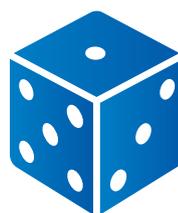




Boletim



SPE

Sociedade Portuguesa
de Estatística

Publicação semestral

primavera de 2022



Liderança Estatística

Liderança estatística no sector privado	
Breve testemunho	Jorge Félix 18
Liderança estatística - uma visão muito pessoal	Manuela Neves 22
A Estatística na investigação científica: Submissão, supremacia ou parceria inseparável?	Maria Eduarda Silva 25
A Excelência como liderança	Carlos A. Braumann 26
Liderança Estatística	Fernando Rosado 29
Liderança da Estatística ou Importância da Estatística?	João A. Branco 34
Ser ou não ser um líder (estatístico)?	Ivette 37
How to be an effective statistician! – With a focus on your leadership skills	Nuno Sepúlveda 39
Mensagem do Presidente da SPE	Alexander Schacht 49
	Miguel de Carvalho 54

Editorial 2
Mensagem do Presidente 3
Notícias 4
Enigmística 9
SPE e a Comunidade 10
Liderança Estatística 18
Ciência Estatística 56
Edições SPE 58

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

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

ISSN: 1646-5903

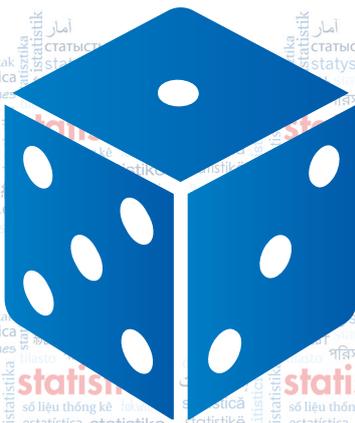
Depósito Legal: 249102/06

Tiragem: Edição digital

Execução Gráfica e Impressão: Gráfica Sobreirens

Editor: Fernando Rosado, fernando.rosado@fc.ul.pt

Sociedade Portuguesa de Estatística desde 1980



SPE

Sociedade Portuguesa de Estatística

<https://www.spestatistica.pt>

Sociedade Portuguesa de Estatística desde 1980

Junta-te à



SPE

Sociedade Portuguesa
de Estatística

*“Se já és sócio da SPE, incentiva os teus colegas,
colaboradores e alunos a juntarem-se à SPE”*

A SPE

- Oferece descontos em congressos e outros eventos organizados pela SPE.
- Oferece distinções e reconhecimento através dos seus prémios.
- Oferece oportunidades para ampliares a tua rede de contactos através da comunidade SPE.
- Oferece aos sócios acesso a um sistema de acreditação internacional.
- Valoriza sócios, comunidade e profissão apostando na educação, formação e inovação.
- Defende a profissão e molda o seu futuro.

Junta-te à SPE:

<https://www.spestatistica.pt/socios/admissao-formulario>

Quota anual

- Regular: 30€
- Estudante: 15€
- Promotor*: 0€
- Recém-licenciado†: grátis!

† Até um ano após terminar a licenciatura.

* Um “sócio promotor” angaria 2 novos sócios por ano.

Editorial

... admirável “mundo novo”?!...

1. Na sequência da pandemia, (já superada?!), a terminologia linguística propõe a pós-pandemia. No decorrer do tempo, durante os últimos dois anos, assistimos ao uso e consolidação mais ou menos generalizados de novas imagens literárias como “novos tempos” ou “nada será como dantes”, numa referência ao que virá. Mas, em tempos de mudanças, (não) terá sido sempre assim? E, rapidamente, chegamos à incerteza e daí à Estatística é um (pequeno) passo.

E, na raia do fundamentalismo, sobre o «mundo novo» que o tempo de nós aproxima, os mais puristas em alguns ramos do saber, nimiamente, preferirão então encaminhar os seus estudos nos domínios de maior fatalismo e prenunciar o caos. Aqui, talvez o que de mais perto pode afetar o ser humano está no mundo da ecologia ou da sua relação com as, assim chamadas, alterações climáticas. Na teoria do caos, foi introduzido por Edward Lorenz, há cerca de cinquenta anos, o denominado efeito borboleta – uma pequena coisa, um *outlier*, uma força dos menores. Este efeito, segundo o autor, é fortemente dependente das condições iniciais e de tal modo que o bater de asas de uma borboleta pode provocar um tufão com a imprevisibilidade sobre onde e como.

Esta breve introdução vem a propósito de uma conferência, na Academia das Ciências de Lisboa, proferida pelos colegas Jorge Buescu e Filipe Duarte Santos, da Faculdade de Ciências da Universidade de Lisboa. À volta de “Ordem no Caos” e de “Descarbonização de Economias” proferiram duas excelentes lições de síntese, às quais tive a oportunidade de assistir e onde, para além da matéria científica, o debate produziu conclusões generalistas sobre a novidade e a sua influência “do novo” na condição humana.

«Admirável Mundo Novo» e «1984» são duas das obras mais populares do século XX; escritas respetivamente, por Aldous Huxley e George Orwell. Para além de uma certa indignação relativamente ao *status quo* elas também representam uma advertência sobre um “futuro” – um mundo novo – onde tudo e todos serão manipulados, controlados e sem privacidade. De algum modo, tudo isto pode ser interpretado como um tufão e também para este poderemos aqui introduzir algum “efeito borboleta”? Conforme a perspetiva onde são inseridas aquelas obras, são classificadas como parábolas sobre a desumanização, com subjugação total do ser humano; ou como distópicas, muito mais do que utópicas e no “mau caminho”. A sua publicação teve, no entanto, uma originalidade. George Orwell solicitou aos editores de «1984» que, antes da publicação, enviassem uma cópia para Aldous Huxley. Esta ação “uniu as duas obras” pois fica em suspenso a decisão de Orwell perante uma análise desfavorável. Surge então uma carta que Huxley enviou para Orwell onde para além da sua apreciação, em síntese e em conclusão, afirmava sentir “que o pesadelo de «1984» está destinado a ser modulado ao pesadelo de um mundo que se assemelha ao que eu imaginei em «Admirável Mundo Novo». Essa mudança irá ocorrer pela necessidade do governo de um modelo mais eficiente”. E conclui-se sobre a necessidade da eficiência.

Numa síntese daquela conferência, Buescu já invocou alguma “modelação no caos” assim chamado de caos determinista. Tudo dependente de condições iniciais e das respetivas variáveis com observações ou dados *outlier*, “ironicamente” concluía que é sempre nos países do terceiro mundo que a borboleta bate as asas para gerar o tufão do primeiro mundo.

Assim, determinismo, modelação, eficiência são palavras-chaves que, para além de outras, se enquadram na Liderança Estatística, o Tema Central deste Boletim primavera.

Estas breves reflexões sobre “dois assuntos” – Liderança e Estatística – que pouco têm em comum mas que podem ser alinhadas num primeiro ponto de partida para uma longa maratona que ambos vão percorrer – como é condição inicial para os mais diversos temas científicos, ao longo do tempo. No início é muito mais difícil pois a consolidação além de exigência para fazer avançar também requer predisposição e esta, por sua vez, admite disponibilidade (muito condicionada pelas diversas agendas pessoais).

Este Boletim primavera 22, na modéstia da sua condição de *Newsletter*, dá um primeiro passo. Que os próximos sejam concretizados em futuro breve! Os leitores têm a palavra.

2. Com o Boletim SPE primavera de 2008 foi iniciada a publicação de uma nova secção a que foi dado o título “**SPE e a Comunidade**”. Com a criação desse espaço editorial, passámos a incluir acrescida matéria científica que podemos situar num objetivo vasto de divulgação da Estatística não só entre os sócios mas também destes para toda a comunidade. No formato editorial do Boletim, que já incluía um Tema Central, as novas páginas acrescentaram diversidade e ampliaram as intenções de divulgação e abertura da SPE.

SPE e a Comunidade apresenta nesta edição uma nova vertente, original, muito prática e que desejamos continuar. Fica o apelo e o convite editorial para a contribuição dos leitores. Iniciamos com a participação de um jovem investigador premiado SPE (Cf. Boletim SPE primavera 21, p. 46).

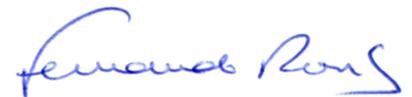
Graças à generosidade da colaboração do nosso colega Ivo Sousa Ferreira – Prémio SPE 2020, este Boletim, naquela secção, regista um “manual de utilização” que podemos intitular “Tudo sobre o R”. É uma contribuição da SPE, um novo serviço, que é de enorme utilidade e com ela certamente toda a comunidade beneficia. Fica desde já o desafio aos leitores e especialistas que, em partilha de informação, desejem continuar esta vertente prática de divulgação de “instrumentos estatísticos”. É um tópico de muito relevo e para o qual o Boletim SPE aguarda por novas contribuições.

Obrigado Ivo!

3. O **Tema Central** deste Boletim primavera pode enquadrar-se “num mundo novo”. É um enorme desafio juntar a liderança e a estatística para formar um título que, tenho a certeza, é ele próprio líder. A partir de uma proposta inicial e após breve troca de ideias, com o colega Nuno Sepúlveda, o coeditor do Tema Central deste número do Boletim, conseguimos juntar as vontades e as generosidades que as agendas dos autores permitiram para consolidar este novo desafio. Neste Boletim temos a felicidade de contar com excelentes colaborações do mundo empresarial que merecem toda a nossa gratidão.

O desafio é grande, também porque a condição de assunto pioneiro dificulta a motivação ou a vontade de abordagem do assunto, para além da disponibilidade de agenda que em conjunto “decidem” as tarefas individuais. Estas, algumas vezes, não deixam finalizar o agendado. E “o prometido fica em suspenso”. Mas a não dúvida sobre a vantagem é razão suficiente para a partilha de “reflexões sobre” e isto para além do interesse na motivação, quiçá formação, dos modernos estatísticos para uma nova sociedade, ou melhor, para uma sociedade de novos tempos. O Boletim SPE aguarda pelos “reflexos das reflexões” ora apresentadas, com a certeza do bem que daí advirá.

Obrigado Nuno!



O Tema Central do próximo Boletim será

Prémios SPE

Mensagem do Presidente

Nota Editorial:

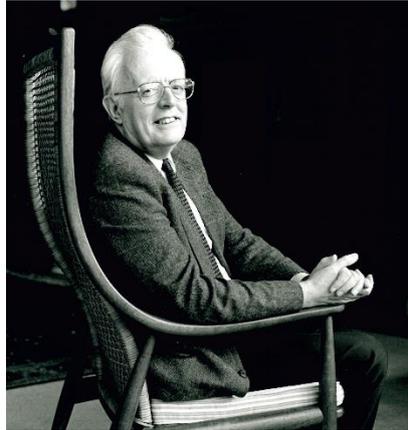
A habitual Mensagem do Presidente está incluída no **Tema Central** deste Boletim SPE primavera 22, na página 54.

Notícias

• Faleceu Sir David Cox (1924-2022) - Sócio Honorário da SPE

Foi um dos pioneiros da moderna Ciência Estatística. Do seu legado salientam-se: regressão logística, o modelo de riscos proporcionais que ele desenvolveu em 1972 e o processo de Cox, também chamado processo de Poisson duplo.

Nuffield College da Universidade de Oxford foi a sua escola onde desempenhou intensa atividade até “ao fim”, como é referido na página oficial (onde foi publicada a foto que juntamos).



O Professor Cox, é nome de uma referência mundial da Estatística que a história já registou. e em todos os campos onde a sua vida decorreu, desde a Academia ao ISI – o Instituto Internacional de Estatística de que foi Presidente e do qual a SPE é uma organização filiada. Também para os estatísticos portugueses, o Professor Cox foi, aos mais diversos níveis, um mestre que para além da formação teórica também ajudou a consolidar a moderna conceção da Estatística.

O Professor Cox, numa das suas últimas deslocações a Portugal, participou no XII Congresso da Sociedade Portuguesa de Estatística realizado em 2004, na cidade de Évora onde foi Conferencista Convidado.

Desse Congresso SPE, que decorreu em ano jubilar, no vigésimo quinto ano da sua história, o texto da Conferência proferida por Sir David Cox foi publicado nas respetivas Atas – *Estatística Jubilar*.



No XII Congresso Anual SPE – Évora, 2004

Para grande honra da Sociedade Portuguesa de Estatística o Professor David Cox aceitou ser Membro Honorário.

O Professor Sir David Cox foi eleito Sócio Honorário da SPE, por unanimidade, na Assembleia Geral realizada após o Congresso.

Fernando Rosado

• In memoriam – Ross Leadbetter (1931-2022)

Foi triste receber a notícia da morte de Ross Leadbetter em 26 de Fevereiro de 2022, com 90 anos.

Ross Leadbetter tinha uma alma enorme e era um dos gigantes da área de *Valores Extremos em Sucessões Dependentes*, tendo sido entrevistado (<https://link.springer.com/article/10.1007/s10687-015-0225-1>) sobre a sua carreira, na prestigiada revista *Extremes*.

O livro que escreveu em 1983, em co-autoria com Georg Lindgren e Holger Rootzén, intitulado '*Extremes and Related Properties of Random Sequences and Processes*' é um marco na área, com cerca de 5000 citações.

Não posso ainda deixar de referir que Ross Leadbetter foi um dos convidados no 'Nato Advanced Statistical Institute', intitulado '*Statistical Extremes and Applications*', que decorreu no Vimeiro, no Verão de 1983, um evento actualmente reconhecido como um marco na afirmação desta área e uma rampa de lançamento da '*Escola de Extremos em Portugal*' a que Ross Leadbetter, com o seu sentido de humor, intitulava de '*Portuguese Gang*'.



A primeira doutorada em *Extremos* em Portugal, a Teresa Alpuim, foi orientada por mim, na FCUL, mas, durante a sua estadia na Universidade de North Carolina, teve o sábio aconselhamento de Ross Leadbetter. Gostaria ainda de referir que Ross Leadbetter honrou a ULisboa ao aceitar, a pedido do DEIO, e durante a Presidência de Kamil Feridun Turkman, a distinção de *Doutor Honoris Causa* da ULisboa, <http://extremes-lisbon2013.weebly.com/doctorate-hc.html>, pois sem dúvida, e tal como escrevi em 2013 na Info-Ciências Digital, <https://ciencias.ulisboa.pt/pt/noticia/26-02-2013/extremistas-da-minha-terra>, '... quando a universidade honra investigadores desta importância está também a honrar-se'. E a EVT 2013 – *Extremes in Vimeiro Today*, <http://ceaul.org/evt2013-extremes-in-vimeiro-today/>, organizada pelas minhas colegas e grandes amigas, Antónia Amaral Turkman, Isabel Fraga Alves e Manuela Neves, para recordar os trinta anos do encontro do Vimeiro em 1983, é outro dos grandes marcos, onde o Ross esteve presente, como se pode ver na fotografia em anexo. O Ross faz-nos sem dúvida grande falta, mas a obra dele estará sempre presente!



Ivette Gomes

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 2021, a REVSTAT – Statisticsl Jornal lançou o Volume 19 – Números 1 a 4, com os artigos listados abaixo (<https://revstat.ine.pt/index.php/REVSTAT/issue/archive>).

Volume 19, Issue 1:

- “Skewed probit regression – identifiability, contraction and reformulation” by Janet van Niekerk and Håvard Rue.
- “Jackknife empirical likelihood inference for the variance residual life function” by Vali Zardasht.
- “Finite mixtures of multivariate skew Laplace distributions” by Fatma Zehra Dogru, Y. Murat Bulut and Olcay Arslan.
- “Choice of smoothing parameter for kernel type ridge estimators in semiparametric regression models” by Ersin Yilmaz, Bahadir Yuzbasi and Dursun Aydin.
- “A stochastic study for a generalized logistic model” by Rafael Luís and Sandra Mendonça.
- “Confidence interval for quantile ratio of the Dagum distribution” by Alina Jedrzejczak, Dorota Pekasiewicz and Wojciech Zielinski.
- “Minimum area confidence region for Weibull distribution based on records” by Junmei Zhou, Fen Jiang and Jin Zhang.
- “Parameter estimation based on cumulative Kullback-Leibler divergence” by Yaser Mehrali and Majid Asadi.
- “Proof of conjectures on the standard deviation, skewness and kurtosis of the shifted Gompertz distribution” by Fernando Jiménez Torres.
- “Mixed double-ranked set sampling: a more efficient and practical approach” by Monjed H. Samuh, M. Hafidz Omar and M. Pear Hossain.

