

# Boletim



# SPE

Sociedade Portuguesa  
de Estatística

Publicação semestral

outono de 2020



## 40 anos SPE: De onde viemos? Onde estamos? Para onde vamos?

<b>Probabilidades e Estatística no Departamento de Matemática da Universidade de Coimbra: uma perspetiva</b> Esmeralda Gonçalves e Paulo Eduardo Oliveira .....	14
<b>Alguns tópicos de investigação em estatística realizada no CIDMA – uma breve discussão</b> Pedro Macedo, Ana Helena Tavares e Marco Costa .....	16
<b>Pseudovalorers-p na construção do conhecimento científico</b> Maria de Fátima Brilhante .....	24
<b>Um olhar sobre o desenvolvimento de modelos de sobrevivência para acontecimentos recorrentes</b> Ivo Sousa-Ferreira, Ana Maria Abreu e Cristina Rocha .....	27
<b>O papel da Estatística na Universidade do Porto</b> Conceição Nunes Rocha e Isabel Silva .....	36
<b>Ser SPE – A vantagem do ónus da Boa Estatística</b> Jessica Silva Lomba .....	41
<b>UM departamento com ESTATÍSTICA/Departamento de Matemática/Universidade do Minho (UM)</b> Ana Paula Amorim e outros .....	44
<b>SPE – 40 anos de “braço dado” com a Estatística</b> Célia Nunes .....	48
<b>Estatística na Universidade de Évora</b> Russell Alpizar-Jara .....	51
<b>Uma caminhada por valores extremos de precipitação na Ilha da Madeira</b> Délia Gouveia Reis .....	55
<b>Alguns desafios para o futuro da SPE</b> Kamil Feridun Turkman .....	58
<b>... quarenta anos de sociedade; quarenta e cinco de investigação...</b> Tiago A. Marques e Lisete Sousa .....	60
<b>Testemunhos de Presidentes da SPE .....</b>	64

Editorial .....	1
Mensagem da Presidente .....	2
Notícias .....	3
<i>Enigmística</i> .....	7
Episódios na História da Estatística .....	8
SPE e a Comunidade .....	10
Ciência Estatística .....	80

### Informação Editorial

**Endereço:** Sociedade Portuguesa de Estatística,  
Campo Grande. Bloco C6. Piso 4.

1749-016 Lisboa. Portugal.

**Telefone:** +351.217500120

**e-mail:** [spe@spestatistica.pt](mailto:spe@spestatistica.pt)

**URL:** <https://www.spestatistica.pt>

**ISSN:** 1646-5903

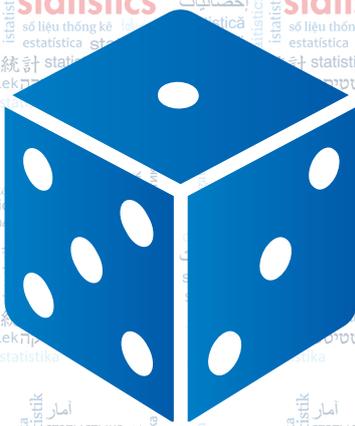
**Depósito Legal:** 249102/06

**Tiragem:** Edição digital

**Execução Gráfica e Impressão:** Gráfica SobreireNSE

**Editor:** Fernando Rosado, [fernando.rosado@fc.ul.pt](mailto:fernando.rosado@fc.ul.pt)

Sociedade Portuguesa de Estatística desde 1980



# SPE

## Sociedade Portuguesa de Estatística



### XXV Congresso

## Sociedade Portuguesa de Estatística

### 2021 Évora



# SPE 2021

XXV Congresso da Sociedade Portuguesa de Estatística  
13 a 16 de Outubro de 2021 | Évora

<http://www.spe2021.uevora.pt>

# Editorial

**... aos quarenta; para onde vamos...**

O “*e-Boletim SPE primavera de 2020*” foi uma surpresa. As mensagens recebidas são confirmatórias de que esse facto teve pouca influência no desempenho que se deseja para que ele cumpra a sua missão.

Após essa “edição com alterações de última hora”, já com a maquete terminada, como noticiado no Boletim primavera 20, p. 6; este Boletim SPE é o primeiro do “novo tempo” também no espaço editorial. A surpresa que a pandemia tem gerado aos mais diversos níveis faz aprofundar a análise dos métodos tradicionais, alguns dos quais com bastante carga rotineira. E, essa análise, ajuda a novas descobertas e a outras opções. E uma delas envolve a incerteza e a resposta “veremos como evolui” surge para as mais variadas decisões adiadas “até ver”.

Desde então tudo se mantém, de tal modo que faz sentido a decisão de que também esta edição do Boletim SPE seja apenas com a versão digital.

2. O *Boletim SPE outono de 2020*, fica associado às comemorações dos 40 anos SPE. Na época especial que vivemos, todos os programas de celebração são “o possível” e exigentes para dignificar suficientemente os momentos a festejar. Assim também com a Sociedade Portuguesa de Estatística.

É um bom momento para, em Tema Central, podermos procurar respostas para uma conjugação de três questões – *De onde viemos? Onde estamos? Para onde vamos?* – sempre oportunas para gerar uma perspetiva de crescimento, como se deseja.

O Mote foi a arte, num quadro de Gauguin. Uma longa tela de quatro metros que também envolve uma “árvore do conhecimento” e pintada em apenas um mês. Com um título desenhado, fora do comum, em maiúsculas no canto do quadro, foi aqui que nos inspirámos com a alteração de “*Quem somos?*” para “*Onde estamos?*”. Referências de que Gauguin indicou que a tela devia ser lida da direita para a esquerda com a conseqüente interpretação dos principais grupos para além de descrever um processo evolutivo na vida foi um estímulo para os nossos objetivos com a presente edição aniversária.

Com aquele lema como guia para os autores, formulámos convites e solicitámos a colaboração daqueles cujas agendas pessoais permitissem e de modo a termos a maior diversidade e representatividade da Ciência Estatística nas Universidades e na Investigação portuguesas. Mais uma vez, a generosidade dos autores conseguiu que fizéssemos “uma viagem pelo país” e que este Boletim descreve. Em data festiva, também presentes os mais recentes Premiados SPE 2019 (Cf. Boletim outono 2019, p. 57-60) e os “Testemunhos de Presidentes da SPE”.

Foi, assim o desejamos, criada uma edição memorial com “passado, presente e futuro na SPE”.

Esta edição teve a preciosa ajuda editorial da Lisete Sousa e do Tiago Marques.

3. A presente época pandémica, diferente a todos os níveis, congregou e estimulou a comunidade científica no sentido apelativo de “pesquisa e sucesso” para as diversas questões que possam conduzir a “uma solução”. Assim, também na comunidade dos estatísticos portugueses. Registam-se seminários e conferências que o confirmam. A SPE tem devolvido diversas atividades, desde logo, “(...) congratulando-se com o empenho da comunidade científica na análise dos dados relativos à pandemia (...) e assim propiciando condições para uma atuação informada das entidades responsáveis” como se refere no sítio da internet.

Eis bons motivos para nos dedicarmos ao assunto; na próxima edição.



O Tema Central do próximo Boletim SPE será:  
*Especial Covid: a Estatística ao serviço da sociedade*

# Mensagem da Presidente

Caros sócios da SPE

O ano atípico e difícil que vivemos impediu-nos de realizar a maior parte das iniciativas que tínhamos planeado para a comemoração dos 40 anos da SPE. Quero agradecer ao nosso caro Editor e à sua equipa, incluindo os autores, todo o esforço que permitiu fazer este boletim assinalando o aniversário da SPE. Apesar das dificuldades que todos sentimos ao longo deste 2020, conseguimos manter os Prémios Estatístico Júnior com a preciosa colaboração do CMUC, a presença na VII Feira da Matemática nos dias 23 e 24 de outubro de 2020 com a colaboração do INE e ter presença na imprensa no Dia Mundial da Estatística. É claro que tudo isto só é possível com a dedicação dos sócios que muitas vezes sacrificam o seu horário de descanso e de convívio familiar.

Esta é a minha última “Mensagem da Presidente”. Foram seis anos de experiências únicas, durante os quais aprendi muito e conheci pessoas excecionais mas durante os quais também ganhei alguns cabelos brancos. O trabalho realizado ficou aquém daquele que eu tinha em mente há 6 anos mas foi o possível nas circunstâncias. E foi o possível principalmente devido às minhas caras colegas de Direção: Cláudia Nunes, Conceição Amado, Esmeralda Gonçalves, Isabel Pereira e Patrícia Bermudez. Sem o seu apoio e dedicação não teria sido possível. Agradeço também o apoio e colaboração dos outros Órgãos Sociais, a saber: a Mesa da Assembleia Geral e o Conselho Fiscal. Quero também agradecer a todos os sócios que de uma maneira ou de outra, colaboraram com a Direção em prol da SPE e da sua coesão, do desenvolvimento da Estatística e da sua afirmação na Sociedade.

Num país da dimensão de Portugal e numa Sociedade da dimensão da nossa SPE, as atividades desenvolvidas, ainda que pareçam poucas, só são possíveis pelo esforço e espírito de autossacrifício de muitos sócios.

Valeu a pena? Como diz o nosso maior poeta, “Tudo vale a pena quando a alma não é pequena”.

Foi uma honra ter representado a SPE e todos os seus sócios.

Cordiais saudações

Maria Eduarda Silva

Porto, 25 de Outubro de 2020

# Notícias

## • III Dia Mundial da Estatística

No dia 20 de outubro de 2020, celebrou-se este ano mais um Dia Mundial da Estatística. O tema escolhido foi: «Ligar o mundo com dados de confiança». Este mote reflete a importância da credibilidade, de dados com autoridade e da inovação nos bens públicos a relevar nos sistemas nacionais de estatística.



Pela primeira vez em

20 de outubro de 2010

e em mais de 100 países, foi celebrado o primeiro Dia Mundial da Estatística.

O segundo Dia Mundial da Estatística decorreu em 2015



Os nossos colegas João Branco e Tiago Marques apresentaram duas publicações nos jornais Observador e Público, respetivamente.

*REVSTAT - Statistical Journal* organizou um *webinar*, como se informa neste Boletim.

Mais notícias e desenvolvimentos sobre o Dia Mundial da Estatística, também nos Boletins outono de 2010 e de 2015 e Boletins primavera de 2011 e de 2016.

Detalhes em <https://www.spestatistica.pt/> e [worldstaticsday.org](http://worldstaticsday.org).



Publicação científica de referência, de acesso aberto com revisão pelos pares, constituída por artigos de elevado interesse científico que contribuem para o desenvolvimento da Ciência Estatística, focada em teorias inovadoras, métodos e aplicações nas diferentes áreas do conhecimento.

Em 2019, a REVSTAT - Statistical Journal lançou o Volume 17 - Números 1 a 4, com os artigos listados abaixo (<https://www.ine.pt/revstat/tables.html>).

**Volume 17, Issue 1:**

- “Some monitoring procedures related to asymmetry parameter of Azzalini’s skew-normal model” by Chenglong Li, Amitava Mukherjee, Qin Su and Min Xie;
- “On the design points for a rotatable orthogonal central composite design” by Christos P. Kitsos;
- “Random environment INAR models of higher order” by Aleksandar S. Nastiaë, Petra N. Laketa and Miroslav M. Ristiaë;
- “Stochastic inequalities for the run length of the EWMA chart for long-memory processes” by Yarema Okhrin and Wolfgang Schimid;
- “Using shrinkage estimators to reduce bias and MSE in estimation of heavy tails” by Jan Beirlant, Gaonyalelwe Maribe and Andréhette Verster;
- “An integrated functional Weissman estimator for conditional extreme quantiles” by Laurent Gardes and Gilles Stupfler.

**Volume 17, Issue 2 (Special Issue on Biometry):**

- “Modeling risk of extreme events in generalized Verhulst models” by M. Fátima Brilhante, M. Ivette Gomes and Dinis Pestana;
- “Modeling large values of systolic blood pressure in the Portuguese population” by C. P. Caetano and P. de Zea Bermudez;
- “Testing conditions and estimating parameters in Extreme Value theory: Application to environmental data” by Helena Penalva, Dora Prata Gomes, M. Manuela Neves and Sandra Nunes;
- “On the parameters estimation of HIV dynamic models” by Diana Rocha, Sónia Gouveia, Carla Pinto, Manuel Scotto, João Nuno Tavares, Emília Valadas and Luís Filipe Caldeiray;
- “Accuracy measures for binary classification based on a quantitative variable” by Rui Santos, Miguel Felgueiras, João Paulo Martins and Liliana Ferreira;
- “Joint modelling of longitudinal and competing risks data in clinical research” by Laetitia Teixeira, Inês Sousa, Anabela Rodrigues and Denisa Mendonça.

**Volume 17, Issue 3:**

- “Classical and Bayesian componentwise predictors for non-compact correlated ARH(1) Processes” by M. Dolores Ruiz-Medina and Javier Álvarez-Liébaná;
- “On weighted Kullback–Leibler divergence for doubly truncated random variables” by Rajesh Moharana and Suchandan Kayal;
- “The Beta Marshall–Olkin Lomax distribution” by Claudio J. Tablada and Gauss M. Cordeiro;
- “The CUSUM median chart for known and estimated parameters” by Philippe Castagliola, Fernanda Otilia Figueiredo and Petros E. Maravelakis;

- “AP-optimum designs for minimizing the average variance and probability-based optimality” by N.M. Kilany and W.A. Hassanein;
- “An Information theoretical method for analyzing unreplicated designs with binary response” by Krystallenia Drosou and Christos Koukouvinos;
- “Prediction intervals of the record-values process” by Amany E. Aly, H.M. Barakat and Magdy E. El-Adll.

#### **Volume 17, Issue 4:**

- “Confidence intervals and regions for the generalized inverted exponential distribution based on progressively censored and upper records data” by Ismail Kinaci, Shuo-Jye Wu and Coskun Kus;
- “A new robust partial least squares regression method based on a robust and an efficient adaptive reweighted estimator of covariance” by Esra Polat and Suleyman Gunay;
- “Reliability aspects of proportional mean residual life model using quantile functions” by N. Unnikrishnan Nair, P.G. Sankaran and S.M. Sunoj;
- “Generalized estimators of stationary random-coefficients panel data models: Asymptotic and small sample properties” by Mohamed Reda Abonazel;
- “Repeated measures analysis for functional data using Box-type approximation - With applications” by Lukasz Smaga;
- “Predictive estimation of population mean in ranked set sampling” by Shakeel Ahmed, Javid Shabbir and Sat Gupta;
- “A review of the Behrens–Fisher problem and some of its analogs: Does the same size fit all?” by Sudhir Paul, You-Gan Wang and Insha Ullah;
- “Forecasting daily exchange rates: A comparison between SSA and MSSA” by Rahim Mahmoudvand, Paulo Canas Rodrigues and Masoud Yarmohammadi.

### **Celebração do Dia Mundial da Estatística**

**20 de Outubro de 2020, 17:00**

#### **WEBINAR - Statistics in times of Pandemics**

A REVSTAT-Statistical Journal organizou um webinar em Celebração do Dia Mundial da Estatística (WSD, <https://worldstatisticsday.org/>), 20 de Outubro de 2020, promovido pelo Centro de Estatística e Aplicações da Universidade de Lisboa (CEAUL, <http://ceaul.org/>) e pela Sociedade Portuguesa de Estatística (SPE, <https://www.spestatistica.pt/>). Este contou com a apresentação científica relacionada com o Artigo

Convidado com Discussão sobre a COVID-19, enquadrado também no lançamento do número especial da REVSTAT - Statistical Journal (<https://www.ine.pt/revstat/tables.html>), revista científica internacional de acesso aberto, editado pelo Instituto Nacional de Estatística (INE, <https://www.ine.pt/revstat/>).

*Statistics in times of Pandemics: The role of Statistical and Epidemiological methods during the COVID-19 emergency* por **Baltazar Nunes** (DE, Instituto Nacional de Saúde Dr. Ricardo Jorge; CISP, Escola Nacional de Saúde Pública, Universidade NOVA de Lisboa)

*A brief appraisal of the COVID-19 Pandemic in Portugal* por **Gabriela Gomes** (University of Strathclyde, Glasgow, United Kingdom; Universidade do Porto)

## • CIDMA da Universidade de Aveiro

No passado dia 8/julho realizou-se, via online, o VIII Workshop of Probability and Statistics group — Interdisciplinarity and applications, inserido nas atividades do grupo de investigação Probabilidades e Estatística do Centro de Investigação e Desenvolvimento em Matemática Aplicada (CIDMA) da Universidade de Aveiro. Nesta edição foi recordado o 40º aniversário da SPE. O programa do encontro pode ser consultado visitando a página <https://sites.google.com/view/workshops-ps-cidma>.

AF

## • Processo Eleitoral 2021-2023

O processo eleitoral para os Órgãos Sociais da SPE; aprovado em Assembleia Geral Extraordinária é o seguinte:

- Abertura do processo eleitoral – 28 de setembro
- Data limite para apresentação de listas candidatas – 3 de novembro
- Divulgação das listas e instruções de voto – 13 de novembro
- Período de votação eletrónica – 28 de novembro às 18:00 horas a 5 de dezembro às 18:00 horas
- Assembleia Geral eleitoral – 7 de dezembro às 12:00 horas

Obs. Horas referidas são de Portugal continental

Naquela reunião extraordinária da Assembleia Geral da SPE no dia 16 de setembro, foram aprovados o novo regulamento eleitoral e as alterações aos estatutos necessárias para dar cabimento ao procedimento eleitoral por via eletrónica. Estes documentos encontram-se na página da SPE para consulta (<https://www.spestatistica.pt/spe/estatutos-e-regulamentos>).

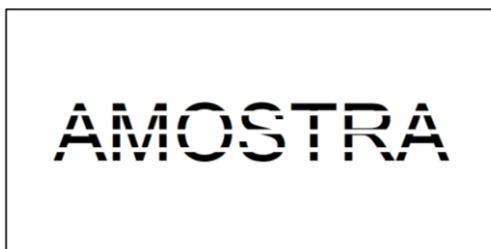
FR

## Enigmística de mefqa

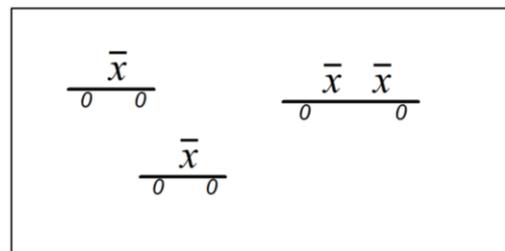
"40 anos SPE: passado, presente e futuro"



No Boletim SPE primavera de 2020 (p. 7):



amostra estratificada



médias móveis

# Episódios na História da Estatística

## Olhar o passado! Melhor, construir o futuro!

Fernando Rosado, [fernando.rosado@fc.ul.pt](mailto:fernando.rosado@fc.ul.pt)

DEIO, Faculdade de Ciências, Universidade de Lisboa

A Sociedade Portuguesa de Estatística – SPE, como sabemos, é uma Associação filiada<sup>1</sup> no International Statistical Institute – ISI; há cerca de vinte anos.

É importante “manter o contacto”.

Toda a pequena comunicação, por mais insignificante que nos pareça, é relevante para a afirmação e construção da Sociedade Mundial da Estatística e o que isso significa numa comunidade que, especialmente nos últimos tempos, está sempre “em ligação” e numa globalidade que, mas também com as devidas cautelas, deve ser incentivada.

A individualidade fortalecida, aumenta o sucesso da intervenção global.

O atleta “forma-se, totalmente, em privado” para “construir o desportista”, participante, em público, na “competição”; nacional primeiro e internacional mais tarde, quando a sua dimensão e o seu estatuto já está “ao nível adequado”.

A SPE, já há vários anos nesse nível, tem agora a acrescida responsabilidade “dos quarenta”.

A partilha da Ciência e dos valores científicos também assim o exige.

O Boletim SPE é um pequeníssimo instrumento que, para esse fim, se pode usar.

Foi nos “pequenos encontros” que se formou e consolidou a SPE de hoje. Neles se iniciou a sua História. Quarenta anos depois, decerto, é frutuoso, considerar uma análise<sup>2</sup> desses “históricos acontecimentos” no caminho da ciência em Portugal.

Tal como na SPE também no ISI, agora com estrutura mundial, as origens<sup>3</sup> estiveram relacionadas com uma série de acontecimentos científicos nacionais: reuniões, encontros, colóquios e congressos.

O ISI “partiu” de um primeiro congresso em Bruxelas, em 1853. Foi formalmente constituído, dois anos depois, num jubileu da *London Statistical Society*, com 81 membros, estatísticos oficiais e da academia. É interessante comparar<sup>4</sup> com “a história da SPE”.

Dos 36 Presidentes do ISI, até hoje, 20 foram da Europa, 8 da América do Norte, 4 da Ásia, 2 da América do Sul e 2 da Austrália; sendo relevante olhar para a sua origem e inclusão ao longo do tempo para concluirmos da génese e da implementação da Ciência Estatística e dos Estatísticos pelo mundo.

O mesmo com a “SPE”: Um “foco inicial” que irradiou!

Atualmente o ISI tem milhares de membros distribuídos por mais de 150 países. Os seus estatutos definem diversos tipos de membros constituintes: membros individuais eleitos inter pares; organizações privadas e organismos dos estados e também membros *in memoriam*. Nestes últimos e numa breve passagem pela sua listagem, “vemos a história”: a história universal da Estatística bem como “a nossa história individual” – os nossos mestres, os nossos colegas. São suportes, que a história fornece e que os

<sup>1</sup> Em <https://www.isi-web.org/index.php/about-isi/who-is-isi/members/organizational> pode verificar-se a estrutura mundial que suporta o ISI; ou em <https://www.isi-web.org/index.php/resources/national-and-regional-statistical-societies#Portugal>.

<sup>2</sup> Por exemplo, nas várias publicações e edições da SPE, <https://www.spestatistica.pt/>.

<sup>3</sup> Pode ser consultada a página <https://www.isi-web.org/index.php/about-isi/what-is-isi/history> do ISI

<sup>4</sup> O que pode ser feito: por exemplo, através do Boletim na página da SPE; no sítio, disponível em versão digital desde 2006. *História da Estatística em Portugal*, de Fernando Pereira de Sousa, Edição INE, 1996, é uma das poucas referências disponíveis. De relevo para este tema, obviamente, são também os textos de vários autores no *Memorial da Sociedade Portuguesa de Estatística* editado em 2005, pela SPE.

modernos meios informáticos, de algum modo, facilitam para se construirem. Tal como no ISI, talvez seja esta<sup>5</sup>, uma boa sugestão para incluir no sítio da SPE.

Mas, para lá dos números, bem o sabemos, estão as pessoas, os membros individuais, “aqueles que...”. Muitos são os exemplos com o maior interesse na história da Estatística e desenvolvidos pelos membros de todas estas associações científicas – e aqui podemos e devemos incluir tanto as internacionais como o ISI, como as nacionais como a SPE.

Neste ano de 2020, o *ISI* está a celebrar<sup>6</sup> o Ano Internacional das Mulheres na Estatística e *Data Science*, desde o passado mês de maio até julho de 2021. Com diversos acontecimentos, vai também ter uma sessão especial no próximo<sup>7</sup> *Congresso ISI*, a realizar em 2021.

No âmbito dessas comemorações e integrada na *História das Mulheres na Estatística*, foi organizada uma homenagem a Florence Nightingale, por ocasião da passagem do 200º dia do seu nascimento. Florence Nightingale nasceu em 12 de maio de 1820. Embora mais conhecida por todo o seu trabalho como enfermeira, ela também se afirmou pelo seu desempenho exemplar na formação e treino de enfermagem durante a guerra da Crimeia. Foi uma pioneira; no desenvolvimento dos registos, tabelas e gráficos como bons instrumentos estatísticos de apoio e decisão para as questões militares nessa época. O seu trabalho de investigação estatística fez dela a primeira mulher *fellow* da *Royal Statistical Society*. Para celebrar o bicentenário do seu nascimento, como reconhecimento e registo de “boa memória”, no Reino Unido, pela **primeira vez** neste ano de 2020, a *Health Foundation* e a *Royal Statistical Society*, instituíram o Prémio *Florence Nightingale para a Excelência na Análise de Dados e Cuidados de Saúde* em trabalhos que estudem e melhorem tempos de espera e os tratamentos ou que otimizem todos os percursos para acesso aos cuidados.

Na celebração dos quarenta anos da SPE, importa também salientar as mulheres da Estatística em Portugal. No nosso país em comparação com a maior parte dos congéneres do mundo é (quase) sempre maior a “percentagem do feminino” nos Departamentos Universitários e na Investigação. Este é um “dado estatístico” relevante e que, desde os meus inícios, senti partilhado, discutido e valorizado nos mais variados fóruns mundiais em que participei. Esse era já um ponto de conversa habitual entre pares que, cada vez melhor, se desejam conhecer; e, para além das suas diferenças culturais. Independente do modelo estatístico considerado, é um *outlier* de que dou testemunho. Em momento aniversário, celebremos esse valor também na SPE. Uma viagem simples e rápida pelos “arquivos” decerto o confirma. Investiguemos e analisemos a História para bem construir o Memorial. E, muitas vezes até, basta olhar à volta.

No Ano Internacional das Mulheres na Estatística, em momento aniversário da SPE, rejubilemos pelas “florencia portuguesas”; também para melhor construir o futuro!

---

<sup>5</sup> Cf. <https://www.isi-web.org/index.php/about-isi/what-is-isi/history>.

<sup>6</sup> <https://www.isi-web.org/index.php/events-and-awards/international-year-of-women-in-statistics-and-data-science>

<sup>7</sup> <https://www.isi2021.org/>

## *Outliers* e imputação nas fontes administrativas

João Lopes, [joao.lopes@ine.pt](mailto:joao.lopes@ine.pt)

João Poças, [joao.pocas@ine.pt](mailto:joao.pocas@ine.pt)

*Instituto Nacional de Estatística*

O uso de dados administrativos nas estatísticas oficiais tem tido uma importância cada vez maior nos principais produtores estatísticos. Este conceito baseia-se no cálculo de indicadores estatísticos oficiais através da reutilização de bases de dados recolhidas para outros fins. Esta reapropriação de dados traz consigo inúmeras vantagens, mas também novos desafios. Um destes desafios é a necessidade de automatização da extração dos dados, a sua limpeza e a integração com outras bases de dados (ver Hand 2018 para uma discussão abrangente sobre as vantagens e desafios inerentes aos dados administrativos).

Nos últimos anos, o INE iniciou um processo de armazenamento de dados da Administração Pública. Recentemente teve acesso aos dados do e-Factura constituídos pelos registos das operações comerciais realizadas em Portugal ao longo de cada mês, tendo um potencial enorme de integração nas estatísticas do Comércio e da Produção do INE. No entanto, a sua utilização requer um cuidadoso tratamento prévio de limpeza, que passa sobretudo por: i) eliminação de *outliers*; ii) procura de *missing data* em séries temporais; iii) e respetiva imputação de dados.

Apesar de haver diversas normas e protocolos estabelecidos para o tratamento de *outliers* (e.g. ASTM 2016), este processo tem uma complexidade específica a cada caso prático (ver Iglewicz e Hoaglin, 1993 para uma análise mais alargada destas técnicas). Contudo, é possível estabelecer os passos gerais na identificação de *outliers*: 1) **segmentação dos dados** de forma a obter distribuições de observações mais homogéneas; 2) **transformação dos dados** para aproximação à distribuição Gaussiana; 3) **cálculo de distâncias** (paramétricas ou não paramétricas) de cada observação ao centro da distribuição; 4) **seleção das observações extremas** através de diversas metodologias (e.g. *rankings* simples ou complexos, testes estatísticos, ou técnicas de *clustering*). No caso da identificação de *outliers* ou *missing values* em séries temporais é necessário um passo extra: o **ajustamento de modelos de séries temporais**, sendo que as observações a analisar passam a ser os resíduos associados ao modelo ajustado (ver Chen and Liu 1993 para um estudo detalhado destas técnicas).

A imputação de dados é também uma área de intenso desenvolvimento estatístico. No INE as técnicas de imputação mais utilizadas (automáticas ou *ad hoc*) baseiam-se no uso de dados temporais (ver Moritz *et al* 2015 para uma comparação de métodos univariados aplicados a séries temporais). Também neste caso é possível definir os principais passos gerais na imputação de séries temporais: 1) **agregação dos dados** de forma a obter séries temporais mais robustas (por exemplo, agregação mensal); 2) **segmentação dos dados agregados** de forma a obter distribuição de observações mais homogéneas; 3) **ajustamento de modelos de séries temporais** (e.g. ARIMA); 4) **imputação de valores** através de previsões do modelo ajustado.

Após o trabalho de identificação de *outliers* e de *missing data* e da sua respetiva imputação, segue-se a fase mais crítica do processo de utilização de dados de fonte administrativa: a sua validação. Esta

validação passa pela comparação entre os valores dos indicadores estatísticos calculados pelos processos clássicos e os calculados através de dados administrativos. Neste sentido, iremos utilizar, numa primeira fase, os principais índices do Inquérito Mensal ao Volume de Negócios e Emprego (IVNE) e, numa fase mais adiantada, a informação anual disponível na base de dados da Informação Empresarial Simplificada (IES). Só após esta validação se poderá integrar o uso de dados administrativos nos processos de produção estatística oficial.

## Referências

- ASTM. 2016. "Standard Practice for Dealing With Outlying Observations". Standard E178 - 16a. Disponível em <https://www.astm.org/Standards/E178.htm> (Acedido em outubro de 2020)
- Chen C e Liu L-M. 1993. "Joint Estimation of Model Parameters and Outlier Effects in Time Series". *J Am Stat Assoc* 88(421): 284-97. DOI: <https://doi.org/10.2307/2290724>
- Hand DJ. 2018. "Statistical challenges of administrative and transaction data (with discussion)". *J R Stat Soc A* 181: 555-605. DOI: <https://doi.org/10.1111/rssa.12315>
- Iglewicz B e Hoaglin D. 1993. *Volume 16: How to Detect and Handle Outliers*. The ASQC Basic References in Quality Control: Statistical Techniques.
- Moritz S, Sarda A, Bartz-Beielstein T, Zaeferrer M e Stork J. 2015. "Comparison of different Methods for Univariate Time Series Imputation in R" Disponível em <https://arxiv.org/abs/1510.03924> (Acedido em outubro de 2020)

# A atividade estatística nacional e a disponibilização de indicadores para o acompanhamento do impacto económico e social da pandemia COVID-19 – exemplos

*Instituto Nacional de Estatística*

## Introdução

Com a declaração do estado de emergência devido ao COVID-19, com efeitos a partir de 16 de março de 2020, e de modo a dar resposta às questões atuais da Sociedade, o INE efetuou alterações em algumas das suas atividades, procurando obter novos indicadores que permitam um melhor acompanhamento do impacto social e económico da Pandemia.

Neste contributo, apresentamos dois exemplos de atividades estatísticas, uma na área das famílias e outra na das empresas: a primeira refere-se ao Inquérito ao Emprego, no qual se acrescentou um conjunto de questões adicionais; e uma outra desenhada especificamente para acompanhar esta nova realidade, o inquérito “Rápido e Excepcional às Empresas – COVID-19”, desenvolvido em colaboração com o Banco de Portugal.

As bases de dados destas duas atividades encontram-se disponíveis para trabalhos de investigação. Estas e outras novas atividades de acompanhamento do impacto económico e social da pandemia podem ser consultadas no portal do INE ([www.ine.pt](http://www.ine.pt)), e em particular no seu espaço dedicado “Especial INE COVID-19”.

## **Inquérito ao Emprego**

O INE passou a incluir questões sobre o “teletrabalho a partir de casa” que foram introduzidas no Inquérito ao Emprego. O intuito principal é o de estimar o número de empregados a trabalhar a partir de casa e, entre estes, quantos o faziam em regime de teletrabalho, de modo a medir a nova realidade provocada pelo confinamento; a inquirição destas novas perguntas permanecerá nos trimestres posteriores até que tal se considere importante avaliar.

A população-alvo deste módulo é composta pela população empregada, estimada em 4731,2 mil pessoas. No 2.º trimestre de 2020, 23,1% do total da população empregada exerceu a sua profissão sempre ou quase sempre em casa na semana de referência ou nas três semanas anteriores, sendo que quase todos os inquiridos indicaram que a razão principal para ter trabalhado em casa se deveu à pandemia COVID-19.

Comparando as horas trabalhadas na semana de referência, não há grande diferença entre trabalhar em casa ou fora de casa. Efetivamente, quem não esteve ausente e trabalhou fora de casa trabalhou em média 36 horas nessa semana e quem não esteve ausente e trabalhou a partir de casa trabalhou 35 horas.

Observou-se ainda que 1 038,0 mil pessoas utilizaram tecnologias de informação e comunicação para poderem exercer a sua profissão em casa, o que representou 21,9% do total da população empregada e 94,8% das que trabalharam sempre ou quase sempre em casa no período de referência. 643,8 mil pessoas empregadas não trabalharam no emprego principal durante o período de referência, nem em casa, nem noutro local, 76,3% (491,5 mil) das quais devido à pandemia COVID-19.

A base de dados das respostas ao Inquérito ao Emprego com as questões adicionais sobre teletrabalho encontra-se disponível para acesso aos investigadores.

## **Inquérito Rápido e Excecional às Empresas – COVID-19**

Para além da iniciativa atrás descrita na área das famílias, o INE iniciou um novo inquérito “Rápido e Excecional às Empresas – COVID-19” no sentido de fazer um acompanhamento do impacto da pandemia nas empresas. Esta operação estatística foi dirigida a empresas de micro, pequena, média e grande dimensão, representativas dos diversos setores de atividade económica. Foi lançado na semana de 6 a 10 de abril de 2020 e permaneceu com uma frequência semanal até ao final de abril. A partir de maio, passou a quinzenal. A última edição disponível refere-se à primeira quinzena de julho de 2020.

Em finais de junho e face à situação que seria expectável sem pandemia, 66% das empresas reportaram um impacto negativo no volume de negócios. O Alojamento e restauração e os Transportes e armazenagem foram os setores com mais empresas a reportarem reduções no volume de negócios (87% e 80%, respetivamente). Ao longo do segundo trimestre, a percentagem de empresas respondentes com redução no volume de negócios, face à situação expectável sem pandemia, decresceu de 80% em abril para 67% em junho.

Também as bases de dados das respostas ao “Inquérito Rápido e Excecional às Empresas – COVID-19” encontram-se disponíveis para acesso aos investigadores.

## **Referências**

Destaque – Informação à comunicação social, INE, 5 de agosto de 2020, “Trabalho a partir de casa – Módulo ad hoc do Inquérito ao Emprego - 2.º trimestre de 2020”

[https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine\\_destaques&DESTAQUESdest\\_boui=445841978&DESTAQUESstema=55574&DESTAQUESmodo=2](https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_destaques&DESTAQUESdest_boui=445841978&DESTAQUESstema=55574&DESTAQUESmodo=2)

Destaque – Informação à comunicação social, INE e BdP, 14 de abril de 2020, “Inquérito Rápido e Excecional às Empresas – COVID-19 – Semana de 6 a 10 de abril de 2020”

[https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine\\_destaques&DESTAQUESdest\\_boui=430126865&DESTAQUESmodo=2&xlang=pt](https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_destaques&DESTAQUESdest_boui=430126865&DESTAQUESmodo=2&xlang=pt)

# Competição Europeia de Estatística

## Equipa portuguesa classifica-se em 2.º lugar na edição de 2020

*Instituto Nacional de Estatística*



A ESC (*European Statistics Competition*) é uma competição organizada pelo Eurostat (o Gabinete de Estatísticas da União Europeia) e diversos Institutos Nacionais de Estatística, com o propósito de promover a literacia estatística entre os alunos (3.º ciclo do ensino básico e secundário) e os professores.

