições podem ser enfraquecidas. Por outro lado, alguns resultados são parciais, não cobrindo todas as sucessões  $k_n$  de interesse. Seria portanto desejável conhecer o comportamento limite também nestes casos.

Note-se ainda que, sendo vários estimadores baseados em estatísticas de ordem, a extensão de resultados actuais de convergência forte a outras situações envolverá naturalmente o estudo das propriedades assintóticas das estatísticas de ordem. Em particular, o desenvolvimento do estudo do comportamento limite forte de estimadores de coeficientes de cauda ou quantis elevados, conduzirá naturalmente, a um desenvolvimento do estudo do comportamento limite forte das  $k_n$ -ésimas estatísticas de ordem superior.

#### Referências

- [1] Anderson, C. W. (1970). Extreme value theory for a class of discrete distributions with applications to some stochastic processes. *J. Appl. Probab.* 7, 99-113.
- [2] Bacro, J. N. e M. Brito (1992). Comportement asymptotique presque sûr de la différence de statistiques d'ordre uniformes. *C. R. Acad. Sci. Paris*, t. 314, Série I, 317-320.
- [3] Bacro, J. N. e M. Brito (1993). Strong limiting behaviour of a simple tail Pareto-index estimator. *Statist. and Dec. Suppl. Issue 3*, 133-143.

- [4] Bahadur, R. R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* 37, 577-580.
- [5] Barne-Delcroix, M. F. e M. Brito (2001). Multivariate stability and strong limiting behaviour of intermediate order statistics. *J. Multivariate Anal.* 79, 157-170.
- [6] Barndorff-Nielsen, O. (1961). On the rate of growth of the partial maxima of a sequence of independent identically distributed random variables. *Math. Scand.* 9, 383-394.
- [7] Barndorff-Nielsen, O. (1963). On the limit behaviour of extreme order statistics. *Ann. Math. Statist.* 34, 992-1002.
- [8] Brito, M. (1986). Sur l'encadrement optimal presque sûr dans un échantillon ordonné. *C. R. Acad. Sci. Paris*, t. 303, Série I, 821-824.
- [9] Brito, M. (1987). Encadrement presque sûr des statistiques d'ordre. Thèse Doctorat, Univ. Paris 6.
- [10] Brito, M. (1990). Sobre o comportamento limite forte das estatísticas de ordem. In M. A. Turkman e M.L. Salema, Eds. *Actas da 1ª Conferência em Estatística e Optimização*, Troia 1990, 113-126.
- [11] Csörgő, M. e P. Révész (1978). Strong approximations of the quantile process, *Ann. Statist.* 6, 882-894.
- [12] Csörgő, M. e P. Révész (1981). *Strong approximations in probability and statistics*. Academic Press.
- [13] de Haan, L. (1970). *On regular variation and its application to the weak convergence of sample extremes*. Mathematical Centre Tracts, Vol. 32, Amsterdam.
- [14] Deheuvels, P. (1974). Majoration et minoration presque sûre optimale des éléments de la statistique ordonnée d'un échantillon croissant de variables aléatoires indépendantes. *Rend. Acad. Nazionale dei Lincei*, S. 8, 6, 707-719.
- [15] Deheuvels, P. (1986). Strong laws for the  $k$ -th Order Statistic when  $k \leq \text{clog}_2 n$ . *Probab. Theory Rel. Fields* 72, 133-154.
- [16] Deheuvels, P. (1989). Strong laws for the  $k$ -th Order Statistic when  $k \leq \text{clog}_2 n$  II. In J. Hüslér and R.D. Reiss, Eds. *Extreme value theory*, Lecture Notes in Statistics, Vol 51, 21-35, Springer, New-York.
- [17] Deheuvels, P., e D. M. Mason (1988). The asymptotic behavior of sums of exponential extreme values. *Bull. Sc. Math. Série II* 112, 211-233.
- [18] Deheuvels, P. e D. M. Mason (1990). Non-standard laws of the iterated logarithm for tail empirical and quantile processes. *Ann. Probab.* 18, 1693-1722.
- [19] Deheuvels, P. e D. M. Mason (1991). A tail empirical process approach to some non-standard laws of the iterated logarithm. *J. Theoret. Probab.* 4, 53-85.
- [20] Erdős, P. (1942). On the law of the iterated logarithm. *Ann. Math.* 43, 419-436.
- [21] Frankel, J. (1976). A note on downcrossings for extremal processes. *Ann. Probab.* 4, 151-152.
- [22] Galambos, J. (1978/1987). *The asymptotic theory of extreme order statistics*. Robert E. Krieger Publishing Company.
- [23] Gather, U. e J. Tomkins (1995). On stability of intermediate order statistics. *J. Statist. Plann. Inference* 45, 175-183.

- [24] Geffroy, J. (1958,1959). Contribution à la théorie des valeurs extrêmes. *Publ. Inst. Statist. Univ. Paris* 7/8, 37-185.
- [25] Geffroy, J. (1965). Sur une condition nécessaire et suffisante de stabilité presque sûre des valeurs extrêmes d'un échantillon. *C. R. Acad. Sci. Paris*, t. 260, 3271-3273.
- [26] Gnedenko, B. V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.* 44, 423-453.
- [27] Goldie, C. M. e R. Maller (1996). A point-process approach to almost-sure behaviour of record values and order statistics. *Adv. Appl. Probab.* 28, 426-462.
- [28] Kiefer, J. (1967). On Bahadur's representation of sample quantiles. *Ann. Math. Statist.* 38, 1323-1342.
- [29] Kiefer, J. (1972). Iterated logarithm analogues for sample quantiles when  $p_n \downarrow 0$ . *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* 1, 227-244, Univ. of California Press, Berkeley.
- [30] Klass, M. J. (1984). The minimal growth rate of partial maxima. *Ann. Probab.* 12, 380-389.
- [31] Klass, M. J. (1985). The Robbins-Siegmund series criterion for partial maxima. *Ann. Probab.* 13, 1369-1370.
- [32] Komlos, J., Major, P. e G. Tusnady (1975). An approximation of partial sums of independent r. variables and the sample d.f. I. *Z. Wahrsch. Verw. Gebiete* 32, 111-131.
- [33] Mori, T. (1976). The strong law of large numbers when extreme terms are excluded from sums. *Z. Wahrsch. Verw. Gebiete* 36, 189-194.
- [34] Naveau, P. (2003). Almost sure relative stability of the maximum of a stationary sequence. *Adv. Appl. Probab.* 35, 721-736.
- [35] Qi, Yongcheng (1997). A problem on stability of order statistics. *J. Statist. Plann. Inference* 64, 21-25.
- [36] Robbins, H. e D. Siegmund (1972). On the law of the iterated logarithm for maxima and minima. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* 3, 51-70, Univ. of California Press, Berkeley.
- [37] Shorack, G. R. e J. Welner (1978). Linear bounds on the empirical distribution function. *Ann. Probab.* 6, 349-353.
- [38] Smirnov, N. V. (1949/1952). Limit distributions for the terms of a variational series. *Trudy Mat. Inst. Steklov* 25, 1 {60 (em russo); traduzido em: *Amer. Math. Soc. Transl.* 67 (1952), 82-143.
- [39] Strassen, V. (1964). An invariance principle for the LIL. *Z. Wahrsch. Verw. Gebiete* 3, 211-226.
- [40] Strassen, V. (1966). Almost sure behavior of sums of independent random variables and martingales. *Proc. Fifth Berkeley Symp.* Vol. 32, Part 1, 51-70, Univ. of California Press, Berkeley.
- [41] Tomkins, R. J. (1996a). Refinement of a zero-one law for maxima. *Statist. Probab. Lett.* 27, 67-69.
- [42] Tomkins, R. J. (1996b). Rates of convergence for the stability of large order statistics. *J. Theoret. Probab.* 9, 841-851.