Volume 19, Issue 2

- “Statistical inference for a general class of noncentral elliptical distributions” by Jimmy Reyes, Diego I. Gallardo, Filidor Vilca and Héctor W. Gómez.
- “The use of control charts to monitor air plane accidents of the Hellenic Air Force” by Vasileios Alevizakos, Christos Koukouvinos and Petros E. Maravelakis.
- “On the estimation for compound Poisson INARCH processes” by E. Gonçalves, N. Mendes-Lopes and F. Silva.
- “Parameter estimation for the two-parameter Maxwell distribution under complete and censored samples” by
- “Talha Arslan, Sukru Acitas and Birdal Senoglu.
- “Semiparametric additive beta regression models: inference and local influence diagnostics” by Germán Ibacache-Pulgar, Jorge Figueroa-Zuñiga and Carolina Marchant.
- “Variance estimation in the presence of measurement errors under stratified random sampling” by Neha Singh, Gajendra K. Vishwakarma and Raj K. Gangele.
- “Concomitants of order statistics and record values from iterated FGM type bivariate-generalized exponential distribution” by H.M. Barakat, E.M. Nigm, M.A. Alawady and I.A. Husseiny.

Volume 19, Issue 3:

- “Calibration of the bulk and extremes of spatial data” by Maria Antónia Amaral Turkman, Kamil Feridun Turkman, Patrícia de Zea Bermudez, Soraia Pereira, Paula Pereira and Miguel de Carvalho.

- “Alternative estimation of the common mean of two normal populations with order restricted variances” by Adarsha Kumar Jena, Manas Ranjan Tripathy and Nabendu Pal.
- “Nonparametric smoothing for extremal quantile regression with heavy tailed data” by Takuma Yoshida.
- “An asymmetric area model-based approach for small area estimation applied to survey data” by Marcelo Rodríguez, Víctor Leiva, Mauricio Huerta, Camilo Lillo, Alejandra Tapia and Fabrizio Ruggeri.
- “Optimal B-robust estimation for the parameters of the Marshall-Olkin extended Burr XII distribution with an application to pharmacokinetics” by Yesim Güney, Senay Özdemir, Yetkin Tuuç and Olcay Arslan.
- “Pseudo-Gaussian and rank-based tests for first-order superdiagonal bilinear models in panel data” by Aziz Lmakri, Abdelhadi Akharif, Amal Mellouk and Mohamed Fihri.

Volume 19, Issue 4:

- “Asymmetric kernels for boundary modification in distribution function estimation” by Habib Allah Mombeni, Behzad Mansouri and Mohammad Reza Akhoond.
- “Production processes with different levels of risk: addressing the replacement option” by Francisco Stefano de Almeida, Cláudia Nunes and Carlos Oliveira.
- “A truncated general-G class of distributions with application to truncated Burr-G family” by Farrukh Jamal, Hassan S. Bakouch and M. Arslan Nasir.
- “An efficient mixed randomized response model for sensitive characteristic in sample survey” by Amod Kumar, Gajendra K. Vishwakarma and G.N. Singh.
- “A regression model for positive data based on the slashed half-normal distribution” by Yolanda M. Gómez, Diego I. Gallardo and Mário de Castro.
- “Comparison of the likelihood ratios of two diagnostic tests subject to a paired design: confidence intervals and sample size” by José Antonio Roldán-Nofuentes and Saad Bouh Sidaty-Regad.



Giovani Silva

• Prémio SPE 2021

O Prémio SPE, é promovido pela Sociedade Portuguesa de Estatística e pretende estimular a atividade de estudo e investigação científica em Probabilidades e Estatística entre os jovens.

O Prémio SPE 2021 – ao qual podiam concorrer trabalhos originais sobre temas de Probabilidades e Estatística, desde que não tivessem sido objeto de qualquer prémio atribuído por outra instituição – não foi atribuído uma vez que não foram recebidas candidaturas no prazo devido.

A candidatura ao Prémio SPE 2022, está aberta até ao dia 15 de setembro de 2022.

A habitual divulgação será feita em breve, na página da Sociedade Portuguesa de Estatística <https://www.spestatistica.pt/>.

FR

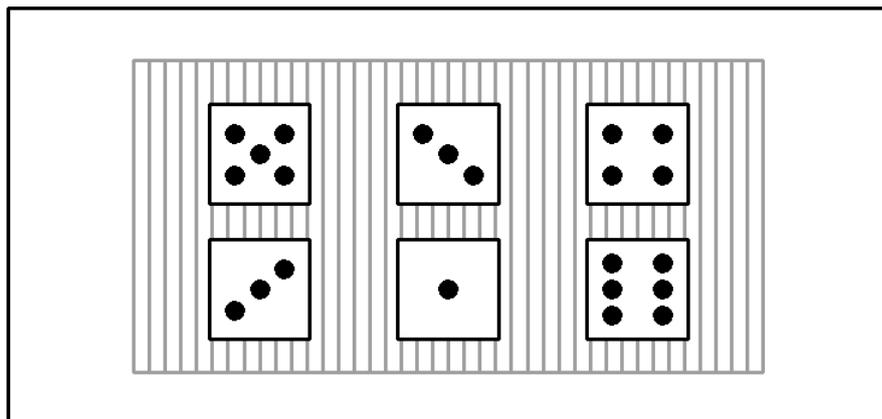
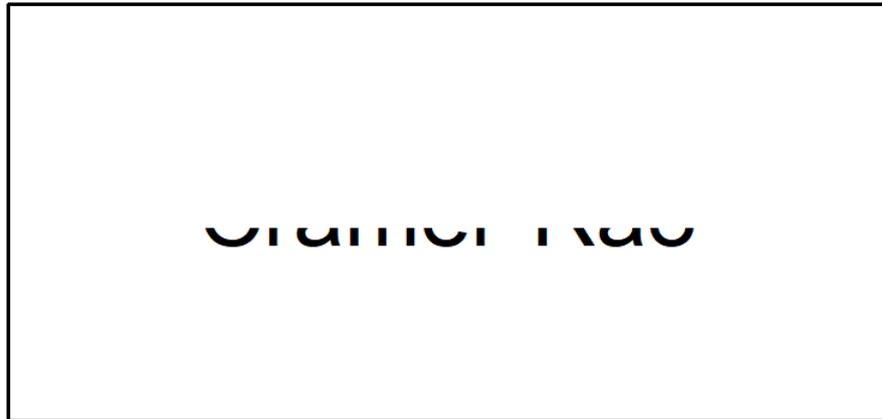
• Prémios Estatístico Júnior

Está em discussão uma alteração aos Prémios Estatístico Júnior. A habitual divulgação será feita em breve, na página da Sociedade Portuguesa de Estatística <https://www.spestatistica.pt/>.

FR



Enigmística de mefqa



Enigmas 35 e 36

No Boletim SPE outono de 2021 (p. 14):

10:12 11:37 18:44 00:13 14:20 22:31 10:57

sucessão cronológica

ERRO

margem de erro

Análise de Sobrevivência em R

Ivo Sousa-Ferreira, ivo.ferreira@staff.uma.pt

*Departamento de Estatística e Investigação Operacional, Faculdade de Ciências,
Universidade de Lisboa, Portugal e CEAUL – Centro de Estatística e Aplicações,
Faculdade de Ciências, Universidade de Lisboa, Portugal*

e

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

A Análise de Sobrevivência é um dos ramos mais antigos da Estatística, com origem no século XVII. Surgiu, essencialmente, como resposta à necessidade de obter métodos estatísticos que permitissem a resolução de problemas na área das Ciências Biomédicas, o que influenciou toda a terminologia utilizada. O seu foco é a análise do tempo de vida, que é definido como o tempo que decorre desde um instante inicial até à ocorrência de um acontecimento de interesse. Este acontecimento é estipulado à partida e pode assumir diversas formas como, por exemplo, a recidiva de uma doença, a avaria de uma máquina, a união matrimonial, a conclusão de um ciclo de estudos ou a dissolução de uma sociedade comercial. Assim, consoante a natureza do acontecimento em estudo, este ramo tem sido aplicado nas mais variadas áreas, nomeadamente na Medicina, Engenharia, Ciências da Educação, Sociologia e Economia. Adicionalmente, a escolha deste tipo de análise estatística tem a particularidade de permitir a inclusão de dados censurados, que surgem quando, por algum motivo, não é possível observar o acontecimento de interesse durante o período em que os indivíduos estão em observação. Para um conhecimento mais aprofundado, consulte-se Lawless (2003).

Os métodos tradicionais de Análise de Sobrevivência encontram-se disponíveis em vários programas de *software* estatístico, sendo que alguns dos mais utilizados são: SPSS, SAS, STATA, S-Plus e R. No entanto, a maioria desses programas é também bastante rudimentar no que diz respeito à análise de dados de sobrevivência, pois oferece pouca diversidade metodológica ao utilizador. Em oposição, o *software* estatístico R destaca-se por ser uma ferramenta bastante abrangente, uma vez que fornece uma panóplia vastíssima de métodos que procuram solucionar os mais variados problemas de Estatística e Probabilidade. Além disso, este programa beneficia do facto de ser gratuito, o que se revela uma mais-valia para toda a comunidade académica, em especial na área da investigação, permitindo que qualquer cientista ou organização possa contribuir para o seu desenvolvimento.

O *software* estatístico R, doravante designado apenas por R, foi concebido no início da década de 90 do século XX, pelos estatísticos Ross Ihaka e Robert Gentleman da Universidade de Auckland, em Nova Zelândia. Desde 1997, tem sido mantido por uma equipa, denominada por R Core Team (2022), a qual é constituída por 20 cientistas e programadores. O seu *download* de acesso livre pode ser feito através do sítio eletrónico da [rede de arquivos do R](#) (CRAN – *Comprehensive R Archive Network*). De entre as diferentes características que apresenta, o R possui um vasto leque de pacotes, em constante crescimento e atualização, o que lhe confere potencialidades acrescidas à sua já poderosa versão base. Porém, a abundância de pacotes (à data do presente texto, encontravam-se disponíveis mais de 19000 pacotes) pode ser vista como um obstáculo na procura rápida e eficiente do(s) pacote(s) mais adequado(s) para um determinado problema. Para ultrapassar essa questão, foram desenvolvidas [CRAN task views](#) sobre tópicos específicos (atualmente, existem 40 *task views*), que visam fornecer orientações para encontrar o pacote mais relevante para executar uma dada tarefa (*task*).

A Análise de Sobrevivência usufrui da vantagem de ter uma *task view* exclusivamente dedicada a si. A [task view Survival](#) é da autoria de Allignol e Latouche (2022), que também têm a responsabilidade de efetuar a sua atualização/manutenção regular. Esta *task view* está organizada em 12 temas centrais: Análise de Sobrevivência *standard*; modelos multiestados; sobrevivência relativa, modelos de efeitos aleatórios, sobrevivência multivariada, modelos bayesianos; aprendizagem automática; previsões e desempenho da predição; análise de potência; simulação; gráficos; e miscelânea. Ao longo destes temas, existe um pacote que é referido com frequência por ser muito generalista no que toca às ferramentas que disponibiliza ao utilizador sobre uma grande variedade de temas – o pacote **survival** (Therneau, 2021).

O pacote **survival**, assim como outros pacotes do R, dispõe de um conjunto de vinhetas que tem como objetivo fornecer tutoriais instrutivos para exemplificar a utilização prática das suas funcionalidades e auxiliar na interpretação e discussão de resultados. A vinheta [A package for survival analysis in R](#) é aquela que, no meu ponto de vista, providencia uma excelente introdução sobre este pacote e aborda mais assuntos relacionados com a análise de dados de sobrevivência. Nessa vinheta, encontra-se uma breve história a respeito da sua origem, onde é referido que as suas funções começaram por ser desenvolvidas para o S-Plus, em 1985. Como ambos os programas, R e S-Plus, têm por base a linguagem de programação S, muitos dos códigos desenvolvidos para o S-Plus podem ser facilmente implementados no R e, por vezes, nem é necessário efetuar qualquer alteração a esses códigos. Deste modo, a transição do S-Plus para o R ocorreu de forma natural, sem o autor indicar a data em que tal aconteceu.

Presentemente, o pacote **survival** é um dos pacotes mais utilizados no estudo da variável tempo de vida e, de acordo com o autor da vinheta supracitada, uma das principais razões para o sucesso deste pacote é o facto de todas as suas funções terem sido escritas para solucionar problemas associados à análise de dados reais, não descurando a fundamentação teórica que sustenta e atribui significado às funções programadas. Como é evidente, o referido pacote não cobre todos os temas deste ramo da estatística, mas seguramente fornece as rotinas e os objetos basilares para a definição de funções mais complexas. Uma prova disso é a existência de mais de 800 pacotes (dependentes) cujas funções são definidas à custa das funções do pacote **survival**. Além do mais, este é um dos poucos pacotes *core* do R, pacotes esses que são instalados em conjunto, e de forma automática, com a versão base do *software*.

Na literatura, é possível encontrar alguns livros que ilustram o modo como o pacote **survival** pode ser usado para modelar dados de sobrevivência e estimar certas quantidades de interesse, como sejam a probabilidade de sobrevivência e o risco relativo. De seguida, passo a indicar 3 livros que, no meu entender, enriquecem o leitor com uma excelente visão geral sobre os princípios fundamentais da Análise de Sobrevivência e as suas aplicações através do R. Um dos livros foi redigido por Therneau e Grambsch (2000) e, ao longo dos seus exemplos práticos, apresenta código do S-Plus que pode ser facilmente implementado no R com recurso ao pacote **survival** (Therneau, 2021). Os outros dois livros são da autoria de Carvalho *et al.* (2011) e de Moore (2016) que, embora abordem maioritariamente as particularidades associadas aos comandos do pacote **survival**, também apresentam código do R referente ao uso de outros pacotes como, por exemplo, o pacote **mstate** (de Wreede *et al.*, 2011) para estimar a função de incidência cumulativa em cenários de riscos competitivos; e o pacote **muhaz** (Gentleman, 2021) para efetuar a estimação da função de risco usando métodos não paramétricos baseados em funções *kernel*. Importa referir que, de entre estes três livros, apenas a obra de Carvalho *et al.* (2011) está escrita em português, aspeto este que pode ser encorajador para os principiantes em R que tenham preferência pela língua portuguesa.

O livro de Moore (2016) faz parte da série de livros [Use R!](#), publicada pela editora Springer. Conforme descrito por esta editora, os livros da referida série destinam-se a apresentar uma discussão acerca do uso do R numa área específica (*e.g.*, biometria, econometria e quimiometria) ou num tópico estatístico em particular (*e.g.*, dados omissos, dados espaciais e dados longitudinais). A série *Use R!* tem no mínimo mais 2 livros que se enquadram no domínio da Análise de Sobrevivência, nomeadamente as obras de Beyersmann *et al.* (2012) e de Willekens (2014). Ambos os livros exemplificam a utilização de diversos pacotes do R para ajustar modelos multiestado, ainda que o primeiro livro dê especial atenção à modelação de causas competitivas. Existem muitas outras editoras bem conceituadas que publicam obras

relacionadas com o R, sendo que algumas dessas obras estão reportadas na página principal deste *software*, mais propriamente na secção de documentação (consulte-se [Books related to R](#)).

Outra forma autodidata de manusear este programa, consiste em aceder a artigos de revistas científicas centradas no desenvolvimento de investigação em Estatística Computacional. Desde 2009, o próprio R tem uma revista indexada de acesso aberto, cujos artigos são sujeitos a revisão por pares. Esta revista, denominada [The R Journal](#), prioriza a publicação de artigos que abordam aspetos computacionais relacionados com as novas ferramentas ou pacotes do R.