Os principais objetivos da ESC são: promover a curiosidade e o interesse dos alunos pela estatística; incentivar os professores a utilizar novos materiais e novos métodos de ensino da estatística, incrementando a utilização de dados estatísticos oficiais e a aplicação do conhecimento estatístico adquirido. A ESC visa também mostrar aos alunos e aos professores o papel da estatística em vários aspetos da sociedade e ainda promover o trabalho de equipa e a colaboração entre os alunos com vista a alcançar objetivos comuns.

A competição tem duas fases: a nacional e a europeia. Os finalistas da fase nacional de cada país poderão vir a participar na fase europeia. Na fase nacional, será utilizada a língua portuguesa; na fase europeia, utilizar-se-á a língua inglesa.

No nosso País, a fase nacional de ESC2020 foi organizada conjuntamente pelo Instituto Nacional de Estatística e pelo Banco de Portugal.

Participaram na terceira edição desta competição de literacia estatística mais de 17 mil alunos de 17 países, distribuídos por duas categorias: A: 16-18 anos; B: 14-16 anos.

A fase nacional da ESC2020, concluída em abril, apurou as duas equipas de cada grupo etário que representaram o seu país a nível europeu. A fase europeia constou da execução de um vídeo subordinado ao tema “Os jovens na Europa”, estabelecendo comparações entre países e ou regiões, com base em estatísticas oficiais.

Ao nível europeu, a equipa “Cirurgiões”, da Escola Secundária Santa Maria do Olival, de Tomar, obteve um honroso 2.º lugar (ver <https://esc-2020.eu/>).



Diogo Pires, João Rodrigues e Manuel Santos

O júri europeu salientou a qualidade geral dos vídeos a concurso e atribuiu à equipa portuguesa um brilhante 2.º lugar, na categoria A, com a sua história ao vivo sobre as dificuldades da personagem principal, o jovem licenciado Ted, no acesso ao mercado de trabalho.

O júri foi presidido pelo Eurostat e incluiu oito especialistas nas áreas de comunicação, jornalismo, educação científica, *gamification* e produção de vídeo.

## Probabilidades e Estatística no Departamento de Matemática da Universidade de Coimbra: uma perspetiva

Esmeralda Gonçalves, *esmerald@mat.uc.pt*

Paulo Eduardo Oliveira, *paulo@mat.uc.pt*

*Universidade de Coimbra, CMUC, DM*

Foi no final dos anos 70 do século vinte que surgiram os primeiros sinais que levaram à consolidação do grupo dedicado ao estudo das áreas de Probabilidades e Estatística associado à Universidade de Coimbra. Manuel Neto Murta, João Antunes Lopes e Lucília Rodrigues são alguns dos professores que foram assegurando disciplinas daquelas áreas, tanto para as formações em Matemática como para as Engenharias.

É naquela década que se estabelecem protocolos de cooperação científica com a Embaixada de França em Portugal que trazem ao Departamento de Matemática professores franceses da área, tal como Patrick Quidel, Raymond Moché e Michel Delecroix, com contributos decisivos para a iniciação à formação específica de Probabilidades e Estatística.

Esta cooperação conduziu ao doutoramento de Nazaré Mendes Lopes em 1985, sob a orientação de Jean Geffroy (Universidade Pierre et Marie Curie, Paris VI). O seu percurso académico revelou-se preponderante na criação, desenvolvimento e afirmação do Grupo de Probabilidades e Estatística do Departamento de Matemática da Universidade de Coimbra.

A influência da escola francesa, pelas mãos de Christian Gouriéroux, Pierre Jacob e Michel Delecroix, esteve também presente na formação e nos trabalhos de doutoramento de Esmeralda Gonçalves, Paulo Eduardo Oliveira, Emília Nogueira, Ana Cristina Rosa e Carlos Tenreiro.

Adicionalmente, as relações científicas que existiam com o Departamento de Estatística e Investigação Operacional da Universidade de Lisboa permitiram enriquecer o Grupo e fortalecer tais relações com os doutoramentos de Helena Ferreira e Graça Temido, orientados por M. Ivette Gomes e L. Canto e Castro. E vemos também surgir o que pode ser designado como segunda geração com os trabalhos de investigação desenvolvidos por Carla Henriques, Cristina Martins, Joana Leite e Filipa Silva, com orientações de Paulo Eduardo Oliveira, Nazaré Mendes Lopes e Esmeralda Gonçalves.

Inferência não Paramétrica, Processos estocásticos, Processos pontuais, Séries Temporais, Teoremas limite e Valores Extremos em Estatística são alguns dos domínios em que se inserem os trabalhos científicos que têm vindo a ser desenvolvidos por este grupo de investigação, a que se acrescentam aplicações variadas no campo da medicina, finanças e astronomia.

A bibliografia sucinta que se inclui abaixo ilustra a produção científica dos últimos anos de alguns dos seus elementos.

Os desenvolvimentos mais recentes de alguns dos temas indicados têm vindo a ser apresentados nas formações de segundo e terceiro ciclos que podem ser seguidas no Departamento de Matemática da UC (Mestrado em Matemática, Mestrado em Métodos Quantitativos em Finanças e Programa Interuniversitário de Doutoramento em Matemática) e têm vindo a contribuir para cativar novos e promissores valores para estas áreas.

Descrevemos neste texto, em passos muito largos e com omissões naturais, uma perspetiva da vivência passada e presente de um grupo de investigação com interesses científicos variados e que tem tido

sempre preocupações de atualização permanente. Estamos certos de que, no futuro, os membros deste Grupo e os seus seguidores serão capazes de responder **Presente** aos desafios que se forem perfilando nos tempos incertos que se avizinham, pois novas ideias e motivações de trabalho certamente surgirão.

### Publicações recentes

- ARAB, Idir, HADJIKYRIAKOU, Milto, OLIVEIRA, Paulo Eduardo (2020) On the behavior of the high order stop-loss transform for convolutions with some applications, *Communications in Statistics - Theory and Methods*, DOI: 10.1080/03610926.2020.1818101
- ARAB, Idir, HADJIKYRIAKOU, Milto, OLIVEIRA, Paulo Eduardo (2020). Failure rate properties of parallel systems. *Advances in Applied Probability*. Vol. 52, 2, pp. 563-587.
- ARAB, Idir, OLIVEIRA, Paulo Eduardo (2019). Iterated failure rate monotonicity and ordering relations within Gamma and Weibull distributions. *Probability in the Engineering and Informational Sciences*. Vol. 33, 1, pp. 64-80.
- ARAB, Idir, OLIVEIRA, Paulo Eduardo (2018). Asymptotic results for certain weak dependent variables. *Theory of Probability and Mathematical Statistics*. Vol. 99, pp. 19-36.
- DIAS, Sandra, TEMIDO, Maria da Graça (2019). Random fields and random sampling. *Kybernetika*. Vol. 55, 6, pp. 897-914.
- DIAS, Sandra, TEMIDO, Maria da Graça (2018). On the maxima of integer models based on a new thinning operator. Oliveira T., Kitsos C., Oliveira A., Grilo L. (Eds.), *Recent Studies on Risk Analysis and Statistical Modeling*. (pp. 213-226). Germany: Springer.
- GONÇALVES, Esmeralda, MENDES-LOPES, Nazaré (2020). Signed compound Poisson integer-valued GARCH processes. *Communications in Statistics - Theory and Methods* 49, 5468-5492, DOI: 10.1080/03610926.2019.1619767.
- GONÇALVES, Esmeralda, MENDES-LOPES, Nazaré (2019). Zero-distorted compound Poisson INGARCH models, In: Steland A., Rafajłowicz E., Okhrin O. (eds), *Stochastic Models, Statistics and Their Applications*, Dresden, March, Springer Proceedings in Mathematics & Statistics, Vol. 294, 305-314.
- GONÇALVES, Esmeralda, MENDES-LOPES, Nazaré, SILVA, Filipa (2019). On the estimation for compound Poisson INARCH processes. *REVSTAT – Statistical Journal*.
- GONÇALVES, Esmeralda, MENDES-LOPES, Nazaré (2018). Zero-truncated compound Poisson integer-valued GARCH models for time series. *Statistics*. Vol. 52, 3, pp. 619-642.
- HENRIQUES, Carla, MATOS, Ana Cristina, PEREIRA, Jorge, AFONSO, Catarina (2018). Complications of fluidotherapy in patients with acute pancreatitis: a contribution. *Millenium - Journal of Education, Technologies, and Health*. Vol. 2, 6, pp. 23-31.
- JESUS, Diogo, MATOS, Ana Cristina, HENRIQUES, Carla, ZEN, Margherita, LAROSA, M., IACCARINO, L., PEREIRA DA SILVA, J. A., DORIA, Andrea, INÊS, Luis Sousa (2019). Derivation and validation of the SLE Disease Activity Score (SLE-DAS): a new SLE continuous measure with high sensitivity for changes in disease activity. *Annals of the Rheumatic Diseases*. Vol. 78, pp. 365-371.
- MACHADO, S., MARQUES, R., NASCIMENTO, E., MATOS, Ana Cristina, HENRIQUES, Carla (2019). Relationship between HbA1c and capillary blood glucose self-monitoring in type 2 diabetics. *Romanian Journal of Internal Medicine*. Vol. 57, 2, pp. 125-132.
- MARQUES, Beatriz Riquito, DINIS, Ana Clara, ROCHA, Gustavo, FLÔR-DE-LIMA, Filipa, MATOS, Ana Cristina, HENRIQUES, Carla, GUIMARÃES, Hercília (2019). Morbidity and mortality in preterm infants less than 29 weeks of gestational age. *Journal of Pediatric and Neonatal Individualized Medicine*. Vol. 8, 1, pp. 1-8.
- TENREIRO, Carlos (2019). A aula de Geometria Descritiva da Faculdade de Matemática e a sua coleção de modelos de Olivier. Coimbra: *Imprensa da Universidade de Coimbra*.
- TENREIRO, Carlos (2019). On the automatic selection of the tuning parameter appearing in certain families of goodness-of-fit tests. *Journal of Statistical Computation and Simulation*. Vol. 89, 10, pp. 1780-1797.
- TENREIRO, Carlos (2018). A new class of boundary kernels for distribution function estimation. *Communications in Statistics - Theory and Methods*. Vol. 47, 21, pp. 5319-5332.

# Alguns tópicos de investigação em estatística realizada no CIDMA – uma breve discussão

Pedro Macedo, *pmacedo@ua.pt*  
*CIDMA, Departamento de Matemática, Universidade de Aveiro*

Ana Helena Tavares, *ahtavares@ua.pt*  
*CIDMA, Escola Superior de Saúde de Aveiro, Universidade de Aveiro*

Marco Costa, *marco@ua.pt*  
*CIDMA, Escola Superior de Tecnologia e Gestão de Águeda, Universidade de Aveiro*

## 1. Introdução

O Centro de Investigação e Desenvolvimento em Matemática e Aplicações (CIDMA), alojado no Departamento de Matemática da Universidade de Aveiro, é, em Portugal, com mais de uma centena de membros e colaboradores, um dos maiores centros de investigação em matemática e aplicações.

Entre outras atividades, o Grupo de Probabilidades e Estatística (GPE), um dos oito grupos de investigação que constituem o CIDMA, desenvolve investigação fundamental em diversas áreas das probabilidades e da estatística, com especial enfoque em séries temporais, estatística multivariada, bioestatística, estatística robusta e máxima entropia. A investigação desenvolvida está, naturalmente, articulada com inúmeras aplicações, em variadíssimas áreas científicas, tais como a educação, a medicina, a genómica, o ambiente e a economia.

Inevitavelmente incompleto e fatalmente enviesado pela perspetiva dos seus autores, este pequeno trabalho destaca alguns dos presumíveis tópicos de investigação promissores no seio do GPE.

## 2. Genómica e padrões genómicos

No contexto atual existe um grande volume de dados genómicos disponível, e é esperado que, com toda esta informação, seja possível construir modelos matemáticos que contribuam para um melhor conhecimento do ADN (sigla de ácido desoxirribonucleico), eventualmente, para uma descrição concisa do genoma. No campo das aplicações estatísticas, a exploração deste tipo de dados exige não só o conhecimento de técnicas estatísticas, mas também de saber interdisciplinar para uma efetiva translação do conhecimento. A área da análise de dados genómicos tem sido visada pelo GPE num contexto dinâmico de interdisciplinaridade, envolvendo investigadores de outras áreas, como a informática, a biologia ou a genética.

A genómica estuda a sequência de ADN de organismos, vírus ou linhas celulares, que pode ser vista como uma longa sequência de quatro símbolos (os nucleótidos), sendo que uma parte significativa do ADN é composta por sequências repetitivas. Acredita-se que repetições de sequências específicas tenham um significado biológico em função dos seus padrões de distribuição. A identificação de características estruturais de novas repetições em todo o genoma poderá fornecer informação sobre funções biológicas.

Um fenómeno que mereceu a atenção de investigadores do GPE é o da simetria em cadeias simples de ADN, conhecido por *single strand symmetry*. Este fenómeno, amplamente confirmado em muitos

organismos, indica que a abundância de uma palavra genômica tende a ser semelhante à abundância da palavra que é seu complemento invertido (termo que não é fundamental aqui explicitar, mas entenda-se como palavras com um determinado tipo de estrutura simétrica). Não há uma explicação consensual para a ocorrência do fenômeno de simetria, mas acredita-se que a ocorrência deste tipo de simetria tenha motivação biológica. Em Afreixo et al. (2013) é efetuada uma avaliação do fenômeno de simetria no genoma humano completo, considerando palavras até tamanho 10. O estudo conclui que o fenômeno de simetria é menos vincado em palavras de tamanhos maiores. Afreixo et al. (2015a) desenvolvem uma metodologia estatística para quantificar a simetria tendo esta medida sido posteriormente utilizada para identificar locais de simetria excepcional no genoma humano (Afreixo et al., 2016).

Uma outra abordagem para estudar as repetições no ADN prende-se com a análise das distâncias a que estas ocorrem, e com a exploração dos seus padrões. Afreixo et al. (2009) desenvolvem uma metodologia de processamento do genoma com base em distâncias entre nucleótidos. A distância entre nucleótidos revelou grande potencial na obtenção de assinaturas genômicas para genomas completos, capazes de discriminar entre espécies e de recuperar relações evolutivas entre estas. Neste trabalho, a assinatura genômica de uma espécie obtém-se por confrontação (erro relativo) do vetor que descreve a distribuição de distância entre nucleótidos de um genoma e a distribuição de distância de referência, nomeadamente, a distribuição de uma sequência onde os nucleótidos são gerados aleatoriamente e independentemente. Enfatize-se que o mapeamento através das distâncias entre nucleótidos permite analisar e comparar sequências de ADN sem recorrer ao alinhamento das mesmas (*alignment free*). Estes métodos são particularmente úteis quando se pretende analisar genomas completos de grande dimensão, como é o caso do genoma humano (com mais de três mil milhões de nucleótidos).

Um passo natural após a investigação de distribuições de distâncias entre nucleótidos é o de se estender a análise a distâncias entre “palavras”, como dinucleótidos, trinucleótidos, etc. Também nestes casos se definem modelos teóricos que descrevem as distribuições de distâncias entre palavras em cenários aleatórios, designados por dados de referência. As propriedades estatísticas de distribuições de distâncias entre dinucleótidos foram estudadas e exploradas utilizando dados de genomas empíricos e dados de referência (Afreixo et al., 2015b). Com base nessas propriedades, os autores propõem um modelo com alta capacidade para discriminar ilhas CpG em sequências de ADN.

Algumas estruturas do ADN, como *stem-loops* e cruciformes, têm demonstrado desempenhar um papel importante no dano e reparação do genoma, instabilidade e regulação genética. Estas estruturas formam-se em locais que contêm palavras que são complementos invertidos. Por esse motivo, o seu estudo leva naturalmente à investigação de propriedades de simetria em sequências genômicas e, em particular, das distâncias entre palavras “simétricas”. Tavares et al. (2017a) efetuaram um estudo exaustivo das distribuições de distâncias, em palavras de tamanho 6 e 7, e identificaram pares de palavras simétricas que ocorrem muito frequentemente a uma distância fixa entre si. A sobrerrepresentação de uma determinada distância, cunhada de “pico” (*peak*), sugere características estruturais do ADN. Ainda no contexto da análise da estrutura do ADN, o trabalho de Bastos et al. (2019) explora e caracteriza a existência de regularidades nos picos de frequências. Os resultados obtidos nestes trabalhos identificam palavras genômicas que são fortes candidatas à formação de estruturas cruciformes e regiões do genoma humano com potencial para a formação dessas estruturas.

A estreita relação entre a frequência das palavras que são complementos invertidos (fenômeno de simetria) não é necessariamente extrapolada para o padrão de distribuição de tais palavras. A semelhança/dissimilaridade entre a distribuição de distâncias de uma palavra e a distribuição de distâncias da palavra que é seu complemento invertido foi explorada em Tavares et al. (2017b, 2018). Os autores propõem uma medida de dissimilaridade entre distribuições, com base na comparação dos seus picos de frequências, que se mostra bastante eficiente na identificação de distribuições de distâncias muito dissimilares.

A associação entre a dissimilaridade de distribuição e a discrepância de frequência também é explorada, e especula-se que pares de palavras simétricas que combinam valores baixos e altos de cada medida podem encobrir características biológicas de interesse.

Naturalmente, a caracterização mais geral das distribuições de distâncias entre palavras genómicas envolve o problema do crescimento exponencial do número de distribuições com o aumento do comprimento da palavra, gerando a necessidade de redução dos dados. Uma solução apontada é o agrupamento das distribuições de distâncias, e consequentemente o agrupamento das palavras genómicas. Tavares et al. (2020) propuseram uma metodologia para o agrupamento de distribuições de distâncias que tem em conta a existência de picos de frequências. Neste procedimento cada distribuição de distâncias é primeiramente decomposta numa curva de *baseline*, através de um método robusto a valores atípicos, e numa distribuição de picos. O procedimento, aplicado em palavras de tamanho 3 a 5 (trinucleótidos, tetranucleótidos e pentanucleótidos), demonstrou ter potencial para identificar palavras com padrões de distância semelhantes. Ao observar a composição das palavras em cada *cluster*, foram encontradas associações com a frequência de certos dinucleótidos que desempenham um papel biológico específico.

Neste contexto de dados genómicos, ainda há um longo caminho a desbravar. As metodologias estatísticas apresentam, sem dúvida, um grande potencial para a identificação de padrões evolutivos e para a previsão das propriedades funcionais representando, assim, um passo adicional na criação de conhecimento sobre sequências genómicas.

### 3. Análise e previsão de séries temporais

A análise de séries temporais e o desenvolvimento de modelos estatísticos para este tipo de dados são temas de investigação que há muito têm suscitado interesse no GPE. É inquestionável que o estudo de dados com uma estrutura de dependência temporal tem vindo a ter cada vez mais interesse. De facto, tem emergido um vasto conjunto de problemas contemporâneos em diversas áreas, por exemplo no Ambiente, na Economia ou na Saúde Pública, aos quais os investigadores nesta área devem responder. Nestas áreas destacam-se a emergência sem precedentes da problemática associada às alterações climáticas, reconhecida por inúmeras instituições internacionais, como a Organização das Nações Unidas através do Painel Intergovernamental para as Alterações Climáticas; a relevância da obtenção de previsões e monitorização dos ciclos económicos, cujo impacto é inquestionável nas políticas internacionais ou nos governos nacionais ou nas políticas da microeconomia e da macroeconomia; ou, mais recentemente, o impacto que os modelos de séries temporais têm tido na área da epidemiologia na atual pandemia provocado pelo vírus SARS-CoV-2.

O GPE tem vindo a desenvolver investigação sobre a modelação de séries temporais e a aplicação de técnicas multivariadas, como análise de componente principais e métodos de agregação, em bases de dados complexas com o foco nas aplicações ambientais (Silva e Freitas, 2020; Scotto et al., 2010; Scotto et al., 2011; Gouveia et al., 2015; Costa e Monteiro, 2016a). Estas aplicações têm tido uma particular relevância nas áreas emergentes do *Big Data*, neste caso no estudo de dados com correlação temporal. A investigação sobre modelos de séries temporais de valores inteiros (não negativos), também designadas por séries temporais de contagens, tem caracterizado de uma forma crescente o GPE, tanto na perspetiva clássica como na perspetiva bayesiana. O GPE tem vindo a estudar a deteção de *outliers*, em particular na modelação de séries temporais não-Gaussianas (Silva et al., 2019), bem como novas variantes dos modelos autorregressivos de médias móveis (ARMA) para séries de contagem com suporte limitado (Gouveia et al., 2018) e a desenvolver os processos autorregressivos de valores inteiros com limiares autoinduzidos (Pereira et al., 2015; Monteiro et al., 2012). O estudo das propriedades estocásticas e o estudo inferencial dos modelos de séries temporais de valores inteiros tem vindo a ser desenvolvido tanto em séries univariadas como bivariadas (Silva et al., 2020; Santos et al., 2019). Alguns desenvolvimentos no âmbito dos modelos de espaços de estados têm permitido a análise de problemas complexos através da modelação de séries temporais baseada em modelos dinâmicos com erros não gaussianos ou com parâmetros estimados (Costa e Monteiro, 2016a, 2016b).

A modelação e a previsão de séries temporais continuam a suscitar interesse ao GPE uma vez que os problemas atuais implicam o desenvolvimento e o estudo de modelos de séries temporais que incorporem múltiplas sazonalidades e erros não gaussianos (Monteiro et al., 2017). Rob Hyndman tem desenvolvido

alguns trabalhos baseados em modelos de suavização exponencial para incorporar as múltiplas sazonalidades, mas, contudo, estes modelos apresentam, ainda, algumas limitações e um desempenho a melhorar em termos da qualidade das previsões produzidas. A detecção e o tratamento de *outliers* em séries temporais continua a constituir um grande desafio aos investigadores uma vez que estes podem sinalizar a ocorrência de eventos raros e pontuais ou podem constituir erros no registo ou na medição. Do ponto de vista da estimação dos parâmetros dos modelos, a presença de *outliers* pode implicar dificuldades na convergência de processos iterativos, principalmente no caso da estimação pela máxima verosimilhança, ou a obtenção de estimativas fora do espaço de parâmetros. Neste contexto, a estimação dos parâmetros dos modelos através de métodos independentes da distribuição de base ou de métodos robustos são áreas promissoras nos próximos anos. O estudo de bases de dados de múltiplas séries temporais tem evidenciado que em muitos problemas a questão mais relevante não é a modelação individual de cada uma das séries, mas, sim, a identificação de padrões comuns, como por exemplo nas suas componentes de tendência sazonal, etc., ou a redução de dimensionalidade com base num critério pertinente para aquele problema concreto. Estes problemas colocam o desafio da aplicação de técnicas multivariadas associadas à modelação de séries temporais especialmente adaptadas às características intrínsecas dos dados ou dos objetivos a atingir (por exemplo, a agregação de séries baseadas em dissimilaridades avaliadas com distâncias adequadas à natureza das séries).

Nas últimas décadas têm sido propostos na literatura modelos e métodos para séries temporais de valores inteiros, que, devido à sua relevância, começam a ter a visibilidade merecida. No contexto dos modelos de contagem para a modelação de séries estacionárias considerando a possibilidade de as contagens serem limitadas ou ilimitadas, destacaram-se duas classes de modelos: a família de modelos baseados no operador de filtragem (INARMA) (com larga predominância do operador de filtragem binomial) e a classe de modelos de regressão para séries de contagem, dos quais se popularizaram os modelos INGARCH e os processos com a estrutura dos modelos lineares generalizados, permitindo uma estrutura de autocorrelação não linear. Ocupando um lugar de menor destaque ainda temos os modelos de Markov escondidos e os modelos ARMA discretos propostos por Jacobs e Lewis (1983).

Os diversos constrangimentos dos modelos de contagem ou de valores inteiros apresentados têm motivado propostas e desenvolvimentos que surgiram mais recentemente na literatura. Um dos constrangimentos de grande parte dos modelos INARMA ou INGARCH apresentados, sobretudo quando o suporte das observações é infinito, é o facto de terem pouca flexibilidade para incorporarem autocorrelações negativas. Mantendo a estrutura de autocorrelação linear, esta restrição tem fomentado a que mais recentemente tenham começado a surgir novos modelos. A título de exemplo, no âmbito de séries com observações de suporte infinito: a utilização de cópulas, de operadores de filtragem binomial sinalizados ou em séries com contagens limitadas através de sobreposição de processos binários. O desenvolvimento de modelos com estrutura de autocorrelação aproximadamente linear permitindo também a existência de autocorrelações negativas irá continuar a nortear novas propostas de investigação.

Outra tendência de investigação que se tem vindo a delinear deu origem a propostas de modelos de contagem que apresentem uma estrutura de dependência não linear. Nomeadamente, e a título ilustrativo, a exploração de modelos de regressão condicionais ou modelos que assumem que os seus parâmetros possam depender de observações anteriores. Ainda neste âmbito, o desenvolvimento de modelos de séries de contagem de memória longa aparenta ser uma área de investigação promissora.

Apesar de ultimamente terem sido desenvolvidas versões multivariadas sobretudo dos modelos INARMA e INGARCH, este é um tópico que deverá ser alvo de novas e promissoras abordagens. O acesso a grandes quantidades de dados possibilitará novos desafios e oportunidades. Por exemplo, o desenvolvimento de modelos multivariados para a monitorização de dados de vigilância na área da saúde pública, permitindo capturar a sobredispersão existente nos dados e lidar com variáveis de natureza diferentes, será um grande desafio que urge dar resposta.

Se os processos que admitem uma representação de espaço de estados, como os modelos ARMA, entre muitos outros, têm sido largamente considerados para modelar muitos fenómenos dinâmicos, o seu

potencial está longe de ser esgotado. As diversas extensões desenvolvidas desde o inicial filtro de Kalman (FK) propostas por Swerling, Kalman e Bucy, nomeadamente para modelos não-lineares, têm permitido a extração de componentes estocásticas não observáveis através de predições ótimas, no sentido do menor erro quadrático médio, quando os parâmetros dos modelos são conhecidos. O estudo do impacto da aplicação do FK quando os parâmetros são substituídos por estimativas, por exemplo, pelo método da máxima verosimilhança, tem sido estudado (Rodríguez e Ruiz, 2012; Costa e Monteiro, 2016a) mas é um tema ainda muito relevante uma vez que, neste caso, os erros quadráticos médios dos preditores de Kalman são subestimados. A flexibilidade dos modelos de espaço de estados dá-lhes um potencial para modelar séries temporais com estruturas dinâmicas complexas, como tendência e múltiplas sazonalidades. Este é um desafio atual nesta área e implica o desenvolvimento de algoritmos computacionalmente eficientes. Como em todas as áreas da modelação estatística, a modelação de séries temporais a partir de modelos de espaço de estados coloca desafios para a investigação de métodos alternativos de estimação dos parâmetros, baseados no método dos momentos generalizados, *bootstrap* ou outras técnicas de reamostragem, etc. As bases de dados atuais, pela sua complexidade, nem sempre permitem a verificação dos pressupostos dos modelos, nomeadamente das distribuições dos resíduos ou da sua estrutura de correlação, usualmente considerada compatível com um ruído branco, pelo que métodos alternativos aos mais aplicados, como o método da máxima verosimilhança, podem constituir avanços importantes na modelação de séries temporais.

#### 4. Máxima entropia e aplicações

A área de máxima entropia, ainda com reduzida expressão no GPE, possui um enorme potencial de investigação. No centro de uma recente abordagem científica denominada *info-metrics* (Golan, 2018), a máxima entropia, pela sua inegável riqueza histórica e cariz multidisciplinar, desempenha um papel preponderante em ciência. A investigação com máxima entropia desenvolvida pelo GPE centra-se em regularização (Macedo et al., 2016; Macedo, 2017), estimação de parâmetros em fronteiras de produção estocásticas e predição de eficiência técnica (Macedo et al., 2014; Macedo e Scotto, 2014; Robaina-Alves et al., 2015; Silva et al., 2019), agregação na análise de grandes volumes de dados (Costa e Macedo, 2019) e seleção de variáveis (Macedo, 2020). Aos três últimos – que serão, previsivelmente, os mais promissores –, poderá ainda adicionar-se, como tópico de crescente interesse no GPE, o *bootstrap* com máxima entropia no estudo de séries temporais.

A estimação de parâmetros em fronteiras de produção estocásticas, com a consequente predição de eficiência técnica, tem sido assegurada, quase exclusivamente, pelo estimador de máxima verosimilhança e suas variantes. Embora a literatura estatística esteja bem consolidada neste tópico, diversas dificuldades subsistem: colinearidade (termo não usado aqui no seu sentido literal), utilização de amostras de reduzida dimensão (uma dificuldade divertidamente designada como micronumerosidade), erros não gaussianos, distribuições estatísticas assumidas para a estrutura de erro composto e modelos indeterminados. Algumas destas dificuldades ditam fortes simplificações na modelação – o que, naturalmente, não constitui um procedimento desejável – ou restringem duramente a abordagem metodológica. Veja-se, por exemplo, a área de regulação do setor elétrico na Europa, na qual, devido às dificuldades anteriormente mencionadas, as fronteiras de produção estocásticas são, recorrente e lamentavelmente, preteridas. Recentemente, Silva et al. (2019) revelam que estas dificuldades não são, de modo algum, um obstáculo inultrapassável, dado que a existência de um modelo mal-posto (no sentido lato do termo) não impede a estimação dos seus parâmetros através de máxima entropia, mesmo quando formas funcionais flexíveis são usadas na especificação de funções distância radiais.

O *magging* (Bühlmann e Meinshausen, 2016) assumiu-se como um procedimento de referência para agregação, ao permitir a análise de grandes volumes de dados não-homogéneos, os quais serão, quiçá, a regra, e não a exceção, em problemas reais. Costa e Macedo (2019) sugerem uma nova abordagem, na qual os pesos são definidos pela informação de cada grupo, informação essa que é medida, por sua vez, através da entropia normalizada. Neste contexto, a estimação de parâmetros é efetuada através dos estimadores de máxima entropia generalizada ou entropia relativa generalizada. Embora muito permaneça por averiguar, alguns resultados de investigação em curso sugerem que esta nova abordagem,

quando comparada com algumas técnicas clássicas de agregação, possui um desempenho similar em predição, mas apresenta um desempenho muito superior em precisão, particularmente em cenários adversos de ruído e de colinearidade severa.

A seleção de variáveis é um tópico que dispensa apresentação e cuja literatura é vastíssima. As abordagens metodológicas são variadas, dependem do tipo de modelação, e incluem, entre inúmeras outras, a família *stepwise*, a regressão de todos os subconjuntos possíveis, a ponderação de modelos bayesianos, o *lasso* e suas generalizações. A indiscutível popularidade do tópico, naturalmente associada às suas implicações, não é alheia, todavia, à crescente capacidade computacional verificada nas últimas décadas. Duas críticas recorrentes aos procedimentos automáticos de seleção são o esforço computacional requerido (a investigação em otimização tem resolvido algumas destas dificuldades) e a ausência de contextualização do problema (parcialmente abordada por algumas técnicas que incorporam informação de contexto). Golan et al. (1996) sugerem o uso da entropia normalizada em seleção de variáveis e apresentam diversas vantagens desta abordagem: facilidade de implementação, incluindo em modelos com muitas variáveis; possibilidade de utilização de informação de contexto, através de suportes de parâmetros ou vetores de informação; utilização de estimadores de contração; possibilidade de implementação em modelos mal-postos; entre outras. No entanto, não obstante estas e outras vantagens, é escassa, na literatura, a implementação desta abordagem. Uma possível justificação poderá estar relacionada com os suportes de parâmetros, necessários ao problema de otimização, pois existe evidência empírica de que diferentes suportes podem implicar diferentes resultados na seleção de variáveis. Num trabalho recente, Macedo (2020) ilustra esta sensibilidade na avaliação do paradoxo de Freedman. Neste caso, onde a entropia normalizada é obtida através dos estimadores de máxima entropia generalizada e entropia relativa generalizada, a sensibilidade verificada poderá, no entanto, não ser impeditiva da correta identificação de um modelo de puro ruído.

Em 1994, Ehsan S. Soofi, no artigo intitulado *Capturing the Intangible Concept of Information*, no *Journal of the American Statistical Association*, refere, na página 1244, que “(Edwin) Jaynes introduced the maximum entropy principle of inference with which many statisticians have some familiarity but for which the statistics community as a whole has not yet developed sufficient appreciation.”. Presume-se que exista, atualmente, um maior reconhecimento, dada a imensa investigação entretanto realizada em torno deste tópico. E, com o protagonismo que a estatística tem granjeado nos últimos anos, a máxima entropia continuará, previsivelmente, a desempenhar um papel importante na ligação da estatística com a ciência da computação e com a teoria da decisão. Dos trabalhos de investigação acima mencionados, e no que aos tópicos aqui discutidos diz respeito, emergem diversas linhas de investigação futura: adequação das metodologias para estimação de parâmetros em modelos mal-postos; avaliação da sensibilidade das estimativas aos suportes de parâmetros e à informação de contexto definida na função objetivo; avaliação da performance da entropia normalizada em diversas circunstâncias de violação de pressupostos clássicos, no âmbito da agregação e da seleção de variáveis.

## 5. Conclusão

A avaliação realizada neste trabalho é, naturalmente, da inteira responsabilidade dos seus autores e não reflete, necessariamente, a apreciação individual dos membros e colaboradores do GPE. Certamente, outros tópicos de investigação serão, se não mais, igualmente promissores. Note-se, além do mais, que, por vezes, são insondáveis e desconcertantes as vagas de popularidade em alguns tópicos da investigação científica. Contudo, considera-se que os tópicos aqui desenvolvidos são áreas de investigação consolidadas no GPE e que, certamente, terão novos avanços no futuro próximo, sem prejuízo de outros tópicos que se encontram em desenvolvimento.

## Agradecimentos

Este trabalho é suportado pelo Centro de Investigação e Desenvolvimento em Matemática e Aplicações (CIDMA), através da Fundação para a Ciência e a Tecnologia (FCT), com referências UIDB/04106/2020 e UIDP/04106/2020.