- [43] Tomkins, R. J. (1999). Regular stability of order statistics. *Statist. Probab. Lett.* 41, 145-151.
- [44] Tomkins, R. J. e H. Wang (1992). Stability theorems for large order statistics with varying ranks. *Statist. Probab. Lett.* 14, 91-95.
- [45] Tomkins, R. J. e H.Wang (1998). Zero-one laws for large order statistics. In: N. Balakrishnan e C. R. Rao, Eds. *Order Statistics: theory and methods*, Handbook of Statistics, Vol. 16, 375-384, Elsevier Science B. V..
- [46] Wang, H. (1997). Generalized zero-one laws for large-order statistics. *Bernoulli* 3, 429-444.
- [47] Wang, H. e R. J. Tomkins (1992). A zero-one law for large-order statistics. *Can. J. Statist.* 20, 323-334.

## **Margarida Maria Araújo Brito**

Licenciatura em Matemática, ramo científico Estatística, pela Faculdade de Ciências e Tecnologia da Universidade de Coimbra em 1981.

Doutoramento em Matemática pela Universidade de Paris VI, Laboratoire de Statistique Théorique et Appliquée, em 1987.

Foi membro do Centro de Matemática Aplicada da Universidade do Porto e é actualmente membro do Centro de Matemática e professora associada do Departamento de Matemática Aplicada da Faculdade de Ciências da Universidade do Porto, tendo colaborado em diversas actividades de administração e coordenação.

Tem participado activamente na formação pós-graduada a nível de mestrado, incluindo a orientação de várias dissertações e a coordenação do Mestrado em Estatística, de que foi proponente.

A sua investigação científica tem focado o comportamento assintótico de estatísticas e processos relevantes na estimação semi e não-paramétrica, com destaque para as estatísticas de ordem, processos empíricos e dos quantis e aplicações na teoria do risco e valores extremos.

# A Evolução da Estatística Bayesiana em Portugal

*Carlos Daniel Paulino*

As bodas de prata que a comunidade estatística portuguesa celebra em 2005 é pretexto para eu me abalançar a traçar uma panorâmica sobre a evolução da Estatística Bayesiana no nosso país. Apesar dos esforços envidados no sentido de tentar transmitir o mais fiel e abrangentemente possível a realidade evolutiva dessa área, é natural que a análise possa reflectir algumas insuficiências que fui incapaz de suprir. As minhas antecipadas desculpas pela eventual omissão de dados indiscutivelmente relevantes.

Julgo ser incontestável afirmar que foi Bento Murteira o pioneiro no estudo e divulgação científico-pedagógica da Estatística Bayesiana no meio universitário português através da sua actividade no então ISE/UTL. Até ao despontar dos anos 80 – e à parte alguns esporádicos seminários do centro que viria a dar origem ao actual CEAUL, onde tal temática foi objecto de discussão que opôs fortemente Tiago de Oliveira e Simões Neto —, ninguém mais em Portugal, tanto quanto é do meu conhecimento, se aventurou no terreno da difusão das ideias e métodos dessa nova (embora setecentista de origem) abordagem à Inferência Estatística, que nos círculos internacionais vinha começando a impor-se paulatinamente, particularmente desde os anos 50, graças ao trabalho de Jeffreys, Savage, Good, Jaynes e Lindley, entre outros. A obra mais arrojada de Bento Murteira, *Estatística: Inferência e Decisão*, que só em 1988 seria publicada, pela profunda reflexão sobre as metodologias estatísticas que constitui, foi no nosso meio um marco impulsor de uma consciência crítica da teoria e prática estatística que, ainda hoje, deve ser leitura assídua e reflexiva de quem faz da Estatística a sua ocupação profissional.

M. Antónia Amaral Turkman foi o primeiro docente universitário português a doutorar-se (em 1980 na *University of Sheffield*) na área (tópico de Predição). Aquando do seu regresso ao DEIO/FCUL, apesar de ambiente adverso, não hesitou em dar o seu contributo na difusão de tópicos de Estatística Bayesiana no primeiro mestrado em Probabilidade e Estatística criado no país pelo seu departamento (em 1981). Esta sua acção teve eco em mim próprio como aluno desse mestrado, ao ter corajosamente (pode-se dizê-lo) decidido concentrar a minha tese de mestrado na área bayesiana (Análise de Regressão) como forma de me inteirar das suas características. A minha adesão à “causa” e a existência de circunstâncias favoráveis levaram a manter-me na mesma escola de

pensamento ao longo do meu processo de doutoramento na Universidade de São Paulo, Brasil, que terminei nos fins da década de 80 (agora no tópico de Dados Categorizados Omissos).

No findar dos anos 80, o meu regresso permitiu-me reforçar a acção de M. Antónia Amaral Turkman na frente portuguesa da luta pela implantação da Estatística Bayesiana na actividade universitária, cada um no seu terreno (DM/IST e DEIO/FCUL, respectivamente), conseguindo ambos expor as principais ideias e métodos dessa abordagem a audiências crescentes, então alargadas a estudantes de licenciatura, e a atrair novos elementos para o espírito aberto e confraternizador da reunião magna quadrienal dos bayesianos de todo o mundo, os *Valencia Meetings*. Rapidamente este duo foi ampliado com os primeiros doutores em Portugal na área bayesiana, sob a supervisão de M. Antónia Amaral Turkman, João Faria (Teoria das Probabilidades Subjectivas), e depois, Isabel Pereira (Séries Temporais), aos quais se seguiram novos mestres Fernando Magalhães (*Screening*) pelo DEIO/FCUL e Paulo Soares (Dados Categorizados Incompletos) pelo DM/IST, vindo o primeiro dos quais a doutorar-se ainda na década de 90 na *University of Sheffield* (Predição). Devo ainda acrescentar que o primeiro doutorado em Portugal na área (Modelos lineares mistos em melhoramento animal) foi Fernando Macedo (DM-UTAD) em 1990, sob a supervisão do nosso colega uruguaio Daniel Gianola (então na *University of Illinois*).

M. Antónia Amaral Turkman e eu próprio continuámos activos e animados na lida de contribuir para o aumento da população portuguesa de estatísticos bayesianos e para a ampliação do campo da investigação bayesiana no país, através da organização de cursos intensivos e de seminários com bayesianos internacionalmente reconhecidos e da dinamização da actividade de investigação, difundindo e promovendo o uso dos novos métodos bayesianos radicados em simulação e implementáveis em *software* entretanto disponível. A publicação em 2003 do nosso livro *Estatística Bayesiana*, com co-autoria de Bento Murteira, e que é, pela sua abrangência, a primeira obra do género em língua portuguesa, poderá ser um útil instrumento para a prossecução desses fins.

A criação de um quadro internacional incontestavelmente mais favorável à aceitação e aplicação do paradigma bayesiano (bem visível a partir de meados dos anos 90) e a adopção no país de uma nova política de incentivo e concepção de doutoramentos facilitaram sobremaneira a consecução dos objectivos norteadores da acção que me unia a M. Antónia Amaral Turkman. E é assim que começa a surgir neste novo século uma nova fornada interna de doutores em Estatística Bayesiana sob supervisão de diferentes colegas, constituída numa ordenação cronológica por Giovani Silva (Análise de Sobrevivência), Patrícia Bermudez (Extremos), Marília Antunes (Séries Temporais), Lisete Sousa (Expressão Diferencial de Genes), Luzia Gonçalves (Mapeamento físico em Biologia Molecular), Paulo Soares (Dados Deficientemente Categorizados), Natércia Durão (*Screening*) e Isabel Natário (Epidemiologia Espacial).