Apesar do R permitir a programação de qualquer função através da sua interface de linha de comando, a sua versão base não incorpora uma interface gráfica do utilizador (GUI) que promova uma interação amigável, por meio de elementos gráficos. Consequentemente, aprender a manusear este programa é, sem dúvida, uma tarefa árdua para os iniciantes em programação. Para ultrapassar esse obstáculo, têm sido desenvolvidas GUI para o R. O *R Commander* é, provavelmente, a GUI mais antiga do R, tendo sido construída no formato de um pacote, designado por **Rcmdr** (Fox, 2005), o que significa que a sua instalação e carregamento são feitos da maneira habitual. Esta interface possibilita a implementação de funções através de um simples clique no seu menu de opções, além de continuar a permitir a escrita de linhas de comando.

Associado ao **Rcmdr** existem ainda os *plugins*, que são, na realidade, pacotes do R que visam estender as funcionalidades dos menus do *R Commander*. Neste momento, existem mais de 40 *plugins* para o **Rcmdr** e alguns deles são bastante úteis no âmbito da Análise de Sobrevivência, onde considero relevante destacar o **RcmdrPlugin.survival** (Fox e Carvalho, 2012) e o **RcmdrPlugin.EZR**. (Kanda, 2013). O primeiro *plugin* acrescenta algumas das funcionalidades do pacote **survival** aos menus do *R Commander*, incluindo o modelo de Cox e os modelos de sobrevivência paramétricos. Segundo Carvalho *et al.* (2011), o **RcmdrPlugin.survival** surgiu da necessidade em apresentar, de uma forma gentil, as potencialidades do pacote **survival** aos médicos investigadores. Relativamente ao segundo *plugin*, este também foi especialmente criado para facilitar a análise de dados em saúde. Note-se que, em inglês, a sigla EZR é lida como “*easy R*”. Em comparação com o *plugin* anterior, o **RcmdrPlugin.EZR** trata uma maior variedade de temas, incluindo a análise da curva ROC (*Receiver Operating Characteristic*) associada a um modelo de sobrevivência, a meta-análise do risco relativo, o cálculo da dimensão da amostra, entre outros.

Para uma breve revisão sobre outros pacotes do R com elevada utilidade neste ramo da estatística, remete-se o leitor para o trabalho de Abreu e Gouveia-Reis (2016), publicado nas atas do XXII Congresso da SPE. Nesse artigo, as autoras referem que a dinâmica do R faz com que qualquer trabalho sobre este *software* seja inevitavelmente incompleto e um pouco desatualizado. Esta é uma realidade com a qual os utilizadores do R têm que lidar, a qual está bem espelhada na evolução do seu número de pacotes, considerando que à data da escrita desse artigo (setembro de 2016) existiam cerca de 9000 pacotes e, em apenas 5 anos e meio, passaram a existir mais do dobro!

Por fim, não poderia deixar de realçar que existem vários investigadores portugueses, distribuídos ao longo de todo o território nacional (sem esquecer os arquipélagos atlânticos), que têm vindo a contribuir de algum modo para o desenvolvimento de pacotes do R no ramo da Análise de Sobrevivência, alguns em colaboração com cientistas de outros países. Na Tabela 1 estão indicados alguns desses pacotes por ordem alfabética, sem diminuir o mérito de outros trabalhos que mereciam igualmente ser aqui reportados.

Agradecimentos

O presente texto foi redigido na vigência da bolsa de doutoramento DFA/BD/6459/2020 e do projeto UIDB/00006/2020 (CEAUL – Centro de Estatística e Aplicações), financiados por Fundos Nacionais através da FCT – Fundação para a Ciência e a Tecnologia.

Tabela 1: Alguns pacotes do R para a análise de dados de sobrevivência, desenvolvidos por (ou com a colaboração de) investigadores portugueses

Designação	Referência	Breve descrição
clustcurv	Villanueva <i>et al.</i> (2021)	Um método para determinar <i>clusters</i> em curvas múltiplas com seleção automática do seu número, tendo por base os algoritmos <i>k</i> -médias ou <i>k</i> -medianas. A metodologia pode ser aplicada para determinar <i>clusters</i> em várias curvas de sobrevivência.
condSURV	Meira-Machado e Sestelo (2016)	Fornecer alguns estimadores não paramétricos e semiparamétricos (os quais são extensões do estimador de Kaplan-Meier) para estimar a função de sobrevivência condicional, na análise multivariada de tempos por indivíduo.
genSurv	Araújo <i>et al.</i> (2021)	Simulação de dados de sobrevivência com uma covariável dependente do tempo e simulação de dados de sobrevivência a partir de um modelo progressivo de doença-morte.
JMbayes2	Rizopoulos <i>et al.</i> (2021)	Permite realizar o ajustamento de modelos conjuntos para dados longitudinais e de sobrevivência, sob a abordagem bayesiana.
joineR	Philipson <i>et al.</i> (2021)	Análise de medições repetidas e de dados de sobrevivência por meio da sua modelação conjunta com efeitos aleatórios. Os dados de sobrevivência são modelados usando o modelo de Cox com covariáveis dependentes do tempo, enquanto o <i>outcome</i> longitudinal é modelado usando um modelo linear de efeitos mistos. A associação é capturada através de um processo gaussiano latente.
smoothHR	Meira-Machado <i>et al.</i> (2013)	Permite o cálculo das estimativas pontuais do risco relativo (e correspondentes intervalos de confiança) associado a uma covariável contínua, considerando que esta tem um efeito não linear que é definido à custa de uma função <i>spline</i> penalizada.
survidm	Soutinho <i>et al.</i> (2021)	Apresenta alguns métodos desenvolvidos recentemente para estimar várias quantidades associadas ao modelo de doença-morte, tais como a probabilidade de transição, a probabilidade de ocupação, a função de incidência cumulativa e a distribuição do tempo de permanência em cada estado.
vsd	Camacho e Abreu (2021)	Visualização de vários gráficos para dados de sobrevivência com censura à direita, incluindo a representação da estimativa de Kaplan-Meier da função de sobrevivência, da estimativa suavizada da função de risco e dos gráficos de floresta referentes aos coeficientes de regressão de um modelo de sobrevivência.

Referências

- Abreu, A. M. e Gouveia-Reis, D. (2016). Análise de sobrevivência e valores extremos em R. Em C. R. Cordeiro, *Estatística: Progressos e Aplicações. Atas do XXII Congresso da Sociedade Portuguesa de Estatística* (pp. 1 - 14). Portugal: Edições SPE.
- Allignol, A. e Latouche, A. (2022). *CRAN Task View: Survival Analysis*. Consultado a 8 de fevereiro de 2022. URL: <https://cran.r-project.org/web/views/Survival.html>
- Araújo, A., Meira-Machado, L. e Faria, S. (2021). *genSurv: generating multi-state survival data*. Pacote do R versão 1.0.4. URL: <https://CRAN.R-project.org/package=genSurv>
- Beyersmann, J., Allignol, A. e Schumacher, M. (2012). *Competing Risks and Multistate Models with R*. New York: Springer.
- Camacho, D. e Abreu, A. M. . (2021). *vsd: graphical shim for visual survival data analysis*. Pacote do R versão 0.1.0. URL: <https://CRAN.R-project.org/package=vsd>
- Carvalho, M. S., Andreozzi, V. L., Codeço, C. T., Campos, D. P., Barbosa, M. T. S. e Shimakura, S. E. (2011). *Análise de Sobrevivência: Teoria e Aplicações em Saúde* (2.^a ed.). Rio de Janeiro, Brasil: Fiocruz.

- de Wreede, L. C., Fiocco, M. e Putter, H. (2011). mstate: an R package for the analysis of competing risks and multi-state models. *Journal of Statistical Software*, 38(7), 1-30.
- Fox, J. (2005). The R Commander: a basic-statistics graphical user interface to R. *Journal of Statistical Software*, 19(4), 1-42.
- Fox, J. e Carvalho, M. S. (2012). The RcmdrPlugin.survival package: extending the R Commander interface to survival analysis. *Journal of Statistical Software*, 49(7), 1-32.
- Gentleman, R. (2021). *muhaaz: hazard function estimation in survival analysis*. Pacote do R versão 1.2.6.4. URL: <https://CRAN.R-project.org/package=muhaaz>
- Kanda, Y. (2013). Investigation of the freely available easy-to-use software ‘EZR’ for medical statistics. *Bone Marrow Transplant*, 48, 452–458.
- Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data* (2.^a ed.). New York: John Wiley & Sons.
- Meira-Machado, L. e Sestelo, M. (2016). condSURV: an R package for the estimation of the conditional survival function for ordered multivariate failure time data. *The R Journal*, 8(2), 460-473.
- Meira-Machado, L., Cadarso-Suárez, C., Gude, F. e Araújo, A. (2013). smoothHR: an R package for pointwise nonparametric estimation of hazard ratio curves of continuous predictors. *Computational and Mathematical Methods in Medicine*, 1-11.
- Moore, D. F. (2016). *Applied Survival Analysis Using R*. New York: Springer.
- Philipson, P., Sousa, I., Diggle, P., Williamson, P., Kolamunnage-Dona, R., Henderson, R. e Hickey, G. (2021). *joineR: joint modelling of repeated measurements and time-to-event data*. Pacote do R versão 1.2.6. URL: <https://CRAN.R-project.org/package=joineR>
- R Core Team (2022). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. Vienna, Austria. URL: <https://www.R-project.org/>
- Rizopoulos, D., Papageorgiou, G. e Afonso, P. M. (2021). *JMbayes2: extended joint models for longitudinal and time-to-event data*. Pacote do R versão 0.1-8. URL: <https://CRAN.R-project.org/package=JMbayes2>
- Soutinho, G., Sestelo, M. e Meira-Machado, L. (2021). survldm: an R package for inference and prediction in an illness-death model. *The R Journal*, 1-20.
- Therneau, T. M. (2021). *survival: A package for survival analysis in S*. Pacote do R versão 3.2-13: <https://CRAN.R-project.org/package=survival>
- Therneau, T. M. e Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model*. New York: Springer-Verlag.
- Villanueva, N. M., Sestelo, M., Meira-Machado, L. e Roca-Pardiñas, J. (2021). clustcurv: an R package for determining groups in multiple curves. *The R Journal*, 13(1), 164-183.
- Willekens, F. (2014). *Multistate Analysis of Life Histories with R*. New York: Springer.



Segundo filho: uma transição incerta

Rita Brazão Freitas¹, Andreia Maciel², Maria Filomena Mendes^{2, 3}

¹ Instituto Nacional de Estatística, Direção Regional de Estatística da Madeira; Professora da Universidade da Madeira.

² Investigadora do Laboratório de Demografia, CIDEHUS-UE, Universidade de Évora.

³ Professora da Universidade de Évora, Escola de Ciências Sociais.

1. Introdução

Portugal permanece com um dos níveis de fecundidade mais baixos da Europa, contrastando com países como a França, Suécia, Roménia, Irlanda e Dinamarca, cujos atuais níveis de fecundidade se encontram mais próximos do limiar da substituição das gerações (2,1 filhos por mulher). Apesar da ligeira recuperação do Índice Sintético de Fecundidade (ISF¹) observada nos últimos anos – 1,12 filhos por mulher em idade fértil em 2013 e 1,42 em 2019 – este indicador ainda permanece baixo e próximo de um filho por mulher. A partir do Inquérito à Fecundidade realizado em 2019 (IFEC 2019), sabemos que a maioria (60,2%) dos residentes em Portugal (mulheres entre 18 e 49 anos e homens dos 18 aos 54 anos) esperavam terminar o ciclo reprodutivo com pelo menos dois filhos, embora a maioria dessas pessoas ou ainda não tivesse entrado na parentalidade (42,7%) ou tivesse apenas um filho (9,8%), indicando que os portugueses continuam a adiar o nascimento dos filhos ou, possivelmente, a renunciar à sua fecundidade final esperada. Sendo ideais, desejos e intenções de fecundidade ainda associados a dois ou mais filhos por família, importa compreender o que conduz a que os portugueses que tencionam ter dois ou mais filhos tenham apenas um. A partir do IFEC 2019, analisamos uma amostra de homens e mulheres que esperam ter uma fecundidade final de dois ou mais filhos e procuramos identificar os fatores que melhor diferenciam aqueles que já transitaram para o segundo ou terceiro filho dos que ainda permanecem com um único filho.

2. Trajetórias da fecundidade em Portugal

A redução dos níveis de fecundidade no país ocorre já desde o final da década de 70 do século passado, tendo sido fortemente influenciada por profundas alterações sociais e económicas ocorridas no quadro da teoria da Segunda Transição Demográfica [6, 9]: alteração no padrão de formação das famílias; difusão e adoção de métodos contraceptivos mais eficazes; massiva inserção feminina no mercado de trabalho; e prolongamento das trajetórias escolares (potenciado pelo aumento das aspirações das pessoas que procuram uma maior autonomia e autorrealização).

Atualmente, as pessoas tendem a adiar o nascimento do primeiro filho até que terminem os estudos, encontrem um companheiro que considerem adequado e/ou se estabilizem no mercado de trabalho [10]. Se por um lado, incertezas económicas podem contribuir para esse adiamento ou até comprometer a decisão de ter um filho adicional [10], por outro, maiores rendimentos e melhores condições de emprego das mulheres, normalmente associadas a níveis de escolaridade mais elevados, podem favorecer a decisão de ter mais um filho [5].

Ainda, a redução dos níveis de fecundidade pode ser vista como uma interação (ou *trade-off*) entre quantidade e “qualidade” (termo habitualmente designado na literatura para se referir a um maior investimento no capital humano e social dos filhos [2]). Considerando que o aumento do número de filhos aumenta também o custo associado à “qualidade” de cada filho [2], muitos pais ponderam a decisão de ter (mais) um filho, sendo o desejo e a intenção de ter filhos ajustados ao longo da vida com o propósito de assegurar uma maior “qualidade” a cada filho [1, 3, 7, 8].

¹ ISF: Número médio de filhos tidos por mulher em idade fértil (15-49 anos completos) num determinado ano, admitindo que as taxas de fecundidade por idade observadas nesse ano se mantenham constantes.

3. Dados e metodologia

A partir da base de dados do IFEC 2019, analisamos uma amostra de 2 979 pessoas que pretendiam ter pelo menos dois filhos e recorremos a modelos de regressão logística com o objetivo de encontrar os fatores que diferenciam as pessoas que tinham apenas um filho (442) daquelas que já tinham transitado para o segundo (2 537). Considerando o esquema de amostragem complexo definido para este inquérito, estima-se que 265,8 mil portugueses não tinham (ainda) transitado para o segundo filho e que 1 281,3 mil já tinham pelo menos dois filhos.

Um modelo de regressão logística pode ser escrito através da expressão: $\pi(X) = E(Y|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)}}$, onde $\pi(X) = E(Y|X)$ representa a média da variável resposta Y condicionada aos valores das p variáveis explicativas (fatores), tal que $X = (X_1, X_2, \dots, X_n)$. Para ajustar os modelos de regressão seguimos a estratégia definida por [4], consideramos diversas variáveis explicativas apontadas na literatura como condicionantes da fecundidade e definimos a variável resposta: 0 – tem dois ou mais filhos; 1 – tem um filho e espera ter dois ou mais.

4. Fatores com impacto na transição para o segundo filho

Os motivos mencionados como muito importantes para a decisão de ter mais filhos não distinguem de forma clara as pessoas que (ainda) não alcançaram a fecundidade de dois filhos daquelas que já a alcançaram, uma vez que esses motivos são os mesmos para os dois grupos de pessoas: a vontade do próprio ou do cônjuge/companheiro, a existência ou perfil do cônjuge/companheiro ou fazer parte do projeto de vida. Diante de um cenário em que as pessoas têm amplo acesso a métodos contraceptivos eficazes e uma maior capacidade de controlar as suas circunstâncias, é indelével que só venham a ter filhos se e quando efetivamente os quiserem. Contudo, se a vontade constitui o primeiro passo para ter filhos, as circunstâncias que poderão levá-los a concretizar a sua vontade dependem de diversos fatores, como por exemplo, aqueles ligados à segurança financeira, social, psicológica e de saúde, sendo essas as áreas onde as políticas públicas podem (e devem) atuar.