## Referências

- Afreixo, V., Bastos, C., Garcia, S., Rodrigues, S., Pinho, A. & Ferreira, P. (2013). The breakdown of the word symmetry in the human genome. *Journal of theoretical Biology*, 335, 153-159.
- Afreixo, V., Rodrigues, J.M.O.S. & Bastos, C. A. C. (2015a). Analysis of single-strand exceptional word symmetry in the human genome: new measures. *Biostatistics*, 16(2), 209-21.
- Afreixo, V., Rodrigues, J. M. O. S. & Bastos, C. A. C. (2016). The exceptional genomic word symmetry along DNA sequences. *BMC Bioinformatics*, 17(1) 59.
- Afreixo, V., Bastos, C. A. C., Pinho, A. J., Garcia, S. P. & Ferreira, P. J. S. G. (2009). Genome analysis with inter-nucleotide distances. *Bioinformatics*, 25(23), 3064-3070.
- Afreixo, V., Bastos, C.A.C., Rodrigues, J.M.O.S. & Silva, R.M. (2015b). Identification of DNA CpG islands using Inter-dinucleotide distances. *Optimization in the Natural Sciences. Communications in Computer and Information Science*, 499, Springer, Cham.
- Bastos, C. A. C., Afreixo, V., Rodrigues, J.M.O.S., Pinho, A. J. & Silva, R.M. (2019). Distribution of distances between symmetric words in the human genome: Analysis of regular peaks. *Interdisciplinary Sciences: Computational Life Sciences*, 11, 367-372.
- Bühlmann, P. & Meinshausen, N. (2016). Madding: maximin aggregation for inhomogeneous large-scale data. *Proceedings of the IEEE*, 104, 126-135.
- Costa, M. & Macedo, P. (2019). *Normalized Entropy Aggregation for Inhomogeneous Large-Scale Data*. In Valenzuela O., Rojas F., Pomares H., Rojas I. (Eds.), Contributions to Statistics, Theory and Applications of Time Series Analysis, Springer, Cham.
- Costa M. & Monteiro, M. (2016a): Bias-correction of Kalman filter estimators associated to a linear state space model with estimated parameters. *Journal of Statistical Planning and Inference*, 176, 22-32.
- Costa, M. & Monteiro, M. (2016b). Discrimination of Water Quality Monitoring Sites in River Vouga using a Mixed-Effect State Space Model. *Stochastic Environmental Research and Risk Assessment*, 30(2), 607-619.
- Golan, A. (2018). *Foundations of Info-Metrics: Modeling, Inference, and Imperfect Information*. Oxford University Press, New York.
- Golan, A., Judge, G. & Miller, D. (1996). *Maximum Entropy Econometrics: Robust Estimation with Limited Data*. Wiley, Chichester.
- Gouveia, S., Scotto, M.G., Monteiro, A. & Alonso, A.M. (2015). Wavelets-based Clustering of Air Quality Monitoring Sites, *Environmental Monitoring and Assessment*, 187(694), 1-14.
- Gouveia, S., Moeller, T., Weiss, C.H. & Scotto, M.G. (2018). A full ARMA model for counts with bounded support and its application to rainy-days time series. *Stochastic Environmental Research and Risk Assessment*, 32 (9), 2495-2514.
- Jacobs, P., & Lewis, P. (1983). Stationary discrete autoregressive-moving average time series generated by mixtures. *Journal of Time Series Analysis*, 4(1), 19–36.
- Macedo, P. (2020). *Freedman's Paradox: A Solution Based on Normalized Entropy*. In Valenzuela, O., Rojas, F., Herrera, L.J., Pomares, H., Rojas, I. (Eds.), Contributions to Statistics, Theory and Applications of Time Series Analysis, Springer, (in press).
- Macedo, P. (2017). Ridge Regression and Generalized Maximum Entropy: an improved version of the Ridge-GME parameter estimator. *Communications in Statistics - Simulation and Computation*, 46(5), 3527-3539.
- Macedo, P. & Scotto, M. (2014). Cross-entropy estimation in technical efficiency analysis. *Journal of Mathematical Economics*, 54, 124-130.
- Macedo, P., Scotto, M. & Silva, E. (2016). Regularization with maximum entropy and quantum electrodynamics: the MERG(E) estimators. *Communications in Statistics - Simulation and Computation*, 45, 1143-1157.
- Macedo, P., Silva, E. & Scotto, M. (2014). Technical efficiency with state-contingent production frontiers using maximum entropy estimators. *Journal of Productivity Analysis*, 41(1), 131-140.
- Monteiro, A., Menezes, R., Silva, M.E. (2017). Modelling spatio-temporal data with multiple seasonalities: The NO2 Portuguese. *Spatial Statistics*, 22, 371-387.
- Monteiro, M., Scotto, M.G. & Pereira, I. (2012). Integer-valued self-exciting threshold autoregressive processes. *Communications in Statistics - Theory and Methods*, 41, 2717-2737.

- Pereira, I., Scotto, M.G. & Nicolette, R. (2015). Integer-valued self-exciting periodic threshold autoregressive processes, *Mathematical Texts*, University of Coimbra, 47, 81-92.
- Robaina-Alves, M., Moutinho, V. & Macedo, P. (2015). A new frontier approach to model the eco-efficiency in European countries. *Journal of Cleaner Production*, 103, 562-573.
- Rodríguez, A. & Ruiz, E. (2012). Bootstrap prediction mean squared errors of unobserved states based on the Kalman filter with estimated parameters. *Computational Statistics & Data Analysis*, 56, 62–74.
- Santos C., Pereira I. & Scotto M. (2019). Periodic INAR(1) Models with Skellam-Distributed Innovations. In: Misra S. et al. (eds) *Computational Science and Its Applications – ICCSA 2019*, Lecture Notes in Computer Science, vol 11621. Springer.
- Scotto, M.G., Alonso, A.M. & Barbosa, S.M. (2010). Clustering time series of sea levels: extreme value approach. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, 136, 215-225.
- Scotto, M.G., Barbosa, S.M. & Alonso, A.M. (2011). Extreme value and cluster analysis of European daily temperature series. *Journal of Applied Statistics*, 38, 2793-2804.
- Silva, A.O. & Freitas, A. (2020). Time Series Components Separation Based on Singular Spectral Analysis Visualization: an HJ-biplot Method Application. *Statistics, Optimization & Information Computing*, 8(2), 346-358
- Silva, I., Silva, M.E. & Torres, C. (2020). Inference for bivariate integer-valued moving average models based on binomial thinning operation. *Journal of Applied Statistics*, 47:13-15, 2546-2564.
- Silva, E., Macedo, P. & Soares, I. (2019). Maximum entropy: a stochastic frontier approach for electricity distribution regulation. *Journal of Regulatory Economics*, 55(3), 237-257.
- Silva, M.E., Pereira, I. & McCabe, B. (2019). Bayesian Outlier Detection in Non-Gaussian Autoregressive Time Series. *Journal of Time Series Analysis*, 40: 631-648.
- Tavares, A. H., Pinho, A. J., Silva, R. M., Rodrigues, J. M. O. S., Bastos, C. A. C., Ferreira, P. J. S. G. & Afreixo, V. (2017a). DNA word analysis based on the distribution of the distances between symmetric words. *Scientific Reports*, 7(1), 728.
- Tavares, A.H., Raymaekers, J., Rousseeuw, Silva, R. M., Bastos, C. A. C., Pinho, A., Brito, P. & Afreixo, V. (2017b). Dissimilar symmetric word pairs in the Human Genome. *Advances in Intelligent Systems and Computing*, 616. Springer, Cham.
- Tavares, A.H., Raymaekers, J., Rousseeuw, Silva, R. M., Bastos, C. A. C., Pinho, A., Brito, P. & Afreixo, V. (2018). Comparing reverse complementary genomic words based on their distance distributions and frequencies. *Interdisciplinary Sciences: Computational Life Sciences*, 10, 1–11.
- Tavares, A. H., Raymaekers, J., Rousseeuw, P. J., Brito, P. & Afreixo, V. (2020). Clustering genomic words in human DNA using peaks and trends of distributions. *Statistics and Computing*, 14, 57–76.



# Pseudovalores- $p$ na construção do conhecimento científico

Maria de Fátima Brilhante, *maria.fa.brilhante@uac.pt*

*Departamento de Matemática e Estatística da Faculdade de Ciências e Tecnologia  
da Universidade dos Açores  
Centro de Estatística e Aplicações da Universidade de Lisboa*

“*Science is a cumulative process.*” (Shoemaker *et al.*, 2003)

Um dos princípios fundamentais em Ciência é o princípio da reprodutibilidade, que se traduz na possibilidade de uma experiência (ou estudo) poder ser reproduzida, sob condições idênticas e de forma independente, por outros investigadores de modo a permitir validar conclusões científicas<sup>1</sup>. Porém, no meio de um conjunto de experiências que foram realizadas com o objetivo de responder a questões semelhantes, há a possibilidade de se ser confrontado com conclusões diferentes ou contraditórias, sobretudo na área das ciências experimentais e quando existem amostras aleatórias em jogo. Por isso, ao longo dos últimos anos vários têm sido os autores que têm chamado a atenção para a importância de se acumular evidência estatística que se encontre disponível e dispersa por vários estudos similares, com o intuito de promover o avanço do conhecimento científico. A propósito, Sir Ian Chalmers, na sua conferência de 2005 “*The scandalous failure of scientists to cumulate scientifically*”, apresentou uma série de exemplos onde a incapacidade dos cientistas acumularem evidência teve consequências trágicas.

Nas situações em que há informação contraditória e/ou dispersa sobre um determinado problema, a Meta-Análise, proposta por Glass em 1976 e que grosseiramente se pode definir como sendo a síntese que resulta da reanálise das análises estatísticas, pode dar um contributo importante no sentido de combinar, harmonizar ou sintetizar toda a informação relevante disponível sobre o problema. Contudo, a Meta-Análise, na sua versão mais clássica, recorre a métodos que se baseiam em modelos estatísticos que têm associados um conjunto de pressupostos que, uma vez violados, descredibilizam o uso destes. Uma forma de contornarmos esta situação é recorrermos a uma abordagem do tipo não paramétrica, que pode consistir, por exemplo, na combinação de valores- $p$ , sendo estes mais vulgarmente conhecidos pela sua designação inglesa “*p-values*”, se o único objetivo da análise for averiguarmos quão compatíveis são os dados com uma hipótese nula global comum.

No entanto, a teoria clássica para a combinação de valores- $p$  assenta na condição destes serem observações independentes de uma variável aleatória com distribuição uniforme padrão. Esse pressuposto falha quando por deficiente formação científica, ou fraude, a equipa insatisfeita com resultados “inconvenientes” repete a experiência até estabelecer o que convém. Consequentemente, valores- $p$  escolhidos como “os melhores” terão então distribuição Beta(1,  $n$ ) ou Beta( $n$ , 1), com  $n \neq 1$ . Por isso, a não uniformidade da amostra de valores- $p$  observados/publicados acarreta dificuldades a nível da combinação e utilização destes, porque, por exemplo, os métodos de combinação conhecidos, entre os quais os populares métodos de Fisher e de Stouffer, deixam de ter as propriedades que lhes são conhecidas. Mas até que ponto valores- $p$  não uniformes, que passamos a referir por pseudovalores- $p$ , são uma realidade que exija um ajustamento dos métodos existentes, ou a proposta de novos métodos por forma a eventualmente acomodá-los na construção do conhecimento científico? A explicação

---

<sup>1</sup> Em situações muito excecionais pode haver progresso científico com base numa única experiência que, realizada uma única vez devido ao seu preço — por exemplo, a produção em acelerador de átomos transurânicos para determinação do seu peso atómico — não se pensa reproduzir. Mas nesses casos excecionais as experiências foram cuidadosamente planeadas e controladas por uma equipa de cientistas, que pôde atestar o rigor e a validade dos resultados.

encontrada por dois estatísticos portugueses para tentar acabar com a famosa controvérsia de Mendel-Fisher, e que muita tinta tem gerado ao longo dos anos, aponta no sentido de ser uma realidade possível.

Como apontamento, a controvérsia de Mendel-Fisher surgiu em 1936 na sequência do artigo “*Has Mendel’s work been rediscovered?*” de Sir Ronald Fisher. Nesse artigo, Fisher levanta sérias suspeitas sobre a improbabilidade dos dados obtidos por Gregor Mendel, considerado o pai da Genética, acusando-o de ter falseado os seus dados, se não diretamente, indiretamente por associação com colaboradores menos escrupulosos. Desde então, muitos autores têm argumentado a favor de Fisher e outros a favor de Mendel, mas sem nunca se ter chegado a um veredicto final. Em 2010 surge uma explicação plausível para os “dados demasiado bons” de Mendel no artigo “*A statistical model to explain the Mendel-Fisher controversy*” de Pires e Branco. Nesse artigo, os autores propõem dois modelos estatísticos para descreverem os dados de Mendel, sendo que o melhor ajustamento foi obtido com o modelo que prevê a possibilidade de Mendel, ou colaboradores, ter repetido as experiências, sempre que os resultados destas não eram compatíveis com as expectativas iniciais. Esta situação põe a descoberto o que uns já suspeitavam há muito, que é a de alguns investigadores, quando descontentes com os resultados obtidos, repetem as experiências na esperança de obterem “melhores resultados”, ou resultados mais consentâneos com o que esperavam obter.

O nosso trabalho de investigação nesta área, em colaboração com Dinis Pestana, Fernando Sequeira e Paulo Semblano, tem incidido sobre o problema da estimação do número de pseudovalores- $p$  numa amostra de valores- $p$ . Para o efeito, foi necessário arranjar um modelo que traduzisse, de certa forma, a realidade e, para isso, inspiramo-nos no trabalho de Pires e Branco. Como a situação mais provável é a de um investigador repetir a experiência apenas uma vez, é expeável que este venha a escolher como “melhor resultado” aquele que conduzir ao valor mínimo ou ao valor máximo de dois valores uniformes. A escolha entre o mínimo e o máximo dependerá das expectativas criadas *a priori* pelo investigador em relação ao resultado da experiência. Curiosamente, esta prática de escolha do “melhor” valor- $p$  foi designada por “*p-fishing*” pelo psicólogo Uri Simonsohn. Note-se que o problema descrito é também formalmente equivalente à situação em que um investigador escolhe o segundo valor- $p$  obtido, porque simplesmente desconfia do primeiro resultado, possivelmente por algo ter “corrido mal” com a primeira experiência. Assim, consideramos para modelo estatístico o que se decidiu apelidar de modelo de Mendel para descrevermos uma amostra de valores- $p$ , a qual pode conter observações não uniformes.

O modelo de Mendel tem um único parâmetro associado  $m$ , com  $m \in [-2, 2]$ , que está diretamente relacionado com a proporção de pseudovalores- $p$  na amostra. A função densidade de uma variável aleatória de Mendel é  $f(x; m) = m(x - 1/2) + 1$  no suporte  $[0, 1]$  (e com idêntica restrição no que se segue). Acontece que se  $m \in [-2, 0)$ , uma variável aleatória de Mendel é uma mistura entre uma variável aleatória uniforme padrão, com função densidade  $f(x; 0) = 1$ , e uma variável aleatória Beta(1,2), com função densidade  $f(x; -2) = 2(1 - x)$ ; ou se  $m \in (0, 2]$ , é uma mistura entre uma variável aleatória uniforme padrão e uma variável aleatória Beta(2,1), com função densidade  $f(x; 2) = 2x$ . (Recordamos que o mínimo de duas variáveis aleatórias uniformes padrão independentes é uma Beta(1,2), enquanto o máximo destas uma Beta(2,1).) Consequentemente, o coeficiente de mistura  $|m|/2$  determina na prática a proporção de valores- $p$  não uniformes na amostra. Já anteriormente Gomes *et al.* (2009) tinham usado as funções densidade que resultam de balancear a função densidade uniforme com centro em (0.5,1), isto é, misturas de uniforme e Beta(1,2) ou Beta(2,1), para replicar uniformes, aumentar amostras e averiguar uniformidade de valores- $p$ .

Num contexto de combinação de valores- $p$ , por exemplo, a estimação do valor  $m$  torna-se crucial por forma a decidirmos que tipo de metodologia estatística se deve empregar, ou desenvolver, caso se conclua a não uniformidade da amostra, isto é, se  $m \neq 0$ . No entanto, quando recorremos a estudos de simulação para atestarmos a eficácia de diferentes métodos na estimação de  $m$  (máxima verosimilhança, momentos, momentos ponderados, espaçamento máximo, etc.), todos estes demonstraram ter desempenhos pouco satisfatórios, principalmente quando a proporção de contaminação da amostra com pseudovalores- $p$  é inferior a 0.5, ou seja, quando  $m \in (-1, 1)$ . Esta situação levanta-nos um problema, porque queremos acreditar que seja apenas uma pequena fração de investigadores que decida repetir a experiência. Assim, a estimação do número de pseudovalores- $p$ , que inicialmente parecia ser um problema simples, veio mostrar-se ser complexo.

Não alheios às críticas proferidas ao longo das últimas décadas sobre o (mau) uso de valores- $p$  em questões científicas, o que levou inclusive algumas revistas tomarem a decisão radical de banir os valores- $p$ , pensamos que estes não deixam de ter a sua utilidade, desde que corretamente interpretados e

aplicados. Partilhamos da ideia inicial de Fisher sobre o seu uso que é permitirem discriminar resultados desinteressantes daqueles que importam realmente investigar mais a fundo. Todavia, na impossibilidade de evitarmos a tentação do “*p-fishing*” por parte de alguns investigadores menos preparados em termos de conhecimento de metodologias de investigação científica, a questão da estimação do número de pseudovalores-*p* numa amostra continua a ser, na nossa opinião, um tópico de investigação interessante e atual. A verdade é que os (pseudo)valores-*p* andam por aí, e enquanto existirem *softwares* estatísticos a “vomita-los cá para fora” e investigadores a não adotarem boas práticas de investigação científica, há que saber lidarmos com eles, pois, quer queiramos, quer não, estão a ser usados na construção do conhecimento científico. Por isso, pensamos que nesta fase se impõe procurarmos encontrar outros métodos de estimação mais eficazes para o parâmetro *m* do modelo de Mendel, que consideramos ser um modelo bastante realista neste contexto, investigarmos pseudovalores-*p* que sejam extremos de mais do que duas observações uniformes, com a complexidade que tal possa vir a trazer em termos de estimação, assim como, por exemplo, estendermos métodos de combinação, como o método de Fisher, para lidarmos com a eventual existência de pseudovalores-*p* numa amostra.

## Referências

- Brillhante, M.F., Pestana, D., Semblano, P., Sequeira, F. (2015). On the Proportion of Non Uniform Reported p-Values. *AIP Conference Proceedings*, **1648**(1), 540003, 1-4.
- Fisher, R.A. (1936). Has Mendel’s work been rediscovered? *Annals of Science*, **1**(2), 115-137.
- Glass, G.V. (1976). Primary, Secondary, and Meta-Analysis of Research. *American Educational Research Association*, **5**(10), 3-8.
- Gomes, M.I., Pestana, D., Sequeira, F., Mendonça, S., Velosa, S. (2009). Uniformity of offsprings from uniform and non-uniform parents. In Luzar-Stiffer, V., Jarec, I. and Bekic, Z. (eds.), *Proc. of the ITI 2009, 31th International Conference on Information Technology Interfaces*, 243-248.
- Pires, A.M., Branco, J.A. (2010). A Statistical Model to Explain the Mendel-Fisher Controversy. *Statistical Science*, **25**(4), 545-565.
- Semblano, P. (2016). *Estudo Crítico dos Valores de Prova*. Tese de Doutoramento, Faculdade de Ciências da Universidade de Lisboa.
- Shoemaker, P.J., Tankard, J.W., Lasorsa, D.L. (2003). *How to Build Social Science Theories*. Sage, Thousand Oaks, CA.



# Um olhar sobre o desenvolvimento de modelos de sobrevivência para acontecimentos recorrentes

Ivo Sousa-Ferreira<sup>1</sup>, [ivo.ferreira@staff.uma.pt](mailto:ivo.ferreira@staff.uma.pt)  
Ana Maria Abreu<sup>2</sup>, [abreu@staff.uma.pt](mailto:abreu@staff.uma.pt)  
Cristina Rocha<sup>1</sup>, [cmrocha@fc.ul.pt](mailto:cmrocha@fc.ul.pt)

<sup>1</sup> *CEAUL e DEIO, Faculdade de Ciências, Universidade de Lisboa*

<sup>2</sup> *Departamento de Matemática, Faculdade de Ciências Exatas e da Engenharia, Universidade da Madeira e CIMA, Portugal*

## 1. Introdução

A partir da segunda metade do século XX, tem existido um interesse crescente em estudar os tempos até à ocorrência de acontecimentos recorrentes, os quais surgem quando se observam vários episódios de um certo acontecimento durante o período de estudo. Situações que originam dados deste tipo ocorrem, por exemplo, em estudos biomédicos, quando se pretende estudar as recidivas de uma determinada doença; em estudos atuariais, quando se procura analisar os sucessivos incumprimentos de empréstimos bancários; e em estudos de fiabilidade, quando se pretende avaliar as repetidas falhas de uma máquina. Nestas situações em que o acontecimento de interesse ocorre mais do que uma vez para um mesmo indivíduo, por vezes a multiplicidade de acontecimentos observados é ignorada e é analisado apenas o tempo até à ocorrência do primeiro acontecimento. Em consequência, parte da informação que poderia vir a contribuir para uma melhor compreensão do problema em estudo é negligenciada. A maior complexidade de fenómenos desta natureza levou a que tenham sido desenvolvidas várias abordagens para analisar este tipo de dados (Cook e Lawless, 2007).

A abordagem que mais tem sido considerada, dada a sua versatilidade, consiste em estender o modelo semiparamétrico desenvolvido por Cox (1972) ao contexto dos acontecimentos recorrentes. De facto, nas últimas quatro décadas têm sido propostas várias extensões do modelo de Cox que visam ter em conta diversos aspetos relevantes em estudos deste tipo. Sendo estes modelos também semiparamétricos, não existem quaisquer premissas relativamente à distribuição dos tempos até à ocorrência dos sucessivos acontecimentos.

Outra abordagem consiste na modelação paramétrica dos tempos de recorrência. Segundo Kwong e Hutton (2003), quando a distribuição do tempo de vida for adequadamente escolhida, os modelos paramétricos são mais eficientes que os modelos semiparamétricos, no sentido em que produzem erros padrão mais pequenos e, consequentemente, estimativas dos coeficientes de regressão mais precisas. Contudo, considerar uma abordagem paramétrica também apresenta desvantagens, pois a modelação do tempo de vida pode ficar comprometida quando há uma má especificação do modelo. Deste modo, se o objetivo do estudo for apenas obter as estimativas do risco relativo e houver incerteza sobre a distribuição do tempo, é natural ser conservador e assumir pressupostos fracos sob uma abordagem semiparamétrica. Ainda assim, Solomon (1984) afirma que há alguma robustez dos modelos no que diz respeito às covariáveis, mostrando que a importância relativa das mesmas é preservada. Outra razão que contribui para que a aplicação dos modelos paramétricos seja menos frequente, está relacionada com o facto de as distribuições disponíveis na maioria dos programas de *software* estatístico não serem, em geral, suficientemente flexíveis para captar a forma como o risco evolui ao longo do tempo. Por exemplo, ao ajustar um modelo de regressão baseado na distribuição de Weibull está-se a admitir que a função de risco pode ser somente monótona crescente, monótona decrescente ou constante. Tal pressuposto pode ser demasiado restritivo, o que tem levado ao desenvolvimento de outras abordagens paramétricas mais flexíveis.

Por estes motivos, tanto quanto é do nosso conhecimento, há pouco trabalho realizado na área da modelação paramétrica no âmbito dos acontecimentos recorrentes. A maioria dos modelos paramétricos existentes baseia-se na premissa de que os tempos podem ser modelados diretamente através de uma estrutura de riscos proporcionais. Neste contexto, Duchateau *et al.* (2003), Ip *et al.* (2015) e Sousa-Ferreira *et al.* (2020a), consideraram a distribuição de Weibull, e Sousa-Ferreira *et al.* (2019) consideraram a distribuição proposta por Chen (2000). Por outro lado, Navarro *et al.* (2012) formularam modelos de sobrevivência baseados nas distribuições de Weibull, log-normal e log-logística, mas com uma estrutura de tempo de vida acelerado. Existe ainda um tipo diferente de modelos paramétricos que têm sido desenvolvidos sob o pressuposto clássico de que o número de acontecimentos observados até um dado instante segue um processo de Poisson não homogéneo (PPNH), para o qual os tempos não são, em geral, independentes. Macera *et al.* (2015) consideraram a forma exponencial-Poisson para a função de taxa marginal subjacente derivada de um PPNH. Noutra perspetiva, Sousa-Ferreira *et al.* (2020b) propuseram modelar o logaritmo da função de taxa cumulativa subjacente proveniente de um PPNH como uma função *spline* cúbica restrita do logaritmo do tempo.

Um dos maiores constrangimentos na análise de acontecimentos recorrentes é a forte possibilidade de os tempos associados ao mesmo indivíduo estarem correlacionados entre si, isto é, ocorrer correlação intra-individual. De acordo com Box-Steffensmeier e De Boef (2006), esta correlação pode ser proveniente de duas fontes: i) dependência entre acontecimentos, ou seja, a ocorrência de um acontecimento afeta o risco de ocorrência dos acontecimentos subsequentes; ii) heterogeneidade individual não observada, a qual se deve à existência de fatores de risco desconhecidos ou não mensuráveis. Assim, os modelos de sobrevivência para acontecimentos recorrentes podem ser classificados de acordo com a forma como lidam com a correlação. A abordagem mais simples baseia-se em corrigir a estimativa usual da variância (e, por vezes, estratificar os indivíduos por acontecimento), originando um modelo de variância corrigida. Outra abordagem consiste em incorporar um efeito aleatório, designado por variável fragilidade, o qual permite modelar a associação entre os tempos correspondentes ao mesmo indivíduo, dando origem a modelos com fragilidade. Por último, existe ainda a abordagem que recorre à obtenção da distribuição condicional dos tempos de recorrência, fazendo com que a relação existente entre os sucessivos acontecimentos deixe de ser um problema (Cook e Lawless, 2007).

O principal propósito deste trabalho é dar a conhecer dois tipos de abordagem para a modelação dos tempos relativos a acontecimentos recorrentes: semiparamétrica *versus* paramétrica. Para tal, começa-se por referir a abordagem semiparamétrica, onde serão consideradas duas extensões do modelo de Cox e, em seguida, a abordagem paramétrica, onde serão considerados dois modelos baseados na distribuição de Weibull e dois modelos baseados na distribuição de Chen. De forma a ilustrar a metodologia considerada, será apresentado um exemplo de aplicação a um conjunto de dados reais. Por fim, serão feitos alguns comentários a respeito destas duas abordagens.

## 2. Metodologia

Admita-se que existem  $n$  indivíduos em estudo e que cada um deles pode sofrer no máximo  $K$  recorrências de um certo acontecimento. Seja  $T_{ik}$  a variável aleatória (v.a.) que representa o tempo desde o início do estudo até à ocorrência do  $k$ -ésimo acontecimento ( $i = 1, \dots, n$  e  $k = 1, \dots, K$ ). Define-se a v.a.  $Y_{ik} = T_{ik} - T_{i,k-1}$  como sendo a duração do intervalo de tempo entre dois acontecimentos consecutivos (*gap time*), com  $0 \equiv T_{i0} < T_{i1} < \dots < T_{iK}$ . Por último, denote-se por  $\mathbf{z}_{ik} = (z_{ik1}, \dots, z_{ikp})'$  o vetor de  $p$  covariáveis referente ao  $k$ -ésimo acontecimento, associado ao  $i$ -ésimo indivíduo, e por  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$  o correspondente vetor de coeficientes de regressão.

Entre as várias extensões do modelo de Cox para analisar acontecimentos recorrentes, os dois modelos que mais têm sido aplicados foram propostos por Andersen e Gill (AG) e Prentice, Williams e Peterson (PWP) (Andersen e Gill, 1982; Prentice *et al.*, 1981). No que diz respeito à formulação dos modelos AG e PWP, para o  $i$ -ésimo indivíduo em estudo, as respetivas funções de risco são dadas por

$$h(y; \mathbf{z}_{ik}) = h_0(y) \exp(\boldsymbol{\beta}' \mathbf{z}_{ik}), \quad y \geq 0 \quad (1)$$

e

$$h(y; \mathbf{z}_{ik}) = h_{0k}(y) \exp(\boldsymbol{\beta}' \mathbf{z}_{ik}), \quad y \geq 0, \quad (2)$$

onde  $h_0(\cdot)$  representa a função de risco subjacente comum a todos os acontecimentos e  $h_{0k}(\cdot)$  é a função de risco subjacente específica do acontecimento  $k$ . Estas extensões são classificadas como modelos semiparamétricos pois, tal como no modelo de Cox, a forma da função de risco subjacente não é especificada, isto é,  $h_0(\cdot)$  e  $h_{0k}(\cdot)$  são funções arbitrárias. Além disso, repare-se que os modelos AG (1) e PWP (2) são formulados segundo uma estrutura de riscos proporcionais, onde se assume que as covariáveis exercem um efeito multiplicativo sobre a função de risco.

Em termos gerais, o modelo AG (2) foi proposto para o caso em que os acontecimentos ocorrem de forma ordenada e apresentam igual risco de ocorrência, pelo que se considera uma função de risco subjacente comum a todos os acontecimentos. Neste modelo, considera-se que todos os indivíduos em estudo contribuem para o conjunto de risco de qualquer acontecimento, seja qual for o número de acontecimentos observados para cada indivíduo. Por essa razão, diz-se que o conjunto de risco é não restritivo.

O modelo PWP (3) também surgiu para analisar acontecimentos ordenados, mas pressupõe que o risco de ocorrência de um acontecimento é afetado pela ocorrência do acontecimento que o antecede. Consequentemente, é necessário estratificar os indivíduos segundo a ordem pela qual os acontecimentos ocorrem. Assim, se for possível observar  $k$  acontecimentos, existirão  $k$  estratos ordenados, sendo que a cada um deles estará associada a função de risco subjacente  $h_{0k}(\cdot)$ ,  $k = 1, \dots, K$ . Note-se que neste modelo tanto pode ser obtida uma estimativa global dos coeficientes de regressão  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$  (alternativa que adotamos de modo a poder ser comparada com a obtida pelo modelo AG), como as estimativas específicas associadas a cada acontecimento  $k$ ,  $\boldsymbol{\beta}_k = (\beta_{k1}, \dots, \beta_{kp})'$ . Relativamente ao conjunto de indivíduos em risco, considera-se que estão em risco para o  $k$ -ésimo acontecimento apenas os indivíduos aos quais já foi observado o acontecimento  $k - 1$ , o que se traduz num conjunto de risco restritivo.

Neste trabalho considera-se que, em ambos os modelos AG (1) e PWP (2), a construção do intervalo de risco é feita segundo o tempo por intervalos (*gap time*), onde o relógio reinicia a contagem do tempo voltando ao instante zero após a ocorrência de cada acontecimento. Assim, a escala de tempo diz respeito ao tempo desde o último acontecimento. Interessa salientar que o modelo AG foi originalmente proposto para analisar o tempo desde o início do estudo, onde a construção do intervalo de risco é feita através dos processos de contagem. A possibilidade de o modelo AG ser formulado para analisar o tempo entre dois acontecimentos consecutivos foi explorada por Kelly e Lim (2000). Para mais detalhes sobre este assunto consulte-se o referido trabalho, onde os autores também examinaram minuciosamente as particularidades que as diferentes extensões do modelo de Cox apresentam.

Tendo por base a formulação dos modelos AG (1) e PWP (2), quando se especifica uma determinada distribuição para o tempo os modelos resultantes são totalmente paramétricos. Vários autores têm seguido esta estratégia para desenvolver novos modelos paramétricos (Duchateau *et al.*, 2003; Ip *et al.*, 2015; Sousa-Ferreira *et al.*, 2019; Sousa-Ferreira *et al.*, 2020a). Nesse sentido, a distribuição de Weibull tem sido a escolha preferida para especificar a forma da função de risco subjacente, provavelmente devido ao facto desta distribuição ocupar um lugar de referência na análise de dados de sobrevivência. Assim sendo, também se optou por utilizar a distribuição de Weibull com parâmetro de escala  $\lambda > 0$  e parâmetro de forma  $\alpha > 0$ , cuja função de risco é  $h(y) = \lambda \alpha y^{\alpha-1}$ ,  $y \geq 0$ . Por conseguinte, para o  $i$ -ésimo indivíduo em estudo, as correspondentes funções de risco dos modelos Weibull não estratificado (WNE) e Weibull estratificado (WE) são definidas por

$$h(y; \mathbf{z}_{ik}) = \lambda \alpha y^{\alpha-1} \exp(\boldsymbol{\beta}' \mathbf{z}_{ik}), \quad y \geq 0 \quad (3)$$

e

$$h(y; \mathbf{z}_{ik}) = \lambda_k \alpha_k y^{\alpha_k-1} \exp(\boldsymbol{\beta}' \mathbf{z}_{ik}), \quad y \geq 0, \quad (4)$$

em que  $\lambda > 0$  e  $\alpha > 0$  denotam os parâmetros de escala e de forma comuns a todos os acontecimentos, enquanto que  $\lambda_k > 0$  e  $\alpha_k > 0$  representam os parâmetros de escala e de forma específicos do acontecimento  $k$ ,  $k = 1, \dots, K$ , respetivamente. Apesar destes dois modelos já permitirem obter

estimativas da função de risco, são inadequados para situações em que a função de risco não seja monótona.

Para fenómenos complexos associados ao ciclo de vida de um indivíduo, é admissível considerar uma função de risco *bathtub-shaped*. Neste contexto, a distribuição proposta por Chen (2000) é uma boa opção, visto que a sua função de risco é dada por  $h(y) = \varphi\theta y^{\theta-1}\exp(y^\theta)$ ,  $\varphi, \theta > 0$  e  $y \geq 0$ , onde  $\theta$  é o único parâmetro que influencia a forma da distribuição. De facto, tem-se que esta função de risco é: *bathtub-shaped* quando  $\theta < 1$  (atingindo o valor mínimo em  $y_{min} = (1/\theta - 1)^{1/\theta}$ ); e monótona crescente quando  $\theta \geq 1$ . Sousa-Ferreira *et al.* (2019) verificaram que esta distribuição pode ser usada para formular um modelo de regressão pertencente à classe de modelos de riscos proporcionais. Deste modo, considerando a distribuição de Chen para especificar a forma da função de risco subjacente em (1) e (2), tem-se que as funções de risco dos modelos Chen não estratificado (CNE) e Chen estratificado (CE) são, respetivamente, definidas por

$$h(y; \mathbf{z}_{ik}) = \varphi\theta y^{\theta-1}\exp(y^\theta + \boldsymbol{\beta}'\mathbf{z}_{ik}), \quad y \geq 0 \quad (5)$$

e

$$h(y; \mathbf{z}_{ik}) = \varphi_k\theta_k y^{\theta_k-1}\exp(y^{\theta_k} + \boldsymbol{\beta}'\mathbf{z}_{ik}), \quad y \geq 0, \quad (6)$$

onde  $\varphi, \theta > 0$  são parâmetros comuns a todos os acontecimentos e  $\varphi_k, \theta_k > 0$  são parâmetros específicos do acontecimento  $k$ ,  $k = 1, \dots, K$ , respetivamente.

Em suma, os modelos paramétricos WNE (3) e CNE (5) permitem analisar situações em que o risco de ocorrência dos acontecimentos não se altera. Por outro lado, como nos modelos WE (4) e CE (6) são consideradas distribuições com parâmetros específicos para modelar os *gap times* de cada estrato/acontecimento, estes modelos abrangem situações em que os acontecimentos têm riscos de ocorrência distintos.