Este conjunto tem naturalmente de ser complementado quer com novos doutorados no estrangeiro como Rui Paulo (Problemas na interface bayesiana/frequencista — *Duke*

*University*), Sofia Dias (Predição — *University of Sheffield*), Bruno Mendes (Aplicação de métodos MCMC a problemas de contaminação de solos — *University of Stockholm*) e Tomé Ferreira (Regressão com novos modelos assimétricos — *University of Warwick*), quer com dezenas de novos mestres no país, entre os quais se incluem Paulo Nogueira (Análise de Sobrevivência), Marina Andrade (Análise Forense), João Filipe Monteiro (Pescas) e Nuno Sepúlveda (Modelos de interacção genética), para só citar alguns dos que ainda não foram nomeados anteriormente. Prevê-se ainda a curto prazo um novo acréscimo da comunidade bayesiana com novos doutorados no país, bem revelador da significância estatística que essa comunidade vem alcançando.

A disseminação em muitas comunidades científicas das aplicações práticas da metodologia bayesiana também se começou a sentir no nosso país. Constituem manifestações concretas dessa actividade o trabalho no sector das Pescas da bióloga Manuela Azevedo (IPIMAR), o envolvimento de vários sectores bayesianos e não bayesianos no terreno da Biologia Computacional e a incursão mais ou menos pontual numa prática bayesiana de vários colegas estatísticos com produção clássica, claramente patente em trabalhos apresentados nos últimos congressos da SPE.

Em suma, se considerarmos a situação que se nos deparava há 25 anos em termos do conhecimento e aplicação do bayesianismo, não posso deixar de exprimir a minha incontida satisfação pelos inúmeros e marcantes avanços que a esse nível foram registados, impulsionando significativamente o processo de desenvolvimento da Estatística em Portugal. Porém, muito há ainda a fazer até porque a Estatística em geral está longe de ter conseguido uma plena afirmação no seio da sociedade civil portuguesa.

Que os jovens bayesianos portugueses tomem consciência dessa situação e das nítidas insuficiências da difusão do paradigma bayesiano em largos sectores profissionais para, em conformidade, saberem delinear as formas de acção mais consentâneas que permitam, com o seu dinamismo e determinação, novos saltos qualitativos na afirmação do pensamento bayesiano. Para tal parece-me imprescindível participarem crítica e inovadoramente na intensificação das suas aplicações práticas aos vários campos da ciência e na iniciação do seu ensino a níveis básicos, tarefa esta tão difícil quanto ingrata que, do meu ponto de vista, ainda não mereceu da comunidade bayesiana internacional a devida mobilização que a sua importância exige.

## **Carlos Daniel Paulino**

É Professor Associado Agregado no Departamento de Matemática do Instituto Superior Técnico e investigador do Centro de Matemática e Aplicações do IST, onde tem desenvolvido a sua carreira científica e pedagógica.

Obteve os graus de Licenciado em Engenharia Química pelo IST em 1975, Mestre em Estatística e Investigação Operacional pela FCUL em 1984 e Doutor em Ciências Estatísticas pela Universidade de São Paulo, Brasil, em 1989.

Na sua actividade docente tem leccionado variadas disciplinas de licenciatura e pós-graduação tais como Métodos Estatísticos, Probabilidade e Estatística, Inferência e Decisão Estatística, Estatística Bayesiana, Análise de Regressão, Tópicos de Análise Multivariada, Análise de Dados Categorizados, Análise de Sobrevivência, Estatística Espacial e Teoria Estatística.

É autor/co-autor de livros e monografias sobre Teoria Estatística, Inferência Estatística, Estatística Bayesiana, Exercícios de Probabilidades e Estatística e Análise de Dados Categorizados.

A sua actividade de investigação tem-se inserido nas áreas de Dados Categorizados, Estatística Bayesiana, Inferência Estatística e Aplicações às Ciências Biomédicas, nas quais tem orientado diversos trabalhos de pós-graduação e publicado, além das obras citadas, diversos artigos em revistas internacionais e nacionais.

# Estatística e *Agronomia* — Percurso e Ensino

## Algumas Notas Históricas

M. Manuela Neves

A contagem de populações humanas, de terras, de cabeças de gado,... já efectuada na mais longínqua antiguidade, visava, sobretudo, responder às necessidades da guerra e dos impostos. Sousa (1995) refere que o primeiro recenseamento de produções agrícolas, feito na China, data do séc. XXIII a.C. Não parece, no entanto, que tais registos se possam considerar “estatísticas”, tal como hoje são entendidas. Em *Reflexões sobre o Passado da Estatística*, Amaral-Turkman (1999) situa no séc. XIV as possíveis origens da Estatística. Muitos autores se têm interessado e continuam a interessar pelas origens e história da Estatística. *A herança de Bernoulli-Bayes-Laplace; um esboço da história da Estatística* de Tiago de Oliveira (1988) e *Aspectos da Realidade Estatística em Portugal* de João Branco (1999) são exemplos de outros dois textos que aconselhamos vivamente a ler.

Ligada inicialmente ao estudo de “assuntos de estado”, a Estatística foi usada nas mais variadas áreas: economia, geografia, história,... A componente aritmética foi-se sedimentando à medida que mais dados sobre assuntos importantes do estado iam sendo recolhidos, sendo aquela vertente a que, durante muito tempo, dominou os interesses da Estatística. No séc. XIX nasce, motivada pelo desenvolvimento dos estudos demográficos e sociais, a Estatística Matemática, resultante da fusão de diferentes correntes anteriores, entre as quais se pode citar o Cálculo das Probabilidades. Foi já no início do séc. XX que Fisher, Neymann e Wald, na senda de Pearson e *Student*, lançaram os fundamentos da Estatística Moderna. Fisher desenvolveu todo o seu notável trabalho durante os anos em que trabalhou na *Estação Experimental de Rothamsted*. Os métodos de análise estatística no domínio agro-pecuário sofreram desde logo forte impulso, permitindo a melhoria da produtividade, o aumento da eficácia, o estudo cuidadoso e metódico das condições de produção, etc..

Se até aqui nos referimos brevemente à história da Estatística, já o seu ensino é mais recente, aparecendo em finais do séc. XIX. Também este tema tem sido objecto de abundante literatura, como testemunha, por exemplo, a lista bibliográfica de Misra *et. al.* (1987). Inicialmente, o ensino da Estatística era integrado no ensino de outras disciplinas

(economia política, geografia, história,...) e sempre numa perspectiva descritiva ou gráfica.

Em Portugal, a Estatística começa a ser ensinada em 1854, sob proposta de Oliveira Marreca à *Real Academia das Ciências de Lisboa*, ver Tiago de Oliveira (1987). No que refere ao ensino da Estatística em *Agronomia*, vamos encontrar, uns anos mais tarde, a obra de alguns cientistas ilustres que, quer pelo pioneirismo quer pela qualidade didáctica, ficará ligada à História da Ciência em Portugal.

## 1. Estatística e *Agronomia*

O termo *Agronomia*, tal como foi considerado por Dagnelie (1995), vai ser aqui usado referindo-se, não só às grandes culturas, como também à horticultura, à produção animal, à água, à floresta, às indústrias agroalimentares, à economia e sociologia rurais e às biotecnologias.

No domínio agronómico a Estatística Matemática estabelece uma ligação entre as ciências exactas e as ciências biológicas, permitindo a interpretação quantitativa dos fenómenos biológicos observados.