Embora os motivos para a decisão de ter mais filhos não pareçam distinguir-se entre os dois grupos (filhos únicos vs. dois ou mais), os resultados mostram que a probabilidade de não ter transitado para o segundo filho está intimamente ligada ao adiamento da idade de entrada na parentalidade. Como seria de esperar, a idade ao nascimento do primeiro filho permanece como um dos mais importantes preditores da fecundidade, sendo os que adiaram o nascimento do primeiro filho mais vulneráveis ao intervalo de tempo ainda disponível para a chegada do segundo filho. A probabilidade de (ainda) não ter transitado para o segundo filho prende-se, também, com desejos de fecundidade mais baixos, com níveis de escolaridade mais elevados e com rendimentos mais baixos. Se por um lado, maiores níveis de escolaridade implicam uma entrada mais tardia no mercado de trabalho e na parentalidade, por outro, níveis de escolaridade elevados proporcionam uma melhor alocação no mercado de trabalho e mais possibilidades de auferir maiores rendimentos. Assim é expectável que, em idades mais tardias, e na eventualidade de virem a auferir maiores rendimentos, aqueles com níveis de escolaridade mais elevados tenham uma maior probabilidade de vir a recuperar os nascimentos uma vez adiados.

Aos fatores sociais e económicos, acresce a possibilidade de os portugueses poderem vir a limitar o número de filhos como forma de proporcionar a estes menos restrições e mais oportunidades, como fator determinante para não ter (ainda) transitado para o segundo filho. Portanto, numa situação de maior constrangimento e incerteza económica, os portugueses poderão não conseguir planear projetos de futuro com um maior nível de segurança e, conseqüentemente, não ter o segundo filho que um dia tencionaram ter. Importa mencionar que, em algumas situações, a atual pandemia (COVID-19) constitui mais uma agravante nesse cenário de incertezas.

Referências bibliográficas

- [1] Becker, G. (1960). An Economic Analysis of Fertility. Demographic and Economic Change in Developed Countries. *National Bureau of Economic Research*.
- [2] Becker, G., Lewis, H. G. (1974). Interaction between Quantity and Quality of Children. Economics of the Family: Marriage, Children, and Human Capital. *National Bureau of Economic Research*, 81-90.
- [3] Freitas, R. (2019). Que caminhos para o futuro da fecundidade em Portugal? O filho único como modelo de descendência final. Évora: Universidade de Évora. [Tese de doutoramento].
- [4] Hosmer, D., Lemeshow, S. (2013). *Applied Logistic Regression*. 3rd ed. New York: John Wiley & Sons.

- [5] Kohler, H.P., Billari, F., Ortega, J. A. (2006). Low Fertility in Europe: Causes, Implications and Policy Options. In *The Baby Bust: Who Will do the Work? Who Will Pay the Taxes?*, 48-109.
- [6] Lesthaeghe, R. (2020). The second demographic transition, 1986-2020: sub-replacement fertility and rising cohabitation – a global update. *Genus*, 76:10.
- [7] Maciel, A. (2015). Baixa fecundidade: adaptação tardia às mudanças estruturais ou consolidação da preferência por famílias de reduzidos padrões. Évora: Universidade de Évora. [Tese de doutoramento].
- [8] Mendes, M. F., Infante, P., Afonso, A., Maciel, A., Ribeiro, F., Tomé, L. P., Freitas, R. (2016). *Introdução ao estudo, Determinantes da fecundidade em Portugal*. Lisboa: Fundação Francisco Manuel dos Santos.
- [9] Surkyn, J., Lesthaeghe, R. (2004). Value orientations and the Second Demographic Transition (STD) in Northern, Western and Southern Europe. *Demographic Research Special Collection*. 3:3, 45-86.
- [10] Testa, M. R., Basten, S. (2014). Certainty of meeting fertility intentions declines in Europe during the ‘Great Recession’. *Demographic Research*, 31:23.

Anexo

Quadro 1. Coeficientes do modelo de regressão logística para mulheres (dos 18 aos 49 anos) e homens (dos 18 aos 54 anos): têm um filho, mas esperam ter mais vs. têm dois ou mais filhos.

Variáveis explicativas	Fecundidade final esperada de 2 ou mais filhos: fecundidade realizada = 1 vs. fecundidade realizada ≥ 2				
	Coeficiente	Erro padrão	Valor p	Odds Ratio	Intervalo de confiança a 95%
Fecundidade desejada (ref. mais de 2 filhos)					
2 filhos	0,38	0,17	0,023	1,46	(1,05-2,02)
menos de 2 filhos	1,35	0,29	< 0,001	3,86	(2,18-6,77)
Opinião sobre se é preferível ter só um filho com mais oportunidades e menos restrições, do que ter mais filhos (ref. discorda)					
concorda	0,51	0,15	< 0,001	1,67	(1,24-2,25)
Situação conjugal do próprio (ref. tem cônjuge)					
tem companheiro	1,18	0,17	< 0,001	3,25	(2,34-4,52)
não tem cônjuge ou companheiro	0,55	0,25	0,029	1,74	(1,04-2,83)
Nível de escolaridade do próprio (ref. até básico)					
secundário	0,79	0,20	< 0,001	2,20	(1,50-3,25)
superior	0,91	0,21	< 0,001	2,50	(1,65-3,79)
Escala do valor aproximado do rendimento mensal líquido do agregado familiar (ref. ≥ 2 201€)					
até 2 200€	0,73	0,20	< 0,001	2,06	(1,40-3,10)
Naturalidade do próprio (ref. outra)					
portuguesa	0,57	0,25	0,021	1,77	(1,11-2,94)
Tipologia da área de residência (ref. APR)					
AMU/APU	0,83	0,35	0,017	2,30	(1,21-4,81)
Idade ao nascimento do primeiro filho	0,14	0,02	< 0,001	1,16	(1,12-1,19)

Notas:

As categorias "0 filhos" e "1 filho" da variável fecundidade desejada foram agrupadas no modelo final devido ao reduzido número de observações.

As categorias "901 a 1 300 euros", "1 301 a 1 800 euros" e "1 801 a 2 200 euros" do escalão do valor aproximado do rendimento mensal líquido do agregado familiar foram agrupadas no modelo final, pois os coeficientes estimados pelo modelo e respetivos Odds Ratio não diferiam significativamente entre si.

APR - áreas predominantemente rurais; AMU - áreas medianamente urbanas; APU - áreas predominantemente urbanas. As categorias "AMU" e "APU" da tipologia da área de residência foram agrupadas no modelo final, pois os coeficientes estimados pelo modelo e respetivos Odds Ratio não diferiam entre si.

Adequabilidade do modelo: AUC = 0,78; R²=21%; valor p (teste de Hosmer) = 0,06.

Fonte: INE, Inquérito à Fecundidade, 2019.

Este texto foi publicado em “*Inquérito à Fecundidade – 2019* (IFEC 2019).



Liderança estatística no sector privado

Jorge Félix, jorge.felix@exigoconsultores.com

Diretor da Exigo

Leslie G. Godfrey¹, Professor Emérito de econometria na Universidade de York em Inglaterra, lecionava estatística com uma clareza inspiradora. Nas suas aulas sobre as propriedades dos modelos estatísticos utilizados em econometria tudo parecia mais simples e fácil. Com o Professor Godfrey os conceitos de rigor e precisão assumiam uma dimensão ainda mais profunda, para além do seu sentido quantitativo estrito.

Vem este introito a propósito da constituição da empresa Exigo que, desde 2001, opera com o lema “Rigor e Precisão”. Como a maioria das pequenas e médias empresas portuguesas a Exigo nasceu micro, com apenas duas pessoas, sendo uma delas estatística. Hoje a Exigo é uma empresa de consultoria, certificada pela norma ISO 9001: 2015 para a prestação de serviços na área da saúde, incluindo estudos de epidemiologia, medição de resultados em saúde, economia da saúde, acesso ao mercado farmacêutico, modelos matemáticos e estatística, inteligência de mercado, escrita científica e desenvolvimento de projetos de ciência e inovação no sector da saúde.

Passados 20 anos, a Exigo continua a operar alicerçada no rigor e precisão e o departamento de métodos quantitativos é um farol de liderança, determinação e afirmação do posicionamento estratégico da empresa. Vários estatísticos, matemáticos, bem como outros colaboradores da área das ciências sociais e das ciências da vida, com formação pós-graduada em estatística, contribuem diariamente para o vasto portefólio de soluções apresentadas ao mercado. A Exigo mantém a base operacional em Portugal, contudo cerca de 25% da atividade é desenvolvida para outros países da Europa (Espanha, França, Reino Unido, Suíça), da América (Brasil, EUA), da Ásia (Coreia do Sul) e do médio oriente (Arábia Saudita, Bahrein, Emirados Árabes Unidos, Omã).

O conhecimento em estatística é fundamental para a diferenciação e competitividade da Exigo no mercado da consultoria científica. De entre as áreas de atuação com maior contributo da estatística e dos estatísticos destacam-se, o planeamento de estudos epidemiológicos, análise estatística de dados de saúde, análise de sobrevivência sobre dados de ensaios clínicos de medicamentos inovadores, comparação indireta de tratamento através de métodos frequentistas e Bayesianos, exploração das propriedades das distribuições estatísticas, exploração da incerteza em processos de análise de decisão e eliciação de preferências sobre atributos, entre outros.

Uma das atividades de maior especialização da Exigo é a avaliação de tecnologias de saúde (ATS). A ATS é um processo multidisciplinar que utiliza métodos explícitos para determinar o valor de uma tecnologia de saúde em diferentes fases do seu ciclo de vida. A ATS tem como objetivo primordial informar a tomada de decisões, e contribuir para melhorar a qualidade, eficiência e equidade na utilização dos recursos afetos à saúde.

A Organização Mundial de Saúde define tecnologia da saúde como a aplicação de conhecimentos e competências organizadas sob a forma de medicamentos, dispositivos médicos, vacinas, procedimentos

¹ Co-autor do teste Breusch-Pagan-Godfrey (1. Breusch, T.S. and Pagan, A.R., *The Lagrange multiplier test and its applications to model specification in econometrics*. Review of Economic Studies, 1980. 47: p. 239-253, 2. Godfrey, L.G., *Misspecification Tests in Econometrics: The Lagrange Multiplier Principle and other Methods*, Cambridge University Press, Editor. 1988.) (por vezes abreviado para o teste Breusch-Pagan) que é baseado no multiplicador de Lagrange, testa se a variância dos erros de uma regressão é dependente dos valores das variáveis independentes. Nesse caso, a heterocedasticidade está presente.

e sistemas desenvolvidos para resolver um problema de saúde e melhorar a qualidade de vida. A ATS é uma área emergente e existe um número muito elevado de possibilidades de intervenção do estatístico. Antes da entrada no mercado dos novos medicamentos, aquando da ATS inicial, os dados dos ensaios clínicos provenientes do programa de investigação e desenvolvimento clínico, representam a principal fonte de informação. A prossecução dos objetivos da ATS implica frequentemente prever resultados clínicos e económicos a longo prazo, ou avaliar o resultado do novo tratamento em comparação com a prática clínica atual, com recurso a métodos de comparação indireta, quando não estão disponíveis ensaios clínicos de comparação direta (*head-to-head*). Também é frequente estimar o impacto do tratamento nos resultados relevantes para o doente (mortalidade, qualidade de vida, efeitos adversos, utilização de recursos). Em termos quantitativos é comum o foco estar na estimativa, em vez da inferência.

O papel importante da estatística na avaliação de tecnologias de saúde está bem estabelecido [3-6]. Neste contexto, o estatístico assume-se como um verdadeiro líder, participa e influencia as decisões dos projetos de ATS. Na Exigo esta é uma realidade quotidiana. Os nossos estatísticos assumem frequentemente a liderança das equipas multidisciplinares que participam em projetos que envolvem ATS. A liderança do estatístico neste contexto é para nós evidente e natural. Tratando-se de uma área eminentemente técnica, o estatístico encontra-se particularmente capacitado, para orientar os restantes membros da equipa no processo laborioso de geração ou síntese da evidência de suporte às decisões em ATS.

Escolhemos dois exemplos inspiradores acerca do papel e importância da intervenção do estatístico na avaliação de tecnologias de saúde: análise do tempo até evento e comparação indireta de tratamentos.

Os ensaios clínicos aleatorizados e controlados são a principal fonte de informação sobre a eficácia e segurança dos medicamentos inovadores à data da sua entrada no mercado. Com estes estudos experimentais, normalmente desenhados com grande rigor metodológico, incluindo a análise estatística pré-especificada, pretende-se gerar evidência da relação de causalidade entre a exposição e o efeito. Sucede que frequentemente a informação gerada nos ensaios clínicos não é suficiente para apoiar o processo de ATS.

Existem vários eventos de interesse utilizados na investigação clínica de novos medicamentos, mas a progressão e a morte são os mais usuais, sendo a sobrevivência livre de progressão e a sobrevivência global dois dos indicadores de resultados mais utilizados. Na Exigo estamos habituados à experiência dos estatísticos na aplicação de métodos não-paramétricos (Kaplan-Meier) e semi-paramétricos (Cox) para estimar os indicadores de resultados referidos. No entanto, no mundo real é comum os medicamentos inovadores serem utilizados por períodos de tempo muito superiores à duração dos ensaios clínicos. Logo, em ATS é usual utilizar os dados de tempo até evento, obtidos em ambiente experimental, para suportar o processo de extrapolação para horizontes temporais de maior duração, e que normalmente coincidem com a esperança de vida. Neste contexto, os métodos paramétricos assumem particular relevo, sendo as distribuições exponencial, weibull, log-logística e gama frequentemente aplicadas [7].

O objetivo destas práticas é simples: gerar uma expectativa acerca do efeito no longo termo. Esta informação é determinante no processo de decisão sobre o valor da inovação.

Outro desafio interessante em ATS é o de aferir o valor relativo de duas opções terapêuticas sem a existência de um ensaio clínico de comparação direta entre elas ou, no extremo oposto, quando existem vários ensaios clínicos a comparar duas ou mais terapêuticas entre si. No primeiro caso o desafio consiste em gerar uma estimativa do efeito relativo através de métodos de comparação indireta, no último, o desafio consiste no agrupamento de informação proveniente de estudos individuais (meta-análise). A fase qualitativa da comparação indireta de tratamentos inicia-se por uma revisão sistemática da literatura, para identificar os estudos relevantes para o problema de decisão em questão. Posteriormente, os resultados são revistos e selecionados a partir dos estudos individuais.

Na fase quantitativa os estatísticos assumem o protagonismo e definem o caminho a seguir. Se for possível criar uma rede de evidência, composta por vários tratamentos e vários estudos, desenvolvem uma meta-análise em rede, que é uma técnica usual para comparar múltiplos tratamentos simultaneamente numa única análise, combinando evidências diretas e indiretas dentro de uma rede de ensaios controlados e aleatorizados. Nesta fase é comum serem aplicados modelos hierárquicos Bayesianos via métodos de Monte Carlo em cadeias de Markov [8].

Ainda assim, se não for possível criar uma rede de evidência, a estatística tem soluções, como nos casos em que é necessário ajustar para as diferenças em covariáveis de base, quando os dados individuais ao nível do doente estão disponíveis apenas para um estudo e dados agregados estão disponíveis para um outro: comparações Indiretas ajustadas por correspondência (*Matching-Adjusted Indirect Comparisons*) [9].