O método de inferência é baseado na teoria assintótica de máxima verosimilhança, assumindo que os *gap times* são independentes. Considere-se uma amostra em que a cada indivíduo  $i$  corresponde o vetor  $(y_{ik}, \delta_{ik}, \mathbf{z}_{ik})$ , onde  $y_{ik}$  é o *gap time* observado e  $\delta_{ik}$  denota a variável indicatriz que caracteriza o estado do  $i$ -ésimo indivíduo relativamente ao acontecimento  $k$ . Além disso, admita-se que os indivíduos estão sujeitos a um mecanismo de censura à direita e que a censura é não informativa. No caso dos modelos semiparamétricos AG (1) e PWP (2), para estimar o vetor de coeficientes de regressão  $\boldsymbol{\beta}$  é necessário adaptar a função de verosimilhança parcial do modelo de Cox ao contexto dos acontecimentos recorrentes (Cox, 1975; Kelly e Lim, 2000). Quanto aos modelos paramétricos WNE (3), WE (4), CNE (5) e CE (6), os estimadores de máxima verosimilhança dos vários parâmetros de cada modelo podem ser obtidos maximizando o logaritmo da função de verosimilhança dado por  $\ell = \log \sum_{i=1}^n \sum_{k=1}^K \{\delta_{ik} \log h(y_{ik}; \mathbf{z}_{ik}) + \log S(y_{ik}; \mathbf{z}_{ik})\}$ , onde  $h(y_{ik}; \mathbf{z}_{ik})$  e  $S(y_{ik}; \mathbf{z}_{ik}) = \exp[-\int_0^{y_{ik}} h(u; \mathbf{z}_{ik})] du$  são, respetivamente, as funções de risco e de sobrevivência do modelo que estiver a ser ajustado.

Embora o método de máxima verosimilhança assegure a obtenção de estimadores consistentes e assintoticamente normais, a existência de correlação intra-individual faz com que o estimador usual da matriz de covariância (baseado na matriz de informação de Fisher) não seja válido para realizar inferência. Na verdade, esta é uma abordagem *naive* que usualmente deflaciona o erro padrão, originando resultados demasiado otimistas. Deve-se então considerar um estimador mais robusto para a variância. Lin e Wei (1989) recorreram aos estimadores *sandwich*, com o objetivo de desenvolverem um estimador robusto para o modelo de Cox, o qual foi generalizado por Wei *et al.* (1989) de modo a ser utilizado nas extensões do modelo de Cox para acontecimentos recorrentes. No caso dos modelos paramétricos, optou-se por aplicar o estimador *jackknife* “one-step” que foi desenvolvido para dados de sobrevivência agrupados (Lipsitz *et al.*, 1994), tal como acontece no âmbito dos acontecimentos recorrentes em que as observações estão agrupadas por indivíduo. Este estimador é assintoticamente equivalente ao estimador *sandwich* e tem a vantagem de ser mais fácil de programar computacionalmente.

A implementação dos modelos foi realizada no *software* estatístico **R** (R Core Team, 2020), versão 4.0.3, sendo que para os modelos semiparamétricos (AG e PWP) foi usado o *package* **survival** (Therneau, 2020) e para os modelos paramétricos (WNE, WE, CNE e CE) desenvolveu-se a sua programação utilizando o método de otimização de Broyden-Fletcher-Goldfarb-Shanno disponível no *package* **maxLik** (Toomet e Henningsen, 2020).

### 3. Aplicação a um conjunto de dados reais

Com o propósito de exemplificar a aplicação dos modelos formulados anteriormente, considerou-se os dados relativos à doença granulomatosa crónica (DGC) reportados em Fleming e Harrington (1991). A DGC caracteriza-se por uma imunodeficiência, o que faz aumentar a suscetibilidade a infeções recorrentes por microrganismos patogénicos como, por exemplo, determinadas bactérias e fungos. Em termos clínicos, é uma condição que se caracteriza pelo desenvolvimento de áreas de inflamação, designadas por granulomas, em diversos tecidos biológicos.

Os dados considerados representam o tempo (em anos) até uma infeção grave em 128 pacientes, dos quais 63 receberam *interferon gamma* (rIFN-g) e 65 receberam um placebo. O tempo máximo de *follow-up* foi de 1.2 anos (439 dias). Na Tabela 1, sumarizou-se a informação relevante sobre a constituição e evolução do conjunto de indivíduos em risco por acontecimento. Como o conjunto de risco diminui gradualmente, nos modelos com estratificação (PWP, WE e CE) os últimos estratos foram agrupados a partir do 4.º acontecimento, de forma a evitar a obtenção de estimativas pouco fiáveis.

Tabela 1: Resumo da informação sobre os dados da DGC

	Número do acontecimento							
	1	2	3	4	5	6	7	8
N.º de indivíduos em risco	128	44	16	8	3	2	1	1
N.º de acontec. observados								
rIFN-g	14	5	1	0	0	0	0	0
Placebo	30	12	7	3	2	1	1	0
Total	44	17	8	3	2	1	1	0
Percentagem de censura (%)	65.6	61.4	50	62.5	33.3	50	0	100

A título ilustrativo, foram incorporadas duas covariáveis nos modelos: o tratamento (0: placebo ou 1: rIFN-g) e a idade (em anos) à entrada no estudo. Os resultados obtidos no ajustamento de cada modelo, relativamente às estimativas dos coeficientes de regressão, encontram-se compilados na Tabela 2.

Tabela 2: Estimativas dos coeficientes de regressão para cada modelo

Covariável/Modelo	$\hat{\beta}_j$	$\exp(\hat{\beta}_j)$	$EP(\hat{\beta}_j)$	$EP_r(\hat{\beta}_j)$	Valor-p
Tratamento					
AG	-1.1143	0.3281	0.2683	0.3208	0.0005
WNE	-1.0522	0.3492	0.2619	0.3225	0.0011
CNE	-1.0775	0.3404	0.2623	0.3285	0.0010
PWP	-0.9069	0.4038	0.2796	0.2881	0.0016
WE	-0.8245	0.4389	0.2730	0.3013	0.0063
CE	-0.8382	0.4325	0.2731	0.3063	0.0062
Idade					
AG	-0.0294	0.9711	0.0131	0.0139	0.0352
WNE	-0.0286	0.9718	0.0130	0.0148	0.0529
CNE	-0.0292	0.9712	0.0130	0.0151	0.0529
PWP	-0.0236	0.9767	0.0134	0.0110	0.0314
WE	-0.0225	0.9777	0.0134	0.0122	0.0644
CE	-0.0228	0.9774	0.0129	0.0125	0.0672

Comparando as estimativas usual e robusta do erro padrão, observa-se que  $EP_r(\hat{\beta}_j)$  é, em geral, superior a  $EP(\hat{\beta}_j)$ , advertindo para a existência de correlação intra-individual, como antecipado. Quando  $EP_r(\hat{\beta}_j)$  é muito inflacionado comparativamente a  $EP(\hat{\beta}_j)$ , significa que há menor variação entre as observações pertencentes a um mesmo indivíduo do que entre as observações de indivíduos diferentes.

Reciprocamente, a obtenção de um  $EP_r(\hat{\beta}_j)$  muito deflacionado sugere que há menor variação entre os indivíduos do que em cada um deles. Ambas as situações constituem indícios de violação do pressuposto de independência entre as observações. A segunda situação apenas ocorreu nos modelos com estratificação, mais precisamente nas estimativas do erro padrão do efeito estimado da covariável idade. Este aspeto não é surpreendente visto que o número de indivíduos em risco diminui com as recorrências, fazendo com que o conjunto de risco de torne cada vez menos heterogéneo.

No que concerne ao efeito estimado de cada covariável,  $\hat{\beta}_j$ , e correspondente estimativa do risco relativo,  $\exp(\hat{\beta}_j)$ , observa-se que nos modelos sem estratificação (AG, WNE e CNE) os resultados obtidos são semelhantes e o mesmo sucede nos modelos com estratificação (PWP, WE e CE). Aplicando o teste de Wald robusto<sup>1</sup>, verifica-se que em todos os modelos apenas o tratamento tem influência significativa sobre o tempo até à ocorrência do acontecimento, ao nível de significância de 0.01.

Para seleccionar o modelo que melhor se ajusta a estes dados, obteve-se o valor do critério de informação de Akaike (AIC) para cada modelo (ver Tabela 3). Qualquer que seja a abordagem considerada, semiparamétrica ou paramétrica, os modelos com estratificação são aqueles que têm um valor do AIC mais baixo, indicando assim um melhor ajustamento aos dados da DGC, de entre os modelos considerados. Em particular, na abordagem paramétrica constata-se que o menor valor do AIC está associado ao modelo com estratificação baseado na distribuição de Chen (modelo CE). Então, para este conjunto de dados, a distribuição de Chen proporciona uma qualidade de ajustamento superior à distribuição de Weibull.

Tabela 3: Critério de informação de Akaike dos modelos ajustados aos dados da DGC

Modelos Semiparamétricos		Modelos Paramétricos			
AG	PWP	WNE	WE	CNE	CE
705.184	541.691	183.072	178.291	181.329	174.771

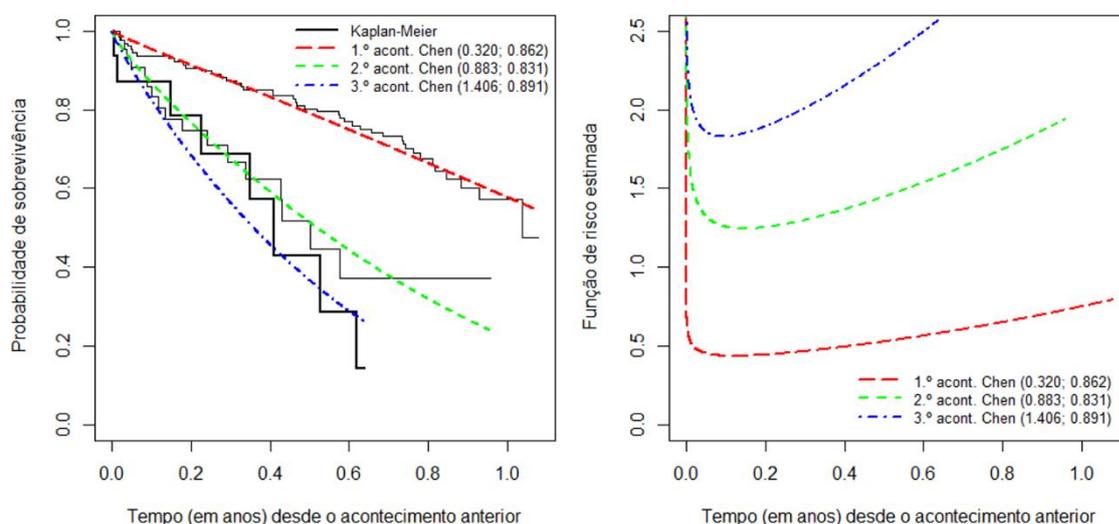


Figura 1: Estimativas da função de sobrevivência, pelo método de Kaplan-Meier e usando a distribuição de Chen, referentes aos três primeiros acontecimentos (à esquerda); e correspondentes estimativas da distribuição de Chen para a função de risco (à direita).

A adequabilidade do modelo CE pode ainda ser avaliada, de modo informal, representando graficamente as estimativas da função de sobrevivência obtidas pelo estimador de Kaplan-Meier e pelo modelo CE nulo (sem incluir as covariáveis), tendo-se obtido o gráfico apresentado à esquerda na Figura 1. Através deste gráfico, verifica-se que as estimativas baseadas na distribuição de Chen são, em geral, bastante próximas das estimativas de Kaplan-Meier, indicando que esta distribuição é uma alternativa paramétrica adequada para modelar o tempo até à recorrência de infeções causadas pela DGC. Além

<sup>1</sup> O teste de Wald robusto é assim designado por considerar a estimativa robusta do erro padrão.

disso, as estimativas específicas do parâmetro de forma da distribuição de Chen para o 1.º, 2.º e 3.º acontecimentos são 0.862, 0.831 e 0.891, respectivamente. Então, como estas estimativas são todas inferiores a 1, as correspondentes funções de risco estimadas são *bathtub-shaped*, tal como é possível observar no gráfico à direita. Assim sendo, a melhoria observada na qualidade do ajustamento dos modelos baseados na distribuição de Chen, pode ser atribuída à capacidade que esta distribuição tem para modelar uma forma mais complexa da função de risco.

#### 4. Considerações finais

Em termos gerais, quando a função de risco subjacente é corretamente especificada, sabe-se que os modelos paramétricos evidenciam maior eficiência que os modelos semiparamétricos. Embora a aplicação dos modelos paramétricos baseados nas distribuições de Weibull e de Chen aos dados da DGC não tenha melhorado a precisão das estimativas dos coeficientes de regressão, estes modelos têm a vantagem de permitir estimar de forma suave a função de risco. Além disso, as estimativas robustas inflacionadas do erro padrão salientaram a importância de ter em conta a correlação intra-individual. Repare-se que os modelos considerados neste trabalho são na verdade classificados como modelos de variância corrigida, pois os parâmetros de cada modelo foram estimados ignorando a correlação, corrigindo posteriormente a estimativa usual da variância.

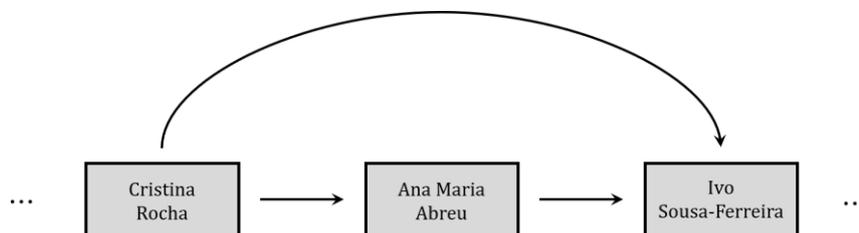
Nos modelos para acontecimentos recorrentes que são extensões do modelo de Cox, a função de risco subjacente não é especificada, o que pode constituir uma limitação.

Com efeito, em certas circunstâncias, a estimação desta função é de extrema importância, em particular na área da saúde, pois contribui para o estudo da evolução de uma doença ao longo do tempo. Na análise de acontecimentos recorrentes, a modelação paramétrica dos tempos é uma abordagem menos frequente, logo propícia a novos desenvolvimentos. Como os modelos paramétricos têm a vantagem de permitir a estimação da função de risco subjacente, atualmente o nosso interesse encontra-se direcionado para o desenvolvimento de novos modelos paramétricos que sejam capazes de descrever estes cenários complexos de forma útil, contribuindo para uma melhor compreensão do processo de recorrência.

Neste âmbito, pretende-se desenvolver modelos de sobrevivência paramétricos capazes de abranger situações em que a heterogeneidade individual não seja totalmente quantificável através das covariáveis observadas. Nesse sentido, será estudada a inclusão de um efeito aleatório (variável fragilidade), com o intuito de modelar essa heterogeneidade. No contexto dos acontecimentos recorrentes, a existência de heterogeneidade não observada é uma fonte de correlação intra-individual que pode fazer com que os indivíduos tenham uma maior (ou menor) propensão para sofrer os acontecimentos (Box-Steffensmeier e De Boef, 2006). Assim, a variável fragilidade irá representar a estrutura de dependência entre os tempos de um mesmo indivíduo, permitindo explicar alguns resultados inesperados ou até mesmo fornecer uma explicação alternativa em certas situações (ver, por exemplo, Duchateau *et al.*, 2003; Ip *et al.*, 2015). Por fim, também se pretende desenvolver modelos para lidar com situações em que haja suspeita da existência de indivíduos curados/imunes na população. Assim sendo, futuramente será investigada a inclusão de uma fração de cura nos modelos desenvolvidos, pois é plausível considerar que, também no contexto dos acontecimentos recorrentes, existam indivíduos para os quais nunca será observado qualquer episódio do acontecimento de interesse (ver, por exemplo, Tawiah *et al.*, 2020).

## Sobre os autores

A maioria dos docentes da área de Probabilidades e Estatística da Universidade da Madeira (UMa) deve a sua formação estatística a docentes da FCUL. Essa formação tem sido profícua e continua a dar frutos. E porque o tema do presente boletim é “*De onde viemos? Onde estamos? Para onde vamos?*”, apresentamos-vos um esquema do nosso percurso em termos de orientação científica:



## Agradecimentos

O presente trabalho foi desenvolvido na vigência da bolsa de doutoramento de Ivo Sousa-Ferreira, atribuída pela Universidade de Lisboa, e foi parcialmente financiado por Fundos Nacionais através da FCT – Fundação para a Ciência e a Tecnologia, no âmbito dos projetos UIDB/00006/2020 (CEAUL – Centro de Estatística e Aplicações) e UIDB/04674/2020 (CIMA – Centro de Investigação em Matemática e Aplicações, grupo de Estatística, Processos Estocásticos e Aplicações).

## Referências

- Andersen, P. K. e Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *The Annals of Statistics*, 10(4), 1100-1120.
- Box-Steffensmeier, J. M. e De Boef, S. (2006). Repeated events survival models: the conditional frailty model. *Statistics in Medicine*, 25(20), 3518-3533.
- Chen, Z. (2000). A new two-parameter lifetime distribution with bathtub shape or increasing failure rate function. *Statistics and Probability Letters*, 49(2), 155-161.
- Cook, R. J. e Lawless, J. F. (2007). *The Statistical Analysis of Recurrent Events*. New York: Springer Science & Business Media.
- Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society (Series B)*, 34(2), 187-220.
- Cox, D. R. (1975). Partial likelihood. *Biometrika*, 62(2), 269-276.
- Duchateau, L., Janssen, P., Kezic, I. e Fortpied, C. (2003). Evolution of recurrent asthma event rate over time in frailty models. *Journal of the Royal Statistical Society (Series C)*, 52(3), 355-363.
- Fleming, T. R. e Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. New York: Wiley.
- Ip, E. H., Efendi, A., Molenberghs, G. e Bertoni, A. G. (2015). Comparison of risks of cardiovascular events in the elderly using standard survival analysis and multiple-events and recurrent-events methods. *BMC Medical Research Methodology*, 15(15), 1-7.
- Kelly, P. J. e Lim, L. L.-Y. (2000). Survival analysis for recurrent event data: an application to childhood infectious diseases. *Statistics in Medicine*, 19(1), 13-33.
- Kwong, G. P. S. e Hutton, J. L. (2003). Choice of parametric models in survival analysis: applications to monotherapy for epilepsy and cerebral palsy. *Journal of the Royal Statistical Society (Series C)*, 52(2), 153-168.
- Lin, D. Y. e Wei, L. J. (1989). The robust inference for the Cox proportional hazards model. *Journal of the American Statistical Association*, 84(408), 1074-1078.
- Lipsitz, S. R., Dear, K. B. G. e Zhao, L. (1994). Jackknife estimators of variance for parameter estimates from estimating equations with applications to clustered survival data. *Biometrics*, 50(3), 842-846.

- Macera, M. A. C., Louzada, F., Cancho, V. G. e Fontes, C. J. F. (2015). The exponential-Poisson model for recurrent event data: an application to a set of data on malaria in Brazil. *Biometrical Journal*, 57(2), 201–214.
- Navarro, A., Morriña, D., Reis, R., Nedel, F. B., Martín, M. e Alvarado, S. (2012). Hazard functions to describe patterns of new and recurrent sick leave episodes for different diagnoses. *Scandinavian Journal of Work, Environment & Health*, 38(5), 447-455.
- Prentice, R. L., Williams, B. J. e Peterson, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika*, 68(2), 373-379.
- R Core Team. (2020). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. Vienna, Austria.
- Solomon, P. J. (1984). Effect of misspecification of regression models in the analysis of survival data. *Biometrika*, 71(2), 291-298.
- Sousa-Ferreira, I., Abreu, A. M. e Rocha, C. (2020a). Modelos de sobrevivência aplicados à análise de acontecimentos múltiplos. Em M. F. Salgueiro, P. Vicente, T. Calapez, C. Marques, & M. E. Silva, *Atas do XXIII Congresso da SPE* (pp. 145-159). Lisboa, Portugal: Sociedade Portuguesa de Estatística.
- Sousa-Ferreira, I., Abreu, M. A. e Rocha, C. (2019). Parametric regression models for recurrent events analysis based on Chen distribution. Em M. Meira-Machado, & G. Soutinho, *Proceedings of the 34th International Workshop on Statistical Modelling* (Vol. 2, pp. 200-205). Guimarães, Portugal.
- Sousa-Ferreira, I., Rocha, C. e Abreu, A. M. (2020b). A flexible marginal rate model for recurrent events with a zero-recurrence proportion. Em I. Irigoien, D.-J. Lee, J. Martínez-Minaya, & M. X. Rodríguez-Álvarez, *Proceedings of the 35th International Workshop on Statistical Modelling* (pp. 417-420). Bilbao, Espanha.
- Tawiah, R., McLachlan, G. J., e Ng, S. K. (2020). Mixture cure models with time-varying and multilevel frailties for recurrent event data. *Statistical Methods in Medical Research*, 29(5), 1368-1385.
- Therneau, T. M. (2020). *survival: A package for survival analysis in S*. Obtido de Package do R versão 3.2-7: <https://CRAN.R-project.org/package=survival>.
- Toomet, O. e Henningsen, A. (2020). Title maximum likelihood estimation and related tools. *Package do R versão 1.4-4*. Obtido de <https://CRAN.R-project.org/package=maxLik>.
- Wei, L. J., Lin, D. Y. e Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association*, 84(408), 1065-1073.



# O papel da Estatística na Universidade do Porto

Conceição Nunes Rocha, *conceicao.n.rocha@inesctec.pt*  
CPES (Power and Energy Systems) - INESC TEC, Porto

Isabel Silva, *ims@fe.up.pt*  
Faculdade de Engenharia, Universidade do Porto & CIDMA

## 1 Introdução

Após a surpresa inicial por recebermos o convite para participar neste Boletim SPE, que agradecemos, e olhando para o tema proposto, *"40 anos SPE: De onde viemos? Onde estamos? Para onde vamos?"*, foi difícil decidir do que falar e por onde começar. Optámos por abordar a investigação em Estatística na Universidade do Porto. Alertamos, desde já, que este trabalho reflete a visão muito particular das suas autoras, que certamente pecará por incompleta e desde já esperamos que qualquer omissão seja perdoada.

Longe vai o tempo em que era suposto os conhecimentos estatísticos dos Engenheiros, dos Psicólogos, dos Químicos, etc., serem suficientes para tratarem os dados e responderem às questões em causa. Na realidade, devido à natureza dos problemas e ao aparecimento de um volume cada vez maior de dados e variáveis, a própria Estatística evoluiu e, como tal, temos mais modelos, mais testes, etc. De uma maneira geral, nas cadeiras de Estatística ou de Análise de Dados lecionadas, os docentes não abordam nem metade dos temas da Estatística (quanto mais dos subtemas), e isso faz com que os estatísticos sejam profissionais essenciais nas equipas que têm por objetivo analisar dados, modelar, prever, etc. No entanto, se antes havia dificuldade em obter amostras que não fossem de reduzida dimensão, hoje para além dessa dificuldade temos também o problema de trabalhar com grandes quantidades de dados. Como, neste último caso, fazer cálculos à mão está fora de questão é reconhecida a importância de um estatístico saber programar, principalmente em R e/ou Python, criando deste modo a necessidade de uma sinergia entre Ciência dos Computadores e Estatística.

Como a nossa colega Conceição Amado (2017) [1] referiu no seu artigo, é cada vez mais importante a *"... necessidade de a Estatística se fortalecer com novas ferramentas teóricas quando a teoria atual se baseia em pressupostos que não são concretizados na prática."*

Adicionalmente,

*"O desenvolvimento de metodologias de análise da fiabilidade da informação será fundamental"*

quando é feito o

*"... tratamento de um enorme volume de dados..."*.

## 2 A Estatística na Universidade do Porto

A Universidade do Porto (U.Porto), fundada em 1911, possui 14 faculdades, uma *business school* e mais de 50 centros de investigação.

Embora a U.Porto não forneça nenhum curso específico em Estatística, é na Licenciatura em Matemática<sup>1</sup> ou nos Mestrados em Matemática<sup>2</sup> ou Engenharia Matemática<sup>3</sup> da Faculdade de Ciências da U.Porto (FCUP), que poderá ser iniciado o estudo da Estatística.

Outras alternativas fornecidas ao nível do 2º ciclo é o Mestrado em Modelação, Análise de Dados e Sistemas de Apoio à Decisão<sup>4</sup> da Faculdade de Economia da U.Porto (FEP), e recentemente o Mestrado em Ciência de Dados<sup>5</sup> da FCUP (primeira edição em 2018/2019) e o Mestrado em Engenharia e Ciência de Dados<sup>6</sup> da Faculdade de Engenharia da U.Porto (FEUP), com primeira edição no atual ano letivo 2020/2021.

Ao nível do 3º ciclo, a investigação em Estatística poderá ser continuada, por exemplo, no Doutoramento em Matemática Aplicada (MAP-PDMA, oferecido pelas Universidades do Minho, de Aveiro e do Porto e suportado por 6 unidades de investigação<sup>7</sup>), no Programa Inter-Universitário de Doutoramento em Matemática (da FCUP em conjunto com a Universidade de Coimbra<sup>8</sup>) ou no Programa Doutoral em Ciência de Dados de Saúde<sup>9</sup> da Faculdade de Medicina da U.Porto.

Parte dos centros de investigação da U.Porto estão integrados nas faculdades ou em laboratórios associados, outra parte foi acolhida em associações de interface em que participam várias entidades públicas e privadas, para além da U.Porto.

Alguns dos centros de investigação da U.Porto são especializados nos mais diversos domínios de conhecimento (desde as Ciências da Vida e da Saúde às Ciências Agrárias, das Engenharias às Belas Artes, não esquecendo as Humanidades) e outros são multidisciplinares. Apresentam dimensões, finalidades e estruturas organizativas muito diversas.

Sediado na FCUP encontra-se o Centro de Matemática da Universidade do Porto (CMUP). O seu Grupo de Probabilidades e Estatística<sup>10</sup>, constituído maioritariamente por matemáticos, desenvolve investigação em várias áreas como teoria ergódica, teoria de valores extremos, processamento de sinal e modelagem estatística e inferência. Além disso, são estudadas aplicações em biomedicina, engenharia, seguros e meteorologia.

Na área das Ciências Sociais, podemos mencionar o Centro de Economia e Finanças da Universidade do Porto (cef.up), sediado na FEP, em que uma das áreas de estudo do grupo de investigação THEOMET<sup>11</sup> (*Theory & Methods*) é a econometria.

Entre os centros multidisciplinares podemos nomear a Unidade de Investigação em Epidemiologia (EPIUnit), no Instituto de Saúde Pública da Universidade do Porto (ISPUP), cujo Grupo de Bioestatística e Medições em Saúde<sup>12</sup> tem como principal objetivo a investigação em bioestatística, através do desenvolvimento de técnicas matemáticas e estatísticas, com aplicações às áreas de Epidemiologia e Saúde Pública, Medicina e Biologia.

<sup>1</sup><https://matematica.fc.up.pt/licenciatura-em-matematica>

<sup>2</sup><https://matematica.fc.up.pt/mestrado-em-matematica>

<sup>3</sup><https://matematica.fc.up.pt/mestrado-em-engenharia-matematica>

<sup>4</sup><https://cursos.fep.up.pt/mestrados/modelacao-analise-de-dados-e-sistemas-de-apoio-a-decisao-mestrado/>

<sup>5</sup><https://mds.dcc.fc.up.pt/>

<sup>6</sup><https://dei.fe.up.pt/mdse/>

<sup>7</sup><https://map-pdma.up.pt/>

<sup>8</sup>[http://www.mat.uc.pt/phd\\_prog/default.php](http://www.mat.uc.pt/phd_prog/default.php)

<sup>9</sup><https://heads.med.up.pt/>

<sup>10</sup><https://cmup.fc.up.pt/main/content/probability-and-statistics>

<sup>11</sup><https://cefup.fep.up.pt/theomet/>

<sup>12</sup><https://ispup.up.pt/research/epiunit/biostatistics-health/>

O Centro de Investigação em Tecnologias e Serviços de Saúde (CINTESIS), embora sediado na U.Porto, é uma grande estrutura que inclui 46 instituições parceiras (29 instituições de ensino superior, 12 hospitais ou instituições de saúde e 5 empresas de saúde) e polos em 6 instituições de ensino superior. A terceira linha temática do CINTESIS é denominada Ciência de Dados, de Decisão & Tecnologias de Informação<sup>13</sup> e é composta por 9 grupos de investigação. Estes grupos são, na sua grande maioria, formados por equipas multidisciplinares de especialistas em áreas como Ciências da Computação, Engenharia, Matemática, Economia e Medicina, cujos objetivos incluem o desenvolvimento de metodologias que permitam contribuir para o processo de tomada de decisão e para a avaliação das intervenções em saúde, assim como para reunir e analisar dados e informação biomédica complexa.

Outra grande instituição de investigação interdisciplinar dedicada à inovação e pesquisa em Ciências da Saúde é o consórcio i3S<sup>14</sup>, que resulta de uma longa colaboração entre o IBMC, o INEB e o IPATIMUP. Entre os cursos oferecidos pela Unidade de Formação Avançada do i3S, podemos encontrar um dedicado à Análise Estatística de Dados Biológicos e outro com dois módulos em Estatística Aplicada.

O INESC TEC<sup>15</sup> é um Laboratório de I&D que agrupa investigadores de diferentes Faculdades da U.Porto, do Instituto Politécnico do Porto, da Universidade do Minho, da Universidade de Trás-os-Montes e Alto Douro e de outras universidades internacionais. Na realidade, o INESC TEC é uma instituição de interface que junta a academia, as empresas, a administração pública e a sociedade. Consequentemente, o seu universo é constituído por investigadores de diferentes origens e com diferentes formações. Dizer que a chave do seu sucesso está no facto de se trabalhar em equipa é pobre. Na realidade, as equipas são dinamicamente formadas com base no reconhecimento das diferentes competências individuais e na mais valia de cada um para se atingir os objetivos associados a cada projeto, construindo deste modo equipas multidisciplinares. Como tal, não é de admirar que matemáticos e estatísticos façam parte dessas equipas. No entanto, há um grupo do INESC TEC onde a sua presença é mais forte; estamos a falar do Laboratório de Inteligência Artificial e Apoio à Decisão (LIAAD)<sup>16</sup>, pertencendo ao *Cluster* de Ciência da Computação, com uma longa experiência em Inteligência Artificial. A investigação do LIAAD foca-se nas áreas de estatística, otimização, matemática e *machine learning*.

### 3 Onde estamos e para onde vamos?

Atualmente, na maior parte dos cursos, já se incluiu no currículo da disciplina de Estatística o uso de softwares como o R, o que no nosso entender, resultou do facto de os estatísticos usarem no seu dia a dia essas mesmas ferramentas e não se imaginarem a fazer uma análise estatística sem esse recurso (ou outro análogo). Por outro lado, na U.Porto temos observado o aparecimento de ciclos de estudos que prometem fazer a integração entre as áreas de Estatística e de Ciência de Computadores na formação dos seus estudantes. Mas apesar da expectativa colocada nestes ciclos de estudo para formar os futuros *Data Scientists*, acreditamos que a seleção dos conhecimentos teóricos que devem fazer parte destes ciclos de estudo está incompleta. Uma análise mais cuidada aos programas dos cursos coloca a descoberto o facto de se disponibilizarem muito poucas disciplinas de Estatística em comparação com as de Ciência de Computadores.

Nos últimos tempos, termos como *Data Science* ou *Data Analyst*, entre outros, têm surgido com frequência, o que levou Diggle (2015) [4] a afirmar que:

*“The rise of data science could be seen as a potential threat to the long-term status of the statistics discipline. I first argue that, although there is a threat, there is also a much greater opportunity to re-emphasize the universal relevance of statistical method to the interpretation of data.”*

<sup>13</sup><http://cintesis.eu/pt/linhas-de-investigacao/>

<sup>14</sup><https://www.i3s.up.pt/>

<sup>15</sup><https://www.inesctec.pt/pt>

<sup>16</sup><http://www.liaad.up.pt/>

Já Olhede *et al.* (2018) [5] referiu que:

*“Much of data science to date has focused on purely predictive “black box” tools rather than classical modeling, inference, and analysis. The role of statistics is to make our understanding of observed phenomena quantitative and precise.”*

Nós concordamos com a nossa colega Paula Brito (2016) [3], que considera que:

*“Data Science - the science of extracting knowledge from data, não é nem pode ser “apenas” Estatística, mas também não se limita à área da Computer Science. E não pode ser apenas a “colagem” de duas abordagens distintas – requer uma visão integrada, que vai buscar contributos importantes àquelas duas áreas, mas também requer competências de comunicação, de técnicas de visualização, de análise de negócio... Se os estatísticos abraçarem o desafio de se abrirem a novas formas de pensar e de comunicar, a Estatística terá um papel relevante, e que se nos afigura imprescindível, para o desenvolvimento sadio da Data Science.”*

E ainda com Regina Bispo (2017) [2]:

*“... a análise de dados obriga à dotação de múltiplas competências que vão de facto além das estritamente associadas à formação clássica em Estatística . Assim, tudo (me) indica que estamos perante uma oportunidade de mudar e evoluir, trabalhando em colaboração e complementaridade com outras áreas do saber.”*

De facto, o termo Estatística continua a aparecer nos editais publicados para contratar bolseiros ou investigadores, por exemplo no INESC TEC. No entanto, cada vez mais os conhecimentos em Estatística aparecem como fator de preferência e menos como formação base. É comum encontrar propostas de trabalho em que se pede que o candidato tenha:

- formação em ciência dos computadores (ou área afim) e bons conhecimentos de estatística,

ou

- formação em matemática ou estatística (ou área afim) e que saiba programar bem em R ou Python.

Estas opções não são equivalentes, nem tão pouco é indiferente qual destas opções é que faz parte do edital, mas isto é um reflexo do pensamento/necessidade da sociedade que a Estatística tem de acompanhar.

Resumindo, na U.Porto temos estatísticos e matemáticos a trabalhar em colaboração com colegas de diversas áreas e em diversas aplicações, fortalecendo deste modo o trabalho científico das equipas. No entanto, acreditamos que o futuro da Estatística na U.Porto, e não só, não se pode limitar a dar formação estatística a graduados em ciência dos computadores ou vice-versa, sendo importante criar um perfil novo. A sociedade necessita de cientistas dos dados, ou seja, formandos com um perfil que preencha todos os requisitos explícitos e implícitos atrás referidos. No nosso entender, o futuro da Estatística na U.Porto passa por criar, de raiz, cursos que formem estes novos perfis científicos.

Assim, ainda há muito a fazer. Voltando ao tema do Boletim e como a nossa colega Conceição Amado (2017) [1] tão bem referiu no seu artigo:

*“Às Sociedades de Estatística será requerido um esforço adicional na direção da promoção de interações no “ecossistema” da Estatística, da Computação (hardware e software), da Análise Numérica e da Otimização.”*

Um bem haja à SPE!

## Referências

- [1] Amado, C. (2017). Espelho meu, espelho meu, que futuro terei eu? *Boletim SPE outono de 2017*, **2**, 32–34.
- [2] Bispo, R. (2017). A tirania dos jargões. *Boletim SPE outono de 2017*, **2**, 38–41.
- [3] Brito, P. (2016). Data Science – um desafio para os estatísticos? *Boletim SPE outono de 2016*, **2**, 40-42.
- [4] Diggle, P.J. (2015) Statistics: a data science for the 21st century. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **178.4**, 793–813.
- [5] Olhede, S.C. e Wolfe, P.J. (2018). The future of statistics and data science. *Statistics & Probability Letters*, **136**, 46–50.