A primeira metade do século XX é marcada por um conjunto de relações muito estreitas entre a *Agronomia* e a Estatística. Estatísticos ilustres desenvolveram uma obra notável no ensino e na investigação agronómicas, na linha dos trabalhos de Francis Galton (1822-1911) e Karl Pearson (1857-1936). Referindo alguns, temos por ordem cronológica de nascimento:

- William Gosset (1876-1937), mais conhecido pelo pseudónimo de *Student*, trabalhou na Guinness – por isso alguns autores classificam a sua actividade na “área agroalimentar” – em Dublin e Londres;
- Ronald Fisher (1890-1962), o iniciador e talvez um dos maiores vultos da Estatística Moderna, foi contratado em 1919 pela *Estação Experimental de Rothamsted*. Quando lhe surgiu a alternativa de ir trabalhar para Londres com Karl Pearson ou para a *Estação Experimental de Rothamsted*, talvez pelas dificuldades de trato mútuo com Pearson, Fisher escolheu ir para Rothamsted. Aqui, onde havia arquivadas mais de cem anos de observações, Fisher procede metodicamente à análise dos dados existentes, introduzindo um grande conjunto de métodos novos, como por exemplo o método da máxima verosimilhança, fazendo o estudo completo de todas as suas propriedades. Pearson era, na altura, defensor do método dos momentos. Mas a Fisher se deve ainda o desenvolvimento de uma boa parte da teoria dos testes, em particular os testes de significância. Alguns autores dizem mesmo que Fisher fez metade da Estatística Moderna. Permanecendo em Rothamsted até 1933, terá sido o primeiro matemático a trabalhar a tempo inteiro num centro de investigação agronómica.
- John Wishart (1898-1956) trabalhou em Rothamsted de 1927 a 1931 e na Escola de Agricultura da Universidade de Cambridge de 1931 a 1953;

- Frank Yates (1902-1994) trabalhou em Rothamsted de 1931 a 1968;
- William Cochran (1909-1980) trabalhou também em Rothamsted de 1934 a 1939, deixando ligadas à *Agronomia* obras notáveis, que todos conhecemos.

Outros grandes nomes têm continuado em Rothamsted a obra destes cientistas, embora não os refiramos nestas notas históricas.

Não se poderá dizer apenas que as competências destes grandes cientistas foram colocadas ao serviço do ensino agronómico, mas também que, partindo muitas vezes da *Agronomia*, eles desenvolveram conceitos como: as distribuições  $t$  (de *Student*), a inferência estatística relativa a médias e aos coeficientes de correlação, as distribuições  $F$  (de Fisher-Snedecor), a análise de variância e de covariância, a amostragem, os planos experimentais, a análise discriminante, etc., utilizados depois em muitas outras áreas do saber.

Poder-se-ia perguntar o que teria suscitado laços assim tão estreitos entre a *Agronomia* e a Estatística. Parece ter sido o facto de os agrónomos terem estado entre os primeiros a ter que “explicar” a variabilidade, por vezes considerável, do seu material experimental, vegetal ou animal e apesar desta variabilidade, a ter que tomar decisões que comprometiam o futuro, muitas vezes a longo prazo. Efectivamente, no séc. XIX, foram organizadas, em matéria de adubação mineral e de rotação de culturas, as primeiras experiências de longa duração, nas quais surgia considerável variabilidade nos resultados. Os conselhos tinham, no entanto, que ser dados a agricultores.

Foi provavelmente a acumulação de resultados das experiências agrícolas organizadas em Rothamsted e a insuficiência dos métodos existentes que levaram Fisher a entusiasmar-se e conduziram o seu espírito genial a inúmeras criações que o imortalizaram.

De facto o que dificulta o trabalho de um experimentador e exige a análise estatística, é a presença nos dados, de efeitos de factores não controlados, como por exemplo em agricultura, pequenas diferenças de fertilidade do solo, variações ligeiras no espaçamento ou na profundidade da sementeira e na constituição genética dos animais ou plantas.

## 2. O ensino da Estatística em *Agronomia*

O Instituto Superior de Agronomia (ISA), escola de referência no ensino agronómico em Portugal, foi criado em 1910, tendo-se instalado no espaço que agora ocupa, a Tapada da Ajuda, onde passou a formar engenheiros agrónomos e silvicultores. Após algumas alterações, actualmente o ISA dispõe das licenciaturas em Engenharia Agronómica, Engenharia Florestal e dos Recursos Naturais, Engenharia Alimentar, Arquitectura Paisagista, Engenharia do Ambiente, Engenharia Zootécnica e Biologia.

Nestas breves notas históricas, pretendemos deixar o testemunho da obra de alguns grandes nomes que, preocupados com os interesses da Escola e dos seus futuros licenciados e que, quer ensinando no ISA quer participando e apoiando trabalhos de investigação, souberam incentivar a necessidade de uma cultura estatística numa escola de *Agronomia*.

Das obras que tivemos oportunidade de consultar, gostaríamos de citar em primeiro lugar uma sebenta <sup>1</sup>, preparada de acordo com as lições de Sebastião e Silva, editada pela Associação de Estudantes do Instituto Superior de Agronomia, em 1954-1955. Doutorado em 1949 na Faculdade de Ciências de Lisboa, após concurso de provas públicas, Sebastião e Silva foi nomeado Professor Catedrático do Instituto Superior de Agronomia. Aí regeu, durante 10 anos, “Matemáticas Gerais” e “Cálculo Infinitesimal e das Probabilidades”. Sempre com perfeita consciência dos interesses desta Escola, objectivos e características próprias dos alunos que a escola pretendia formar, escreveu os textos recolhidos na colectânea *Cálculo das Probabilidades*. Cientista internacionalmente reconhecido, escreveu um texto de uma actualidade, rigor e didáctica, que merece ser consultado. Encontrámos a edição original na Biblioteca do ISA<sup>2</sup>, mas felizmente, foi reeditado pelo Serviço de Educação da Fundação Calouste Gulbenkian, na colecção Textos Didácticos, Vol. I, 1999.

Esta sebenta contém uma primeira parte “Introdução ao Cálculo das Probabilidades” com 126 páginas. Começa por ser introduzida a noção de frequência de um atributo e de um acontecimento sem, nesta primeira fase, se apresentar a definição formal de qualquer uma daquelas entidades. Desde logo há o cuidado de apresentar uma grande variedade de exemplos adequados às áreas do saber em *Agronomia*. Num dos primeiros exemplos (retirado de um trabalho de Daniel Nagore, *Biometria, nociones sobre este método de investigacion en genetica*, Ministério da Agricultura, Madrid, 1941) que se refere a dados recolhidos de uma subvariedade de trigo, faz a representação gráfica (histogramas e polígonos) das distribuições das variáveis em estudo. A independência e associação de atributos, construção e interpretação de tabelas de contingência são apresentados de seguida com uma introdução intuitiva aos testes de significância e testes de hipóteses.

Nesta altura introduz a função que designa “ $\chi^2$  dos desvios” referindo pressupostos a adoptar na sua utilização, indicando como tirar conclusões sobre uma experiência cujos resultados se encontrem numa tabela de contingência.

Já nesta fase dá breves sugestões sobre a importância de um papel mais interveniente que deve ser reconhecido à Estatística, escrevendo:

... *o papel da estatística não se limita à interpretação dos resultados experimentais: ela pode contribuir, com grande vantagem, para o planeamento da experiência.*

A secção seguinte é agora dedicada às Probabilidades. Apresenta uma introdução na qual mais uma vez nos deparamos com as elevadas qualidades pedagógica e didáctica de Sebastião e Silva. A noção de incerteza, com exemplos ilustrativos no âmbito das ciências biológicas e sociais, a axiomatização da probabilidade, a definição de probabilidade condicional e independência, seguem-se numa correcta e clara sequência. Passa ao caso

<sup>1</sup> Publicação académica de apontamentos das lições de um Professor Universitário.