Na Exigo estamos focados na identificação de problemas de decisão e nas soluções, que contribuam para a afetação mais eficiente dos recursos escassos do setor da saúde. Um dos principais pontos fortes dos estatísticos é compreender bem os dados, serem capazes de identificar padrões inerentes, e propor soluções metodológicas consistentes, que minimizem a incerteza subjacente ao processo de decisão. Outro fator de diferenciação que encontramos nos estatísticos é a capacidade de propor investigação adicional, como forma de gerar informação mais relevante e/ou precisa, quando as implicações de uma decisão incorreta superam os seus benefícios potenciais. Esta postura é muito apreciada no setor da saúde, uma vez que a maioria dos recursos disponíveis provêm de fundos públicos. Isto é, são fundos com um valor social importante e sobre os quais a sociedade, em termos genéricos, possui juízos de valor muito fortes.

Muitos projetos desenvolvidos na Exigo são relevantes no contexto nacional e podem condicionar a relação dos indivíduos com o sistema de saúde. Por exemplo, o acesso a medicamentos inovadores em Portugal, é gerido e condicionado pelo INFARMED - Autoridade Nacional do Medicamento e Produtos de Saúde, I.P., através do Sistema Nacional de Avaliação de Tecnologias de Saúde (SiNATS). Neste âmbito são tomadas decisões acerca do financiamento público de medicamentos inovadores, que posteriormente influenciam as opções de tratamento disponíveis no Serviço Nacional de Saúde. Os estatísticos da Exigo ao liderarem as equipas de investigação, ou participarem nestas equipas, contribuem decisivamente para a geração de evidência sobre o valor terapêutico e económico destas tecnologias e, deste modo, mesmo que indiretamente, também contribuem para a saúde e bem-estar da população Portuguesa.

Podemos igualmente identificar o valor e contributo dos estatísticos ao nível da gestão eficiente dos serviços de saúde. Em 2010/2011, uma equipa de elementos da Exigo, incluindo vários estatísticos, esteve envolvida numa consultoria ao Tribunal de Contas de Portugal [10]. Esta consultoria foi realizada em colaboração com académicos do Departamento de Estatística e Investigação Operacional, da Faculdade de Ciências da Universidade de Lisboa, Lisboa. A Auditoria ao sistema de pagamentos e de formação dos preços pagos às unidades hospitalares do Serviço Nacional de Saúde, foi muito compreensiva, mas simultaneamente muito objetiva e rigorosa em relação aos métodos estatísticos utilizados para suportar as conclusões. Com o contributo dos estatísticos foi possível ao Tribunal de Contas emanar um conjunto abrangente de recomendação aos hospitais do SNS, à Administração Central do Sistema de Saúde, I.P. e à Ministra da Saúde da altura. Algumas destas recomendações ainda hoje afetam o quotidiano da relação dos portugueses com as estruturas públicas de saúde.

Em conclusão, a atividade dos estatísticos do setor privado, nomeadamente em empresas de consultoria como a Exigo (setor da saúde), envolve liderança efetiva, participação ativa e proativa na tomada de decisão dos projetos, muito para além da mera consultoria estatística. A integração e envolvimento dos estatísticos em equipas multidisciplinares é determinante junto do poder decisivo, e tem impacto no atributo mais importante para a vida dos portugueses, a sua saúde.

Referências bibliográficas

1. Breusch, T.S. and Pagan, A.R., *The Lagrange multiplier test and its applications to model specification in econometrics*. Review of Economic Studies, 1980. 47: p. 239-253.
2. Godfrey, L.G., *Misspecification Tests in Econometrics: The Lagrange Multiplier Principle and other Methods*, Cambridge University Press, Editor. 1988.
3. Alshreef, A., Latimer, N., Tappenden, P., et al., *Statistical Methods for Adjusting Estimates of Treatment Effectiveness for Patient Nonadherence in the Context of Time-to-Event Outcomes and Health Technology Assessment: A Systematic Review of Methodological Papers*. Med Decis Making, 2019. 39(8): p. 910-925.

4. Hagiwara, Y. and Shiroiwa, T., *Estimating Value-Based Price and Quantifying Uncertainty around It in Health Technology Assessment: Frequentist and Bayesian Approaches*. Med Decis Making, 2022: p. 272989X221079554.
5. O'Hagan, A. and Stevens, J.W., *Bayesian methods for design and analysis of cost-effectiveness trials in the evaluation of health care technologies*. Stat Methods Med Res, 2002. 11(6): p. 469-90.
6. White, S.J., Ashby, D. and Brown, P.J., *An introduction to statistical methods for health technology assessment*. Health Technol Assess, 2000. 4(8): p. i-iv, 1-59.
7. Jackson, C., Stevens, J., Ren, S., et al., *Extrapolating Survival from Randomized Trials Using External Data: A Review of Methods*. Med Decis Making, 2017. 37(4): p. 377-390.
8. Spiegelhalter, D.J., Myles, J.P., Jones, D.R., et al., *Bayesian methods in health technology assessment: a review*. Health Technol Assess, 2000. 4(38): p. 1-130.
9. Signorovitch, J.E., Sikirica, V., Erder, M.H., et al., *Matching-adjusted indirect comparisons: a new tool for timely comparative effectiveness research*. Value Health, 2012. 15(6): p. 940-7.
10. Tribunal de Contas Portugal, *Auditoria ao sistema de pagamentos e de formação dos preços pagos às unidades hospitalares do Serviço Nacional de Saúde*. 2011.



Breve testemunho

Manuela Neves, *manela@isa.ulisboa.pt*

CEAUL & ISA

O convite do Prof. Fernando Rosado para deixar algumas palavras sobre os desafios que me foram surgindo ao longo da minha carreira e o que deles gostaria de partilhar não é tarefa fácil quando, ao necessitar de me sentar para pôr as ideias em ordem, surge uma tarefa competitiva. É que, tal como afirmei há uns meses, tenho que fazer opções: dar colo, ajudar nos almoços/jantares, mudar fraldas, adormecer... ou vir aqui a correr escrever mais algumas linhas. As que aqui deixo foram escritas nestas condições.

Estudei Estatística há muitos, muitos anos, na Faculdade de Ciências ainda na Rua da Escola Politécnica, vagueei por salas onde agora está instalado o Museu de História Natural.

No 2º ano do meu curso as aulas foram lecionadas pelo Prof. Galvão de Melo, com a sua serenidade, a sua gentileza e sendo um desafio para nós.

Recordo alguns dos seus ensinamentos, que ao longo da minha vida ainda fui transmitindo aos meus alunos.

Nessa altura a Estatística estava muito arredada dos estudantes dos outros níveis de ensino, era lecionada apenas a nível universitário.

Na década de 1970, o *International Statistical Institute* (ISI) começou a dar mais atenção à promoção da educação estatística nas escolas e universidades. A Comissão de Educação desenvolveu esforços com a criação de grupos de trabalho. Dois dos mais importantes foram o *Teaching Statistics at School Level* (TOTSAS), liderado inicialmente por Vic Barnett e o *International Conferences in Statistical Education* (ICOTS).

O grupo TOTSAS criou um boletim informativo - *International Statistical Education newsletter* - que foi distribuído a professores universitários e de outros graus de ensino.

Em tempos, cheia de curiosidade sobre o salto indispensável no ensino da Estatística li textos do volume “*Studies in mathematics education. The teaching of statistics*”, editado por Robert Morris em 1989. Muitas das ideias que na altura surgiram como inovadoras no contexto do ensino, continuam atuais nos tempos que vivemos.

Recordo que por essa altura participei em ações de formação de Probabilidades e Estatística para professores do ensino secundário, que realizei nas Escolas Secundárias de Bocage (Setúbal), de Paço de Arcos, da Cidade Universitária e de Alvide (Cascais), a convite da Sociedade Portuguesa de Matemática (de 1990 a 1992).

No 1º Congresso Anual da Sociedade Portuguesa de Estatística, realizado no Vimeiro em junho de 1993, os estatísticos portugueses reuniram-se para refletir sobre a **Estatística e o Futuro e o Futuro da Estatística em Portugal**. Foi sob este tema que saíram as Atas desse Congresso. Quero realçar o debate que então decorreu sobre o ensino da Estatística no ensino básico e secundário. Tal sessão foi preparada pela Profª Eugénia Graça Martins, que tem continuado dedicada à causa da qualidade do ensino da Estatística, a nível básico e secundário. Com ela tenho partilhado ao longo de vários anos a experiência de analisar e premiar os trabalhos candidatos ao Prémio Estatístico Júnior, promovido pela Sociedade Portuguesa de Estatística.

Há ainda um longo caminho a percorrer para o reconhecimento adequado da importância do ensino rigoroso da Estatística em todos os graus de ensino a que, em meu entender, deve ser dada muito mais atenção.

Ensinei Estatística e Probabilidade a vários cursos de licenciatura, quer em contexto mais metodológico e fundamental (como as aulas que dei na Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa à licenciatura em Matemática) quer como uma ferramenta em várias Licenciaturas e Engenharias no Instituto Superior de Agronomia (ISA).

Ensinei também várias disciplinas no âmbito da Estatística em cursos de mestrado e atualmente colaboro na lecionação de uma unidade curricular de doutoramento.

Há cada vez mais dificuldade em se aceitar que os conceitos teóricos devem sustentar o “edifício que se está a construir”.

Considero que as “reformas” em curso, sob o chapéu da facilidade crescente de utilização dos computadores com os muitos e variados pacotes que tudo fazem, devem exigir um sentido maior de responsabilidade e conhecimento da correta utilização de métodos que, podendo ser mais ou menos sofisticados, dependem de pressupostos muitas vezes não verificados e que por isso são usados indevidamente. A acessibilidade de muitos pacotes pode favorecer a má aplicação de métodos estatísticos.

Voltei a rever partes daquele livro e a minha experiência docente e as minhas ideias sentem-se muito próximas das preocupações lá expressas. Jacobsen começa por dizer “... *Pegue num jornal diário e sublinhe cada palavra ou expressão em que a matemática seja necessária para sua compreensão ou interpretação. Encontra alguma equação? Alguma questão geométrica ou trigonométrica?... Em vez disso encontra tabelas, gráficos e palavras como “médias”, “tendências”, “projeções”, “estimativas”, “correlações”, “improváveis”...*” Todas essas palavras são do domínio da Estatística.

Há termos e expressões que já entraram no quotidiano, mas tal não significa que haja conhecimento adequado do seu significado. Não conhecer e por isso não perceber e não conseguir interpretar o significado de um termo estatístico é persistir na iliteracia estatística.

Uma pergunta pertinente - quem deve ensinar estatística agora que ela surge em tantas disciplinas escolares?

As dificuldades não estão agora tanto em calcular medidas, mas em interpretá-las. O recurso à estatística para interpretar, prever os fenómenos do mundo real não se pode reduzir à aplicação mecânica de certas técnicas padronizadas.

O investigador de áreas nas quais necessita de utilizar estatística deve trabalhar em cooperação com o estatístico que estará em condições de procurar um modelo que melhor se ajuste à situação em análise ou eventualmente criar novos modelos.

A tradução das conclusões estatísticas para o contexto e linguagem específica da área em estudo deve ser feita pelo utilizador, com o apoio do estatístico.

A Estatística tem renovado o seu ensino pela inclusão das ferramentas que agora tem ao seu dispor, mas não pode deixar de transmitir o rigor e justificação dos pressupostos e restrições sobre as quais muitos modelos são construídos.

Assustam-me, mas devo ser “um(a) velho(a) do Restelo”, tantos movimentos que vejo atualmente nalgumas das nossas escolas procurando reduzir o ensino da Estatística à mera utilização de ferramentas.

Referências:

- Gani, J. (1988). New Challenges to the ISI Education Committee. In: J. Swift (eds.), *Proceedings of the Second International Conference on Teaching Statistics*, Victoria, British Columbia, Canada, II-16 August 1986. Victoria, B.C., University of Victoria.
- Barnett, V. (1982). *Teaching Statistics in Schools Throughout the World*. Voorburg (Netherlands), International Statistical Institute. 250 pp. Cockcroft.

Studies in mathematics education - The teaching of statistics (1989), Volume 7 Edited by Robert Morris, United Nations Educational, Scientific and Cultural Organization.

A Estatística e o Futuro e o Futuro da Estatística (1994). Editada por Dinis Pestana, Antónia Amaral Turkman; João Branco, Luísa Loura, Ana Pires. SPE e Edições Salamandra.



Liderança estatística - uma visão muito pessoal

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

Faculdade de Economia, Universidade do Porto & CIDMA

Quando o nosso caro Editor escolheu Liderança Estatística como tema central deste boletim e pediu colaboração, confesso que fiquei algo perplexa pois nunca me havia cruzado com a expressão. No entanto, uma busca dos termos “Statistical Leadership” e “Leadership in Statistics”, no incontornável Google, leva-nos a verificar que o tema foi já abordado no ano longínquo de 2004 [1]. O artigo mais recente que encontrei data de 2021, é da autoria da Helen MacGillivray [2] e aborda as diversas dimensões que esta designação contém. A liderança estatística exerce-se em equipas multidisciplinares na investigação, na indústria e serviços, nos negócios, na administração pública e ensino a todos os níveis e em todas as áreas. O que se segue é a minha interpretação pessoal do tema, fruto necessariamente da minha experiência essencialmente académica e depois de alguma reflexão.

A liderança estatística assenta em saber ouvir, saber questionar, saber explicar, saber comunicar. E nem sempre é fácil. Quando eu digo (e faço-o com alguma frequência mas estou a tentar emendar-me) “dados são dados”, não estou a ter uma atitude de liderança pois cada conjunto de dados tem uma história que é preciso conhecer e entender para que a sua análise seja de interesse. Por esta razão, eu peço aos estudantes (Mestrado e Doutoramento) que escolham para os respetivos projetos temas que sejam do seu interesse (com séries temporais, claro!) pois assim terão maior interesse em mergulhar no problema e na história dos dados. O outro ponto que eu considero importante é que uma resposta aproximada ao problema certo é preferível a uma resposta exata ao problema errado, onde problema certo significa aquele que tem de ser resolvido e problema errado qualquer outro que possa ser parecido. É por isso precisa alguma flexibilidade e sentido prático nas análises. Quanto à comunicação, ela é fundamental em todas as etapas do estudo, para que entremos na linguagem do problema a resolver e ajudar à tomada de decisão com base nos resultados.

Do meu ponto de vista, nós que somos académicos e votamos grande parte da nossa vida à formação quer de estatísticos quer de utilizadores da estatística, temos responsabilidade acrescida. Aos futuros profissionais estatísticos temos de transmitir os valores e boas práticas que lhes permitirão serem líderes estatísticos. Aos futuros profissionais das outras áreas, e aqui estou a referir-me ao nosso ensino de serviço a outras áreas, temos de dar a conhecer o mundo mágico da estatística porque serão afinal eles quem no futuro estará (ou não) disponível para aceitar liderança estatística.

Referências

- [1] Ronald D. Snee and Roger W. Hoerl (2004). Statistical Leadership, STATISTICS ROUNDTABLE, <http://rube.asq.org/quality-progress/2004/10/statistics-roundtable/statistical-leadership.html>
- [2] Helen MacGillivray (2021). Leadership Across the Diversity of Statistics, in Leadership in Statistics and Data Science, Amanda L. Golbeck Editor, Springer.



A Estatística na investigação científica: Submissão, supremacia ou parceria inseparável?

Carlos A. Braumann, braumann@uevora.pt

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

Quero começar por agradecer aos Editores deste número do Boletim o convite para dar um contributo sobre o relevante tema da liderança estatística. Trata-se de uma temática de que só agora ouvi falar, mas onde reconheço conexões com alguns episódios da minha experiência pessoal. Como leigo da matéria, limitar-me-ei a falar de alguns exemplos dessa experiência e das reflexões certamente muito amadoras que me suscitaram, contando com a benevolência do leitor amigo.