## Ser SPE – A vantagem do ónus da Boa Estatística

Jessica Silva Lomba, *jslomba@fc.ul.pt*

*CEAUL - Centro de Estatística e Aplicações, Faculdade de Ciências, Universidade de Lisboa  
Nova School of Business and Economics, Universidade Nova de Lisboa*

Convidada a escrever esta reflexão, fui confrontada com três realidades. Em primeiro lugar, fui forçada a reconhecer a minha própria ignorância (fica aqui ponto assente) no que respeita todo o caminho que foi percorrido antes de mim, todo o passado desta Sociedade à qual passei a pertencer há sensivelmente dois anos. Por outro lado, foi preciso olhar para o contexto atual e aceitar que os tempos de pandemia em que vivemos não têm precedente e que todos temos um papel a desempenhar aqui, nesta comunidade e fora dela, e agora, como agentes ativos na luta pela saúde e pela informação do público. E finalmente, encontro-me, e convido o leitor, a ponderar como poderemos influenciar e avançar, individualmente e como grupo, a causa da consagração e aceitação da Estatística como disciplina nobre e indispensável para todos os desafios comunitários desta nova era.

Como decerto não será novidade para o leitor, à data da sua criação em 1980, a nossa sociedade embandeirava o nome de Sociedade Portuguesa de Estatística e Investigação Operacional, tendo passado em 1991 a designar-se apenas Sociedade Portuguesa de Estatística. Não que eu fosse sequer nascida na altura, mas em retrospectiva leio neste facto uma clivagem de objetivos e interesses – afinal de contas, a Estatística e a Investigação Operacional são duas áreas bem distintas, com uma fronteira que pode parecer, aos mais fundamentalistas, bastante óbvia. Vistas bem as coisas, já que se lutou pela oportunidade de responder por mais do que o generalista “matemático”, ditaria a lógica que se desejassem também aqui as específicas bandeiras de “estatístico” e “investigador operacional”, não vá haver a mistura das águas. Reitero que sou desconhecadora dos motivos factuais que levaram a esta fragmentação, assim como de muito do que se passou antes e depois deste movimento, mas não seria de todo algo a estranhar que assim a história tivesse ocorrido. Mas porquê parar por aqui? Porque não fazemos agora as malas e vamos viver cada um para o seu lado, num T0 ao pé da praia, extremistas, analistas de sobrevivência, bioestatísticos...? Temos sempre a escolha de nos focarmos no que nos separa e procurar o conforto de manada, em que somos iguais aos que nos rodeiam.

Esta inclinação para a segmentação das áreas é, de resto, algo generalizado em todos os campos científicos; longe vão os tempos do amigo Gauss, que, sendo hoje rotulado de “matemático, astrónomo e físico”, contou as suas contribuições em tópicos como a teoria dos números, a Estatística, análise matemática, geometria diferencial, eletrostática e ótica, e assim se sentia na altura satisfeito com ser chamado simples e abrangentemente de cientista (H0 – hipotiso eu). Naturalmente, nem todos podemos ter competências ao nível “gaussiano” para mais do que um restrito tópico (muitos de nós, nem sequer para um), nem seria comportável que todos se especializassem num pouco de tudo. Mas não estamos nós, enquanto estatísticos, afiliados ou simpatizantes, numa posição privilegiada para interagir e colaborar, integrar e reunir o que deve ser integrado e reunido? Não temos nós o poder de ser a ponte, o poder de fortalecer os argumentos e conclusões, o poder de lançar luz sobre variadíssimas questões?

Claro está que muitos de nós já agem sobre este privilégio e responsabilidade de forma imersiva, atuando na Estatística aplicada a X, Y e Z, procurando conexões com diversos setores, e fazem-no, a meu ver, bem. Este é o meu ponto fundamental acerca do presente e necessariamente do futuro; esta é, acredito eu, a grande responsabilidade da SPE, também por intermédio da responsabilidade individual dos seus elementos – a responsabilidade de abertura a outras áreas, a outros setores fora da academia, a outras causas além do avanço científico por si só. E a Estatística só terá com isso a ganhar.

Todos já sentimos vergonha alheia, incredulidade, frustração, ao ver na comunicação social ou em artigos especializados de medicina ou sociologia, por exemplo, análises estatísticas feitas a martelo (que

é como quem diz “o *software* tem um botão que faz isso”), direcionadas às conclusões especificamente desejadas, ou simplesmente interpretadas à luz dos interesses próprios – ou pior, do sensacionalismo. Não é incomum nestes contextos que se encontre um teste de comparação de médias aplicado a dados recolhidos numa escala de Likert – que bom seria ter um estatístico à mão!

Em ano de celebração do quadragésimo aniversário da SPE, coincide a celebração quinquenal do dia Mundial da Estatística, sob o tema “*Connecting the world with data we can trust*”. Estando a importância desta questão de qualidade dos dados já bem estabelecida, queria ir aqui mais longe e sugerir que não são apenas os dados em que temos de poder confiar, mas também muito em quem os manuseia e como, em quem os lê, trata, e, fundamentalmente, interpreta e divulga. É neste ponto que devemos e podemos agir agora. Um dos efeitos desta pandemia tem sido a normalização do vocabulário estatístico usado por pivôs de notícias, que tenha passado a ser corriqueiro falar de modelos, previsões, curvas epidemiológicas e do agora famoso “*r*” acima de 1. Ótimo! Que se entenda lá fora a importância que a Estatística tem no estudo destes fenómenos. Mas que a informação seja dada de forma correta e fidedigna, por quem de facto entende de que se trata o “*r*” e qual a fiabilidade de uma previsão dada por um modelo. Que se pratique Boa Estatística. E que se entenda que é necessário que os especialistas das várias áreas trabalhem em conjunto, unam esforços pelo objetivo comum que neste caso é o mais nobre que pode existir – a saúde pública.

Neste e em tantos outros aspetos, vemos a necessidade de unir os esforços e os talentos dos elementos da nossa relativamente pequena comunidade com aqueles que trabalham sob outros interesses. Friso que não pretendo defender que todos nos viremos para as variadas aplicações, que nos associemos a um campo de estudo e que estejamos o resto das carreiras a calcular verosimilhanças e analisar tabelas ANOVA. Na verdade, quantas questões teóricas e metodológicas podem surgir dos problemas aplicados? Quanta inovação começa não pela busca pura do avanço numa determinada direção teórica, mas sim da procura de soluções para perguntas colocadas em contextos reais? E trabalhar fora da comunidade traz para os estatísticos, além de outras, a vantagem da disponibilidade e variedade de dados. Assim, é nossa responsabilidade e está nos nossos melhores interesses termos uma Sociedade aberta, virada às empresas, virada aos centros de investigação de outros campos, virada ao Estado e também às instituições de solidariedade social. Uma Sociedade sem medo das misturas e da globalização, sem amor único à bandeira, onde não seja preciso dizer “eu não sou estatístico, sou matemático” ou “eu não sou biólogo, sou estatístico”, como justificação para seja o que for, pois estaremos todos a colaborar dentro das nossas valências, mas para um mesmo objetivo. Que estejamos numa procura constante de terreno comum, da interseção de interesses, de colaborações que certamente beneficiarão toda a gente e colocarão a Boa Estatística no seu devido lugar de destaque. Uma SPE envolvida nas iniciativas e eventos dos centros de investigação e empresas, que esteja presente e se faça ver, em Portugal e internacionalmente, tem muito a ganhar. Estarmos abertos, ativos e recetivos levará a que não só tenhamos de procurar estas ligações, mas que comecemos nós a ser procurados e que o trabalho estatístico de qualidade seja valorizado.

E para que essa comunidade aberta floresça, é essencial que se atraia talento e que se enfatize a importância e versatilidade da nossa disciplina. Este esforço começa, a meu ver, numa profunda reflexão acerca do ensino da Estatística. Além da possibilidade de atrair jovens para este estudo, um ensino bem estruturado traz a vantagem de uma melhor compreensão por parte da sociedade civil dos trabalhos estatísticos difundidos na comunicação social, das suas implicações e limitações: pensemos, por exemplo, no entendimento do público relativamente às estatísticas oficiais.

O apoio e cativação dos jovens é algo que tem estado nos objetivos da SPE, através de iniciativas como o programa “A Estatística Vai À Escola”, a atribuição dos prémios “SPE” e “Estatístico Júnior” e a concessão de bolsas de participação nos congressos bienais a estudantes. Todos estes esforços são, sem dúvida alguma, importantes e louváveis, passos na direção certa. Mas seria bom conseguir fazer mais, transversalmente aos níveis de ensino. Este ponto carece, na minha opinião, de um sério debate entre todos os membros e simpatizantes, pois será neste movimento que a Sociedade se poderá renovar e prosperar.

Um ponto que me parece fulcral seria a atualização, renovação e modernização do ensino da Estatística, que, convenhamos, começa logo na desvantagem de ser uma área da Matemática, essa disciplina em geral tão pouco acarinhada. Além da difícil tarefa de envolver e cativar os alunos, é necessário dar-lhes ferramentas adequadas aos tempos correntes, mostrar a verdadeira utilidade da disciplina além da velha questão dos frangos. Da minha experiência, este ensino desatualizado, em conteúdo mas sobretudo em

forma, desmotivada e desinteressada – deixo aqui o convite à discussão, que não será de todo consensual ou unânime. A níveis de escolaridade intermédios, até que ponto não seria mais apelativo para os alunos que aprendessem a utilizar software, a programar, a par da aprendizagem dos conceitos fundamentais? Talvez seja tempo perdido estar 20 minutos de aula a passar listas de peso e altura para a calculadora ou a desenhar histogramas nos cadernos de quadriculas.

Mesmo no ensino superior, a abordagem à Estatística, nos cursos de foco matemático assim como outros, deveria, na minha opinião, ser repensada (mais que não seja, para que deixemos de ser acusados de “adivinhação” pelos colegas do determinístico com conhecimento menos do que superficial da Estatística). Tornar aparentes os méritos de encontrar um intervalo de confiança, além de demonstrar como se calculam os seus extremos. Focar casos de estudo além dos exercícios puramente académicos. O papel da SPE aqui poderá ser duplo: fomentar a discussão do tópico entre os seus distintos membros, e ser um elo entre a vertente escola da academia e as empresas, instituições e unidades de investigação de onde os casos reais brotam. Ser uma base onde possam convergir bancos de dados, fóruns de discussão, seminários de divulgação, entre outros, o que por si só têm o potencial de atrair novos membros.

A nível avançado, poderíamos falar das dificuldades de financiamento doutoral e pós-doutoral na nossa área, que são dissuasão suficiente de que jovens interessados a persigam. Idealmente, a Sociedade estaria em posição de financiar alunos, futuros membros, mas todos sabemos que não é algo ao alcance de um estalar de dedos. Contudo, esta não é a única forma de apoiar estes jovens investigadores; a divulgação e suporte de projetos como os *blogs* e *webinars* criados pela Federação de Sociedades Nacionais de Estatística (FENStatS), na figura da *Young Statisticians Europe*, é um exemplo que não deveria passar despercebido.

Em suma, queremos que ser membro da SPE seja uma evidente vantagem para quem o decide ser. E desejamos que muitos a ela decidam pertencer. Queremos reforçar as fileiras, com novas ideias e posturas. E para este fim, penso que teremos de ser missionários da Boa Estatística em todos os setores da sociedade e tomar para ela o seu lugar de destaque. No entanto, a dinamização, as iniciativas, as discussões, dependem da nossa ação individual enquanto sócios, e todos somos pessoas ocupadas. Cabe agora a cada um a reflexão não sobre “o que a Sociedade pode fazer por mim, mas o que posso eu fazer pela minha Sociedade”.



# UM departamento com ESTATÍSTICA

## Departamento de Matemática Universidade do Minho (UM)

Ana Paula Amorim, *apamorim@math.uminho.pt*, Emilia Athayde, *mefqa@math.uminho.pt*,  
Cecília Castro, *cecilia@math.uminho.pt*, Marta Ferreira, *msferreira@math.uminho.pt*,  
A. Manuela Gonçalves, *mneves@math.uminho.pt*, Raquel Menezes, *rmenezes@math.uminho.pt*,  
e M. Conceição Serra, *mcserra@math.uminho.pt*

*CMAT, Departamento de Matemática, Universidade do Minho*

Susana Faria, *sfaria@math.uminho.pt*, Luís Machado, *lmachado@math.uminho.pt*  
e Inês Sousa, *isousa@math.uminho.pt*

*CBMA, Departamento de Matemática, Universidade do Minho*

e

Carla Moreira, *d8434@math.uminho.pt*

*CMAT, Universidade do Minho*

## 1. Introdução

Trinta anos se passaram desde o início da formação do que viria a ser o grupo de Probabilidades e Estatística do Departamento de Matemática (DMAT) da Escola de Ciências da Universidade do Minho (ECUM). Este grupo tem hoje uma dezena de elementos. A Emilia Athayde, a Cecília Castro e a Ana Paula Amorim foram as primeiras a entrar. Seguiram-se a A. Manuela Gonçalves, o Luís Machado, a Susana Faria e a Conceição Serra. A Marta Ferreira e a Raquel Menezes fecharam o século. A Inês Sousa abriu a nova década que a Carla Moreira fechou em 2019, ao entrar para o Centro de Matemática (CMAT) da ECUM.

O DMAT, que se reparte desde longa data entre os dois polos da universidade (sitos em Gualtar, Braga e em Azurém, Guimarães), passou por uma fase de divisão em dois departamentos, um em cada polo, durante a qual foi criada uma licenciatura em Estatística Aplicada e um mestrado em Estatística, ambos em Azurém.

O Programa Doutoral em Matemática Aplicada – programa conjunto das universidades do Minho, Aveiro e Porto (MAP-PDMA), tem contado com a nossa colaboração, não só na lecionação de UC e orientação de projetos, como também na supervisão de teses. Da mesma forma, os outros cursos de doutoramento em que o DMAT é promotor ou parceiro, têm contado igualmente com a colaboração do grupo.

Além disso, a Licenciatura em Estatística Aplicada e o Mestrado em Estatística requerem o envolvimento ativo e contínuo de todos nós, e as solicitações na área de Estatística das Escolas e Institutos da UM têm crescido bastante e acabam por exceder a capacidade do nosso grupo. É claro que esperamos, e ansiamos, que este também cresça, pelo menos na mesma proporção, nesta nova década que está prestes a iniciar.

## 2. Projetos de ensino

Desde muito cedo que tem havido grande vontade em proporcionar formação em Estatística aos estudantes da UM. A Licenciatura em Matemática Aplicada (Ramo Estatística e Ramo Investigação Operacional) entrou em funcionamento no ano letivo 2003/2004. A sua criação foi sequência de uma proposta do Departamento de Matemática para a Ciência e Tecnologia com a colaboração do

Departamento de Produção e Sistemas e do Departamento de Sistemas de Informação da Universidade do Minho. O curso funcionou nas instalações da ECUM, no Campus de Azurém, em Guimarães. Esta licenciatura tinha no seu plano de estudos um estágio, com carácter profissional, ou, em opção, um projeto, fornecendo já uma orientação prática bastante vincada.

Quando, em 2007/08, se implementou o modelo de Bolonha, surgiu a oportunidade de oferecer o curso de Licenciatura em Estatística Aplicada (LEAp) na ECUM, no Campus de Azurém, em Guimarães, com um ciclo de estudos com seis semestres curriculares dos quais cinco são totalmente letivos e o último com um estágio/projeto curricular, estruturado de acordo com as diretrizes do Tratado de Bolonha, o que levou à extinção da Licenciatura em Matemática Aplicada (LMA). Um dos principais objetivos da LEAp é formar profissionais que dominem técnicas variadas da Estatística e preparados para aceder a profissões que requeiram conhecimentos e domínio de metodologias estatísticas, competências em análise de dados com elevado volume e/ou complexidade, permitindo a sua integração em equipas multidisciplinares para o apoio à tomada de decisões. O estágio, ou projeto, realizado no último semestre, confere também a esta licenciatura um carácter fortemente profissionalizante. Este ciclo de estudos conta com a participação de outros departamentos da Universidade do Minho na lecionação de várias UC por especialistas de outras áreas, nomeadamente da Economia e da Informática, promovendo, desta forma, o carácter interdisciplinar da Estatística. O curso tem sofrido várias remodelações, sempre com vista a acrescentar valor e oportunidade, com adequação ao tempo, incluindo atualmente UC com carácter fortemente marcado pela era da Ciência de Dados.

O Mestrado em Estatística teve também início no ano letivo de 2007/2008, inicialmente como um projeto em colaboração com a Universidade de Santiago de Compostela. O plano de estudos deste curso tinha UC que poderiam ser realizadas na Universidade de Santiago de Compostela, tal como a UM poderia receber alunos. O curso desde 2009/2010 abriu anualmente, tendo o número de candidatos crescido ao longo do tempo e preenchido a totalidade das vagas disponíveis nos últimos anos. Em 2012 o curso sofreu a sua primeira reestruturação do plano de estudos, tendo sido terminada a parceria curricular com a Universidade de Santiago de Compostela, uma vez que nessa altura havia já um número de alunos inscritos na UM que não tornava possível preencher mais vagas com alunos de outras instituições.

Em 2018 estabeleceu-se uma parceria com a Universidade de Palermo, Itália, com a possibilidade da obtenção de um diploma com dupla titulação (Universidade do Minho e Universidade de Palermo). Em 2020 o curso sofreu uma reestruturação no plano de estudos, passando a designar-se Mestrado em Estatística para Ciência de Dados. As dissertações dos estudantes podem resultar de estágios em empresas, sendo muitas, e muito variadas, as que têm vindo a receber os nossos alunos ao longo destes mais de 10 anos. São exemplos empresas privadas como AGEAS, NOS, SONAE, Continental Mabor, e empresas públicas como Banco de Portugal, Câmaras Municipais, Hospitais, Instituto Português de Oncologia (Porto) e Instituto Nacional de Estatística. Ainda no âmbito da dissertação, alguns estudantes têm colaborado em projetos de investigação.

O grupo esteve também envolvido na criação de um mestrado na área da estatística em parceria com o DPS, o Mestrado em Estatística de Sistemas, que decorreu durante dois anos letivos numa única edição. O grupo tem colaborado, ao longo dos anos, na lecionação de várias UC na área da estatística em diversos cursos da UM. Em particular, cursos da responsabilidade do Departamento de Matemática, como a Licenciatura em Matemática, a Licenciatura em Ciências da Computação e alguns mestrados na área da matemática.

### **3. Investigação**

Ao longo do tempo o grupo tem manifestado interesse em várias áreas da Probabilidade e Estatística. A área da Probabilidade tem vindo a ser utilizada em particular para modelar fenómenos físicos e biológicos. Na Estatística, o foco tem sido dirigido para a modelação de fenómenos ambientais e relacionados com a saúde. As áreas de investigação nas quais os membros do grupo estão ativamente envolvidos são análise de sobrevivência, estatística espacial e temporal, modelos longitudinais, teoria de valores extremos, processos de ramificação, análise multivariada, modelos de mistura, modelos mistos generalizados, modelos aditivos generalizados e inferência não paramétrica. Este trabalho de investigação tem sido maioritariamente desenvolvido no âmbito do Centro de Matemática da UM, mais recentemente também no Centro de Biologia Molecular e Ambiental da UM, sendo alguns dos membros do nosso grupo colaboradores de centros externos à UM, tais como o CEAUL e o CEMAT.

Uma forte componente da investigação produzida pelo grupo tem vindo a ser financiada por diversos projetos nacionais através do orçamento da FCT. Em 2009, seria financiado o primeiro projeto FCT “Métodos Estatísticos em Investigação Médica: Estudos Espaciais, Longitudinais e de Sobrevivência”, que teve na sua génese o objetivo de desenvolver e aplicar modelação conjunta de dados longitudinais e de sobrevivência em dados das ciências da saúde. Este projeto agregou algumas das linhas principais de investigação do grupo, reforçando a cooperação entre os seus elementos, assim como com instituições externas à UM com as quais já havia colaborações no passado. Em 2010 obtivemos a aprovação para financiamento de novo projeto FCT, estendendo as suas áreas de aplicação da saúde às ciências do ambiente, sob o tópico “Modelos conjuntos para processos espaço-temporais e respetivo desenho amostral”. No âmbito deste projeto, destaca-se a colaboração com o Centro de Biologia Ambiental da FCUL, que viabilizou a utilização de dados de biomonitorização de poluição atmosférica em Portugal Continental recolhidos em diferentes campanhas nacionais, que vieram a ser adotados como exemplos de motivação em vários artigos de contribuições metodológicas. Em 2013 foi aprovado para financiamento um projeto de 12 meses no âmbito dos projetos exploratórios da FCT com o título “Modelação Conjunta de Dados Multivariados Longitudinais e de Sobrevivência” em parceria com o Hospital de Braga. Neste projeto foi possível fazer uma recolha exaustiva de dados regionais do cancro da mama e desenvolver uma aplicação de monitorização de marcadores tumorais.

O grupo viu dois projetos aprovados no processo de candidaturas promovido pela FCT em 2017, maioritariamente financiado por fundos europeus. Um desses projetos é subordinado ao tema “Melhorar a estimação espacial e o delineamento amostral recorrendo a amostragem preferencial em pescas e aplicações biológicas”, tendo como instituições parceiras a NOVA.id.FCT, o IPMA, a Faculdade de Economia da UP e a FCIências.ID. É amplamente reconhecida a importância da modelação da distribuição de espécies de peixes comerciais e a sua representação através da construção de mapas, pois estes permitem uma avaliação simples de diferentes áreas marinhas, proporcionando também medidas apropriadas de incerteza associadas às estimativas. Devido à quantidade limitada de dados e às dificuldades inerentes à modelação dos dados das pescas, ainda existem muitos problemas em aberto na modelação espacial e espaço-temporal destes recursos marinhos. Este projeto, agora a decorrer e com término expectável em 2022, tem na sua génese o objetivo de contribuir para o estudo da biodiversidade e sustentabilidade de espécies marinhas.

O segundo projeto aprovado, com o título “Long Time Health: Modelos Conjuntos para Dados Longitudinais e Tempo-até-Evento para as Ciências da Saúde”, é um projeto em colaboração com o Instituto de Saúde Pública da Universidade do Porto e foi apresentado como uma continuação do trabalho desenvolvido no projeto exploratório terminado em 2015. Neste projeto, que se deverá prolongar até 2022, dá-se continuação ao desenvolvimento de modelos estatísticos para dados longitudinais motivados por dados clínicos e das ciências biológicas.

Para além dos projetos FCT mencionados, coordenados por elementos do grupo, tem-se mantido desde sempre a nossa participação em inúmeros projetos nacionais e internacionais, coordenados por investigadores externos ao grupo, que abordam temas de investigação atuais e alinhados com os nossos interesses. Estas e outras colaborações têm envolvido entidades de renome tais como as Universidades de Valencia, Santiago de Compostela e Vigo, Lancaster University, University College Dublin, Chalmers University Technology, Leiden University e Université de Lille. Adicionalmente, temos assinado contratos de colaborações com instituições públicas e privadas, tais como o Departamento do Mar e Recursos Marinhos do IPMA, os Hospitais de Braga e de Guimarães e empresas como a Congelagos – Transformação e Comércio de Produtos Alimentares AS, que têm demonstrado ser importantes para motivação das contribuições metodológicas da autoria ou co-autoria dos elementos do grupo.

É importante realçar que a dinâmica imposta pelos diversos projetos de investigação tem permitido viabilizar financiamento para a contratação de bolseiros com diferentes tipologias, nomeadamente investigadores doutorados, assistentes de investigação ou mesmo bolseiros de iniciação científica. O grupo, consciente da importância da formação avançada de novos investigadores, conta até ao momento com um total de 7 projetos de doutoramento terminados, 5 em desenvolvimento, 5 post-docs e cerca de duas dezenas de bolseiros BI, BIC e BM. No que respeita à divulgação do trabalho desenvolvido, podemos indicar mais de uma centena de publicações em revistas ISI ou Scopus, assim como a organização e a participação em diversas conferências nacionais e internacionais.

#### 4. Outras atividades

O dinamismo do grupo está bem patente na organização de eventos científicos nacionais e internacionais na área da Estatística nos últimos anos.

A nível internacional, em outubro de 2005, organizou o "I Congresso de Estatística e Investigação Operacional da Galiza e Norte de Portugal / VII Congresso Galego de Estatística e Investigación de Operacións - CEIO'05" em Guimarães, em colaboração com a Sociedade Galega para a Promoción da Estatística e da Investigación de Operacións. Realça-se ainda, em setembro de 2012, a organização da "VI International Workshop on Spatio-Temporal Modelling (METMAVI)" em Guimarães e, em julho de 2013, do "I Encontro Português de Biometria e o I Encontro Luso-Galaico de Biometria" em Braga. Mais recentemente, em julho de 2019, destaca-se a organização do congresso "International Workshop on Statistical Modelling – IWSM2019", em Guimarães, em parceria com a Statistical Modelling Society. Já a nível nacional, sobressaem as participações dos membros do grupo na organização do "III Congresso Anual da SPE" em junho de 1995, em Guimarães, na organização do "VII Congresso Anual da SPE" em outubro de 1999, em Ofir, e na organização das "XX Jornadas de Classificação e Análise de Dados – JOCLAD 2013" em abril de 2013, em Guimarães. Foram também organizados desde 2012 vários *workshops* e encontros de curta duração.

No âmbito de orientações conjuntas de dissertações de mestrado ou de estágios de licenciatura, os membros do grupo têm estabelecido protocolos com várias Empresas/Instituições dos mais variados sectores, nomeadamente, Instituto Nacional de Estatística, Hospital de Braga, Hospital de Guimarães, Instituto Português de Oncologia, Administração Regional de Saúde Norte e Centro, SONAE, NOS, Farfetch, Continental Mabor, Banco de Portugal, Millenium BCP e AGEAS, entre outros.

Todos os anos, com o objetivo de divulgar e promover a Estatística, os membros deste grupo participam na organização de atividades dirigidas ao público em geral ou destinadas a estudantes e/ou professores dos ensinos básico e secundário. Dessas atividades realçam-se as "II Jornadas UM Futuro com Ciência", o "UPA – Universidade Portas Aberta", o "Verão no Campus", a "Noite Europeia dos Investigadores" e o Programa de Visitas da ECUM ao longo de todo o ano letivo. Durante vários anos, colaboraram na atividade "iSci - Interface Ciência" promovida pela ECUM, que permitia a interação dos alunos dos vários ciclos de estudo com o ambiente empresarial. A exposição itinerante "Explorística" esteve patente na Biblioteca Lúcio Craveiro da Silva em Braga, em 2013.

Destaca-se ainda, no âmbito da comemoração do Dia Mundial da Estatística, em 20 de outubro de 2015, a contribuição do grupo na organização de uma exposição sobre a vida e obra de estatísticos incontornáveis no desenvolvimento da Estatística, apresentada na Biblioteca Lúcio Craveiro da Silva, réplica da comemorativa do Ano Internacional da Estatística (2013) que decorreu na Galeria do Paço da UMinho, em Braga.

Por fim, a nível de divulgação científica, salienta-se o empenho do grupo na organização de seminários, com uma periodicidade mensal, e de *workshops*, com uma periodicidade anual, que contaram com a participação regular de investigadores nacionais e estrangeiros.

#### 5. E o futuro?

Queremos continuar a assumir o compromisso com as necessidades de formação na região e dar resposta às expectativas dos atuais e futuros estudantes apostando na formação nesta área incontornável e em constante mudança. Como investigadores pretendemos contribuir para o desenvolvimento de novos modelos e métodos para dar resposta a problemas em aberto. O tratamento matemático destes modelos e métodos, o estudo das suas propriedades teóricas e a análise do seu comportamento através de estudos de simulação e em aplicações, ocupam uma parte importante da nossa atividade de investigação. A implementação computacional dos novos métodos e a sua difusão na comunidade científica também fazem parte da nossa missão como investigadores em Estatística. Com o fim de proporcionar partilha de conhecimento, continuaremos a colaborar com investigadores de outras áreas científicas, a estabelecer parcerias com empresas e organizações e a promover atividades de divulgação e interação com a sociedade.



# SPE – 40 anos de “braço dado” com a Estatística

Célia Nunes, *celian@ubi.pt*

*Departamento de Matemática e Centro de Matemática e Aplicações, Universidade da Beira Interior*

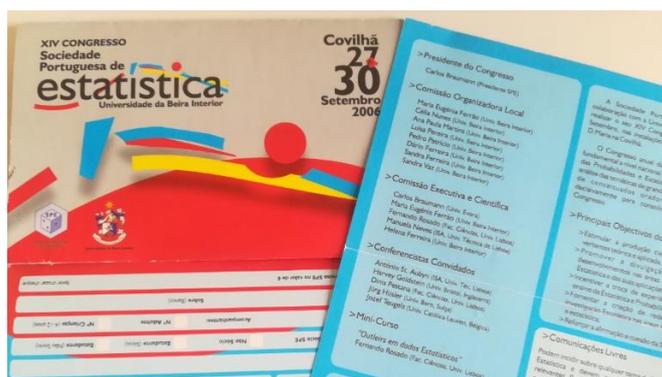
É impossível falar da SPE sem fazer referência aos seus congressos. Assim, quando recebi o amável convite por parte do Professor Fernando Rosado para escrever este texto, ocorreu-me de imediato falar sobre o assunto.

Os Congressos da SPE, que até 2013 foram anuais e que a partir dessa data passaram a bienais, têm sido, segundo a Professora Manuela Magalhães Hill “um veículo insubstituível de promoção da Estatística e das Probabilidades no nosso país e a nível internacional”, Hill (2006). Já no Boletim de outono de 2016, o professor Carlos Braumann referiu que os congressos “para além de constituírem uma vivida e desejada celebração da comunidade e do seu labor já consagrada nos seus hábitos, permitiram um melhor conhecimento mútuo e um incentivo ao desenvolvimento científico e aperfeiçoamento profissional”, Braumann (2016). A professora Manuela Neves define-os como “a reunião magna da nossa sociedade, a oportunidade de iniciação de “jovens” estatísticos, de partilha e de aquisição de novas experiências dos utilizadores da Estatística, o momento de encontro dos “menos jovens”, o reencontro de “amigos jovens ou menos jovens””, Neves (2016).

Em 2006, tive o prazer de fazer parte da comissão organizadora do XIV Congresso da SPE e do corpo editorial das atas do encontro, em conjunto com a Professora Maria Eugénia Ferrão e o Professor Carlos Braumann, o então presidente da “nossa” Sociedade. Para mim, e certamente para os restantes membros da comissão organizadora, foi um gosto acolher na cidade da Covilhã e na Universidade da Beira Interior toda a nossa comunidade científica. E julgo que posso dizer “toda” sem correr o risco de exagerar, já que o evento contou com 195 participantes. Foi sem dúvida uma excelente oportunidade para mim, uma então “jovem” estatística.

Na altura, fiquei a saber que em 1981 se tinha realizado na Covilhã e no Fundão o II Colóquio da Estatística e Investigação operacional, organizado pela então Sociedade Portuguesa de Estatística e Investigação Operacional e pelo Instituto Universitário da Beira Interior. Tal evento levou o Professor Carlos Braumann a escrever no prefácio das atas do congresso de 2006, algo que me enche de orgulho: “Não foi só um voltar a casa, foi o sentirmo-nos em casa na imponente paisagem serrana, com a hospitalidade e o acolhimento da Comissão Organizadora, do Departamento de Matemática e do Sr. Reitor da Universidade da Beira Interior”, Ferrão et al. (2007). E sentirmo-nos em casa não é afinal a melhor sensação do mundo?

Na opinião da Professora Manuela Neves, o sucesso dos congressos “resulta de que eles são preparados “com o coração!””, Neves (2016). Não podia estar mais de acordo! Aliás, esse é por certo um dos motivos pelo qual ainda guardo religiosamente no meu gabinete alguns *flyers* do encontro (ver figura abaixo).



Após a realização de 24 congressos e com a entrada nos “entas” da SPE, há sem dúvida uma questão que se nos coloca: E para onde vamos?

Nestes últimos 40 anos, a estatística “evoluiu” bastante a nível internacional. Como não podia deixar de ser, entre nós, essa evolução também se fez sentir, resultando num crescente reconhecimento da nossa comunidade científica não só a nível nacional, mas também a nível internacional. Essa evolução, muito se deve por certo à SPE, como a Sociedade que promove e cultiva o estudo da Estatística em Portugal. Apesar de um dos objetivos da Estatística ser o de “estimar ou possibilitar a previsão de fenómenos futuros”, os desafios que iremos encontrar pela frente constituem uma incerteza até mesmo para nós, estatísticos. No entanto, ao pensarmos no futuro da Estatística, algo que não pode ser completamente dissociado do que será também o futuro da “nossa” Sociedade, estou certa de que a muitos de nós nos ocorrerão termos como “*Data Science*”, “*Big Data*” ou “*Machine Learning*”.

Em outubro de 2015, num comunicado à imprensa, David Morganstein, o então presidente da ASA (*American Statistical Association*), evidenciou a importância da Estatística na Ciência dos Dados: “*Through this statement, the ASA and its membership acknowledge that data science encompasses more than statistics, but at the same time also recognize that statistical science plays a critical role in the fast-growing field. It is our hope the statement will reinforce the relationship of statistics to data science and further foster mutually collaborative relationships among all key contributors in data science*”, ASA (2015). No ano seguinte, a Professora Paula Brito escreveu um artigo no Boletim de outono de 2016, onde, de uma forma bastante interessante, faz a interpretação do porquê e como estas “temáticas” constituem um desafio para os estatísticos, Brito (2016). Sobre *Data Science* a colega escreveu: “*the science of extracting knowledge from data, não é nem pode ser “apenas” Estatística, mas também não se limita à área da Computer Science. E não pode ser apenas a “colagem” de duas abordagens distintas – requer uma visão integrada, que vai buscar contributos importantes àquelas duas áreas, mas também requer competências de comunicação, de técnicas de visualização, de análise de negócio... Se os estatísticos abraçarem o desafio de se abrirem a novas formas de pensar e de comunicar, a Estatística terá um papel relevante, e que se nos afigura imprescindível, para o desenvolvimento sadio da Data Science*”. Já o Professor Fernando Rosado considera que “o entusiasmo à volta dos *big data* é bom para a estatística e os estatísticos devem colaborar com os especialistas para conjuntamente poderem avançar na procura de resultados úteis para os problemas complexos que estas estruturas estão a colocar”, Rosado (2016). Ainda sobre esta “temática”, o interessante artigo da Professora Conceição Amado, publicado no Boletim de outono de 2017, merece a nossa especial atenção, Amado (2017).

Também para mim, o futuro da Estatística passa, em parte, por uma cada vez maior conjugação e partilha de saberes em diferentes áreas e domínios. De entre todas, a ligação entre a Estatística e a Ciência dos Computadores será por certo uma das mais entusiasmantes. Estou certa de que se fará sentir cada vez mais a necessidade do desenvolvimento de novas metodologias estatísticas, parte delas fortemente fundamentadas através da teoria das probabilidades, por forma a fazer face à resolução de problemas em grandes volumes de dados. A título de conclusão, há algo que julgo podermos depreender desta discussão, é que parte da “receita” para melhor encararmos o futuro que se avizinha deverá passar por “*being Open-Minded*”.

Focando-nos no futuro da “nossa” Sociedade, com mais ou menos desafios que todos nós iremos encontrar pela frente, julgo que podemos considerar como “acontecimento certo” que a entrada nos “entas” da SPE não lhe irá trazer “falta de visão”, e cabe-nos a todos nós continuar a contribuir para que tal nunca aconteça.