<sup>2</sup> Uma nota de agradecimento à Biblioteca do ISA, que com a maior solicitude nos facultou a consulta desta obra e de outras que são seu património.

bivariado, a que chama *Sistema de duas experiências*, faz o estudo da distribuição de um par, define a condicional e a noção de independência; generaliza ao caso de *Sistemas de várias experiências*. São então consideradas a *distribuição binomial*, *multinomial* e é feita ainda uma breve referência à lei *normal*.

Na segunda parte destes apontamentos, que ocupa 70 páginas, Sebastião e Silva começa por tratar as *Distribuições de uma variável contínua* — e dado que até ali nunca foi apresentada a definição de variável aleatória, ela é então introduzida, naturalmente com pouca formalização. São estudados os momentos de uma variável aleatória, passando então ao caso de duas variáveis reais, com a introdução dos conceitos de correlação e regressão.

A distribuição normal é estudada com todo o pormenor. São muito ricos os exemplos apresentados, perfeitamente enquadrados. Sobre a justificação de se considerar como aproximadamente normais algumas variáveis escreve,

*... Note-se que os agrupamentos biológicos tais como as espécies, as raças e as variedades constituem populações praticamente infinitas, que vêm do passado e se prolongam no futuro, com uma certa constância de caracteres. Entre as constantes biométricas de uma tal população, figuram precisamente, os valores médios e os desvios padrões de atributos quantitativos....*

É tratada a propriedade reprodutiva da distribuição normal e é estudada a convergência da distribuição binomial para a normal. Termina esta parte com o estudo da distribuição qui-quadrado.

Coligido ainda nesta sebenta existe um aditamento publicado em 1957-1958, no qual, ao longo de 58 páginas é apresentado o estudo da regressão e da correlação. É estudado o ajustamento pelo método dos mínimos quadrados e o ajustamento com mudanças não lineares de variáveis.

Termina esta colectânea apresentando as distribuições de *Student* e *Fisher*, suas aplicações e ainda as noções de intervalo de confiança e intervalo de tolerância.

Sebastião e Silva cita como referências bibliográficas as obras de *Castelnuovo* (1933), *Finney* (1953), *Cramér* (1951), *Yule e Kendall* (1937), *Varenes e Mendonça* (1950), *Aitken* (1944), *Hald* (1952), *Lanotte* (1957), *Moroney* (1954) e *Sixto Rios* (1952).

Como vemos, de entre as referências citadas por Sebastião e Silva encontra-se uma sebenta de *Pedro de Varenes e Mendonça*<sup>3</sup>, então catedrático de “*Mecânica Racional e Teoria Geral de Máquinas*”. Como refere o próprio, em fins de Janeiro de 1950 foi encarregado de reger provisoriamente a disciplina “*Cálculo Infinitesimal e das Probabilidades*”. Reconhecendo não ser fácil para os alunos o estudo através dos livros então publicados e porque tinha tido já

*... assíduo contacto com a Estatística Matemática de maior importância na Investigação Agronómica... ,*

<sup>3</sup> À Professora Amarilis Varenes e Mendonça um muito obrigada por nos ter facultado cópias dos trabalhos de seu pai e a leitura do único exemplar que se conseguiu encontrar da referida sebenta.

decidiu redigir umas notas, para o que *dispôs de seis semanas, preenchidas por muitos outros afazeres*, que foram publicadas pela Associação de Estudantes do Instituto Superior de Agronomia em 1950, num total de 85 páginas.

Numa introdução muito interessante faz uma interpretação com pormenores didácticos de cunho pessoal, do conceito de acontecimento e de casos igualmente prováveis. Por exemplo,

... *sendo o lançamento de uma moeda exemplo clássico habitualmente introdutório de um capítulo de probabilidades, não é propriamente o lançamento real de uma moeda real que devemos imaginar, mas sim o lançamento ideal de uma moeda ideal.*

À definição de acontecimento e propriedades dos acontecimentos segue-se a de probabilidade. A noção de probabilidade condicional é introduzida no teorema da probabilidade composta, onde é demonstrado que (atente-se na notação)

*A probabilidade  $p$  de  $A$  é dada por*

$$p = P(A) = P(A_2)_{(A_1)} P(A_3)_{(A_1, A_2)} \dots P(A_m)_{(A_1, A_2, \dots, A_{m-1})}$$

*sendo  $A$  um acontecimento que consiste na realização de todos os acontecimentos  $A_i$  ( $i = 1, 2, \dots, m$ ).*

Surge a definição de independência de acontecimentos e segue-se a introdução do conceito de *variável casual ou aleatória* ainda em contexto discreto.

Introduz o conceito de função geradora, deduzindo a esperança matemática e os momentos. Termina a parte de probabilidades com o Teorema de Tchebicheff.

Faz o estudo das principais distribuições – binomial, multinomial, de Poisson, normal – deduzindo as suas propriedades, as bases das noções de convergência, a distribuição qui-quadrado de Pearson. Impressionante a preocupação pedagógica da demonstração<sup>4</sup> “da gênese do qui-quadrado, como soma de quadrados de  $n$  variáveis aleatórias independentes, normalmente distribuídas com valores médios nulos e desvios padrões unitários” – que, para ajudar a compreender o raciocínio, efectua para  $n=2, 3$  e finalmente  $n$ . Termina derivando a distribuição da variável  $z$  de Fisher e  $t$  de Student seguida de uma breve abordagem à distribuição normal bivariada e de umas breves noções de amostragem casual.

Porém, antes da publicação desta sebenta, encontramos trabalhos publicados por Varennes e Mendonça na área da probabilidade/estatística. Em *Das distribuições mais usadas em provas de significação* (1940), Varennes e Mendonça começa por apresentar com detalhe a distribuição qui-quadrado de Pearson, com demonstração rigorosa do teorema que estabelece a distribuição da soma dos quadrados de  $M$  variáveis  $x_1, x_2, \dots, x_M$ , normais estandardizadas e independentes.

<sup>4</sup> Referida algumas vezes no texto de Sebastião e Silva.

Deduz a distribuição conjunta de  $(\bar{x}, s)$  numa população normalmente distribuída. Com uma vasta referência às obras de Fisher, apresenta as distribuições *t-Student* e *z* de Fisher.

Em *Ortogonalidade e Análise de Variância* (1942), trabalho este publicado nos Anais do Instituto Superior de Agronomia<sup>5</sup>, Varennes e Mendonça faz uma revisão dos conceitos de ortogonalidade e de mudança de base. Segue-se uma generalização do Teorema de Fisher de 1925 e a sua aplicação ao caso de blocos casualizados. Ainda no mesmo ano, Varennes e Mendonça publica *Da terminologia estatística portuguesa*, onde escreve

... a maioria dos trabalhos é publicada em inglês e ingleses são os termos que Fisher foi obrigado a introduzir para designar os novos conceitos que criou e os que modificou.

Preocupado então com a não existência de termos portugueses equivalentes, apresenta uma lista, com a

... esperança que esta pequena contribuição possa fornecer uma base para trabalho futuro e para ... o progresso da ciência.

Afinal, a comunidade científica portuguesa reconhece ainda hoje a necessidade da existência de um glossário estatístico — trata-se de um desafio a que a Sociedade Portuguesa de Estatística terá que dar resposta!