Muitos projetos de investigação científica, diria mesmo quase todos (mas não todos), exigem tratamento estatístico adequado, quer ao nível do planeamento, quer da execução e recolha de dados, quer da análise e conclusões, quer ainda da divulgação dos resultados e projetos de futuro aprofundamento da investigação. No decorrer do projeto, as análises preliminares e intermédias podem ainda, com o auxílio da Estatística, apoiar processos de reajuste ou de prossecução de avenidas alternativas ou complementares. Se o projeto visa dar apoio a um processo de decisão com consequências sobre a vida de pessoas, comunidades ou empresas e serviços, o estudo das consequências das possíveis decisões pode ainda envolver a utilização de modelos que permitam a análise de cenários ou condições alternativas, modelos cuja validação e avaliação do grau de incerteza dos resultados exige normalmente métodos probabilísticos e estatísticos.

Estes factos são do conhecimento da comunidade científica e são reconhecidos pelas revistas científicas e outras publicações conceituadas na avaliação dos artigos científicos submetidos para publicação.

Mas será que desse conhecimento se extraem, por parte dos investigadores e das entidades que financiam a investigação, as naturais consequências sobre a necessidade de garantir um acompanhamento estatístico competente e permanente em todas as etapas do projeto: antes, durante e depois? Quantas equipas de investigação conhecemos em que um ou mais investigadores com competências estatísticas adequadas integrou de pleno direito a equipa em todas estas etapas? Algumas certamente, mas serão porventura muito mais as outras, com consequências que podem ser muito negativas sobre a qualidade e/ou a eficiência da investigação, o bom aproveitamento dos financiamentos nela aplicados e a bondade das decisões que dela possam decorrer.

Nos programas de formação superior de onde saem os futuros investigadores das várias áreas científicas já é normal haver pelo menos uma cadeira de iniciação à Estatística e, portanto, é comum os projetos de investigação contarem com investigadores com tal formação. Isso é importante para eles terem os conhecimentos básicos de Estatística e poderem compreender e dialogar com os estatísticos propriamente ditos. O problema surge quando se convencem de que essa formação é suficiente e que estão habilitados a desempenhar com qualidade as competências estatísticas necessárias. Alguns, mais sábios e por isso mais humildes, saberão que tal não é suficiente e terão adquirido eles próprios uma formação estatística mais avançada ou integrarão nas suas equipas pessoas com formação estatística adequada. E os outros?

Parece-nos, assim, importante que saibamos veicular aos nossos formandos as sérias consequências da arrogância de considerar a Estatística como meia dúzia de receitas que se dominam com pouco esforço e de a investigação dispensar o apoio de especialistas estatísticos. Mas há outros riscos que podem ser tanto ou mais sérios.

O mais típico é fazer toda a investigação sem qualquer apoio estatístico especializado e depois pedir o apoio estatístico competente: “Eis aqui a investigação que fizemos e os dados recolhidos e precisávamos que fizesse o tratamento estatístico.” Tem a virtude de evitar erros graves que, sem esse apoio, poderiam ocorrer, mas não evita possíveis consequências negativas sérias que seriam evitáveis se o acompanhamento estatístico tivesse ocorrido durante todo o processo.

Esta é uma situação muito generalizada. Tenho vários exemplos da minha vida académica em que investigadores de variadas instituições (de ensino superior ou de outra natureza) me contactaram *a posteriori*. Vou-me referir a dois desses exemplos alheios à minha instituição em que essa atitude prejudicou gravemente a investigação, com desperdício do esforço nela envolvido e com desapontamento dos investigadores sobre as expectativas que tinham criado. Fá-lo-ei, naturalmente, poupando o leitor dos pormenores e com a devida cautela para manter o anonimato das pessoas envolvidas e suas instituições.

Um dos exemplos refere-se a um extenso, diria até enciclopédico, estudo de caracterização das centenas de castas de determinada espécie vegetal de interesse económico, com informações completas sobre cada casta em termos de crescimento, floração, produção, características morfométricas, etc., que decorreu durante 3 anos consecutivos (naturalmente, com características climáticas variáveis de ano para ano). Só havia um pequeno problema: a amostra anual de cada casta tinha tamanho um. Embora isso não invalide a utilidade deste estudo, remete-o para um carácter muito mais preliminar do que porventura seria a intenção dos investigadores.

O outro exemplo refere-se ao estudo da influência das visitas escolares a museus na melhoria da avaliação dos estudantes de certa disciplina. O estudo envolvia um número elevado de estudantes, creio (já não me recordo bem dos pormenores) que do 2º ou 3º ciclo do ensino básico. Após a realização de uma prova escrita de avaliação da disciplina versando determinada matéria nela lecionada, os estudantes iam visitar um museu relacionado com a matéria e realizavam depois outra prova de avaliação sobre a mesma matéria. A ideia era comparar as classificações das duas provas por um método estatístico adequado (e daí o pedido de apoio *a posteriori* de um estatístico) para ver se havia uma melhoria significativa da classificação após a visita ao museu, sustentando assim a tese de que tais visitas podiam contribuir para melhorar o sucesso escolar. Havia de facto uma melhoria significativa e de magnitude assinalável da classificação. A questão era se tal se devia à visita ao museu ou ao facto de os alunos estarem a fazer uma prova de avaliação pós-visita que era igualzinha à que tinham feito antes da visita. Será que não haveria sempre uma melhoria significativa ao repetirem uma prova cujas perguntas já conheciam, houvesse ou não visita ao museu? E note-se que esta dúvida nada tem a ver com a Estatística, é antes uma questão de senso comum, mas os investigadores estavam tão seguros da sua tese que nem sequer a puseram. O que tem a ver com a Estatística era a necessidade de ter havido um delineamento experimental adequado que permitisse isolar a influência da visita ao museu na classificação (por exemplo, grupos de controlo que não tivessem feito a visita e/ou fizessem provas diferentes). Nenhum delineamento adequado foi utilizado e, portanto, a dúvida, para quem a tem, terá de ficar para já sem resposta.

O que poderá estar por trás desta atitude de só obter apoio estatístico *a posteriori*? Será por presunção de que tudo se pode fazer com a Estatística, ou porque ela é realmente poderosa (e é mas não faz milagres) ou porque ela pode sempre “torturar os dados” até que eles confessem o que queremos? Será porque só então se apercebem de possíveis dificuldades e lá vai de chamar o 112 para ver se ainda se salva o doente. Ou será que não se apercebem de qualquer dificuldade, mas simplesmente acham que fica bonito, está na moda e dá mais credibilidade integrar alguma componente estatística, qual flor na lapela que até pode facilitar a publicação dos resultados numa revista científica conceituada?

Haja ou não haja apoio estatístico continuado ou só *a posteriori*, também não é inédito haver estudos com pouca relevância científica em que se aproveita a credibilidade que a Estatística conquistou para

criar uma cortina de fumo sobre a irrelevância do estudo, quando não mesmo sobre os seus erros. Todos já ouviram falar dos prémios IgNOBEL. Nalguns casos explora-se, consciente ou inconscientemente o facto de, se tivermos muitas variáveis, mesmo sem qualquer relação entre elas, e as correlacionarmos duas a duas, vamos sempre encontrar algumas correlações significativas que, principalmente se forem surpreendentes, podem ser incorretamente exploradas.

Claro que também já tive experiências muito positivas. A mais recente tem a ver com o estudo do crescimento de bovinos mertolengos e de raça alentejana, no âmbito da Ação E (modelos de crescimento) do Grupo Operacional PDR2020-1.0.1-FEADER-031130 - GOBovMais – Productivity improvement in the system of bovine raising for meat (financiado pelo PDR 2020). Essa Ação decorre no CIMA sobre a coordenação do Colega Gonçalo Jacinto e a colaboração da Colega Patrícia Filipe e minha, em estreita parceria com as associações de produtores das raças envolvidas, a ACBM e a ACBRA, e colegas zootécnicos.

Nesta Ação houve oportunidade de desenvolver modelos estocásticos de crescimento dos animais em ambiente sujeito a variabilidade aleatória e de os utilizar para otimizar o lucro dos produtores. Para a conceção dos modelos e estimação dos parâmetros usaram-se dados já disponíveis na base de dados GENPRO, facultados pelas associações de produtores. Mas, dada a informação disponível abranger essencialmente idades jovens, o que dificulta a estimação de certos parâmetros, houve necessidade de conduzir experiências de crescimento até idades mais avançadas. Para tal usaram-se os modelos e estimativas provisórias dos parâmetros em estudos de simulação que deram indicações sobre a efetiva necessidade de novos dados em termos de animais envolvidos, frequência das suas pesagens e idades abrangidas. Com isso, foi possível desenhar as experiências de crescimento de forma a permitir a estimação dos parâmetros com a qualidade necessária e com contenção dos custos associados. A estatística esteve presente ao longo de todo o processo. Mas não em posição de supremacia nem de submissão, antes de diálogo colaborativo. Por exemplo, quando se verificaram dificuldades nas experiências devido a problemas de agressividade entre os animais de certas idades nelas envolvidos, foi fácil encontrar um compromisso e ajustar o desenho das experiências.

Há, com efeito, vantagens claras na participação da estatística em todas as fases do projeto. Não falaria tanto na necessidade de liderança estatística, preferindo antes falar da necessidade de uma verdadeira parceria estatística. Quando ela não existe, a Estatística é certamente uma vítima, mas, de facto, as principais vítimas são a investigação em causa e os investigadores que nela participam, bem como as áreas científicas em que ela decorre e os financiadores e os possíveis beneficiários dos seus resultados. Como membros da SPE, como atores na formação estatística de futuros investigadores de múltiplas áreas científicas, como parceiros de investigação de projetos científicos nos mais variados domínios, como membros interventivos da comunidade académica e da sociedade, podemos contribuir para difundir a necessidade da Estatística como parceira inseparável.

Agradecimentos: O autor integra o CIMA - Centro de investigação em Matemática e Aplicações, da Universidade de Évora, projeto UID/04674/2020 financiado pela Fundação para a Ciência e a Tecnologia.



A Excelência como liderança

Fernando Rosado, *fernando.rosado@ciencias.ulisboa.pt*

*Departamento de Estatística e Investigação Operacional
Faculdade de Ciências, Universidade de Lisboa*

Introdução

Liderar é dirigir. E, assim e desde logo, é optar por uma determinada direção. Liderar, acima de tudo, trata de relações humanas. A natureza e o exercício da liderança, desde sempre foram assuntos de “investigação” para o ser humano, ao longo da história; de onde podemos concluir sobre a evolução de: “o quê” e o “porquê” ou mesmo o “para quê” quando abordamos essa temática. O herói, o líder e o ídolo são “ideias” para temas de debate profundo – com as suas semelhanças definidoras e até pelas diferentes variantes na imagem que nos provocam. Nas últimas décadas o herói passou a líder que (em terminologia política também) pode ser eleito e demitido. Nesta perspetiva (quase), em especial no sentido político, pode entender-se o “ser líder” como “uma profissão”. Não avancemos, por aqui.

Dirigir é ter poderes para decidir e também a responsabilidade de administração ou de gestão. Para tal a Estatística, a Ciência Estatística, a todos os níveis é um apoio à decisão. A decisão, a teoria da decisão e bem assim a Estatística em conjugação com a liderança geram Liderança Estatística. Este é, na realidade, um tema que, no contexto de uma sociedade, científica ou não, tem reconhecida importância. A teoria da decisão é um tema que, no contexto português, tem sido muito pouco abordado¹.

Para aprofundar devidamente a questão, não sou especialista nessa habilidade (ou jeito) para influenciar a fim de atingir objetivos, como muitas vezes e em termos muito gerais se interpreta a liderança. Muito menos o sou para uma intervenção ou qualquer relato a nível histórico, quiçá científico, sobre as diversas teorias para a construção de uma boa liderança. Assim, para além desta breve introdução resta-me tentar transmitir alguma reflexão sobre a componente Estatística como construtora do tema central que neste Boletim é um desiderato editorial.

Os mais seniores, partilharam uma época em que, muito particular em Portugal, ainda se colocava a dúvida sobre o próprio significado e principalmente sobre o (enorme) valor da Estatística como Ciência. Uma consulta das referências históricas e bibliográficas – que vão desde as mais conceituadas revistas científicas até às mais modestas publicações – permitirá encontrar, nos mais diversos campos, um suporte para calcular o valor da Estatística como Ciência de Apoio à Decisão. Mas, não basta para essa avaliação. É fundamental a consolidação. A administração de uma comunidade, através das suas instituições de governo e comércio, depende muito das informações estatísticas e essa dependência aumenta à medida que o negócio intervém, cada vez mais, no planeamento da vida económica e social.

Os propagandistas, gestores e dirigentes administrativos que utilizam (e que às vezes deturpam) as estatísticas são bastante numerosos e a eles podem juntar-se os mais diversos utilizadores – desde os estudantes de ciências sociais aos políticos. Todos empregam factos e métodos estatísticos para construir bases para políticas. Tais factos e métodos também têm um lugar muito importante no desenvolvimento da sociologia e economia como ciências, são relevantes para os experimentadores, na maior parte dos ramos da biologia e são fundamentais para aqueles que trabalham nas ciências mais exatas como a física, a química ou a engenharia. É a importância da Estatística.

Um pouco de história – pioneiros em Portugal

A Estatística tornou-se conhecida no século XX como um instrumento matemático para analisar os dados e afirmou-se como ciência na sequência da contribuição inovadora desenvolvida por grandes

¹ Cf., por exemplo, *Estatística: Inferência e Decisão*, de Bento Murteira. Imprensa Nacional Casa da Moeda, 1988.

cientistas do século XIX que foram introduzindo o “pensamento estatístico” nas diversas áreas do saber. A (criação da) Estatística é, portanto, o culminar de uma “tradição de pensamento científico”. Quando e como, despertou Portugal para essa realidade (estatística) mundial?

Situemo-nos na década de 70 do século passado.

Portugal vive “momentos novos” nos mais diversos caminhos e também no campo da ciência. Com especial incidência a partir dos anos 60, muitos portugueses migravam no seu país, procurando melhores condições de vida ou, (pelo menos,) trabalho. Alguns, não poucos, partiam para países terceiros onde “a certeza” de “melhor vida” ajudava a enfrentar esse desafio. Este estado de espírito português também arrastava e desafiava os jovens para novos caminhos. Alguns descobriam assim o percurso para dizer “não à guerra colonial”. Muitos jovens de Portugal, terão aproveitado esta “motivação” que lhes trouxe a feliz consequência de progressão (também) no saber. Na sua grande maioria eram oriundos das grandes cidades e das zonas “mais evoluídas”. Ser jovem no interior do país e com ambições de “estudos avançados” implicava a migração - eventualmente de toda a família - para uma das três cidades com universidade. Para essa minoria, na maioria dos casos, os percursos eram bastante sinuosos (e aleatórios?) tornando difícil atingir o “grande objetivo familiar” – completar “um curso superior”. Nesta época, na língua portuguesa, as palavras mestrado e doutoramento tinham um significado não muito bem definido e, sempre se relacionavam com graus científicos “do estrangeiro”. No início dos anos 80 começaram a ser criados os Mestrados em Probabilidades e Estatística. Doutoramento em Portugal era um acontecimento raro e a especialidade de Probabilidades e Estatística não existia. Muito se avançou nos últimos 40 anos. Nesses anos – de pós-guerra colonial – Portugal iniciava-se pois, para um lugar na Ciência. É uma excelente referência temporal para se iniciar um memorial que – como “aquele outro” de enorme sucesso – também envolveu “homens e formigas” que levaram “isto daqui para ali porque as forças não dão para mais, e depois vem outro homem que transportará a carga até à próxima formiga, até que, como de costume, tudo termina num buraco...” (citando *Memorial do Convento* de José Saramago). Na década em referência, alguns obreiros, despertavam, tentavam transformar. Uns, com a ajuda de bolsas de estudo peregrinavam lá para fora, às vezes para bem longe, por outras universidades dando passos fundamentais. Outros, poucos e em muito menor número, avançaram “dentro de portas”, com diferentes dificuldades. Referindo uns e outros, de pioneiros estamos a falar!