## Referências:

- Amado, C. (2017). Estatística - “Espelho meu, espelho meu, que futuro terei eu?”. Em Rosado, F. (Ed.). O Tema Central da Estatística – um novo olhar. *Boletim SPE outono de 2017*, Edições SPE.
- ASA (2015). ASA issues statement on role of statistics in data science. In a policy statement, October 1, 2015. <https://phys.org/news/2015-10-asa-issues-statement-role-statistics.html>
- Braumann, C. (2016). Reflexões Estatísticas. Em Rosado, F. (Ed.). O tema central da Estatística. *Boletim SPE outono de 2016*, Edições SPE.
- Brito, P. (2016). *Data Science* – um desafio para os estatísticos? Em Rosado, F. (Ed.). O tema central da Estatística. *Boletim SPE outono de 2016*, Edições SPE.

- Ferrão, M.E., Nunes, C. and Braumann, C.A. (2007) (Eds.). Estatística: Ciência Interdisciplinar. Actas do XIV Congresso Anual da Sociedade Portuguesa de Estatística. Edições SPE. ISBN 978-972-8890-12-4
- Hill, M.M. (2006). Notícias: XV Congresso. Em Rosado, F. (Ed.). Ensino e Aprendizagem da Estatística. *Boletim SPE outono de 2016*, Edições SPE.
- Ferrão, M. M. (2016). Estatística Educacional: a propósito dos Modelos de Valor Acrescentado. Em Rosado, F. (Ed.). O tema central da Estatística. *Boletim SPE outono de 2016*, Edições SPE.
- Neves, M. (2016). Uma reflexão enviesada (!) sobre a “nossa” SPE. Em Rosado, F. (Ed.). O tema central da Estatística. *Boletim SPE outono de 2016*, Edições SPE.
- Rosado, F. (2016). O Futuro da Estatística. Em Rosado, F. (Ed.). O tema central da Estatística. *Boletim SPE outono de 2016*, Edições SPE.



# Estatística na Universidade de Évora

Russell Alpizar-Jara, [alpizar@uevora.pt](mailto:alpizar@uevora.pt)

*Centro de Investigação em Matemática e Aplicações, Instituto de Investigação e Formação Avançada  
Departamento de Matemática, Escola de Ciências e Tecnologia, Universidade de Évora*

A Sociedade Portuguesa de Estatística está em festa! É o momento para celebrar esta importante data onde se comemora o 40º aniversário de consolidação de uma Sociedade ao serviço da Comunidade. A SPE muito tem contribuído **para promover, cultivar e desenvolver, em Portugal, o estudo da Estatística, as suas aplicações e ciências afins**. Era neste ano precisamente, que esperávamos comemorar, na cidade de Évora, património da humanidade, o XXV congresso da SPE em homenagem ao seu 40º aniversário. É do interesse desta Sociedade **unir todos os estatísticos e juntar pessoas que trabalham em diferentes áreas da estatística nas universidades, no sector privado e na administração pública**. Infelizmente, a situação pandémica que enfrentamos obrigou-nos a adiar este importante evento. No entanto, há muitas outras formas de festejar e é nesse sentido que se realizam as várias contribuições deste Boletim. As linhas anteriores, sublinhadas em itálico, fazem parte dos objetivos desta Sociedade [<https://www.spestatistica.pt/spe/missao-e-objetivos>].

Foi-me solicitado pela equipa editorial do Boletim, fazer um apanhado das atividades e contribuições mais importantes do grupo de Évora, no âmbito do desenvolvimento da Estatística em Portugal e a sua projeção na Comunidade. Ciente de que haverá pessoas mais indicadas para elaborar este testemunho, aceitei o desafio, agradecendo a confiança dos editores e desculpando-me desde já por qualquer omissão ou má interpretação dos factos mais relevantes, pois a minha perspetiva não deixa de ser uma opinião baseada na vivência pessoal como docente da Universidade de Évora e como sócio da SPE, entidades às quais tenho um grande orgulho de pertencer e servir.

Faz pouco mais de 21 anos que imigrei para esta linda e inigualável terra Alentejana que agora considero a minha casa: Évora! Após um longo percurso, pelos Estados Unidos da América, onde completei todos os meus estudos de pós-graduação, e após ter emigrado do meu país natal, Costa Rica, considero-me um legítimo Alentejano adotado. Por esta razão, qualquer referência ao passado, remonta na realidade a um passado recente, por mim vivenciado! No entanto, farei também um breve relato da Estatística que encontrei na altura e que, como muitos dos leitores sabem, foi fortemente acarinhada e valorizada desde os primórdios da Universidade de Évora.

## Os primórdios

Tal como descrito na breve nota histórica do Departamento de Matemática (DMAT) da Universidade de Évora [<https://www.dmat.uevora.pt/departamento/historia>], desde a fundação do Departamento em 1975, então designado Secção de Matemática, o primeiro responsável direto foi o nosso colega José Rodrigues Dias, que mais tarde veio a ser substituído por António Samuel do Carmo Moral. José Rodrigues Dias foi também o primeiro Presidente eleito do DMAT. O segundo Presidente eleito foi o nosso colega Carlos Braumann. Ambos são atualmente Professores Eméritos na Universidade de Évora. Assim, não é por acaso que a Estatística na Universidade de Évora sempre esteve muito bem representada, não só institucionalmente pelos contributos destes dois colegas no que se refere à investigação e à docência na área da Estatística, mas também pelos vários cargos desempenhados a nível dos órgãos de gestão institucional. Na oferta formativa do DMAT, nas licenciaturas, mestrados e no doutoramento, sempre houve uma forte componente nas áreas das Probabilidades, Estatística e

Investigação Operacional. Igualmente, a Estatística sempre esteve presente em todas as áreas de formação da Universidade de Évora. Naturalmente, o desenvolvimento da Estatística nesta instituição também não tinha sido possível sem o apoio e colaboração de outros colegas estatísticos vindos de várias zonas do país, em particular, das Universidades de Lisboa e Nova de Lisboa, entre outras. Como resultado destas interações, vários colegas do atual corpo docente tiveram formação inicial nesta casa e, posteriormente, concluíram os seus doutoramentos nesta e noutras instituições nacionais ou estrangeiras.

Em 1993 foi criado o Centro de Investigação em Matemática e Aplicações (CIMA) da Universidade de Évora [<http://www.cima.uevora.pt/>]. Este Centro veio impulsionar a investigação em Matemática e nas suas aplicações, convertendo-se no único Centro de investigação nesta área na parte sul de Portugal continental, sendo muito importante para a projeção local, regional e nacional. Um dos fundadores e primeiros diretores deste Centro foi o nosso colega Carlos Braumann, tendo desde então existido um grupo de investigação dedicado à área de Processos Estocásticos, Estatística e Investigação Operacional, o qual coordenou durante vários anos.

### **O passado (mais recente)**

Em 1999, o Departamento de Matemática oferecia 3 licenciaturas: Matemática Aplicada (com uma área de especialização em Estatística), Ensino da Matemática e Engenharia Informática. As 3 formações pré-Bolonha tinham uma forte componente nas áreas de Probabilidades e Estatística. Existia também um programa de Mestrado em Matemática Aplicada e o Programa de Doutoramento em Matemática (ambos com áreas de especialização em Estatística). O Departamento tinha cerca de 55 docentes, a grande maioria Assistentes a fazer doutoramentos na altura. Este corpo docente incluía colegas nas áreas das Ciências da Computação, que posteriormente passaram a formar o Departamento de Informática da Universidade de Évora. Ao longo destes anos tem havido várias reestruturações na oferta formativa do Departamento, nalguns casos para adequação ao processo de Bolonha. Algumas destas formações, como a Licenciatura em Matemática Aplicada, o Mestrado em Matemática e Aplicações e o Mestrado em Matemática para o Ensino, viram a certa altura as suas atividades suspensas no âmbito das políticas institucionais de redução de cursos que não tivessem um número mínimo de vagas preenchidas. Convém destacar que na última edição do Mestrado em Matemática e Aplicações foi adicionada a modalidade deslocada em Santo Tomé e Príncipe, sendo na altura a diretora da Comissão Executiva e de Acompanhamento a nossa colega Dulce Gomes. Neste país, o curso foi frequentado por alguns técnicos superiores e por professores do ensino secundário e superior, que após conclusão do mestrado ocuparam postos de chefia.

Tem sido visível a participação dos colegas do Departamento de Matemática na organização de vários eventos científicos na área da Estatística. Convém realçar a organização do XII Congresso da SPE que teve lugar em Évora em 2004, sendo a Comissão Organizadora Local presidida pelo colega Carlos Braumann, que veio a ser Presidente da SPE no período de 2006-11 (ver depoimento neste volume). Este congresso teve 254 participantes e 165 comunicações científicas, sendo a maioria delas orais e outras em poster. Contou com os ilustres oradores estrangeiros David Cox, Bradley Efron e Daniel Peña e o patrocínio de várias entidades, com particular relevo o Instituto Nacional de Estatística e a Câmara Municipal de Évora. Como resultado deste importante evento, foi publicado o livro de Atas com um total de 819 páginas, composta com artigos que seguiram um rigoroso processo de revisão por pares. Mais recentemente (2016), foram organizadas na Universidade de Évora as Jornadas da Associação Portuguesa de Classificação e Análise de Dados (CLAD), sendo a Comissão Organizadora Local presidida pelo colega Paulo Infante. Vários colegas do DMAT têm participado também na organização de atividades científicas diretamente ou indiretamente ligadas com a investigação em Estatística, como por exemplo: *Workshops on Estimating Animal Abundance; Sampling and Experimental Designs with Applications; Statistics, Mathematics and Computation; Computational Data Analysis and Numerical Methods; and Dynamical Systems Applied to Biology and Natural Sciences*, entre outros.

Para chegar aos mais jovens, o Departamento tem promovido várias palestras de divulgação nas Escolas da região (Ensino Básico e Secundário) e tem publicado crónicas de divulgação matemática nos jornais regionais, muitas na área de Estatística, dirigidas ao público em geral. Os colegas do grupo de Estatística

têm aderido fortemente a estas iniciativas e já colaboraram conjuntamente com o projeto *A Estatística vai à Escola* (AEVAE) da SPE. Os colegas Dulce Gomes e Paulo Infante, também integraram ou integram as equipas do Radical Estatística e/ou Explorística, projetos da SPE.

## O presente

Como resultado das reestruturações do processo de Bolonha e da avaliação dos cursos pela A3ES, temos atualmente em funcionamento, no DMAT, as seguintes formações:

- Licenciatura em Matemática Aplicada à Economia e Gestão, resultado de um esforço interdepartamental com os departamentos de Gestão e de Economia da Universidade de Évora, e cuja Comissão Executiva e de Acompanhamento é dirigida pelo colega Paulo Infante. Este ano entraram 53 estudantes nesta licenciatura, sendo cerca de 20%, alunos internacionais, e tendo-se ultrapassado as 40 vagas propostas para admissão.

- Mestrado em Modelação Estatística e Análise de Dados. Este ano tivemos 22 candidaturas e 18 admitidos, tendo-se inscrito 16 alunos, dos quais 5 são estrangeiros provenientes de países de língua portuguesa. A Diretora da Comissão Executiva de Acompanhamento deste Mestrado é a colega Anabela Afonso. Os colegas Manuel Minhoto (anterior Diretor) e Dulce Pereira também integram esta Comissão. A colega Patrícia Filipe foi adjunta desta comissão por vários anos, até muito recentemente.

- Doutoramento em Matemática nas modalidades presencial e e-learning, sendo, nesta última modalidade, pioneiros no país. Este ano foram admitidos 15 estudantes, a maioria estrangeiros vindo principalmente de países de língua portuguesa. A colega Manuela Oliveira tem sido adjunta da Comissão Executiva e de Acompanhamento deste Programa Doutoral por vários anos.

Aludindo às palavras do Sr. Diretor do DMAT (o colega Fernando Carapau) na página do departamento: *“O Departamento de Matemática de hoje, com olhos no futuro, está alicerçado no passado. Existindo desde 1975, formou centenas de Professores de Matemática e centenas de Matemáticos, que hoje se encontram nas mais diversas atividades profissionais e nas mais variadas regiões do país e estrangeiro. Mais recentemente tem recrutado vários alunos e bolseiros dos programas de ERASMUS, ampliando também assim a sua projeção internacional.”* Muitos profissionais a que o colega refere, foram na área de especialização da Estatística, e também uma grande quantidade de estágios, trabalhos de fim de curso, teses de mestrado e de doutoramento se realizaram nesta área.

Como resultado da reestruturação dos Centros de Investigação em Portugal, o Centro de Investigação em Matemática e Aplicações alargou-se formando um polo na Universidade da Madeira, contando também neste local com estatísticos como o colega José Luís da Silva, responsável pela Coordenação científica do polo e da linha de investigação em Estatística, Processos Estocásticos e Aplicações do CIMA.

A investigação na área da Estatística no CIMA é fundamentalmente aplicada. O desenvolvimento de novas metodologias é motivado pelos problemas nas áreas de aplicação e pela interação dos investigadores em equipas interdisciplinares. Desenvolve-se trabalho em várias linhas de ação, mas com grande incidência nas áreas de Processos Estocásticos e Séries Espaço-Temporais, Controlo de Qualidade, Análise de Sobrevivência, Equações Diferenciais Estocásticas, Estatística Multivariada e Delineamento Experimental, Amostragem, e mais recentemente Estatística de Extremos com a contratação da colega Lígia Henriques-Rodrigues. A aplicação destas metodologias nas Ciências da Vida tem sido o principal foco de atenção, mas recentemente, grande parte das aplicações também se têm centrado nas Ciências Sociais e da Saúde. Vários exemplos destas aplicações e da projeção social do Centro ao serviço da Comunidade podem ser encontrados nos seus relatórios científicos. Nestes relatórios também se pode apreciar o impacto do grupo de Évora nas publicações científicas, geralmente derivadas dos projetos em que se integram, sendo muitas delas em coautoria com investigadores de outros centros e no seio da interdisciplinaridade.

## O futuro

A reestruturação da oferta formativa é um processo dinâmico, à procura de servir a Comunidade através da Academia. Recentemente foi aprovada pelos órgãos institucionais o relançamento da Licenciatura em

Matemática, sendo o grupo de trabalho liderado pelo colega Gonçalo Jacinto. Também está na mesa a proposta de reestruturação do Mestrado em Matemática. Ambas as formações têm uma forte componente na área disciplinar da Estatística e constituem uma fileira que culmina num 3º ciclo com o Programa de Doutoramento em Matemática. Aproveitaremos esta reestruturação para fortalecer o recrutamento de alunos nacionais e estrangeiros, alguns provenientes dos países de língua portuguesa. Desejamos também continuar a estreitar as colaborações com a Universidade de Extremadura, Espanha, dada a sua proximidade geográfica.

Temos que apostar num Ensino misto, com uma forte componente à distância. Urge também fortalecer o corpo docente e substituir os colegas que se vão reformando ou que saem da instituição. Esta reposição geracional é imprescindível para dar continuidade ao legado que este grupo tem conseguido ao serviço da Estatística e da Sociedade.

Recentemente o Conselho Científico do CIMA aprovou a extensão do Centro para a criação de um polo no Instituto Superior de Engenharia de Lisboa (ISEL). Assim, a tendência futura do Centro é expandir os seus horizontes sempre na procura duma maior interação com membros da Academia noutras localidades do país. O futuro aponta para a necessidade de criar pontes e desbloquear barreiras entre áreas do conhecimento, como, por exemplo, Estatística Algébrica, Estatística nos Sistemas Dinâmicos, Inteligência Artificial e *Big Data*, entre outras.

Os colegas do DMAT e do CIMA, certamente continuarão a dar as suas contribuições e a participar nas várias atividades e eventos científicos em prol do desenvolvimento futuro da Estatística em Portugal e a sua projeção a nível internacional.

Esperamos por todos vós em Évora, no XXV Congresso da SPE de 13-16 outubro 2021, mais perto do seu 41º aniversário! Nesta oportunidade, formam parte da Comissão Organizadora, as colegas Dulce Gomes, Lígia Henriques-Rodrigues e Patrícia Filipe, que recentemente migrou para o ISCTE, mas continua a ser membro integrado do CIMA.

#### ***Nota de agradecimento:***

Deixo aqui um agradecimento especial a todos os colegas docentes e investigadores, funcionários e alunos, que ao longo destes anos fizeram ou fazem parte do DMAT e/ou do CIMA da Universidade de Évora, e que têm enriquecido a Ciência Estatística, quer pelos seus ensinamentos, quer pela investigação que realizam e as suas contribuições na vida da comunidade Portuguesa, quer simplesmente pela simpatia e apreciação que têm por esta importante área do conhecimento! São muitas as pessoas que têm passado por esta *Alma mater* e deixaram os seus contributos. É graças a todos eles que a nossa instituição projeta essa visibilidade a nível regional, nacional e internacional. Bem hajam!

#### ***Aviso legal:***

Este texto é da total responsabilidade do autor e não reflete a opinião de nenhum dos colegas nele referido, nem da instituição.



# Uma caminhada por valores extremos de precipitação na Ilha da Madeira

Délia Gouveia Reis, *delia.reis@staff.uma.pt*

*Departamento de Matemática, Faculdade de Ciências Exatas e da Engenharia, Universidade da Madeira, Portugal*

*e*

*CEAUL – Centro de Estatística e Aplicações, Faculdade de Ciências, Universidade de Lisboa*

## Introdução

De onde viemos? Onde estamos? Para onde vamos? A Sociedade Portuguesa de Estatística, com os seus 40 anos de existência, conta com muitos membros e diversas caminhadas pelas áreas das Probabilidades e da Estatística. O tema deste boletim incentivou-me a recordar algumas etapas da minha curta e modesta caminhada pela área de valores extremos. Tive a sorte desta minha caminhada ter tido a orientação da Professora Sandra Mendonça, mas também o suporte e o incentivo de outros membros da Sociedade Portuguesa de Estatística como as Professoras Ana Abreu, Ivette Gomes e Rita Vasconcelos e dos Professores Dinis Pestana e Sílvio Velosa. Todos os congressos nos quais tive a oportunidade de estar presente são marcos importantes neste meu simples trajecto. No XIX congresso da Sociedade Portuguesa de Estatística, realizado na Nazaré em 2011, como nos outros após esse congresso, tive a oportunidade de aprender bastante e de sentir a alegria dos saudáveis convívios. Assim, descrevo em seguida as etapas da minha caminhada por valores extremos de precipitação na Ilha da Madeira que cruzaram com esses congressos, numa forma singela de agradecimento pelo que me foi dado a conhecer por todos os membros da Sociedade Portuguesa de Estatística.

## Valores extremos de precipitação na Ilha da Madeira

A Ilha da Madeira, apesar de ser uma ilha vulcânica com uma área de apenas 737 km<sup>2</sup>, apresenta diferentes regiões relativamente aos extremos de precipitação, em consequência da sua complexa orografia e do pronunciado desnível das suas vertentes (e.g. [10]). Os extremos de precipitação constituem um factor importante na ocorrência de cheias rápidas e fluxos de lamas e de detritos, que marcam o passado da ilha [11]. Dos eventos ocorridos num passado mais recente, foi o de 20 de Fevereiro de 2010 o mais devastador, tendo provocado 45 mortes e danos significativos em muitas infraestruturas [2].

A distribuição de Gumbel é um dos modelos extremos mais aplicados na modelação de extremos hidrológicos, especialmente quando o objetivo é quantificar o risco associado à precipitação extrema. No entanto, existem estudos que sugerem que a distribuição de Gumbel pode subestimar os maiores valores extremos de precipitação, o que leva a uma preferência pela distribuição generalizada de valores extremos (GEV) (e.g., [8]). O estudo de valores extremos por meio da metodologia de Gumbel apresenta a desvantagem de considerar apenas um máximo por bloco, o que pode limitar esse estudo. Uma metodologia alternativa para a metodologia de Gumbel é a metodologia PORT (*peaks over random threshold*), cujo foco está nas  $k$  observações da amostra que excedem um nível aleatório (e.g. [4]). Neste contexto, o método para testar a suposição de que a função de distribuição  $F$  para  $n$  variáveis aleatórias independentes pertence ao domínio de atracção de uma distribuição de valores extremos apresentado em [1] foi aplicado aos valores máximos mensais de precipitação diária da época das chuvas de sete estações pluviométricas [5]. Neste estudo elaborado em [5], foram também aplicados os procedimentos

estatísticos para a escolha estatística de domínios de atração para máximos analisados em [9] considerando os valores disponíveis. Os resultados da análise realizada indicam o valor de  $k$  a utilizar e o sinal de cada índice de valores extremos para cada localização considerada.

O segundo cruzamento entre a minha caminhada pelos valores extremos e os congressos da Sociedade Portuguesa da Estatística foi no Porto em 2012. A dependência entre variáveis tem um papel fulcral na análise multivariada de extremos. A grande vantagem na abordagem ao estudo dessa dependência por meio das funções cópula consiste na possibilidade de descrever e modelar a dependência entre variáveis, independentemente das respectivas funções de distribuição marginais. Na prática, a aplicação das funções cópula pode assentar na estimativa de uma medida de associação, como, por exemplo, o tau de Kendall (e.g. [12]). Assim, com o intuito de quantificar tal dependência no caso de extremos de precipitação na ilha da Madeira, foi calculado no estudo em [7] uma estimativa para o valor do tau de Kendall e o valor-p para o correspondente teste de independência [3]. Nesse estudo foram analisados os máximos anuais de precipitação utilizados, em mm, de 17 estações e postos udométricos com vista a um posterior ajuste à uma família de cópulas de extremos envolvendo a família de Marshall-Olkin, cujos parâmetros podem ser escritos em função do tau de Kendall.

Outra metodologia alternativa para a metodologia de Gumbel é a metodologia POT (*peaks over threshold*), na qual estava focada a minha atenção na altura do XXI congresso da Sociedade Portuguesa de Estatística realizado em Aveiro. Em [6], foi efectuada uma procura de valores para o limiar de séries de valores diários de precipitação de 12 estações meteorológicas e postos udométricos por meio de uma metodologia assente numa análise gráfica. Apesar dos valores obtidos neste estudo, como também no estudo apresentado em [5], serem resultantes do enfraquecimento das hipóteses de independência e de homogeneidade, os resultados já indiciam que a caracterização dos extremos de precipitação na ilha da Madeira deverá ter em conta factores como a vertente e a altitude onde estão localizadas as estações e a sua proximidade ao mar.

### Considerações finais

A precipitação na ilha da Madeira é condicionada por diferentes factores [10], tais como altitude, relevo e orientação das vertentes, de tal forma que a quantidade de precipitação pode ser diferente à mesma altitude, mas em vertentes distintas, sendo em geral maior na vertente norte do que na sul. Os resultados obtidos em [7] com os extremos de precipitação analisados, para as estações localizadas tanto na vertente norte como na vertente sul, indiciam que deve ser dada atenção especial a diferentes factores, incluindo a altitude, a distância entre estações, a vertente onde estas estão localizadas, o relevo envolvente e a proximidade ao mar. No entanto, a dependência nos extremos de precipitação não parece ser apenas caracterizada por estes factores, dada a existência de estações associadas mas distantes umas das outras, em diferentes vertentes, com altitudes bastantes diferentes e com distâncias distintas em relação ao mar. Uma completa caracterização dos extremos de precipitação na ilha da Madeira, tanto à nível univariado como multivariado, é ainda a meta a atingir desta caminhada.

### Referências

- [1] Dietrich, D., de Haan, L., & Hüslér, J. (2002). Testing extreme value conditions. *Extremes*, 5(1):71-85.
- [2] Fragoso, M., Trigo, R.M., Pinto, J.G., Lopes, S., Lopes, A., Ulbrich, S. & Magro, C. (2012). The 20 February 2010 Madeira flash-floods: Synoptic analysis and extreme rainfall assessment. *Natural Hazards Earth System Sciences*, 12, 715-730.
- [3] Genest, C. & Favre, A. C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, 12(4):347-368.
- [4] Gomes, M.I., Fraga Alves, M. & Neves, C. (2013). *Análise de Valores Extremos: Uma Introdução*. Edições SPE.
- [5] Gouveia, D., Guerreiro Lopes, L. & Mendonça, S. (2014). Application of the theory of extremes to the study of precipitation in Madeira Island: Statistical choice of extreme domains of attraction. In

- Pacheco, A., Santos, R., Oliveira, M.R., Paulino, C.D. (eds.): *New Advances in Statistical Modeling and Applications*, Studies in Theoretical and Applied Statistics Societies, 187-195, Springer.
- [6] Gouveia-Reis, D., Guerreiro Lopes, L. & Mendonça, S. (2014). Aplicação do método dos excessos de nível a valores extremos de precipitação na Ilha da Madeira. In Pereira, I., Freitas, A., Scotto, M., Silva, M.E. & Paulino, C.D. (eds.): *A Ciência da Incerteza - Actas do XXI Congresso da Sociedade Portuguesa de Estatística*, 43-50. Edições SPE.
- [7] Gouveia-Reis, D., Guerreiro Lopes, L. & Mendonça, S. (2013). Um estudo da dependência entre extremos de precipitação na Ilha da Madeira. In Maia, M., Campos, P. & Silva, P.D. (eds.): *Estatística: Novos Desenvolvimentos e Inspirações - Actas do XX Congresso da Sociedade Portuguesa de Estatística*, 167-174. Edições SPE.
- [8] Koutsoyiannis, D. (2004). On the appropriateness of the Gumbel distribution for modelling extreme rainfall. In Brath, A., Montanari, A. & Toth, E. (eds.): *Hydrological Risk: recent advances in peak river flow modelling, prediction and real-time forecasting. Assessment of the impacts of land-use and climate changes.*, 303-319. Castrolibero.
- [9] Neves, C. & Fraga Alves, M.I. (2007). Semi-parametric approach to the Hasofer-Wang and Greenwood statistics in extremes. *Test*, 16(2):297-313.
- [10] Prada, S., Sequeira, M. M., Figueira, C. & da Silva, M. O. (2009). Fog precipitation and rainfall interception in the natural forests of Madeira Island Portugal. *Agricultural and Forest Meteorology*, 149(6-7):1179-1187.
- [11] Quintal, R. (1999). Aluviões da Madeira; séculos XIX e XX. *Territorium*, 6, 31-48.
- [12] Salvadori, G., De Michele, C., Kottegoda, N.T. & Rosso, R. (2007). *Extremes in Nature: An Approach Using Copulas*. Springer.



## Alguns desafios para o futuro da SPE

Kamil Feridun Turkman, *kfturkman@fc.ul.pt*

*CEAUL&FCUL-ULisboa*

A actual Sociedade Portuguesa de Estatística (SPE) que data de 1991, herdou da Sociedade Portuguesa de Estatística e Investigação Operacional (SPEIO), criada em Novembro de 1980, a sua história e os seus membros da comunidade Estatística. A SPEIO nasceu, naturalmente, após o esforço militante do Professor Tiago de Oliveira, durante os anos 60 e 70, para o reconhecimento da Estatística em Portugal como uma área científica independente da Matemática. Na década de 80 observou-se um grande crescimento e desenvolvimento das universidades em Portugal com a contratação de um corpo docente jovem e bem qualificado em todas as áreas científicas. Nessa altura o objectivo da SPEIO era fazer crescer o interesse pela Estatística e Investigação Operacional junto da comunidade científica. O grupo de jovens doutorados e outros ainda não doutorados que nessa altura se associou à SPEIO era um grupo dinâmico, entusiasta e decidido em fazer crescer a sociedade. Entretanto surgiu a Associação para o Desenvolvimento da Investigação Operacional (APDIO) e foi com naturalidade que se deu a separação, dentro da SPEIO, entre a Estatística e a Investigação Operacional, em certa medida decidida numa reunião em Tróia em 1990. A SPE foi sendo consolidada graças ao muito esforço e dedicação das várias direcções. Os congressos anuais foram sempre um momento importante, não só do ponto de vista científico como social, e quem desde o início neles participou sabe bem como foram importantes para a consolidação da investigação Estatística em Portugal. O número de participantes foi sempre aumentando e a qualidade dos trabalhos apresentados foi sempre melhorando chegando actualmente a níveis de qualidade internacional. Nesse aspecto a SPE cumpriu a sua missão e ambição inicial. Mas não chega. A SPE precisa de uma nova injeção de gente jovem e dinâmica que a coloque num patamar adequado aos tempos em que vivemos.

Durante a última década observou-se uma mudança drástica daquilo que é pedido à Ciência Estatística e o paradigma da Estatística tal como o conhecemos tem de sofrer necessariamente com as alterações que se impõem para dar resposta às solicitações. O grande volume dos dados e a sua natureza tipicamente não estruturada, forçou as metodologias de análise de dados a aproximarem-se cada vez mais às Ciências da Informática e com isso observou-se o nascimento de uma nova área designada por Ciência de Dados a qual se está a consolidar rapidamente. Ao reconhecer este facto, a Royal Statistical Society (RSS) tomou todos os passos para que a Estatística continue a ser parte indispensável desta nova área científica. Assim a SPE também tem de se preparar para esta realidade. Vejo este desafio como um dos mais importantes desafios que se coloca à SPE, assim como promover a importância do uso correcto da Estatística dentro da sociedade utilizando sempre todos os meios e oportunidades que surjam. Outro desafio muito importante é a pressão que a SPE deve colocar nos dirigentes académicos para o rejuvenescimento dos corpos docentes na área da Estatística. Embora o envelhecimento do corpo docente e o número muito limitado de lugares a concurso para docentes seja generalizado a todas as áreas científicas, observa-se que este problema é particularmente agudo nas áreas de Estatística. Urge alterar esta situação a breve trecho.

As várias direcções iniciam o seu mandato com muita vontade de fazer responder aos muitos desafios que encontram, alguns dos quais descritos acima. Infelizmente esta vontade rapidamente se tende a transformar numa frustração, pois grande parte destes desafios não pode ser concretizado sem o apoio activo dos membros da sociedade. Infelizmente existe a tendência mais ou menos generalizada dos sócios, de que uma vez votada uma nova direcção, a responsabilidade de levar cabo todos os objectivos

da SPE é da completa responsabilidade da direcção e a participação dos sócios restringe-se à crítica à actuação da direcção. Parece-me fundamental que os sócios da SPE assumam a necessidade de contribuir de forma construtiva para atingir os objectivos da mesma. É certamente mais relevante apresentar críticas construtivas e soluções do que apenas criticar. Neste aspecto, mais uma vez, gostaria de dar o exemplo da RSS, cujo prestígio como instituição acaba por prestigiar os seus membros, e vice versa, num contínuo de reforço positivo que a transformou decisivamente em anos recentes numa sociedade que não pode ser ignorada, tendo uma palavra a dizer sempre que se discutem temas relacionados com a Estatística. Quando a direcção solicita a membros para contribuir para uma questão específica, estes aceitam isso como um reconhecimento das suas competências e disponibilizam-se de imediato para a tarefa atribuída. Consequentemente, hoje em dia, a RSS é uma das sociedades mais intervenientes e respeitadas na comunidade. Assim a participação activa dos sócios é uma peça fulcral para que a SPE possa atingir os seus objectivos.

Em conclusão: A SPE precisa de uma nova injeção de gente jovem e dinâmica que a coloque num patamar adequado aos tempos em que vivemos. Precisa de mostrar que existe fora da Academia. Precisa de se virar para a Sociedade Civil e mostrar a sua importância. Precisa de se renovar e isso só é possível com a participação activa dos sócios. Acredito que a nova geração, liderada por jovens entusiásticos como a Jéssica, que partilhou comigo esta avaliação do passado e do futuro da estatística no âmbito das comemorações da SPE, terá a energia necessária, associada a uma compreensão do passado fundamental, para poder elevar a estatística ao lugar de destaque que merece na sociedade actual. Até porque esse lugar é dela, quer o reconheçam, quer não.



## ... quarenta anos de sociedade; quarenta e cinco de investigação...

Tiago A. Marques, *tiago.marques@st-andrews.ac.uk*

*University of St Andrews,  
Departamento de Biologia Animal, Faculdade de Ciências, Universidade de Lisboa,  
Centro de Estatística e Aplicações, Faculdade de Ciências, Universidade de Lisboa*

e

Lisete Sousa, *lmsousa@fc.ul.pt*

*DEIO, Faculdade de Ciências, Universidade de Lisboa,  
Centro de Estatística e Aplicações, Faculdade de Ciências, Universidade de Lisboa*

Quando recebemos o convite para, enquanto CEAUL (Centro de Estatística e Aplicações), apresentar uma contribuição para este número especial do boletim SPE, a primeira coisa em que pensámos foi como deveríamos abordar o tema? Quem são a Lisete e o Tiago? São dois cidadãos portugueses. Ambos pais. Ambos docentes da FCUL, ambos estatísticos, atualmente em diferentes departamentos, mas em tempos no mesmo - o DEIO. São também ambos membros da SPE e membros do CEAUL - única unidade de investigação em Portugal exclusivamente focada na Estatística. Por fim, são atualmente elementos da equipa que coordena o CEAUL. Então quem está a escrever este artigo? Esta é a visão desse emaranhado de identidades que, como tal, poderá a espaços ser difícil separar. No entanto, em última instância vamos fazer uma apreciação na perspetiva de quem está no momento ao leme de um barco que assume ter, para o bem ou para o mal, alguma responsabilidade naquele que possa ser o futuro da Estatística em Portugal no século XXI.

De onde viemos? Onde estamos? Para onde vamos? Foi este o mote e o desafio para o Boletim de Outono da SPE 2020! Decidimos pegar neste mote e usá-lo como espinha dorsal desta contribuição, fazendo uma apreciação do passado, uma descrição do presente e os objetivos para o futuro, certos de que, de uma forma ou de outra, quer queiramos ou não, quer consigamos aproveitar o impulso ou não, a Estatística será fundamental no século XXI e no futuro da humanidade.

### **De onde viemos?**

Era uma vez um estatístico português de seu nome Tiago de Oliveira, homem de enorme cultura e inteligência, que tinha sonhos... e concretizou vários. Um deles, foi a criação do Centro de Estatística e Aplicações da Universidade de Lisboa, em 1975, que coordenou até 1981. Muitas vezes referido como o *berço da investigação em Probabilidade e Estatística* em Portugal, é, ainda hoje, o maior grupo de Investigação em Estatística no nosso país, e teve um papel de relevo na construção e fortalecimento das bases para a Investigação em Estatística nacional. A promoção de encontros científicos em diversas áreas da Estatística, do ensino da Estatística e da Investigação, têm sido o motor não só do crescimento significativo da investigação da Estatística em Portugal, mas também da afirmação social da Estatística.

Concretizado um sonho, Tiago de Oliveira parte para outro sonho e preside à primeira Direção da SPE em 1980 até 1989 (na altura SPEIO). Sobre a nossa Sociedade, a SPE, remetemos o leitor para os excelentes artigos que compõem este boletim.

O CEAUL faz, este ano, 45 anos de existência! É quase meio século de existência durante o qual se assumiu como o maior grupo de investigação nacional dedicado à Estatística, quer nas suas vertentes mais fundamentais quer nas áreas de aplicações.

O sonho de um homem tornou-se o sonho de muitos. Perpetuamos esse sonho, enquanto estatísticos que fomos no passado, que somos no presente e seremos no futuro.