Não resistimos à tentação de deixar aqui alguns dos termos propostos por Varennes e Mendonça, escolhidos propositadamente por diferirem dos que hoje usamos:

*Bias* (of an estimate) – excentricidade

*Bias* (of an experiment) – incorrecção

*biased estimate* – avaliação mal centrada

*Null hypothesis* – hipótese anulável

*Design* (of an experiment) – traçado ou projecto

*Estimate* ou *estimation* – avaliação

*Test* – prova; exemplos: *test of significance* – prova de significação,

*t-test* – prova de *t*

*Unbiased estimate* – avaliação bem centrada.

Mas se até aqui demos ênfase a obras escritas para o ensino, outras encontrámos que deixaram a marca da utilização da estatística no domínio agrónómico. Em 1957, Célio Pestana escreve o seu Relatório Final do curso de Engenheiro Agrónomo, com o título *Notas sobre a Estimação e Significância dos Parâmetros de Modelos Lineares Normais*.

<sup>5</sup> Este trabalho foi apresentado no Congresso Luso Espanhol para o Progresso das Ciências – Porto e a sua tradução, *Orthogonality and Analysis of Variance*, publicada na *Portugaliae Mathematica* (1942), Vol 3, 4, 234-252.

Não podemos deixar de realçar que se trata de um trabalho preparado sob a orientação de Sebastião e Silva e com o apoio de Pedro Braumann a quem agradece as ... *preciosas referências bibliográficas, algumas conversas tão proveitosas como agradáveis e um sem número de esclarecimentos.*

Deixemos aqui as notas de abertura da introdução do relatório

*... Por circunstâncias de ordem prática fui levado a tomar contacto com certos problemas da Estatística, particularmente com alguns aspectos do delineamento experimental. ... Baseei os primeiros estudos no que podia chamar-se "literatura de manual" e achava-me satisfeito por dominar a aplicação de algumas técnicas. Mas, a natural tendência para uniformizar num esquema vasto e simples o que então, por ignorância, se me afigurava desligado e particular, levou-me a meditar....*

Para a elaboração do relatório teve que estudar algumas questões de Álgebra, de Cálculo das Probabilidades e de Estatística em geral. Diz

*... As deduções são originais ou, pelo menos expostas de um modo original. E se dou importância a demonstrações que poderia referir para a literatura, é por pensar que só a sua compreensão permite encarar os assuntos em toda a sua beleza e utilidade <sup>6</sup>.*

Além de docente do ISA, Célio Pestana colaborou activamente na orientação de muitos trabalhos em que o delineamento experimental e a análise estatística eram fundamentais.

Mas, outras obras surgiram quando também em Portugal começa a reconhecer-se que, a investigação em *Agronomia* está fortemente ligada à necessidade de usar metodologias estatísticas.

Mário Rodrigues de Carvalho publica em 1946 na Colecção "A Terra e o Homem", a obra *A Estatística na Experimentação Agrícola*. Vemos que se trata de um trabalho contemporâneo aos clássicos *Statistical Methods for Research Workers* (1944) e *The Design of Experiments* (1947) de Ronald Fisher, bem como *Sampling Methods for Censuses and Surveys* (1949) de Frank Yates.

Mário de Carvalho, faz um historial muito interessante de como se fazia inicialmente a experimentação agrícola:

*... No seu início a experimentação agrícola limitou-se a comparar um tratamento com outro.... Não vai longe o tempo em que, para se saber por exemplo o valor de uma variedade de trigo, se semeavam dois talhões de igual área levando um a*

<sup>6</sup> Como passaria hoje em dia esta mensagem?

*forma desconhecida e outro o trigo da região com o qual se desejava fazer o respectivo confronto. Havia o cuidado de escolher um terreno o mais homogéneo possível onde se marcavam os dois talhões, cada um deles recebia igual adubação e a sementeira era efectuada no mesmo dia.*

*Como se sabia que as condições climatéricas diferiam de uns anos para os outros, ..., repetia-se esse mesmo ensaio três anos, ao fim deles tirava-se as médias aritméticas e sobre elas concluía-se à cerca do valor dos novos tipos.*

Como o autor refere foi um livro escrito para os que não sabiam estatística, mas sentiam a necessidade da sua aplicação a muitos dos trabalhos de investigação e experimentação que pretendiam desenvolver ou estavam a realizar.

Mário Rodrigues de Carvalho edita ainda em 1988 um outro livro *A Estatística Aplicada à Experimentação Agrícola*, com uma orientação ligeiramente diferente da usada no livro anterior.

Ao longo do seu percurso como escola de referência no ensino agronómico, o Instituto Superior de Agronomia percebeu entretanto, como dizia já Sebastião e Silva, a necessidade de existir uma disciplina *Estatística e Delineamento Experimental*, que a reforma de 1986 veio a integrar na maioria das licenciaturas da Escola. No entanto em 1999 uma nova reforma retirou esta disciplina do plano de estudos. É sem dúvida um retrocesso no reconhecimento que durante décadas cientistas eminentes tão bem testemunharam, mas que é o resultado (acredito!!) da tendência que estamos a atravessar nas nossas escolas – a de retirar dos currículos de muitos cursos, as disciplinas de índole matemática. Acreditamos que em futuro muito próximo esta lacuna, que passou a existir nos planos curriculares do ISA, seja preenchida.

Nas últimas décadas, num período que inclui já o nascimento da Sociedade Portuguesa de Estatística, a Estatística viu os seus métodos teóricos e práticos serem fortemente influenciados pelo desenvolvimento fulgurante dos meios de cálculo: estamos na era da computação. Muitos cálculos extremamente pesados, metodologias válidas apenas sob certos pressupostos ou resultados obtidos após complicados e aproximados desenvolvimentos, puderam encontrar na *força da computação* uma ajuda preciosa. Consideramos, todavia, que esta fase só pode ser vista com uma maior responsabilização na utilização das ferramentas agora colocadas à disposição dos investigadores; a fundamentação teórica das técnicas e metodologias usadas é insubstituível.

Em 1976, António St. Aubyn, então professor na Faculdade de Ciências de Lisboa, é convidado a leccionar no Instituto Superior de Agronomia. Desde então tem sido um defensor acérrimo da cultura estatística em *Agronomia*. Consciente das mudanças que a era da computação estava inevitavelmente a introduzir, deixará o seu nome definitivamente ligado à criação de um Centro de Informática no Instituto Superior de Agronomia, contribuindo para um novo período da história do ensino no ISA. Foi ele o criador e primeiro director do Centro e, quando ainda outras escolas estavam a começar de introduzir *software* estatístico de ponta, já António St. Aubyn o tinha no ISA.