Nos “idos anos oitenta” surgiu pois o despertar português para a investigação científica – nessa época incipiente (também) em estatística. Alguns históricos estatísticos portugueses – na sua maior parte incentivados por Tiago de Oliveira – juntavam-se à diáspora lusitana. Foram, com sucesso, até outras universidades aprender; para até nós trazer ciência estatística. Outros – e neste grupo me integro – pelas razões mais diversas, decidiram ficar e, acumulando ensino e investigação, também ajudaram a sedimentar novos cursos universitários ajudando a Matemática Aplicada a “dar à luz a Estatística” nas Universidades Portuguesas e aqui, é justo referir o protagonismo da Faculdade de Ciências da Universidade de Lisboa. Uns e outros, alguns anos mais tarde, congregavam esforços para alcançar “novas perspetivas” científicas em Portugal, também na ciência estatística que, entretanto, vinha sendo implantada a partir de pioneiros como Tiago de Oliveira, Bento Murteira e outros.

Portanto, “aconteceu o acaso” que seria a génese da moderna Estatística em Portugal. Um primeiro grande fruto, em 1980, foi a fundação da Sociedade Portuguesa de Estatística – SPE, (durante alguns anos incluindo também a Investigação Operacional) com a designação de Sociedade Portuguesa de Estatística e Investigação Operacional – SPEIO². No início da década de 90, Ivette Gomes liderou o grupo que sedimentaria a SPE – associação científica de onde a Investigação Operacional se tinha separado pois entretanto tinha sido criada a APDIO – Associação Portuguesa para o Desenvolvimento da Investigação Operacional, onde os investigadores dessa área se congregaram.

Decisão – o fundamental

Chegamos assim, à Teoria da Decisão Estatística. É nela que se baseia e onde (cada vez mais) está a génese da “profissão estatístico”. E, por consequência a possível liderança (pelo menos numa equipa)

Sobre este tema – de facto, teoria fundamental para tudo e para todos – respigamos algumas ideias basilares de Murteira (1988) (Cf. Nota 1). “Apesar de o homem ser chamado diariamente a tomar

² Mais desenvolvimento em Rosado, F. (2005) – *Memorial da Sociedade Portuguesa de Estatística*.

Escolar Editora, p. 139 e seguintes. Disponível em <https://www.spestatistica.pt/publicacoes/publicacao/memorial-da-sociedade-portuguesa-de-estatistica>

decisões só muito recentemente os problemas que estas suscitam começaram a ser tratados segundo uma ótica científica" (*ib*; p. 97). Historicamente, "a teoria da decisão estatística deve-se essencialmente a A. Wald que seguindo a tradição de Neyman-Pearson alargou consideravelmente os horizontes abertos por estes, tirando partido do desenvolvimento da teoria dos jogos realizada por von Neumann e Morgenstern. O grande mérito de Wald (...) consistiu na contribuição para uma argumentação onde (...) em termos gerais, os procedimentos clássicos são casos particulares da decisão estatística" (*ib*; p. 108-9). No entanto, para evitar confusões, importa que se clarifique que "(...) a teoria que vai desenvolver-se diz respeito à decisão individual e não de grupo. (...) a teoria que vai estudar-se não pretende substituir o decisor – mas, sim fornecer um conjunto de regras que auxiliem o decisor (...). Em termos gerais pode dizer-se que se está perante um problema de decisão quando se torna imperioso escolher ou optar entre, pelo menos, dois cursos de ação" (*ib*; p. 97). Mas, (haverá ainda quem considere que) a Estatística não é uma teoria? Poderemos admitir que se trata de "um instrumento ou alfaia cujas aplicações mais relevantes se situam, naturalmente, no domínio da investigação científica"? "Nos problemas de inferência estatística ou de decisão estatística trabalha-se, quase sempre, no quadro de um modelo probabilístico ou, pelo menos, com uma forte componente probabilística" (*ib*. p. 23).

Mas, o culminar da decisão exige uma travessia (do deserto, algumas vezes) onde as observações, os dados, a amostra são ingredientes de primordial importância. Chegamos a *Data Science*.

A necessidade de *outliers*³

Juntando esta componente, os modelos são muito importantes (fundamentais!) na investigação científica. Além disso, (também) na modelação surge de capital importância a amostra e cada um dos seus constituintes – (em particular) as observações e a sua dimensão. Na pesquisa de *outliers* numa amostra é, pois, de um problema de investigação que se trata. E, como tal, vai gerar (ou criar a necessidade de se apoiar em, pelo menos) uma Teoria! Alguns cientistas tomam Teoria como sinónimo de hipótese. Mas, a Teoria é distinta da hipótese e da ciência como sistema total. Porque, no método científico, a hipótese é uma fase anterior à Teoria e é parte integrante da ciência, teórica ou aplicada. A Teoria opõe-se à *praxis* ou ação mas são complementares. Como sabemos, o método científico atravessa várias fases: observação e experiências, hipótese e formulação da lei geral ou Teoria. Esta é pois uma hipótese já confirmada pela experiência e que faz parte integrante da Ciência. Mas há vários tipos de Teoria e de Ciências. No entanto, só são possíveis dois tipos basilares de Teoria: as dedutivas e as indutivas. Nas dedutivas, existe um conjunto de proposições válidas ou verdadeiras (teorema) que se constrói a partir de um grupo de proposições primitivas (axiomas) pela aplicação de certas regras de inferência. Nas indutivas temos um conjunto de proposições verdadeiras ou prováveis (teoremas, axiomas e definições) que se elabora a partir de vários casos particulares por um processo de inferência imediata e generalizante. Para muitos pensadores a conclusão da indução é apenas uma probabilidade. E, a "probabilidade de uma lei" cresce com o número de casos que a confirmam. Se, por diversos testes "confirmamos um *valor discordante*" temos o conjunto, isto é, o suporte das regras que podem formar – uma ou a – teoria dos *outliers*.

Uma das formas mais gerais de definir Estatística é aquela que a considera um método de proceder à escolha de alternativas de ação em face da incerteza, pelo recurso à recolha e interpretação de dados sobre os fenómenos em estudo. Assim, nesta generalidade, a Teoria dos *Outliers* em Dados Estatísticos torna-se uma mais-valia para a Ciência Estatística.

Em conclusão – uma ou a – teoria dos *outliers* deve fornecer um conjunto de regras que auxiliem o decisor além de construir instrumentos que permitam avaliar a qualidade da decisão. Com esse desiderato devemos prosseguir! Mas, para crescer e se desenvolver, a teoria deve ser transmitida a outrem que (afortunadamente?) a expandirá com a sua contribuição!

A Fortuna⁴ / O Acaso decide!?

Admitamos uma análise de *outliers* em dados estatísticos. Dividamos – separemos – os dados em estudo em duas partes (dois e só dois grupos?): "os eleitos", que admitimos em maior número – a maior parte – e "os suspeitos". Estes (sempre presentes?!) estão em muito menor quantidade pois vulgarmente⁵ consideramos apenas um ou dois valores discordantes. Não há qualquer motivo assinalável para aquela

³ Breve síntese deste tema. Desenvolvimento em Rosado (2006), *Outliers em Dados Estatísticos*, Edições SPE

⁴ A deusa romana que, tal como a congénere grega Tyche, fazia a seu capricho, a felicidade e a infelicidade.

⁵ Também por razões científicas!

escolha. No entanto, ela é (quase sempre) feita. A confirmação está na (cada vez maior) utilização dos testes de *outliers* nos mais diversos domínios científicos e pelos diferentes especialistas também nas aplicações, inclusive dos pacotes estatísticos. Todos desejam melhorar a qualidade dos seus trabalhos e conclusões através de uma "purificação dos dados". Mas, poderão os suspeitos conter mais e melhor informação que os eleitos? Qual é a causa que faz eleger estes em detrimento daqueles? Porque é que os suspeitos – a quem uma obscura fama esconde – não são os eleitos das análises estatísticas mais eloquentes? É que se trata de um assunto muito importante este de mostrar⁶ quais são "os verdadeiros"; embora possa não ser deles que nos vêm todas as (melhores!) pistas e, também, fazer ver todas as fragilidades das conclusões. Divididos os dados estatísticos entre eleitos e suspeitos, devemos questionar "quem é" ou "porque é" eleito. Quem atribui essa "condição estatística"?

O Acaso⁷ é a única coisa que não acontece por acaso.

A Estatística⁸ é muito antiga mas tem uma história curta. Só entrou nas academias de todo o mundo⁹ no segundo quartel do século XX e o principal arquiteto foi Fisher – justamente apelidado de fundador da moderna ciência estatística. Com certeza, a Fortuna é senhora; é ela que, mais conforme o seu capricho do que conforme a justiça, assegura a todos "a eleição" ou, pelo contrário, "a obscuridade".

Mas, quantos eleitos e quantos suspeitos?

Se a Fortuna decide quem são os obscuros – aqueles que têm "a força" – porque não consegue essa honra para si própria – é caso para pensar? Porque sofreu uma fortuna adversa? Nesse caso ela, que nobilita os outros mas a si própria não pode nobilitar, é adversária de si própria! Os eleitos devem estar sempre em maioria? E, merecem mais atenção? Muitas vezes são escolhidos em razão da maior importância ou da contribuição para o estudo. E de entre os escolhidos, quais são "os mais fracos"?

São todos igualmente bons?

Qual é então a razão (ou a causa!) que obriga alguns eleitos a terem menor importância (por que estes devem existir!).

E a comparação entre os "eleitos menores" e "os "suspeitos"? Estes, já chamados *outliers*, podem ter muito mais valor. É um *outlier* que abre o acesso a uma análise estatística mais profunda. Essa pode ser a origem de um trabalho de excelência. É a primazia em qualquer estudo de *outliers* que pode fazer a diferença entre um estatístico e um utilizador da estatística. Assim também na investigação.

São os *outliers* – a quem uma obscura fama esconde – quem confere vida aos dados. Com certeza que o acaso é, quem mais assegura a cada estatístico a notoriedade ou a obscuridade. Não se julgue mais digno de honra, aquele que é agrupado nos eleitos. Mais fortes, são aqueles a quem o acaso – a mãe natureza – deu a condição de possuir muito mais informação (estatística). Para os eleger crie-se (pelo menos) uma teoria! A Fortuna / O Acaso também decide na investigação!?

A experiência pode tornar a investigação numa paixão. A investigação científica de base, tanto como a aplicada, constitui uma expressão significativa do domínio do homem sobre a natureza. Por sua vez, a investigação é *leitmotiv* na procura da Verdade. Esta pesquisa, por seu lado, estimula a grande reflexão (de sempre!) que deseja o concílio entre a razão e a fé que, muitas vezes, se questiona. Elas são as duas asas com as quais o espírito humano se eleva para a contemplação da verdade. Se a investigação científica proceder segundo métodos de rigor absoluto e se mantiver fiel ao seu objeto próprio não há oportunidade para qualquer dissensão. E, se a pesquisa é feita nos pequenos pilares – os de segunda linha – ainda mais fácil é a compreensão mútua dos "contendedores"; que não chegam a metas (supostamente) contraditórias.

Todas as grandes teorias começam com um pequeno passo e, muitas vezes, com base em dados (estatísticos). Pelo seu envolvimento em diversas matérias, a Estatística é transversal na Ciência e, por isso, a Investigação (em) Estatística é um grande pilar.

Na Investigação (em) Estatística, tal como noutra ciência, busca-se a verdade. A verdade? A procura é um caminho contínuo que se constrói com pequenos passos – indecisos e frágeis no princípio – mas,

⁶ É sempre a procura da Verdade que está em causa!

⁷ A propósito de *outliers*, revisitemos este tema. A frase é de Almada Negreiros (Cf. p. 125 de *Matemática e Cultura*. Furtado Coelho et al. 1992. Edições Cosmos). Já foi tema de conferência de Tiago de Oliveira (*ib*; p. 125-49). A estatística casa bem com o acaso e ambos criam necessidade. É um tema recorrente que também já gerou o título para uma edição SPE – *Estatística com Acaso e Necessidade*; Atas do XI Congresso Anual.

⁸ Neste contexto, leia-se a "pequena excursão" apresentada por Tiago de Oliveira (*ib*; p. 125-8).

⁹ Sobre este assunto são importantes os artigos de Efron em *Statistical Science* 1998, 13:95-122 e de Rao, em Rao e Székely

firmes quando a experiência e a solidez de conhecimento o permite. No entanto, “não há caminhos há que caminhar” (como diz a sabedoria da poesia) pois “dos caminhos apenas ficam os rastos...” e, assim também na Estatística.

Caminheemos!

À medida que se caminha – mais aqui e menos ali – surgem perguntas laterais sobre “objetivos”, “interesse” ou “valor” da “ciência produzida”. No final está sempre a “procura da verdade” e qual o seu valor. Decerto, as pequenas contribuições são sólidas porque facilmente “se controlam”, “assimilam” e “arrumam”. Assim, ocupam o devido lugar! As grandes, por sua vez, poderão ser mais frágeis pela vulnerabilidade nos pequenos suportes em que se apoiam. E são estes que, formando um todo, criam uma teoria. Como na Estatística, nas diversas áreas do conhecimento também existem *outliers* – desde a física até à metafísica. Em todas elas, grandes interrogações se podem formular sobre o tema Investigação. Pesquisando *outliers*, onde quer que se encontrem, criamos Ciência – em particular, Estatística. Por consequência, abrimos (mais) um caminho na pesquisa da Verdade.

É a força dos menores!¹⁰

Conclusão

Consolidada a importância da Estatística em todos os ramos e domínios da Sociedade – científica e não só – é fundamental que esse valor seja utilizado em favor da humanidade.

E, para isso, há que implementar e deixar atuar os meios e a metodologia natural. Uma sociedade governa-se por e com lideranças competentes.

A competência é condição *sine qua non* para uma boa liderança. Esta requer decisões. Aquela, por sua vez, pode ser revelada por um currículo¹¹.

A conjugação destes dois termos, devidamente implementada, seguramente, é um bem para a Humanidade.

“Competência, currículo, (capacidade para) liderança”, formam um “triumvirato” – em que cada triúmviro sustenta e faz crescer os outros dois.

Deste trio, apenas o currículo é aquele que “mais podemos controlar”.

A recente pandemia fez surgir ainda mais e consolidar ainda melhor a necessidade de clarificação na “definição” de “os especialistas” – em particular no campo da Estatística e da modelação. Controlado o risco do aparecimento de especialistas que de facto não o são – pois sem currículo nada feito – a sociedade científica deve aproveitar este elã para (ainda) mais afirmar a Estatística. Tudo isto passa por um bom currículo, o que é equivalente a um reforço dos planos de estudo para obtenção das qualificações nos diversos graus académicos para uma maior, uma excelente eficácia nos grandes e novos desafios que a sociedade moderna enfrenta.

Liderar é decidir. E toda a boa decisão exige trabalho em equipa. A Ciência Estatística, cada vez mais, é um dos elementos fundamentais. A qualidade dos seus profissionais, os estatísticos e o seu valor, como em quase tudo, está em relação direta com a excelência do seu desempenho.

A excelência, sem dúvida, é a melhor opção para esse objetivo. O futuro, que se constrói no presente, muito beneficiará.

Para consolidar e como reflexão, é muito importante a partilha de todas as contribuições por mais modestas que sejam ou que, como tal, possam ser consideradas.

O todo é muito mais que a soma das partes, as pequenas contribuições.



¹⁰ Rosado, F. (2014) – *Outliers: The Strength of Minors*. New Advances in Statistical Modeling and Applications (A. Pacheco, R. Santos, M. R. Oliveira e C. D. Paulino Editors) p. 17-27. Springer. *Studies in Theoretical and Applied Statistics*.

¹¹ Um currículo, mais do que um portfolio para determinada ação ou iniciativa é uma “história de vida” e que permite, nas mais diversas variáveis (detalhes) validar a maior parte da competência que se deseja apresentar.