### **Onde estamos?**

Volvidos 45 anos da criação do CEAUL e 40 da SPE, a Estatística é uma ciência transversal a todas as áreas do saber. Nesse aspeto, talvez a par da informática, representam as duas disciplinas que são absolutamente ubíquas, não existe ciência moderna sem elas. A Estatística encontra-se em pouco por todo o mundo e em rápida mudança. Lá fora têm sido dados passos no sentido de harmonizar a relação entre a Estatística e as Ciências da Computação, porque, de facto, o mundo mudou muito em 40 anos. Quando a SPE nasceu, eventualmente o principal problema de um estatístico era não ter dados suficientes para conseguir justificar os seus resultados com passes mágicos – para alguns quase uma prova da existência divina tão notável e de quão transversais são as suas implicações – como o Teorema Limite Central. As contas faziam-se à mão ou numa calculadora de bolso. Hoje estamos na era do famoso *big data*, essa *buzz word* que depois ninguém sabe ao certo que é, em que temos à disposição as mais variadas ciências e tecnologia incrível cheia de sensores sofisticados que conseguem recolher informação a várias centenas de milhares de vezes por segundo. Ou seja, atualmente é quase inútil saber Estatística – pelo menos num contexto de aplicações concretas – se não conseguir primeiro processar essa avalanche de dados. Por isso, é preciso encontrar formas de juntar estas duas realidades, e desenvolver a área da Estatística Computacional. Na Estatística, em Portugal, tal como em outras áreas do saber, estamos provavelmente um pouco atrasados. Nem toda a internet do mundo, que nos posicionou numa aldeia global, consegue em 20 anos colmatar de forma quase instantânea o atraso histórico deste belo jardim à beira mar plantado, atraso esse que é estrutural e com variadas explicações históricas. Se é verdade que hoje em dia é tão fácil comunicar com um colega Japonês, Canadiano ou Chileno, é também verdade que as diferenças de sistemas e mentalidades ainda persistem, e, por isso, as mudanças acontecem a diferentes velocidades. Estamos num momento crucial, um daqueles momentos em que só mais tarde é possível perceber que afinal era um momento chave e que a direção tomada foi a acertada, ou a errada. O futuro o dirá. Infelizmente, nos momentos cruciais a avaliação da direção ideal a tomar é bastante mais complicada e, como tal, estas considerações não passam de conjeturas e opiniões.

À mulher de César não basta ser (honestamente), deve parecer, ou seja, não podemos fechar-nos na nossa concha, é preciso mostrar quem somos e ganhar visibilidade externa. Implementámos uma renovação da imagem e fizemos uma aposta em iniciativas e projetos que possam projetar a imagem do CEAUL. Há uma necessidade de balancear aquilo que é investigação fundamental, necessária para garantir que o Centro continua a ser uma referência nacional, com participações em atividades de impacto na sociedade civil com alguma projeção mediática. A contribuição para o esforço nacional na abordagem às questões da pandemia, onde a análise estatística é certamente fundamental, é disso um bom exemplo. Entre outras contribuições, desenvolvemos esforços no sentido de disponibilizar dados atualizados sobre a pandemia ([https://github.com/CEAUL/Dados\\_COVID-19\\_PT](https://github.com/CEAUL/Dados_COVID-19_PT)), utilizamos esses dados de forma simples e compreensível por um público alargado (<http://ceaul.org/covid19/evolucao/>), e participamos em projetos de investigação que visam ser mais um contributo não só para lidar com esta pandemia mas também para prevenir outras que possam surgir no futuro (e.g. o projeto COVIDTECT). Para complementar, fizemos um esforço para aparecer nos órgãos de comunicação social no âmbito desta temática, com contribuições que passaram pelo Jornal Público, Observador ou RTP3, por exemplo. Empenhamo-nos em angariar novos membros e lentamente, mas com persistência, tentamos também que alguns dos nossos membros assumam e desenvolvam novas iniciativas.

### **Para onde vamos?**

“Statistics is a data science for the 21st century” (Diggle, 2015) e nós queremos estar presentes nesse processo. O futuro é incerto e a Estatística sofre os males do seu próprio sucesso. Não existe desemprego

entre os jovens estatísticos e, de uma forma geral, aqueles que são bons são rapidamente absorvidos por um tecido empresarial que precisa, como de pão para a boca, de quem os possa ajudar a lidar com o volume avassalador de dados que são produzidos. Nós sofremos na pele, recentemente e em concreto, esse revés. Na sequência de um projeto muito interessante sobre segmentação de clientes com um gigante empresarial nacional, o bolsheiro do CEAUL que estava no projeto foi contratado pela empresa e assim, de certa forma, foi perdido para a Ciência. Ou seja, numa parceria com um privado que inclui a contratação de um jovem investigador, se as coisas correm bem, corremos o risco de perder massa crítica essencial para se trabalhar no sentido de construir um futuro promissor. Esta é uma desvantagem clara em relação a outras áreas, como por exemplo a Biologia, em que existem poucas saídas profissionais para além da academia e, como tal, os melhores são facilmente mantidos em Ciência. Não é, por isso, de estranhar que existam centros de investigação como o MARE, CE3C, CIBIO ou CESAM, colossos da primeira liga ao nível de um Porto, Sporting ou Benfica, e nós nos tenhamos que contentar em ser um Atlético que se tenta manter à tona de água nas distritais. Enquanto não houver desemprego na Estatística, e não se antevê que isso venha a acontecer tão cedo (felizmente!), será impossível atingir esse tipo de estrutura. Assim sendo, as universidades têm de ser estrategicamente inteligentes e criar mecanismos para garantir que estruturas como o CEAUL não desaparecem. Não deixa de ser paradigmático que uma universidade focada na investigação, como é a Universidade de Lisboa na figura em concreto da Faculdade de Ciências, precisa de um Centro como o CEAUL. Ignorá-lo é dar um estratégico tiro nos pés. A maior parte das pessoas pensa que qualquer um pode fazer Estatística, que sabendo implementar uma ANOVA e um teste t temos o suficiente para conseguir publicar na *Nature* os trabalhos que fazem. Essa é, naturalmente, uma visão redutora e inocente, porque é como pensar que “lá porque sei pregar pregos e usar um berbequim, posso construir uma casa”, e acalantar a esperança que essa casa vai ganhar um prémio Valmor. Não vai acontecer. A Estatística do século XXI não é sobre saber qual o procedimento a executar para rejeitar ou não uma hipótese nula. É sobre como podemos usar as estratégias de modelação estatística existentes para otimizar, para cada problema em questão, a análise a fazer e, assim, garantir inferências robustas. Uma boa introdução a esse tema é, por exemplo, o livro, recente mas a caminho de se tornar um clássico, de McElreath (2020): *Statistical Rethinking*. Tudo isto será ultrapassado se, de facto, estatísticos e investigadores das diferentes áreas trabalharem em conjunto. É esse o nosso objetivo dentro da Academia. Conseguir assumir e divulgar esse papel: estarmos disponíveis para quem, do outro lado, perceber que essa é a forma inteligente de dar um passo de gigante em direção ao futuro.

Uma possibilidade que temos vindo a explorar é procurar estabelecer laços entre o CEAUL e empresas com interesse em I&D e, assim, garantir que é possível desenvolver ciência também em ambiente empresarial. A noção de *dual appointments* (e.g. Cattaneo et al., 2018) não existe em Portugal, mas parece ser uma opção interessante a explorar. No concreto, pessoas da área da Estatística teriam obrigações formais noutras áreas, quer em diferentes áreas da academia, quer em empresas. Isso permitiria acesso a um conjunto de novas oportunidades e uma fertilização cruzada de ideias que poderia ser determinante. Esta ideia não é nova, e nada como citá-la verbatim de um dos grandes nomes da Estatística, o ex-presidente da Royal Statistical Society, Peter Diggle, no seu *presidential address* de 2015: “I would like to see every research-led university in the UK create a statistics institute. Each statistician on the university’s staff would have a dual appointment, to the institute and to an appropriate second discipline, be it mathematics, computer science or any one of the natural, biomedical or social sciences.”. Esta parece-nos uma ideia que deveria ser explorada. Uma instituição como a Faculdade de Ciências certamente poderia desenvolver esforços nesse sentido.

O CEAUL tem desenvolvido esforços para que a importância da Estatística seja reconhecida, e isso passa também por promover a Literacia Estatística na sociedade civil. Estamos a celebrar em 2020 o Dia Mundial da Estatística, a 3.<sup>a</sup> edição desta efeméride sob os auspícios das Nações Unidas. Desta vez sob o lema “Connecting the world with data we can trust”. Esta noção é fundamental e tão atual como nunca. Nunca como hoje assistimos a campanhas de desinformação globais, cada vez mais baseadas em aspetos quantitativos. Uma percentagem ou uma probabilidade conferem a uma qualquer afirmação nos media um grau de credibilidade, até de certeza, muitas vezes assente na incapacidade do cidadão comum de desconstruir o seu significado. É fundamental que os cidadãos tenham conhecimentos mínimos de Estatística para serem capazes de distinguir o trigo do joio. Atrevemo-nos a dizer que a nossa capacidade

de sobrevivência poderá estar dependente dessa capacidade, individual mas naturalmente societal, de compreender e interpretar dados. Os muitos dados que nos inundam a velocidades e quantidades nunca antes vistas. E cuja interpretação e utilização depende de saber interpretar a sua variabilidade, identificar os seus vieses desconstruir os seus confundimentos. Como a presidente da Royal Statistical Society disse claramente no seu presidential address de 2019 (Digby, 2019), the “ambition for the Society’s membership to go from hundreds to thousands has been fulfilled. The need now is for millions, even billions, of statistically literate people.”. A importância da Estatística está por isso assegurada nos anos que se seguem. Resta-nos esperar que consigamos estar à altura dos desafios que se avizinham

## Referências

- Ashby, D. (2019) Pigeonholes and mustard seeds: growing capacity to use data for society *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 182: 1121-1137
- Cattaneo, M., Horta, H. & Meoli, M. (2018). Dual appointments and research collaborations outside academia: evidence from the European academic population. *Studies in Higher Education* 44: 2066-2080
- Diggle, P. J. (2015). Statistics: a data science for the 21st century. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 178: 793-813
- McElreath, R. (2020). *Statistical rethinking: a Bayesian course with examples in R and Stan*. Chapman & Hall/CRC



## Três desafios para a Estatística: uma visão pessoal

Maria Eduarda Silva, *mesilva@fep.up.pt*

*Faculdade de Economia, Universidade do Porto & CIDMA*

Presidente da SPE (2015 – 2020)

A Sociedade Portuguesa de Estatística celebra, dentro de dias, os seus 40 anos. Quando a SPE foi constituída, eu não conhecia ainda o termo ou, se o conhecia, não lhe atribuía qualquer significado especial. O meu interesse pela Estatística nasceu por volta de 1984, quando frequentava o 4º ano da Licenciatura e o fascínio pelas Séries Temporais em 1988 pela mão de T. Subba Rao. Não vou aqui escrever sobre os desenvolvimentos da Estatística e, conseqüentemente da SPE, nos últimos 40 anos. Vou apenas apresentar uma visão muito pessoal dos desafios que a Estatística e conseqüentemente a SPE, enfrenta.

A Estatística, podendo ser definida como a ciência da incerteza, fornece a tecnologia que permite transformar dados em informação. A Estatística é fundamental no desenvolvimento e monitorização de políticas públicas e serviços públicos e na tomada de decisão quer no sector público quer no sector privado. Permite o desenvolvimento científico e tecnológico em todos os domínios do conhecimento com um papel interdisciplinar único. As estatísticas oficiais são produzidas de acordo com um código de conduta que assegura a sua independência, fiabilidade e credibilidade, sendo, portanto instrumentos indispensáveis de cidadania.

No entanto, verifica-se que os cidadãos têm uma perceção muito limitada da estatística, mostrando uma desconfiança crescente nas estatísticas. A evidência baseada em informação factual está-se a tornar menos relevante na discussão pública, dando origem à era da pós-verdade, expressão eleita palavra do ano 2016 (sim, já há 4 anos!) pelo dicionário Oxford. Na origem desta situação estão práticas menos correctas no uso de evidência estatística, nomeadamente no que respeita à apresentação, comunicação e interpretação da informação. Para além de estatísticas relevantes, fiáveis, acessíveis, precisamos de promover a confiabilidade da informação. É necessário dotar os cidadãos de todas as idades de: conhecimentos que lhes permitam reconhecer o valor dos dados e a importância de sermos capazes de os usar; capacidades que lhes permitam compreender o que é apresentado e distinguir informação válida de representações falaciosas de dados. A crise atual associada à pandemia COVID-19, veio reforçar a necessidade de dados fiáveis e confiáveis, de estatísticas de qualidade e de competências em “literacia de dados”.

De facto, esta crise que vivemos em 2020 sublinha a necessidade de apoiar a investigação em Estatística, que tem potencial para criar benefícios económicos e sociais. Põe-se a questão de qual é a estrutura onde esta investigação pode ser desenvolvida de forma eficaz e eficiente. Uma característica essencial da investigação em Estatística é a de se desenvolver no seio das várias ciências. A estatística matemática é importante no desenvolvimento de metodologias e requer que seja apoiada. Mas o foco da investigação em Estatística encontra-se na resolução de uma multitude de problemas, muitas vezes interdisciplinares em áreas como a saúde, o clima ou ciências sociais. A resolução destes problemas requer muitas vezes o desenvolvimento de metodologias no contexto das áreas de aplicação. O financiamento da investigação

A transformação digital está a decorrer a diferentes velocidades em diferentes países e setores e está a tornar a Ciência de Dados indispensável para empresas serviços e até administração pública. Os termos Machine Learning (ML) e Artificial Intelligence (IA) (Aprendizagem Máquina e Inteligência Artificial) tornam-se sinónimos de conversão de dados em valor num contexto de “big data” e de dados não estruturados. Reconhecidamente muitos destes dados, sobre os quais os algoritmos são desenvolvidos, são enviesados e levam a tomadas de decisão injustas. Não tendo dúvidas de que o papel da Estatística é fundamental na Ciência de Dados sinto que é necessário que ele seja reconhecido e esse é um grande desafio.



## Novas e velhas ideias para depois dos 40

Carlos Daniel Paulino, *daniel.paulino@tecnico.ulisboa.pt*

*CEAUL e Instituto Superior Técnico, Universidade de Lisboa*

Presidente da SPE (2012 – 2015)

Atingida pela SPE a idade adulta da razão como quadragenária, interessa em especial refletir sobre qual a projeção multilateral da Estatística no Portugal hodierno como domínio académico, científico e profissional. É o que se procura fazer sinteticamente neste texto como ponto de partida para o traçado de linhas de ação no sentido de afirmação e fortalecimento da comunidade estatística portuguesa.

A larga maioria do pessoal estatístico académico está dispersa, na sua atividade como docentes e investigadores, por diversos departamentos e unidades de investigação com a denominação de Matemática, de âmbito científico mais amplo ou mais restrito.

No quadro das instituições universitárias e científicas conhecem-se apenas um único (!) departamento de Estatística (em conjunto com a área da Investigação Operacional) e um único (!) centro de Estatística, ambos na Universidade de Lisboa (UL), Faculdade de Ciências (FC). Somente dois (!) cursos de 1º ciclo em Estatística Aplicada, lecionados na UL-FC e Universidade do Minho, Escola de Ciências, com cerca de 60 vagas no total, garantem uma formação básica consistente na área. Os correspondentes e bastos cursos de Matemática integram geralmente duas ou três disciplinas de Probabilidade e Estatística. Uma formação avançada de 2º e 3º ciclos em Estatística é obténivel de programas de mestrado e doutoramento lecionados em número um pouco maior de escolas e universidades.

Interesses egocêntricos manifestados aqui e ali, em universidades, escolas e unidades de ensino e investigação, têm dificultado ou mesmo obstado a uma autonomização filiacional e avaliativa de quem se insere na ampla área científica da Estatística (com as suas diversas subáreas e vertentes teóricas e aplicadas) e pretende integrar-se no meio que mais lhe apraz e considera racionalmente mais justo.

As entidades empregadoras de pessoal superior (INE, bancos, seguradoras, empresas de consultoria, sondagens e prospeção de mercado, etc.) para fins de análise estatística, seja ela elementar ou mais ou menos avançada, recorrem por vezes a quadros que não são qualificados como profissionais de Estatística (matemático(a)s, economistas, financistas, engenheiro(a)s, etc.), mesmo quando tal não se afigura justificável. Esta prática denota frequentemente sérias incompreensões sobre o que é o âmbito e nível da aprendizagem estatística dos formandos nesses cursos quando em contraponto com uma especialização em Estatística.

É certo que a criação da SPE há 40 anos e a sua atividade crescente ao longo do tempo desde então, num quadro de franco incremento da qualificação do pessoal universitário, contribuiu para se avançar numa lenta mas decidida afirmação da área da Estatística no panorama do ensino e da ciência num meio em clara expansão. Todavia, a situação corrente acima descrita comprova que ainda falta envidar recursos, velhos e novos, necessários para atingir um estado de reconhecimento social da Ciência Estatística ao nível do que se passa, por exemplo, em países que nos são cultural ou historicamente próximos, estejam eles geograficamente a pequena ou grande distância.

Para tal propósito, a SPE deve manter e diversificar as atividades relacionadas com os seus objetivos estatutários. Nesse sentido, começa-se por propor que a SPE estimule e apoie iniciativas que possam partir de unidades de investigação em Estatística ou de outrem e que envolvam pessoal de domínios com interesses afins, visando uma cooperação conjunta entre grupos de ambas as partes, com possível

realização de encontros específicos. Domínios como a Epidemiologia e a denominada Ciência de Dados constituem exemplos de grande interface com a Estatística, a qual pode propiciar um ambiente favorável à produção de resultados proveitosos para os dois lados (estatísticos versus epidemiologistas ou peritos de computação), se houver uma íntegra vontade bilateral de colaboração e a junção de distintos conhecimentos for bem conduzida na análise de grandes ou não tão grandes volumes de dados.

A bem-sucedida experiência de organizar bioestatísticos numa Secção de Biometria dentro da SPE, que se tem revelado bem atuante, deveria tentar alargar-se a outros setores. Grupos locais de docentes e investigadores atuando num dado domínio de aplicação probabilístico-estatística (cito, a título meramente ilustrativo, Econometria, Atuariado, Psicometria, Ambientometria, *inter alia*), que manifestem interesse em estreitar relações com outros grupos congéneres, podem e devem propor a sua inserção ou reorganização na SPE de acordo com a sua extensão, o que facilitará decerto a sua interação estatística e a organização dos seus encontros temáticos.

A projeção social da comunidade estatística impõe que a sua sociedade se organize internamente para intervir ativamente na denúncia fundamentada de dislates estatísticos contidos em textos saídos em órgãos de informação. A pandemia em curso, ao induzir a apreciações diversas sobre o processo de contágio por parte de opinantes setores, alguns até bem opiniáticos, tem constituído uma boa fonte de absurdos. Apraz-nos salientar que alguns destes já foram contestados por textos oportunos elaborados por alguns colegas da SPE e que estão disponíveis na página da SPE na secção sobre a COVID-19.

Apenas para exemplificação de alguns sérios equívocos mencionamos as seguintes noções que transparecem de alguns documentos publicados:

-- a curva epidémica diária da COVID-19 (ao longo de um dado período) ou o associado cronograma de mortes traduzem uma aproximação empírica da função densidade do nº diário de novos casos de contágio ou de mortes, respetivamente.

[Apetece questionar assim: como é que um cronograma (primitivo ou suavizado), pela evolução temporal desse nº de casos ou de mortes que graficamente retrata, pode alguma vez ser encarado como um diagrama de frequências associadas a uma amostra aleatória da correspondente variável de interesse?.....]

-- a sensibilidade de um teste de diagnóstico é a proporção de pessoas com infeção em um conjunto de testes com resultado positivo e a especificidade do mesmo teste é a proporção de indivíduos não infetados dentre os testados negativamente.

[É caso para perguntar: o que seriam então os valores preditivos positivo e negativo ou, equivalentemente, as probabilidades de um indivíduo ser falsamente positivo ou falsamente negativo?]

O nosso idioma nacional é uma das principais línguas europeias de comunicação universal e, mesmo que o não fosse, tem o direito inalienável de ser considerado uma língua inteira. Língua oral e escrita. Língua rústica e urbana. Língua pluridialetoal e pluricontinental. Língua multirracial e multinacional. Língua vulgar e culta. Língua literária e, como não podia deixar de ser, científica. Compete a nós, membros da comunidade estatística de fala portuguesa, praticar e desenvolver o português nas ciências estatísticas, por vezes em cooperação com profissionais dos domínios de aplicação (*e.g.*, Quimiometria, Astroestatística, Demografia e Jurimetria).

E porque vem a calhar, transcrevem-se estes versos do poeta quinhentista António Ferreira,

*Floresça, fale, cante, oiça-se e viva/ a portuguesa língua, e já onde for, /  
senhora vá de si, soberba e altiva! / Se até aqui esteve baixa e sem louvor, /  
culpa é dos que a mal exercitaram, / esquecimento nosso, e desamor!*

Um instrumento da maior relevância para tal desiderato foi criado há mais de 10 anos e sempre ampliado e melhorado desde então, tornando-se rapidamente num dos melhores glossários *anglo-et al.* de termos estatísticos – basta para o efeito consultar os vários glossários disponíveis na página do ISI (*International Statistical Institute*). Este documento de referência deve-se ao trabalho em equipa de uma comissão conjunta luso-brasileira constituída no seio da SPE e ABE e pode ser consultado através da página

*www.spestatistica.pt/glossario* e também na secção de *Endereços úteis* da página do *Ciberdúvidas da Língua Portuguesa*.

A consulta do glossário e a sua divulgação pública devem ser uma obrigação de todo o estatístico falante e escrevente do português. Além disso, através dos seus órgãos diretivos, a SPE deve reforçar no seu seio a aplicação de uma política de defesa do português como língua de ciência por parte dos seus falantes nativos, nomeadamente nos encontros científicos de âmbito nacional que promove e nas correspondentes atas que edita. Obviamente que nada disto interfere negativamente com a necessária publicação internacional de produção técnico-científica de todos nós, quase sempre na língua franca contemporânea.

Muitas das ações que têm vindo a ser realizadas pela SPE ao longo dos últimos anos devem naturalmente ser continuadas. A efetivação de algumas delas, bem como de algumas das medidas aqui propostas, poderá defrontar-se com possíveis dificuldades decorrentes da real situação presente e daquela que sairá da extinção da corrente pandemia, que hoje ninguém arrisca caracterizar. Se pensarmos nos efeitos que diversos confrangimentos por que temos passado poderão originar, talvez este alerta não seja despropositado de todo. Entretanto, mantenhamo-nos precavidos e resilientes tanto quanto possível!



# Comemorando os 40 anos da SPE

Carlos A. Braumann, *braumann@uevora.pt*

*Departamento de Matemática, Escola de Ciências e Tecnologia, Universidade de Évora*

Presidente da SPE (2006 – 2012)

Parabéns à Sociedade Portuguesa de Estatística (SPE) pelos seus 40 anos ao serviço da Estatística e da comunidade. De parabéns estão também o Fernando Rosado e a sua equipa, Tiago Marques e Lisete Sousa, pela realização deste número especial do Boletim. Já em 2005 tinha sido iniciativa do Fernando Rosado, então Presidente, o livro “Memorial da Sociedade Portuguesa de Estatística”, comemorativo dos 25 anos da SPE, e a ele devemos desde há muitos anos a constante e qualificada comemoração, no Boletim da SPE, dos avanços da Estatística e das atividades e contributos da comunidade estatística nacional e suas instituições.

Pediram-me um depoimento como antigo Presidente da SPE. É com especial satisfação que o presto. Na SPE sinto-me em casa como sócio fundador e participante ativo de todos os Congressos, feliz por a ter podido servir como Presidente da Direção e da Mesa da Assembleia Geral e profundamente grato pelo Prémio Carreira que me atribuiu em 2019 e me tornou seu sócio honorário. E recordo com saudade o meu Pai, Pedro Bruno Teodoro Braumann, um dos 11 subscritores da escritura de constituição da Sociedade em 28 de novembro de 1980.

Uma comemoração desta natureza é uma viagem no tempo, um registo dos progressos passados e um olhar sobre o presente e o futuro. No que se refere ao passado, pareceu-me importante que se registasse de forma especial o período que se seguiu ao “Memorial”. Os depoimentos dos representantes de diferentes instituições certamente focarão os contributos que elas deram para o progresso da Estatística, pelo que entendi caber-me, como antigo Presidente entre janeiro de 2006 e janeiro de 2012 (ou seja, nos dois triénios que se seguiram imediatamente à comemoração dos 25 anos), focar os aspetos de natureza societária, deixando aqui um registo do que de mais relevante a SPE, como força coletiva dos seus sócios, realizou nesse período.

Os órgãos sociais têm naturalmente um papel importante para além da gestão corrente, quer no lançamento de iniciativas quer na dinamização da Sociedade, pelo que quero deixar registada a composição dos órgãos eleitos, que foi idêntica nesses dois triénios:

- Mesa da Assembleia Geral: João Branco (Presidente), Maria do Carmo Miranda Guedes e Paulo Infante.
- Direção: Carlos Braumann (Presidente), Paula Brito (Vice-Presidente), Luísa Canto e Castro Loura (Secretária), Russell Alpizar-Jara (Secretário-Adjunto) e Manuela Souto Miranda (Tesoureira). Nota: Na descrição das atividades aqui feita, só se indicam coordenadores ou responsáveis quando tais funções não recaíram na Direção ou em elementos seus.
- Conselho Fiscal: Manuela Neves Figueiredo (Presidente), Paulo Rodrigues (Secretário) e Carlos Marcelo (Relator).

Foi um trabalho intenso em que me apraz constatar, com justiça mas também com amizade e enorme gratidão, o espírito de missão e a competência com que estes meus colegas serviram a SPE. E convém não esquecer o Secretariado da SPE, tão competente e empenhadamente assegurado pela D. Margarida Silva, a quem estou muito grato e a quem tanto a SPE deve.

Mas, obviamente, o papel fundamental na vitalidade e progresso da SPE cabe aos sócios, com a participação na vida da Sociedade, as iniciativas que lançam ou propõem e a aceitação e empenho no exercício de variadas funções e tarefas ao serviço da Sociedade. Bem hajam!

Registemos então as realizações mais relevantes da SPE nos dois triénios referidos.

## **Congressos e atividade editorial**

A atividade que mais se destaca na vida da SPE é o seu Congresso, a sua festa, a celebração coletiva dos progressos da Estatística, o convívio dos sócios, o espírito de corpo e de colaboração. Desde o I Congresso em 1993, após a antiga SPEIO ter passado a SPE, até 2013, os Congressos foram anuais e, apesar dessa frequência, fortemente participados e com elevado número de comunicações relevantes, muitas delas publicadas nas Atas após processo de revisão por pares, sinal inequívoco da pujança da comunidade estatística nacional. Estes dois triénios não foram exceção, com uma média superior a 200 participantes e 150 comunicações orais ou em poster por Congresso. Ao empenho das Comissões Executivas e Científicas, dos ilustres oradores convidados nacionais e estrangeiros, dos responsáveis dos minicursos, dos avaliadores dos artigos e dos editores das Atas, dos patrocinadores (com especial relevo para o INE – Instituto Nacional de Estatística) e, especialmente, dos sócios e das Comissões Organizadoras, muito se deve este sucesso.

Os Congressos no período 2006-2011 foram (indica-se entre parêntesis a instituição organizadora e o Presidente da Comissão Organizadora):

- XIV na Covilhã em 2006 (Univ. da Beira Interior, Maria Eugénia Ferrão)
- XV em Lisboa em 2007 (ISCTE, Manuela Magalhães Hill)
- XVI em Vila Real em 2008 (Univ. de Trás-os-Montes e Alto Douro, Irene Oliveira)
- XVII em Sesimbra em 2009 (FCT – Univ. Nova de Lisboa, Manuel Esquível)
- XVIII em S. Pedro do Sul em 2010 (Univ. de Coimbra e Inst. Polit. de Viseu, Paulo Oliveira)
- XIX na Nazaré em 2011 (Inst. Superior Técnico e Inst. Polit. de Leiria, António Pacheco).

De registar o intercâmbio com a SIS (Sociedade Italiana de Estatística) que permitiu que, nos últimos anos deste período, cada sociedade organizasse uma sessão no congresso da outra.

A atividade editorial da SPE neste período esteve ligada aos congressos, cada um com o livro de resumos e o manual do minicurso, impressos gratuitamente pelo INE, e com as atas como repositório do progresso da ciência estatística e suas aplicações.

As Atas dos três primeiros congressos deste período continuaram a ser publicadas pela SPE no ano seguinte ao Congresso. Neste período foram publicadas as atas dos congressos de 2005 (“Ciência Estatística”, ISBN: 978-972-8890-09-4), 2006 (“Estatística. Ciência Interdisciplinar”, ISBN: 978-972-8890-12-4), 2007 (“Estatística da Teoria à Prática”, ISBN: 978-972-8890-17-9, impressas excepcional e gratuitamente pelo INE) e 2008 (“Estatística: Arte de Explicar o Acaso”, ISBN: 978-972-8890-20-9). Nos três últimos congressos deste período, fruto do acordo de 4 sociedades estatísticas (SEIO de Espanha, SFdS de França, SIS de Itália e SPE) com a Springer e da decisão da Assembleia Geral da SPE, elas vieram a ser publicados mais tarde em livro e em língua inglesa (série “Studies in Theoretical and Applied Statistics. Selected Papers of the Statistical Societies” da Springer), ganhando visibilidade internacional: “Advances in Regression, Survival Analysis, Extreme Values, Markov Processes and Other Statistical Applications” (XVII Cong., ISBN: 978-3-642-34903-4), “Recent Developments in Modeling and Applications in Statistics” (XVIII Cong., ISBN: 978-3-642-32418-5) e “New Advances in Statistical Modeling and Applications” (XIX Cong., ISBN: 978-3-319-05322-6).

Os manuais de minicurso publicados neste período foram, como sempre, notáveis textos em língua portuguesa com influência no ensino superior e na investigação: “Outliers em Dados Estatísticos” (Fernando Rosado, 2006), “Introdução aos Métodos Estatísticos Robustos” (Ana M. Pires e João A. Branco, 2007), “Análise de Dados Espaciais” (M. Lucília Carvalho e Isabel C. Natário, 2008), “Análise de Sobrevivência” (Cristina Rocha e Ana Luísa Papoila, 2009), “Uma Introdução à Estimção Não-Paramétrica da Densidade” (Carlos Tenreiro, 2010) e “Análise de Dados Longitudinais” (M. Salomé Cabral e M. Helena Gonçalves, 2011).

Alguns dos manuais de minicurso, apesar das tiragens generosas, tiveram elevada procura extra-congresso e esgotaram, pelo que se acordou com o INE a impressão de reedições dos manuais esgotados que para tal tivessem o consentimento dos autores. Saíram neste período as segundas edições revistas e aumentadas dos manuais “Séries Temporais, Modelações Lineares e Não Lineares” (Esmeralda Gonçalves e Nazaré Mendes Lopes, 2008) e “Controlo Estatístico da Qualidade” (M. Ivette Gomes, Fernanda Figueiredo e M. Isabel Barão, 2010).

## **Participação no International Statistical Institute (ISI) e na organização da sua 56ª Sessão**

A SPE, como membro do ISI e da organização satélite International Association for Statistical Education (IASE), participou (através do presidente ou seu delegado) na reunião das sociedades estatísticas

nacionais realizadas nos congressos do ISI (as reuniões magnas mundiais da estatística) de 2007, 2009 e 2011, e colaborou ativamente com estas instituições, com destaque para o glossário estatístico e o apoio ao International Statistical Literacy Program (incluindo a comparticipação nas despesas de deslocação à sessão de 2009 em Durban da comitiva portuguesa de jovens e professores).

Permito-me realçar um grande sucesso, o congresso ISI 2007 (56ª sessão do ISI), realizado em Lisboa em agosto de 2007, com cerca de 3000 participantes (mais de 500 portugueses). Foi coorganizado pelo INE, SPE e CLAD e presidido por Paulo Gomes, tendo Ivette Gomes presidido à Comissão de Programa Local em representação da SPE e integrado a Comissão Coordenadora do Programa (internacional). O Presidente e vários outros sócios da SPE integraram a Comissão Organizadora Local e a Comissão de Programa Local. Houve o cuidado de articular este evento com o XV Congresso da SPE em termos de calendário e possibilidade de inscrição conjunta.

Apoiámos ainda a reunião científica COMPSTAT 2008 (Porto, 24-29 agosto) da IASC – International Association for Statistical Computing, sociedade satélite do ISI, que teve grande sucesso. A Paula Brito (futura Presidente do IASC em 2013-15) presidiu às Comissões Organizadora e Científica de Programa.

### **Apoio à comunidade científica, seminários e Prémio SPE**

Em termos de apoio à comunidade científica, continuaram os Seminários SPE com Novos Doutores (com apresentação de resultados das suas teses de doutoramento) e iniciaram-se, com o objetivo de incentivar o intercâmbio entre instituições com o patrocínio da SPE, os Seminários SPE, realizados em diversas instituições com oradores de outras instituições. Em 2007 arrancaram os temáticos Encontros de Estatística SPE/CIM (Centro Internacional de Matemática) com a colaboração do INE, por vezes envolvendo também outras entidades, que se realizaram regularmente em vários locais versando diversos temas de grande atualidade.

A SPE apoiou diversos eventos e outras iniciativas científicas em termos de divulgação e em termos de patrocínio institucional e financeiro, tendo neste caso os sócios da SPE reduções na taxa de inscrição.

Outro incentivo, o Prémio SPE, foi naturalmente continuado neste período, tendo o trabalho premiado pelo júri sido apresentado pelo(s) autor(es) no congresso do respetivo ano, designadamente por Sandra Dias (2006), Lígia Carla Henriques Rodrigues (2007), Nuno Sepúlveda (2008), Miguel de Carvalho e Paulo Canas Rodrigues (2009) e Gonçalo Jacinto (2010). Em 2011, o júri reconheceu o valor dos resultados mas não atribuiu o prémio por considerar que os trabalhos careciam de revisão aprofundada.

### **Comissões Especializadas e jSPE**

Logo em 2006 foram criadas duas importantes Comissões Especializadas:

- a de Educação (Luísa Canto e Castro Loura, Maria Eugénia Graça Martins) que organizou os Prémios Estatístico Júnior e a revisão de manuais escolares (serviço aberto às editoras, a que aderiu a Porto Editora), coordenou as atividades ligadas ao ensino básico e secundário (como, por exemplo, a colaboração com o Ministério da Educação na revisão e implementação de programas ou a colaboração com a SPM – Sociedade Portuguesa de Matemática na sua Comissão do Ensino Básico e Secundário e na certificação de manuais escolares) e, a partir de 2010, ficou responsável pelo Clube Júnior;
- a de Nomenclatura Estatística (Daniel Paulino, que presidiu, Dinis Pestana, João Branco e, temporariamente, Lucília Carvalho e Alfredo Egidio dos Reis), que, tendo por base o glossário em língua inglesa do ISI e tendo conseguido obter a colaboração da ABE – Associação Brasileira de Estatística, elaborou o Glossário estatístico em língua portuguesa, o qual ficou disponível nas páginas web da SPE e do ISI e foi sendo atualizado e expandido por esta Comissão.

Em 2010, a Assembleia Geral aprovou a criação e o regulamento da jSPE - Secção de Jovens Estatísticos na SPE, cuja Comissão Instaladora foi constituída por Paulo Canas Rodrigues e Miguel de Carvalho e iniciou a atividade com uma sessão no XVIII Congresso Anual da SPE.