## Referências Bibliográficas

- [1] Amaral-Turkman, M. A. (1999). Reflexões sobre o “Passado da Estatística”. *Boletim Informativo da Sociedade Portuguesa de Estatística* 2 e 3, 31-36.
- [2] Branco, J. A. (1999). Aspectos da Realidade Estatística em Portugal. *Boletim Informativo da Sociedade Portuguesa de Estatística* 1, 19-23.
- [3] Cochran, W. G. (1945). Training at the professional level for statistical work in agriculture and biology. *J. Amer. Stat. Ass.*, 40, 160-166.
- [4] Dagnelie, P. (1995). Statistique, Biométrie, Agronomie: Approche Historique. *C.R. Acad. Agric. Fr.*, 81, n° 8, 33-38.
- [5] Fisher, R. (1944). *Statistical Methods for Research Workers*. Oliver and Boyd LTD. (9ª Edição).
- [6] Fisher, R. (1947). *The Design of Experiments*. Oliver & Boyd (4ª Edição).
- [7] Misra, S. C., Sahai, H., Gore, A. P. and Garrett, J. K. (1987). A bibliography on the teaching of probability and statistics. *The Amer. Stat.*, 41, 284-310.
- [8] Pestana, C. G. A. (1957). *Notas Sobre a Estimação e Significância dos Parâmetros de Modelos Lineares Normais*. Relatório Final do Curso de Engenheiro Agrônomo. Instituto Superior de Agronomia.
- [9] Rodrigues de Carvalho, M. J. (1946). *A Estatística na Experimentação Agrícola*. Coleção “A Terra e o Homem”, Livraria Sá da Costa.
- [10] Sebastião e Silva, J. (1954-55). *Cálculo das Probabilidades*. Edições AGROS. Associação dos Estudantes do Instituto Superior de Agronomia.
- [11] Sousa, A. P. F. (1995). *História da Estatística em Portugal*. Instituto Nacional de Estatística.
- [12] Yates, F. (1949). *Sampling Methods for Censuses and Surveys*. Charles Griffin & Company Limits.
- [13] Tiago de Oliveira, J. (1987). Um projecto de Matemática Aplicada – O DEIOC. In *Collected Works/Obras* (org. Tiago de Oliveira, J. C., 1995) Vol II, 151-165, Pendor.
- [14] Tiago de Oliveira, J. (1988). A herança de Bernoulli-Bayes-Laplace; um esboço da história da Estatística. In *Collected Works/Obras* (org. Tiago de Oliveira, J. C., 1995) Vol II, 151-165, Pendor.
- [15] Varennes e Mendonça, P. (1940). Das distribuições mais usadas em provas de significação. *Revista Agronómica*, XXVIII, 1-20.
- [16] Varennes e Mendonça, P. (1942). Da terminologia estatística portuguesa. *Anais do Instituto Superior de Agronomia*, Vol XIII.
- [17] Varennes e Mendonça, P. (1942). Ortogonalidade e Análise de Variância. *Anais do Instituto Superior de Agronomia*, Vol XIII.
- [18] Varennes e Mendonça, P. (1950). *Noções de Cálculo das Probabilidades*. Instituto Superior de Agronomia.

## **Maria Manuela Costa Neves Figueiredo**

É Professora Catedrática no Departamento de Matemática do Instituto Superior de Agronomia (ISA).

Licenciou-se em Matemática Aplicada - Estatística e Computação (1976) pela FC/UL. Doutorou-se em Matemática, especialidade de Estatística (1990) pela FCT/UNL e realizou provas de Agregação em Matemática pelo ISA/UTL (2003).

Tem leccionado disciplinas de licenciatura no domínio da *Estatística*, *Probabilidades* e *Análise* e tem colaborado em vários cursos de mestrado no ISA e exteriores ao ISA.

Além da publicação de textos de apoio às aulas de licenciatura e mestrado, tem publicado vários trabalhos na sua área de investigação, *Estatística de Valores Extremos*, *Simulação* e *Estatística Computacional*, além de outras publicações de índole mais aplicada, resultado de colaboração noutras áreas científicas.

Actualmente é Presidente do Departamento de Matemática do ISA, membro da Assembleia de Representantes do ISA e Membro da Direcção da Sociedade Portuguesa de Estatística.

Foi Presidente da Comissão Organizadora do VIII Congresso da SPE (2000).



# Sobre o Estudo de Séries Temporais e Previsão

## Breve Nota Pessoal

*Nuno Crato*<sup>1</sup>

A própria designação é polémica. Há quem fale em «séries temporais» ou «crono-séries» e quem prefira «sucessões cronológicas». Mas nenhuma das expressões é perfeita. A primeira tem o inconveniente de usar a palavra «séries», que tecnicamente designa sucessões de somas parciais. A última, a desvantagem de excluir implicitamente os casos de tempo contínuo. Há alguns anos levávamos esta discussão a sério. Hoje estamos apenas preocupados com o conteúdo da disciplina. E como todos nós lemos as revistas internacionais, naturalmente em língua inglesa, «séries temporais» soa a «time series» e sai-nos mais naturalmente.

Foi um pequeno avanço, mas os verdadeiros avanços nos últimos 25 anos foram muito marcados. Seria bom fazer a sua história e muito gostaria de o poder fazer. Mas apenas posso, neste momento, dar um pequeno testemunho pessoal, quase forçado por um insistente e amável convite do editor, mas orgulhoso por poder prestar este breve testemunho na altura em que a nossa Sociedade comemora um quarto de século.

O estudo moderno das séries temporais começa nos princípios do século XX com várias técnicas e modelos parcelares. Entre essas técnicas salienta-se, naturalmente, a regressão, a análise de autocorrelação e o periodograma de Schuster. Entre os modelos, salienta-se a autoregressão de Yule e as médias móveis de Slutsky (V. Klein, 1997). Estamos a 80 anos ou mais desses marcos pioneiros, mas nessa altura notava-se já a tentativa de ver os choques aleatórios como processos de geração dos dados observados, afastando a visão tradicional de erros residuais como insuficiências da observação. O próprio Slutsky falava do «papel predominante na natureza das somas móveis» e de uma «conexão entre flutuações aleatórias e ondas mais ou menos regulares» (Slutsky 1937).

Sabe-se que se teve de esperar por Kolmogorov e por Khinchin para ter definições operacionais de processos estacionários, por Wiener e, mais uma vez, por Kolmogorov, para ter uma teoria linear de processos estacionários à segunda ordem, e por vários

---

<sup>1</sup> Agradeço à prof. Antónia Turkman algumas informações preciosas que me forneceu. As muitas insuficiências e possíveis erros deste texto são, como é evidente, da minha inteira responsabilidade.

investigadores, já depois da guerra, para começar um estudo operacional das séries temporais. A grande síntese aparece em 1970, com a célebre obra de Box e Jenkins (1970). A partir desse momento, os modelos lineares ARIMA tornam-se a referência de base dos modelos de séries temporais e o método de análise preconizado por estes autores passa a chamar-se «método de Box-Jenkins».

Em meados dos anos 80, quando comecei a interessar-me por análise de séries temporais e por previsão, os modelos ARIMA estavam pois estabelecidos na literatura estatística e em alguma literatura econométrica. Estavam a aparecer os modelos de espaço de estados. O filtro de Kalman começava a sair da área do controlo e da engenharia para se divulgar entre estatísticos e econometristas, que têm sempre andado a par e passo nestas áreas. Mas tanto quando pude na altura perceber, isso era o que se passava no mundo. Em Portugal, ou no cantinho de Portugal que eu conhecia, esses avanços eram ainda desconhecidos.

E quando eram conhecidos eram olhados com alguma desconfiança. A ideia de modelos de diferenças estacionárias não era bem vista. Faltava-lhe uma variável explicativa... Continuava a ser difícil ultrapassar a ideia de uma aleatoriedade puramente residual para passar a vê-la como geradora de processos observados. A econometria baseada em modelos de parâmetros fixos continuava dominante.

No entanto, quem trabalhava no meio empresarial estava confrontado com a necessidade de estabelecer modelos de previsão que funcionassem melhor que uma simples extrapolação gráfica. Comigo, isso aconteceu numa empresa onde arranjei um primeiro emprego e onde fui encarregue de estabelecer modelos de previsões de preços para a construção civil. Na altura, a inflação estava ainda pelos dois dígitos e havia muito dinheiro em jogo na revisão de preços das empreitadas de obras públicas. Os modelos econométricos de parâmetros fixos ou de decomposição clássica não serviam. As previsões que davam eram piores que as que qualquer técnico poderia fazer usando apenas a intuição. Não se conseguia prever preços com modelos de tendência fixa a que se somavam componentes aleatórias.