### **Funcionamento**

Várias melhorias no funcionamento foram sendo introduzidas, como a regularização da situação fiscal, a nova opção de pagamento das quotas por débito em conta ou a melhoria contínua e posterior reestruturação da página web. Esta passou a ser mais informativa sobre a SPE e eventos estatísticos, integrou o Glossário e, graças à introdução da versão em inglês que tínhamos acordado com o ISI, ganhou acessibilidade internacional. A comunicação com os sócios e a divulgação de eventos recorreu quer à página web quer ao e-mail direto. Fizeram-se reedições atualizadas do folheto de divulgação da

SPE, em português e inglês, quer para apoio e divulgação do ISI2007, quer posteriormente. No âmbito da revisão da Classificação Nacional de Profissões, deu-se parecer sobre a profissão de estatístico.

A SPE esteve representada em diversos encontros científicos, em reuniões de discussão do Plano de Ação para a Matemática e em encontros e sessões destinadas a homenagear individualidades da área da Estatística, de que se destaca a comemoração em 2008 do 80º aniversário da Professora Doutora Maria de Fátima Fontes de Sousa (sócia nº 1 da SPE).

Em termos financeiros, as receitas provinham essencialmente dos resultados líquidos dos Congressos (a maior fatia), das quotas dos sócios e de patrocínios. Apesar de o Fundo de Apoio à Comunidade Científica da FCT (Fundação para a Ciência e a Tecnologia) só nos ter patrocinado no início deste período por a FCT ter interrompido o financiamento de sociedades com boa saúde financeira, os patrocínios (muitos em espécie) de várias outras entidades ao congresso e a outras atividades permitiram que, para além de manter a atividade regular, se pudesse investir numa considerável expansão das atividades da SPE. O que foi feito sem prejudicar a já favorável situação líquida da SPE (53041,77 Euros no início de 2006), antes a melhorando substancialmente (no início de 2012 era de 100258,72 Euros em termos efetivos e 119572,32 Euros em termos contabilísticos, considerando neste último valor o subsídio antecipado de 19313,60 Euros do Ciência Viva para a Explorística e a Radical Estatística).

### **Boletim**

Um instrumento central na vida da sociedade é o Boletim da SPE. Sob a responsabilidade do Fernando Rosado, este ganhou uma nova dimensão, sendo os seus dois números anuais (Primavera e Outono) disponibilizados aos sócios em versão papel e a todos na página web.

### **Primeiro Dia Mundial da Estatística (20 de outubro de 2010)**

A SPE associou-se às comemorações deste Dia, em colaboração com o Centro de Estatística e Aplicações da Universidade de Lisboa, com uma mostra, coordenada pela Maria Antónia Amaral Turkman, da atividade na área da Estatística dos Centros e Grupos de Investigação portugueses.

### **Comissão Nacional de Matemática**

Graças a diligências que iniciámos em 2008, a SPE participou como observadora em reuniões da CNM – Comissão Nacional de Matemática que, posteriormente, integrou por unanimidade a SPE como membro efetivo, qualidade em que já participou na reunião de 26 de novembro de 2010.

### **Cooperação institucional, criação da FENStatS**

Conforme se pode constatar ao longo da leitura deste texto, o funcionamento da SPE muito beneficiou das parcerias e colaborações institucionais em atividades conjuntas que foi estabelecendo desde a sua criação, algumas delas estabelecidas ou reforçadas neste período.

Destacam-se os sócios institucionais e parceiros privilegiados (INE e Banco de Portugal) e as instituições de que a SPE é sócia (ISI, IASE e CIM). Regista-se a colaboração com outras sociedades em diversas atividades: SPM, com intensa colaboração, Royal Statistical Society, CLAD – Associação Portuguesa de Classificação e Análise de Dados, com a qual acordámos a reciprocidade nas taxas de inscrição nos eventos científicos, ABE, com idêntico acordo e colaboração no Glossário, e Associação de Professores de Matemática. A divulgação ao público beneficiou da colaboração com o portal Cienciapt.NET.

A aposta na juventude teve o patrocínio e colaboração da Porto Editora e do Ciência Viva.

Uma aposta que se fez neste período foi a cooperação com sociedades europeias congêneres. Começou-se com a SIS – Sociedade Italiana de Estatística e a organização por cada sociedade de uma sessão no congresso da outra. Estendeu-se depois ao acordo com a Springer por parte de 4 sociedades (SEIO – Sociedad de Estadística e Investigación Operativa, SFdS – Société Française de Statistique, SIS e SPE) para a publicação por aquela casa editora de Selected papers de congressos (com um corpo de “referees” partilhado) e de monografias científicas. Finalmente, estas 4 sociedades e mais outras 6 sociedades fundadoras (CSS – Česká Statistická Společnost, DStatG – Deutsche Statistische Gesellschaft, ÖSG – Österreichische Statistische Gesellschaft, RSS – Royal Statistical Society, SDS – Statistično Društvo Slovenije, VvS+OR – Vereniging voor Statistiek en Operations Research, da Holanda) constituíram em 18 de Dezembro de 2011 a FENStatS – Federation of European National Statistical Societies.

### **Juventude**

Reconhecendo a necessidade de assegurar o futuro da Estatística como instrumento fundamental de apoio ao desenvolvimento económico e social, de suporte à atividade científica e de literacia estatística

ao serviço de uma participação informada dos cidadãos na vida da sociedade, deu-se neste período um grande impulso à sensibilização da juventude para a importância da Estatística.

Resolveu-se por isso dar continuidade aos Prémios Estatístico Júnior, que em boa hora a anterior Direção promoveu no âmbito das comemorações em 2005 do 25º aniversário da SPE. Estes prémios passaram a funcionar anualmente a partir de 2007, graças ao patrocínio da Porto Editora. Os trabalhos premiados pelo júri nas várias categorias (3º ciclo do ensino básico, ensino secundário e, a partir de 2010, também cursos de educação e formação de adultos) foram apresentados, e os respetivos prémios entregues, durante o Congresso Anual da SPE.

Mais tarde, em 2010, foi criado, sob a égide da Comissão Especializada de Educação, o Clube Júnior, com local próprio na página web. Nesse âmbito, foram criados dois grupos de trabalho para as seguintes atividades a realizar junto dos jovens:

- **Radical Estatística** (Paulo Rodrigues, que coordenou, Dulce Gomes, Bruno de Sousa e, nas fases iniciais, Paulo Infante), destinada a realizar várias atividades lúdicas junto dos jovens a culminar num campo de férias estatístico, tendo o primeiro campo decorrido já em 2012 com muito sucesso.
- **Explorística** (Pedro Campos, que coordenou, inicialmente Andreia Hall, Bruno de Sousa, Emília Oliveira, Jorge Nuno Silva e Maria Eugénia Graça Martins, mais tarde Conceição Rocha e Paulo Infante), destinada à criação de uma exposição itinerante de Estatística para o Centro Ciência Viva. Esta, concluída no mandato da Direção seguinte, suscitou grande interesse das escolas e incentivou outras sociedades estatísticas estrangeiras na promoção de iniciativas análogas.

Ambas as iniciativas mereceram apoio financeiro considerável do programa Ciência Viva (parcial no caso da Radical Estatística e integral para a fase de lançamento no caso da Explorística), tendo os contratos sido assinados em 2011.

O grande sucesso destas iniciativas foi justamente premiado a nível internacional pelo International Statistical Literacy Program (ISLP) e IASE. A “Radical Estatística” conquistou o 2º lugar do Best Cooperative Project Award in Statistical Literacy 2013, prémio entregue na 59ª Sessão do ISI em Hong Kong. A “Explorística” foi vencedora do Best Cooperative Project Award in Statistical Literacy 2015, prémio entregue na 60ª Sessão do ISI no Rio de Janeiro.

Procurei dar uma visão dos expressivos progressos da SPE no período de janeiro de 2006 a janeiro de 2012, que, graças à atividade e empenho dos sócios, visaram dar pleno cumprimento aos seus objetivos e realizar as atividades preconizadas no artº 1º dos Estatutos, construindo sobre o brilhante legado que receberam. Com avanços e recuos próprios de uma instituição plural, a SPE está de parabéns neste seu 40º aniversário pois tem continuado sempre a bem servir a Estatística e suas aplicações, a comunidade estatística e a literacia estatística. De parabéns estão também os seus sócios e os órgãos sociais pelo trabalho desenvolvido.

A recente pandemia veio mais uma vez mostrar a importância da Estatística e da literacia estatística. Em boa hora a SPE criou a secção “COVID-19” na sua página web. Nela se apresentam contributos interessantes.

Infelizmente, a iliteracia estatística afeta ainda alguns académicos. Os realmente perigosos são os que ignoram a sua ignorância estatística e até a arvoram quando escrevem na comunicação social, podendo induzir em erro os leitores ofuscados pela sua pretensa sapiência. Um desses escritos, entre outros dislates, “aplica” os conceitos estatísticos de assimetria e achatamento/curtose a uma curva epidémica (curva dos novos casos detetados em função do tempo) como se esta fosse um polígono de frequências de uma amostra, usando-os para comparar países quanto ao maior ou menor êxito no achatamento (conceito epidemiológico e não estatístico) da curva epidémica. Na secção “COVID-19” podem encontrar a denúncia polida e correção destes dislates num texto da Presidente da SPE e num artigo de Lisete Sousa, Tiago Marques e Antónia Turkman, exemplos de cidadania e apoio à literacia estatística.

A Estatística e a SPE continuam, como se pode ver, bem vivas, apesar de algumas limitações temporárias na forma de funcionamento devido aos atuais condicionamentos.

Da importância da Estatística no futuro ninguém tem dúvidas de que ela só poderá crescer e até sabemos que novas áreas de investigação e aplicação estão neste momento em grande desenvolvimento. Com o tradicional dinamismo da comunidade estatística portuguesa e o seu continuado apoio à atividade da SPE, esta continuará certamente de boa saúde a servir cada vez melhor os seus nobres objetivos.

Bom trabalho a todos nós!

## A SPE como caminho

Fernando Rosado, *fernando.rosado@fc.ul.pt*

*DEIO, Faculdade de Ciências, Universidade de Lisboa*

Presidente da SPE (2000 – 2006)

*De onde viemos?*

A SPE é um projeto científico, assim posso testemunhar, que tive o enorme prazer de ver nascer e crescer.

A SPE é uma organização à qual Portugal deve a introdução e o incremento da moderna Ciência Estatística.

A SPE ofereceu-me uma possibilidade, um complemento, no caminho da minha realização científica, pessoal e profissional.

O nascimento da SPE permitiu criar uma escola estatística e uma profícua geração de eminentes estatísticos portugueses de renome nacional e internacional.

A SPE surgiu quando a Estatística em Portugal pouco mais era do que uma secção da Matemática Aplicada da Faculdade de Ciências da Universidade de Lisboa. O estatístico não existia. A Estatística era “apenas” uma espécie de matemática aplicada.

Acompanhei a SPE desde o seu nascimento e tive o enorme prazer de a poder servir durante toda a minha vida científica e profissional; muito em especial quando “respondemos presente” perante o desafio de encabeçar a equipa que durante seis anos conduziu os destinos da SPE.

Do livro Memorial da Sociedade Portuguesa de Estatística, publicado em 2005, transcrevo e, com emoção, recordo:

“É também (o) 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 Carvalho, a Manuela Neves, o Manuel João Morais (no primeiro mandato), a Zilda Mendes e a Paula Brito (no segundo mandato) trabalhei formando uma equipa que soube dar resposta a todos os projetos 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”.

Passados “alguns anos”, com emoção, renovo esta gratidão!

Foi uma honra poder ter dado continuidade ao projeto sonhado pelo Professor Tiago de Oliveira – Presidente da SPE (1981 – 1989); e por todos os seniores académicos que há 40 anos elevaram o pilar que hoje festejamos e do qual muito nos orgulhamos.

### *Onde estamos?*

Passados “os inte” e “os inta”, a SPE entra agora nos “enta”. Simbolicamente, são “idades” onde é interessante e também com proveito fazer uma retrospectiva, “um olhar para a vida e para os resultados” do projeto científico envolvido pela e da responsabilidade da Sociedade Portuguesa de Estatística. Nos “inte” cresceu, nos “inta” consolidou e, com a ajuda de ambas, na próxima podemos formular votos de que mais cresça e mais se consolide, com todos os novos projetos assentes nas bases que todos ajudámos a criar.

Sou o sócio 24 que, convicta e apaixonadamente, se tem dedicado a continuar e a ajudar a nascer o *Boletim SPE* cujos pais são as dezenas de autores e autoras que, generosamente, o geram semestralmente.

Neste testemunho em boletim aniversário, cumpre-me agradecer a todos “meus colegas Presidentes” a generosidade e a dedicação com que desempenharam os seus cargos; cada um em diferentes épocas científicas, académicas e até administrativas – todas com os seus desafios específicos.

À Presidente Maria Eduarda Silva, que agora termina essas funções, manifesto os meus agradecimentos pela dedicação à causa SPE e, decerto a sua “paixão SPE”, vai fazer continuar com a sua intervenção de excelência, em prol da Ciência e dos Estatísticos portugueses, com que sempre nos presenteou.

Neste momento de “aniversário simbólico”, ao relembrar outros tempos, é justo recordar e agradecer à “Margarida” – nome de referência e grande motor administrativo na SPE – que durante bastantes anos foi uma dedicada funcionária e a quem carinhosamente quando, para concretizar a referência, dizíamos: a “Margarida da SPE”. E, a todas que continuaram essas funções fundamentais de secretariado até ao presente; à Dra Isabel Maurício que, na nova época desafiadora a tantos níveis, desempenha e dá continuidade a este projeto.

### *Para onde vamos?*

Decerto guiados pelas novas gerações de estatísticos portugueses que devemos apoiar e que já temos o prazer de ver brotar nos sítios certos, a SPE continuará com força aumentada.

Os escritos, as ideias, as opiniões e as sugestões (algumas bastante práticas e concretas) que a experiência dos mais seniores permite transmitir – em todos os fóruns e em particular neste Boletim – podem ser pontos de partida para um aprofundamento e consolidação, um guia para as novas gerações. Agora, consolidada, já nos quarenta; agora que já não é obrigada a ser jovem, “espetacular” para tudo e para todos, o que se espera é que a SPE seja ela própria.

Agora, tem o desafio de transmitir o seu legado e fomentar nas novas gerações o interesse pelo uso do seu acervo histórico; para que a continuidade científica permita uma maior afirmação dos estatísticos e da Estatística portuguesa tanto a nível nacional como internacional.

Assim também, se sentirá realizado o trabalho que tantos dedicados cientistas sonharam para a Estatística em Portugal.

Um caminho!



## Nos 40 anos da SPE: uma breve reflexão

João A. Branco, *jbranco@math.tecnico.ulisboa.pt*

*CEMAT, Instituto Superior Técnico, Universidade de Lisboa*

Presidente da SPE (1994 – 2000)

Embora com humildade, foi com muito prazer que aceitei o convite do editor do Boletim da SPE, Professor Fernando Rosado, para contribuir, com um pequeno texto, para este número especial do Boletim da SPE, cujo tema central se integra nas celebrações do 40º aniversário da SPE.

A primeira parte do texto é constituída por algumas notas, que apesar de dispersas e incompletas, pretendem convencer o leitor que a SPE nasceu quando tinha que nascer, isto é, quando estavam conseguidas as condições para que isso acontecesse.

A segunda parte conta aspectos da minha vivência com a SPE.

### *A seed can start a garden*

Para mim, a semente mais produtiva foi colocada na FCL (FCUL) e o seu semeador, e perspicaz visionário dos eventuais frutos, foi sem dúvida o Professor José Tiago de Oliveira. Desses frutos, os mais apreciados estão prontos a colher sucessivamente em: i) 1964 - licenciaturas em matemática aplicada (que incluíam matérias de probabilidade e estatística), ii) 1969 - Centro de Matemáticas Aplicadas, que em 1975 passou a Centro de Estatística e Aplicações, o actual CEAUL, iii) 1980 – Sociedade Portuguesa de Estatística e Investigação Operacional (SPEIO), donde em 1991 emergiu a SPE, iv) 1981/1982 – mestrado e primeiras licenciaturas em probabilidades e estatística.

Não vivi esta sequência fantástica que ocorre num tempo especial de conturbação política a caminho do final de um sombrio e demasiado longo regime. Mas nada é definitivo e nesse final começam a surgir sinais de abertura, talvez forçada, como é o caso da profunda reforma do ensino, de Veiga Simão, que incluiu a expansão do ensino superior acompanhada de uma maior capacidade e vontade de atribuição de bolsas de estudo para universitários, alguns dos quais acabaram por fazer os seus doutoramentos nas áreas da probabilidade e estatística. E logo a seguir a surpresa do 25 de Abril que nos devolveu a esperança que faltava! Fico muitas vezes a imaginar, com curiosidade, a agitação, a capacidade, o esforço, as alegrias e certamente as vicissitudes deste grandioso empreendimento que teve origem na FCUL.

Mas como o sol não é o universo, não podemos esquecer todas as outras instituições e estudiosos solitários que antes deram a sua contribuição para que a estatística em Portugal viesse a conseguir a relevância e o dinamismo que levou à constituição da SPE. A longa marcha que conta o andamento desse lento progresso está documentada em vários escritos visando a história da estatística em Portugal, podendo ser vistos no Boletim da SPE e em outras publicações que não vou citar por limitações de espaço. E não é tarde nem é cedo para eu acrescentar à lista do que é conhecido mais uma instituição e o nome de um homem, só pelo facto de, ao acompanhar essa marcha, não ter encontrado as suas referências. Trata-se da Estação Agronómica Nacional (EAN) e do seu departamento de Estatística Experimental. Foi aqui, na valiosa biblioteca da EAN, que encontrei o primeiro número da revista *Biometrika*. Havia ali interesse no estudo e aplicação de métodos estatísticos, em particular de métodos de classificação automática e de delineamento experimental. Lembro-me de ter conhecimento de contactos de John Gower com a EAN e também de James Rolhf que a partir de 1971 visitou Portugal, cerca de uma dúzia de vezes, para fazer seminários na EAN e cursos de verão na Fundação Calouste Gulbenkian, em Oeiras, sobre taxonomia numérica, biometria e

computação gráfica, com base no software NTSYSpc da sua própria autoria. O homem é Gustavo de Castro, matemático, interessado no cálculo das probabilidades e estatística matemática, deixou obra escrita sobre estes temas, além de muitos outros. É dele a frase "... a Estatística Matemática é jovem e em crescimento e, assim ninguém pode dizer em poucas palavras o que é e, talvez ninguém saiba...". Era conhecido por manter seminários permanentes no LNEC e CTT sobre aquelas matérias contribuindo, à sua maneira, para a divulgação das probabilidades e estatística no país. São estas, e outras actividades antecedentes à existência da SPE, umas que já foram ditas e outras que ainda estão por dizer, que paulatinamente foram construindo a oportunidade da sua criação.

### **Episódios de fim de século**

Durante 16 anos pertenci aos órgãos sociais da SPE, mas foi no período de 1994 a 2000 que me envolvi de veras com toda a actividade da SPE. Senti que foi um período marcado por amorosismo mas com muita dedicação e gosto em servir a SPE. Não havia experiência mas havia sonhos partilhados, especialmente com os colegas da Direcção: como atrair mais sócios, como melhorar a comunicação com os sócios, como conseguir o reconhecimento da importância da estatística e da SPE na sociedade em geral, como lutar pelo ensino da estatística ao nível pré-universitário, como dinamizar a organização de seminários e diversificar o seu local de realização, como criar secções de especialização dentro da SPE, como ligar a SPE a outras organizações com interesse no estudo e utilização da estatística, como tornar o Boletim atractivo, como elevar a qualidade dos nossos congressos e como resolver várias outras preocupações. Muito ficou por fazer mas deram-se alguns passos. O número de sócios manteve-se a crescer sucessivamente ao longo do período e foi conseguida a consolidação da organização do congresso anual. O congresso tornou-se num ícone da SPE e paralelamente à concentração exigida pelos trabalhos a atmosfera de cordialidade e a dinâmica de juventude que se instalava fazia nascer um ambiente de verdadeira festa. Na altura eu conhecia quase todos os sócios mas não me tinha apercebido da real composição da sua população, até que um dia, na sequência de uma notícia sobre um congresso do ISI, fui levado a olhar melhor para o conjunto, e não só.

Por altura do 50º WSC do ISI em Beijing, em Agosto de 1995, foi criado o *ISI Committee on Women in Statistics* com o objectivo de fazer subir a extremamente baixa proporção de membros femininos do ISI. Falando de mulheres lembrei-me de 3, cujos nomes encontrei na literatura enquanto estudava em Oxford, e que me impressionaram pelas suas vidas e pelo seu envolvimento com a estatística: Florence Nightingale (the lady with the lamp) e F.N.David (Florence Nightingale David) e também Gertrude Cox, a do clássico Cochran and Cox ( *Experimental Designs*, 1950) que era, naquela tempo, a bíblia nesta área. E, claro, lembrei-me das tantas colegas que apareciam no congresso da SPE. Aproveitei então a deixa para informar (Branco, 1997) que em Portugal o panorama mostrava o contrário das preocupações do ISI, com o sector feminino em franca maioria.

### **Notas Finais impregnadas de agradecimento**

Os sonhos de fim de século antes referidos, cujo alcance afinal é o engrandecimento da SPE e da estatística através da SPE, são sonhos que certamente outras ou todas as outras Direcções da SPE tiveram. E a prova disso é verificar que aquilo que não se conseguiu concretizar durante o período indicado foi sendo feito por outras Direcções ou está a ser feito pela actual Direcção. Cito apenas dois casos para justificar a minha convicção.

Uma das concretizações que atingiu um grau de verdadeiro profissionalismo é o Boletim da SPE. O Professor Fernando Rosado tem sido, desde 2006 (15 anos!), o editor permanente do Boletim, publicação que tem vindo a tornar-se cada vez mais valiosa e mais apreciada pelo interesse dos tópicos e experiência dos autores contribuintes, propostos e seleccionados pelo editor, pelo rigor da edição de cada número do Boletim e pela pontualidade na sua distribuição. É um trabalho que o Fernando faz com muito fervor e muita paixão e eu aproveito para deixar aqui escrito o meu reconhecimento, a minha admiração e o meu agradecimento por tão importante serviço prestado à nossa Sociedade. E para se perceber o que está em causa basta perguntar o que seria a SPE sem os seus dois mais firmes suportes, isto é, sem o Congresso e sem o Boletim.

Como já foi referido, uma ambição legítima, para a divulgação da estatística e para a compreensão da sua importância como ferramenta na resolução dos mais variados problemas da vida real, é a de promover a visibilidade da SPE no seio da sociedade em geral. Não houve, no passado, condições para pôr em marcha esse promissor projecto. Mas as condições mudaram e foi com grande satisfação que recentemente vi a SPE mostrar-se na imprensa portuguesa, dando sinais da sua existência e da sua capacidade de intervenção em problemas que lhe digam respeito. A pandemia que bruscamente assolou o mundo inteiro atingiu-nos também com um tsunami (paperdemia) de artigos e comentários em revistas científicas, meios de comunicação social e redes sociais. A onda chegou a Portugal e foram muitos, matemáticos, estatísticos, epidemiologistas e outros, os que se entusiasmaram a divulgar as suas interpretações e previsões sobre a “curva e o achatamento da curva”, um tema que se tornou viral nos vários meios de comunicação social. A importância destas contribuições não é pequena pois podem influenciar comportamentos e certamente servir de base para decisões sobre orientações a adoptar com vista a enfrentar a doença. Infelizmente os resultados de várias dessas contribuições são obtidos usando a estatística de forma incorrecta, ou porque os métodos são inadequados ou porque os dados não são confiáveis. Fez bem a Presidente da SPE, Professora Maria Eduarda Silva, ao alertar publicamente para o perigo de más práticas da utilização da estatística e fizeram bem os especialistas, sócios da SPE, que secundaram o aviso reagindo com as correcções técnicas dos erros encontrados num artigo publicado num certo jornal. A nota da Presidente e o artigo dos sócios da SPE, relativos a estas intervenções, estão disponíveis em <https://www.spestatistica.pt/noticias/noticia/covid-19-e-estatistica>.

Numa altura em que a estatística está sob ataque e a própria palavra sofre um apagamento por muitos que usam os seus métodos de forma encapotada, devemos, todos, incluindo a SPE e seus associados, estar atentos e lutar contra o abuso da má utilização da estatística.

Nesta data, em que se celebram 40 anos de existência da SPE, damos os parabéns ao mentor que promoveu a criação da SPE e a todos que têm contribuído para o bom desempenho da sua missão. E parabéns a você, SPE!

#### **Referências:**

Branco, J. A., Women in statistics: a counterexample? *Royal Statistical Society News*, 24(9), 7, 1997.



## Sociedade Portuguesa de Estatística – Passado, Presente e Futuro?

M. Ivette Gomes, *migomes@fc.ul.pt*

*CEAUL e DEIO, Faculdade de Ciências, Universidade de Lisboa, e  
Academia de Ciências de Lisboa*

Presidente da SPE (1990 – 1994)

Quando fui eleita *Presidente da Sociedade Portuguesa de Estatística e Investigação Operacional* (SPEIO), em 1990, a *Associação Portuguesa de Investigação Operacional* (APDIO), tinha ganho uma dinâmica e relevância que evidenciavam o papel residual da *Investigação Operacional* na SPEIO. Por outro lado a experiência de 10 anos mostrava que importava alterar e simplificar os estatutos de 1980.

Foi esta a tarefa prioritária no meu primeiro ano de mandato. Vários sócios se envolveram muito entusiasticamente no processo, e de entre eles destaco a Maria Antónia Amaral Turkman, que tem continuado muito activa no desenvolvimento da *Sociedade*, e o saudoso António Raúl Simões Neto, principal autor da reformulação dos estatutos, que transformaram a SPEIO em SPE, a *Sociedade Portuguesa de Estatística* que actualmente conhecemos.

Sem descurar o reconhecimento internacional da SPE, nomeadamente pelo *International Statistical Institute*, procurei também impulsionar o seu papel catalizador na investigação científica na área da *Estatística* em Portugal. Para tal tive a colaboração de uma equipa – Dinis Pestana, Luísa Canto e Castro, Isabel Fraga Alves, Ana Pires, entre outros colegas – que imaginou os *Congressos Anuais da Sociedade Portuguesa de Estatística*, ocorrendo o primeiro no Vimeiro em 1991.

A história da SPE está esquematicamente traçada na sua página WEB, mas a meu ver de forma algo incompleta. Conviria haver uma digitalização exaustiva dos *Boletins da SPE*, dos *Livros de Resumos e Atas dos Congressos*, dos *Mini-Cursos*, que se tornaram um evento marcante pré-congresso, e uma referência clara aos cientistas internacionais, e mesmo nacionais, que nos honraram, sendo oradores convidados nesses Congressos.

A nível geral, parece-me que um dos problemas essenciais com que nos confrontamos continua a ser o de renovar o stock da espécie “estatístico”, a fim de esta se manter diferenciada das outras espécies, e não se extinguir. Tem pois de se manter o prestígio da carreira, desenvolvendo investigação de qualidade. E é também essencial intervir em áreas chave: vida pública em geral, política local e global, interface com áreas que já capturaram a atenção do grande público. A instituição de mais prémios, que incentivem a qualidade da investigação *Estatística* e a participação de “jovens cientistas” em reuniões internacionais de prestígio, é também desejável. Deverá ainda ser ambição da SPE estar na génese de organismo independente que possa pronunciar-se profissionalmente sobre questões de utilidade pública, tais como a qualidade de sondagens, fidedignidade de estatísticas oficiais, autenticidade do controlo da qualidade, entre outras.

Gostaria também de ver incentivadas intervenções da SPE numa comunidade mais alargada de estatísticos que se exprimem, também, em *Português*. As crises, que são também oportunidades, talvez permitam que o formato dos congressos passe a comportar, para além de *Congressos Anuais* e da desejável componente presencial, uma abertura à *Participação à Distância*, incentivando laços com Angola, Brasil, Cabo Verde, Macau, Moçambique e São Tomé e Príncipe.

## • Artigos Científicos

- Caeiro, F., Gomes M.I., Henriques-Rodrigues, L. and Cabral, I. (2020). Minimum-variance reduced-bias estimation of the extreme value index: A theoretical and empirical study. *Computational and Mathematical Methods*, 2: e1101  
DOI: 10.1002/cmm4.1101
- Gomes, M.I., Caeiro, F., Figueiredo, F., Henriques-Rodrigues, L. and Pestana, D. (2020). Reduced-Bias and Partially Reduceb-Bias Mean-of-order-p Value-at-Risk Estimation: a Monte-Carlo Comparison and an Application. *J. Statist. Comput. and Simul.* 90:10, 1735-1752.  
DOI: 10.1080/00949655.2020.1746787
- Gomes, M.I., Caeiro, F., Figueiredo, F., Henriques-Rodrigues, L. and Pestana, D. (2020). Corrected Hill versus partially reduced-bias value-at-risk estimation. *Communications in Statistics: Simulation and Computation* 49:4, 867-875.  
<https://dx.doi.org/10.1080/03610918.2018.1489053>
- Papança, F. (2019). A Estatística na Organização Curricular da Escola de Guerra Durante e Após a 1ª Guerra Mundial. Encontro Anual da Investigação e Desenvolvimento em Ciências Militares (2018) no Instituto Universitário Militar. Poster do projeto na Academia Militar; na Exposição do Cinamil - Dia da Academia Militar 2019.  
Texto publicado na Plataforma EasyChair .Preprint Nº 1942.
- Penalva, H., Gomes, M.I., Caeiro, F. and Neves, M. (2020). A couple of non reduced bias generalized means in extreme value theory: an asymptotic comparison. *Revstat—Statist. J.* 18:3, 281-298.  
[https://www.ine.pt/revstat/pdf/REVSTAT\\_v18-n3-3.pdf](https://www.ine.pt/revstat/pdf/REVSTAT_v18-n3-3.pdf)
- Penalva, H., Gomes, M.I., Caeiro, F. and Neves, M. (2020). Lehmer's mean-of-order-p extreme value index estimation: a simulation study and applications. *J. Applied Statistics* 47:13-15 (Advances in Computational Data Analysis), 2825-2845.  
<https://doi.org/10.1080/02664763.2019.1694871>

## • Capítulos de Livros

**Título:** *Papel do controlo estatístico da qualidade em indústria e serviços*

**Autoras:** Gomes, M.I. and Figueiredo, F.

**Livro:** *Memórias da Academia das Ciências de Lisboa, Classe de Ciências.*

Ano: 2020. Tomo XLVI, Vol. 1, pp. 63-71.

Comunicação apresentada à Classe de Ciências na Sessão de 5 de Janeiro de 2017.

**Título:** *Generalização do estimador de Hill, baseada na média de Lehmer: Resultados adicionais*

**Autores:** Cabral, I., Caeiro, F. and Gomes, M.I.

**Livro:** *Atas do XXIII Congresso da Sociedade Portuguesa de Estatística*

**Editores:** M.F. Salgueiro, P. Vicente, T. Calapez, C. Marques e M.E. Silva

Ano: 2020. Edições SPE, pp.129-144.

ISBN: 978-972-8890-46-9

**Título:** *Extremal index estimation – application to financial data*

**Autora:** Maria Cristina Souto Miranda

**Livro:** *Handbook of Research on Accounting and Financial Studies*

**Editores:** Luís Farinha, Ana Baltazar Cruz e João Renato Sebastião

Ano: 2020. Edição: IGI Global. Pag. 96-130

ISBN13: 9781799821366; |DOI: 10.4018/978-1-7998-2136-6

## • Teses de Mestrado

**Título:** Modelo de previsão de vendas para otimização do reaprovisionamento do produto às lojas

**Autora:** Ana Julieta Oliveira Morais, *leta.morais@gmail.com*

**Orientadora:** Adelaide Freitas

**Título:** Novas Abordagens na Detecção de *Outliers* em Dados Composicionais

**Autora:** Marta Luísa Santos Maltez, *martamaltez@ua.pt*

**Orientadora:** Adelaide Freitas

## • Livros

**Título:** *Bioestatística. Síntese teórica e Exercícios*

**Autores:** Adelaide Freitas e João Pedro Cruz

Ano: 2020. Edições Sílabo.

ISBN: 978-989-561-039-6

**Título:** *Probabilidades e Estatística. Teoria, Exemplos & Exercícios*

**Autor:** Manuel Cabral Morais

Ano: 2020. Editora: IST Press – Instituto Superior Técnico.

ISBN 978-989-8481-78-8





# Índice

Editorial .....	1
Mensagem da Presidente .....	2
Notícias .....	3
<i>Enigmística</i> .....	7
<b><i>Episódios na História da Estatística</i></b> .....	8
<b><i>SPE e a Comunidade</i></b>	
Contributos do Instituto Nacional de Estatística .....	10
<b><i>40 anos SPE: De onde viemos? Onde estamos? Para onde vamos?</i></b>	
Probabilidades e Estatística no Departamento de Matemática da Universidade de Coimbra: uma perspetiva <i>Esmeralda Gonçalves e Paulo Eduardo Oliveira</i> .....	14
Alguns tópicos de investigação em estatística realizada no CIDMA – uma breve discussão <i>Pedro Macedo, Ana Helena Tavares e Marco Costa</i> .....	16
Pseudovalorers-p na construção do conhecimento científico <i>Maria de Fátima Brilhante</i> .....	24
Um olhar sobre o desenvolvimento de modelos de sobrevivência para acontecimentos recorrentes <i>Ivo Sousa-Ferreira, Ana Maria Abreu e Cristina Rocha</i> .....	27
O papel da Estatística na Universidade do Porto <i>Conceição Nunes Rocha e Isabel Silva</i> .....	36
Ser SPE – A vantagem do ónus da Boa Estatística <i>Jessica Silva Lomba</i> .....	41
UM departamento com ESTATÍSTICA/Departamento de Matemática/Universidade do Minho (UM) <i>Ana Paula Amorim e outros</i> .....	44
SPE – 40 anos de “braço dado” com a Estatística <i>Célia Nunes</i> .....	48
Estatística na Universidade de Évora <i>Russell Alpizar-Jara</i> .....	51
Uma caminhada por valores extremos de precipitação na Ilha da Madeira <i>Délia Gouveia Reis</i> .....	55
Alguns desafios para o futuro da SPE <i>Kamil Feridun Turkman</i> .....	58
... quarenta anos de sociedade; quarenta e cinco de investigação... <i>Tiago A. Marques e Lisete Sousa</i> .....	60
<b><i>Testemunhos de Presidentes da SPE</i></b>	
Três desafios para a Estatística: uma visão pessoal <i>Maria Eduarda Silva</i> .....	64
Novas e velhas ideias para depois dos 40 <i>Carlos Daniel Paulino</i> .....	66
Comemorando os 40 anos da SPE <i>Carlos A. Braumann</i> .....	69
A SPE como caminho <i>Fernando Rosado</i> .....	74
Nos 40 anos da SPE: uma breve reflexão <i>João A. Branco</i> .....	76
Sociedade Portuguesa de Estatística – Passado, Presente e Futuro? <i>M. Ivette Gomes</i> .....	79
<b><i>Ciência Estatística</i></b>	
<i>Artigos Científicos</i> .....	80
<i>Capítulos de Livros</i> .....	80
<i>Teses de Mestrado</i> .....	57
<i>Livros</i> .....	58