O problema, no entanto, é que nenhum modelo de tendência fixa, aquilo a que hoje chamamos «estacionário em torno de uma tendência», poderia fazer previsões razoáveis. A média de crescimento não fornecia boas previsões para o crescimento no passo seguinte, tal como a recta ou curva ajustada não tinha em conta um necessário peso maior para as observações mais recentes.

No entanto, alguma coisa existia e tinha existido. Em 1953 Bento Murteira publicou a sua tese de doutoramento, orientada por Maurice Kendall em 1950-51 na London School of Economics (Murteira, 1953). A tese deduzia propriedades de modelos autoregressivos e o trabalho deu origem a uma publicação na *Biometrika* (Murteira, 1951). No estilo pedagógico que lhe é peculiar, Bento Murteira prosseguiu as suas explicações em alguns trabalhos publicados nos *Anais do ISCEF* (V. AAVV 1994). Em fins da década de 60, Tiago de Oliveira introduzia séries temporais na cadeira de Estatística Matemática dada aos alunos de Matemática Aplicada da Faculdade de Ciências da Universidade de Lisboa.

Nos princípios da década de 80, estava eu a fazer um mestrado e, numa das aulas, perguntei se não íamos falar de Box-Jenkins. Parecia que não. Um dos colegas achou mesmo tão curioso o nome que desenhou um ringue de boxe, com um tal Jenkins lá dentro — devia ser disso que se tratava: o boxe do Jenkins...

Numa cadeira, para estabelecer um modelo Monte Carlo para os preços futuros que tinham várias componentes, simulei passeios aleatórios com tendência e agreguei os resultados. Não foi bem visto: onde estavam as variáveis explicativas?

Noutra altura, num intervalo das aulas, numa conversa com professores, falou-se de métodos de alisamento exponencial e de modelos ARMA. Houve quem se manifestasse contra a «econometria de encher chouriços», numa referência ao carácter *ad hoc* destes modelos, que contrastam com a fundamentação teórica das equações econométricas clássicas.

Poucas semanas depois, Andrew Harvey apareceu em Lisboa e deu uma série de palestras no ISEG. Lembro-me que começou a primeira dizendo qualquer coisa como «Without studying time series, no one can understand what we are doing in econometrics today». Para mim foi um alívio.

Entretanto, as coisas começavam a mexer. Em alguns cursos de gestão e no mestrado de Investigação Operacional entretanto lançado no Instituto Superior Técnico em Lisboa, começou a estudar-se «Métodos de Previsão», que era o nome dado ao estudo prático de séries temporais, ou seja, um pouco de regressão, métodos de alisamento e umas referências aos ARIMA de Box-Jenkins. Na Faculdade de Ciências da Universidade de Lisboa, Feridum Turkman é contratado para leccionar uma disciplina individualizada de séries temporais — talvez, como tal, a primeira no nosso país. Entretanto, Bento Murteira — mais uma vez ele — faz umas folhas de métodos de previsão em que explica sistematicamente os modelos ARIMA. Em Coimbra, termina o seu doutoramento Nazaré Mendes-Lopes (1982 e 1985) e, com Esmeralda Gonçalves, que estuda generalizações dos processos autoregressivos e modelos de memória longa (Gonçalves 1987), surge um núcleo de investigação da área de séries temporais.

Estamos em 2005 e os anos 80 parecem já distantes. Estão, de facto. Felizmente. Hoje temos doutorados especializados em séries temporais em todas as grandes universidades. Nos congressos internacionais, nomeadamente no International Symposium on Forecasting, onde em princípios dos anos 80 aparecia ocasionalmente um ou outro português, os nossos investigadores marcam uma presença regular. De tal forma que em 2000 esse simpósio foi realizado no nosso país com organização de uma equipa de investigadores das universidades de Lisboa e do Porto. Nesse evento inscreveram-se mais de 40 portugueses. Organizam-se entretanto eventos com larga participação. De 1997 a 2001 realizou-se anualmente uma semana no Convento da Arrábida intitulada «Recent Developments in Time Series» por onde passaram Robert Engle, Peter Brockwell e outros investigadores de elevada craveira internacional. Em Faro, no Porto, Coimbra e Lisboa há frequentemente encontros e seminários em que os mais activos investigadores mundiais participam. O progresso foi pois imenso.

Talvez esta seja apenas uma visão pessoal, mas o grande impulso no estudo das séries temporais no nosso país parece-me devido a dois factores, nem sempre ambos presentes, mas no conjunto complementares: o contacto com a vida empresarial e com outros problemas reais, como seja a análise de séries climáticas, e o contacto com a investigação internacional através de palestras de visitantes ou com visitas prolongadas ao exterior para doutoramento. Há hoje problemas, é certo. A investigação, tanto teórica como aplicada, nesta área como noutras, não se desenvolveu como deveria. Mas os problemas já não são, certamente, aqueles com que há 25 anos nos defrontávamos.

### Referências Bibliográficas

- AAVV (1994). *Jubileu Bento Murteira — ISEG, 17 de Junho de 1994*, Lisboa.
- Box, George E.P. e Gwilym M. Jenkins (1970). *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco.
- Klein, Judy L. (1997). *Statistical Visions in Time: A History of Time Series Analysis, 1662–1938*, Cambridge University Press, Cambridge, U.K.
- Gonçalves, Esmeralda (1987). Une généralisation des processus ARMA, *Annales d'Economie et de Statistique* **5**, 109–145.
- Murteira, Bento (1951). Note on the variate differences of autoregressive series, *Biometrika* **38**, 479–480, agora in AAVV (1994).
- Murteira, Bento (1953). *Algumas Propriedades dos Processos Auto-Regressivos*, Lisboa.
- Nazaré Mendes-Lopes (1982). Problèmes d'estimation dans les processus ponctuels chromatique, Thèse de Doctorat de 3ème cycle, Université Pierre et Marie Curie (Paris VI).
- Nazaré Mendes-Lopes (1982). Análise estatística dos processos pontuais cromáticos, Tese de Doutoramento em Ciências, Universidade de Coimbra, 1985.
- Slutsky, Eugen H. (1937), The summation of random causes as the source of cyclic processes, *Econometrica* **5**, 105–146 (trad. do original russo de 1927).

## **Nuno Crato**

É Professor Associado com Agregação no Instituto Superior de Economia e Gestão, em Lisboa. Doutorou-se em Matemática Aplicada com Howard M. Taylor, em Delaware, nos Estados Unidos em 1992. Ensinou no ISEG, na Universidade dos Açores, no Stevens e no New Jersey Institutes of Technology. É investigador no Centro de Matemática Aplicada à Previsão e Decisão Económica (Cemapre), em Lisboa.

Tem trabalhado em séries temporais, previsão e aplicações de processos estocásticos em finanças, computação e problemas climáticos. Tem trabalhos publicados em diversas revistas internacionais da especialidade, nomeadamente *Statistical Papers*, *Journal of Econometrics*, *Economic Letters*, *International Journal of Forecasting*, *Communications in Statistics* e *Journal of Forecasting*.

É membro da Sociedade Portuguesa de Estatística, da American Statistical Association e de outras associações científicas. Foi presidente do International Symposium on Forecasting em 2000. É actualmente presidente da Sociedade Portuguesa de Matemática e membro dos corpos gerentes do Fórum Internacional de Investigadores Portugueses, de que é sócio fundador.

Em paralelo com o seu trabalho académico, está empenhado na divulgação científica. Colabora regularmente na imprensa e é autor de várias obras de divulgação. A Sociedade Europeia de Matemática atribuiu-lhe em 2003 o Primeiro Prémio do concurso Public Awareness of Mathematics pelo seu trabalho de divulgação da matemática.