Liderança Estatística

João A. Branco, jbranco1802@gmail.com

*Departamento de Matemática
Instituto Superior Técnico
Universidade de Lisboa*

1. Introdução

Quando em certa altura da história da humanidade houve a necessidade de viver em grupo notou-se que por vezes alguns indivíduos passavam a ser olhados e seguidos como exemplos e guias do próprio grupo, possivelmente em consequência das suas decididas intervenções e atitudes com vista a resolver problemas relacionados com as necessidades do grupo. Terá sido este o início de um longo caminho de séculos até surgirem pela primeira vez as palavras líder e liderança. Porém, o interesse em estudos sobre as ideias de líder e liderança só veio a verificar-se no século XX, tendo conhecido um grande crescimento a partir dos anos 90 até ao presente. A investigação e ensino deste tema está hoje muito presente na área académica e a literatura produzida neste período é extremamente vasta (Bass, 2008). Em tempos modernos e de globalização, os especialistas consideram que a liderança é vital para o funcionamento das organizações e da sociedade e que a necessidade de líderes existe em muitos aspectos da actividade humana. No endereço <https://www.apollotechnical.com/leadership-statistics/> podem consultar-se várias estatísticas que revelam a importância que as empresas dão ao tópico da liderança. Mas afinal o que é liderança e quais são as características que um líder deve possuir?

É difícil definir liderança de forma rigorosa e a prova é que há quase tantas definições como especialistas. Dada a nossa fraca cultura nesta área, elegemos uma ideia mais óbvia que consideramos ser útil para a segunda parte destas notas despretensiosas. Liderança é a capacidade de influenciar, motivar e conduzir um grupo de potenciais seguidores com vista a atingir um determinado objectivo. Entre as qualidades de um grande líder encontramos integridade, confiança, carisma, assim como capacidades de comunicação e de inspirar os elementos do seu grupo, de inovação, de criatividade e de abertura para a mudança. É preciso salientar que um líder (*does the right things*) não é um gestor (*does the things right*), uma vez que o líder não se limita a gerir o que existe mas prepara a sua acção para enfrentar a mudança que deseja efectuar. No século XIX pensava-se que o líder é alguém que nasceu para ser líder. No século XX outras teorias surgiram e uma delas considera que as características de um líder podem ser adquiridas e desenvolvidas. Isto é importante por revelar que as características de um líder podem ser aprendidas e consequentemente também ensinadas. É o que justifica a proliferação de cursos e programas universitários sobre liderança e sobre como formar um líder.

Na prática, sem as exigências de rigor da definição de liderança, não parece difícil identificar e apreciar lideranças, bem como reconhecer grandes líderes nos mais diversos campos, como no campo religioso, no campo militar, político, desportivo e muitos outros. A história também nos diz que um grande líder pode ser o salvador para uns, mas um vilão para outros. Especialmente em situações de crise, catástrofe ou conflito, olhamos para as lideranças como fonte de soluções e esperanças. As flagrantes crises da pandemia Covid-19 e a invasão da Ucrânia vieram mostrar como podem ser diversas as lideranças das várias instituições e personalidades envolvidas na reacção a esses fenómenos. E quando nos pomos a imaginar as intrépidas viagens marítimas que os portugueses fizeram no passado não podemos deixar de admitir que foram grandes líderes e lideranças que levaram as suas armadas de caravelas a descobrir o que havia para além do desconhecido.

Liderar é uma actividade multidisciplinar, no sentido em que o grupo de seguidores é formado por membros com diversidade de conhecimentos e de especialidades e certamente de condição humana. O

nosso interesse é especular sobre o interesse que a liderança em estatística pode trazer para o desenvolvimento da estatística.

2. Liderança Estatística

A liderança estatística não parece ter sido uma preocupação dos estatísticos ao longo dos tempos. Talvez porque os estatísticos tenham optado por uma vivência menos conspícua, de professores, investigadores e consultores. Muitos dos métodos estatísticos são identificados pelo nome do seu autor (função discriminante de Fisher, coeficiente de correlação de Pearson, teste de Wilcoxon, regressão de Cox e tantos outros) o que leva a imaginar que são produto de trabalho solitário, sem o envolvimento de equipas e seguidores que as lideranças consagram. Fala-se de grandes estatísticos, como grandes pioneiros e grandes cientistas, mas não se identificam grandes líderes. Porém, à luz do que foi definido anteriormente, não será difícil encontrar verdadeiros líderes em estatística. Para exemplificar lembramos três estatísticos que muito admiramos e que no nosso entendimento perfazem amplamente as condições de grandes líderes: Prasanta Mahalanobis (humanista, contribuição notável para a implantação e o desenvolvimento da estatística na Índia – pai da escola de estatística indiana, fundou o Instituto de Estatística Indiano, prémio internacional Mahalanobis em sua homenagem), John Tukey (estatístico de muitas ideias, visionário, inventor, pioneiro, deu início a uma nova era no desenvolvimento da estatística), Edwards Deming (o guru da qualidade, grande líder, popularizou a estatística, responsável pela melhoria da qualidade dos produtos no Japão, milagre industrial japonês – prémio Deming em sua homenagem). Todos estes estatísticos são autores de uma produção científica de grande relevo, trabalharam na universidade e em instituições exteriores à universidade e conduziram grupos de colaboradores que vieram a fazer escola e dar continuidade aos seus ensinamentos. Com isto queremos dizer que apesar de não haver uma cultura de liderança os líderes existem na área da estatística. É claro que alguma forma de liderança em estatística sempre existe, quer ao nível do ensino, em que o professor conduz os seus alunos, quer ao nível da investigação em que o investigador responsável conduz os colaboradores do seu, geralmente pequeno, grupo. Fica de fora a actividade de consultor de ocasião, que de acordo com as práticas que geralmente têm sido seguidas não encontra condições para a emergência de liderança. Nestas condições é importante trabalhar para que o consultor solitário possa vir a ser líder de grupo. E o caminho que leva a este objectivo já foi apontado anteriormente. O pretendente a líder deve aprender a ser líder, isto é, deve fortalecer-se com as características necessárias de forma a influenciar o grupo para o uso convincente das melhores práticas estatísticas na procura dos objectivos dos projectos em estudo. Na prática deve contar com as fontes de formação disponíveis que o possam ajudar.

Uma área em que as condições de trabalho podem proporcionar, com alguma naturalidade, a emergência de liderança estatística é a área dos ensaios clínicos, cuja temática nos é familiar em virtude da nossa ligação ao ensino desta matéria. O ensaio clínico é um processo longo e complexo que tem em vista avaliar a eficácia e segurança de um tratamento ou medicamento sujeito a restrições orçamentais e de exposição dos doentes ao risco. Como um aspecto fundamental do ensaio clínico é a análise dos dados observados em doentes dir-se-ia que esta é a tarefa a realizar por um estatístico credenciado. Em verdade a intervenção do estatístico é bem mais extensa, é muito mais do que apenas analisar dados. A sua presença é necessária em todas as fases do processo, onde se incluem: a preparação antes do ensaio, o planeamento, a execução, a análise dos dados e a interpretação e divulgação dos resultados. As decisões que é preciso tomar no decurso do ensaio dependem muito da opinião preciosa do estatístico, opinião que ele constrói a partir da visão global que mantém ao longo do andamento do estudo. Esta omnipresença coloca o estatístico numa posição de grande oportunidade para estabelecer contacto com os restantes elementos da equipa responsável pelo ensaio clínico, uma equipa multidisciplinar que envolve investigadores, gestores, médicos, enfermeiros e outro pessoal de apoio. Liderança requer um conhecimento profundo dos objectivos do ensaio e um envolvimento com os elementos da equipa, não só do ponto de vista das suas competências como também do ponto de vista humano. O estatístico terá que fazer perguntas, para depois dar respostas claras que possam: i) mostrar as verdadeiras potencialidades da estatística, ii) ajudar a melhorar a cultura estatística dos elementos do grupo iii) estimular a atracção e apreço dos elementos do grupo pela estatística, iii) promover a ideia de que não é possível realizar um ensaio clínico nem fazer investigação médica sem o recurso à estatística. O sucesso a caminho da liderança passa por jogar com as características da liderança já anunciadas e neste caso

particular o saber comunicar a acção dos métodos usados e a interpretação dos seus resultados e ser tolerante ao esclarecer os elementos do grupo, devem ser especialmente valorizados. Vemos assim que a sabedoria da liderança pode promover a estatística. Imaginamos que semelhantes contextos aos dos ensaios clínicos possam existir no mundo empresarial e noutros ambientes, mas isso nem sempre acontece. A liderança faz sentido quando é necessária e há condições para que ela exista.

3. Discussão

As fortes e rápidas transformações que têm vindo a ocorrer nos vários sectores da sociedade terão acelerado o interesse e o desenvolvimento do tema liderança. A expectativa é que a acção de bons líderes se traduza numa mais-valia para as empresas e demais sectores de actividade que enfrentam mudanças provocadas pela introdução de inovações induzidas pela revolução tecnológica. Assim acontece com a estatística onde a introdução da liderança, novidade chegada há muito pouco tempo, tem a aparente finalidade de ajudar a reconhecer o valor da estatística e aumentar a sua visibilidade. A literatura sobre liderança em estatística encontrada é muito escassa. No trabalho que achámos mais relevante (Gibson, 2019) encontramos, na lista de 57 referências, apenas 2, ambas de 2012 e com duas páginas cada uma, que incluem nos seus títulos a palavra liderança, aplicada à estatística num caso e à bioestatística no outro. Mais duas contribuições com interesse são (Snee and Hoerl, 2004) e (Golbeck, 2021). E recentemente realizou-se no dia 14 de Fevereiro 2022 o encontro “*ISI side event; Statistical Leadership – The Role of National Statistical Institutes*” por ocasião da 53ª Sessão da Comissão de Estatística das Nações Unidas. Este acontecimento teve por objectivo contribuir para o desenvolvimento das lideranças ao nível dos Institutos Nacionais de Estatística.

Não há dúvidas, como se viu no exemplo do ensaio clínico, quanto à contribuição da liderança para promover o valor da estatística e a sua visibilidade, desde que a sua implementação seja possível e faça sentido. Mas convenhamos que a generalização desta parceria não deve ser fácil de implementar pois o ambicionar ser líder é de cunho pessoal e há também o historial da atitude dos estatísticos que parece revelar pouca atracção por liderança.

Contudo, como refere W. Edwards Deming: “*It is not necessary to change. Survival is not mandatory*” Quanto à visibilidade da estatística não temos que nos preocupar pois a estatística está em todo o lado, nos vários países, nas instituições, nas nossas casas e nas nossas vidas todos os dias. Os estatísticos é que andam por aí e fazem falta onde não estão, mas poderiam estar. Ao não deixarem a sua zona de conforto permitiram uma fragmentação da estatística, bem aproveitada por aqueles que se mantiveram atentos às transformações que vieram com as novas tecnologias, que é prejudicial à disciplina. Outros males têm atingido a estatística que permanece constantemente preocupada com o seu futuro (Branco, 2012). No nosso entender um repensar profundo sobre a maneira como a estatística é ensinada poderia trazer benefícios para todos. Por agora a estatística continua a ser abusada e é acusada de má reputação. Os estatísticos devem lutar pela irradicação de tão injusta classificação para que um dia não seja mais possível ouvir impropérios que tanto nos entristecem:

“*La estadística, otra mas que nos egaña*” (provérbio espanhol)

4. Referências

- Bass, B. M. (2008). *The Bass handbook of leadership: Theory, research, and managerial applications* (4th ed.). Free Press, New York.
- Branco, J. A. (2016). O Futuro da Estatística. *Boletim da SPE, Outono*, p. 49-58.
- Gibson, E. W. (2019). Leadership in Statistics: Increasing Our Value and Visibility. *The American Statistician*, **73:2**, 109-116.
- Golbeck, A. L. (2021) (ed.). *Leadership in Statistics and Data Science: Increasing Our Value and Visibility*. Springer Nature Switzerland AG.
- Snee, R. D. and Hoerl, R. W. (2004). Statistical Leadership. *Quality Progress*, October, 83-85.



Liderança da Estatística ou Importância da Estatística?

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

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

O panorama das *Ciências* em fins do século XIX, “positivisticamente” retratado na famosa classificação de Augusto Comte (1798-1857) — *Matemáticas* (plural interessante, que deve conter a *Probabilidade* e a *Estatística* incipiente da altura), *Astronomia*, *Física*, *Química*, *Biologia* e *Sociologia*, sendo a *Filosofia* o resíduo da indagação sobre o mundo que não se enquadra nesse naipe de ciências. Nessa classificação hierárquica (em que com exceção das *Matemáticas*, cada *Ciência* para além de incorporar os princípios das que a precedem ganha estatuto autónomo por necessitar de leis e princípios independentes) não havia lugar, então, para *Economia*, *Psicologia*, *Genética* ou mesmo *Medicina*. Creio que atualmente quase ninguém conseguiria recitar sem falhas a listagem labiríntica das centenas de *Ciências* arroladas em

https://en.wikipedia.org/wiki/Index_of_branches_of_science

ou mesmo a lista mais breve em

https://en.wikipedia.org/wiki/Branches_of_science.

É neste contexto que a *Genética* começa a afirmar-se, com Francis Galton (1822-1911) e o seu protegido Karl Pearson (1857-1936), sendo também uma área de formação de Sir Ronald Fisher (1890-1962). *The Grammar of Science*, de Karl Pearson (1927) tem forte impacto na modernização das ideias sobre metodologia da investigação científica, nomeadamente por propor a correlação como conceito mais abrangente do que o de causalidade, e teve decerto importância no reconhecimento de que ciências experimentais deviam usar o acaso como aliado e não como inimigo, sendo a amostragem (Kiaer, 1895) um recurso inevitável na investigação em *Ciências Experimentais*, e essencial no *Controle da Qualidade* que é uma das bases da indústria moderna.

E é indubitável que um dos maiores progressos da *Ciência* na primeira metade do século XX foi a criação do planeamento experimental, um dos títulos de glória de Sir Ronald Fisher. A minha memória já não consegue fornecer-me pistas credíveis sobre quem, à sombra de gigantes como Karl Pearson, Sir Ronald Fisher, e William Edwards Deming (1900-1993), futurava que a *Estatística* seria líder das *Ciências* na segunda metade do século XX, desiderato que é um possível bastardo de *Mathematics, Queen and Servant of Sciences*, junção das duas obras de Eric Temple Bell (1883-1960) em 1931 e em 1937.

Claro que antes dos anos 50 não seria previsível que a dupla hélice de Watson e Crick (1952), dariam um protagonismo então inesperado à *Biologia*, partilhado pelas *Ciências da Computação* (confesso que não tenho conhecimentos suficientes para saber se a frase deve ficar como está, ou se devia pôr antes *Ciências da Vida*, *Genética*, *Informática*, *Robótica*, *Telemática*, não estando a injustiçar *Ciências do Ambiente*, *Energia* e *Climatologia*, e tenho que assumir a benevolência de quem ler).

Tudo isto (e muitas mais coisas) me levam a crer que tal como o mundo a *Ciência* é composta de mudança, e que a ideia de liderar em *Ciência* não me atrai. A meu ver o que é saudável é ser um parceiro credível no empreendimento coletivo que é o progresso científico (e apetece-me, aqui, recomendar a leitura de [A History of Probability and Statistics and Their Applications before 1750](#) seguido de [A History of Mathematical Statistics from 1750 to 1930](#) de A. Hald, que tão lucidamente inscreve o desenvolvimento da *Estatística* no diálogo com as outras *Ciências*).

Penso por isso que a *Estatística* tem dois aliados (profissionais e amadores) e dois inimigos (profissionais e amadores), e que o importante muito mais do que liderar, é incentivar os aliados, e tentar “converter” quem deslustra a *Estatística* pela demonstração de que os abusos da *Estatística*, nomeadamente pelo uso acríptico de *software* que não deixa de responder a perguntas inadequadas ou mal formuladas, são tão frequentes e insípidos que nem têm o atrativo de merecer uma gargalhada. Devemos na realidade estar atentos e lutar contra o abuso da má utilização da Estatística.

Contudo, tal como ‘o sonho é uma constante da vida’, também o acaso é uma constante da vida. E sendo a *Probabilidade* e a *Estatística* as *Ciências do Acaso*, penso que serão fundamentais no século XXI e em todo futuro da humanidade, uma vez que se trata de *Ciências* transversais a todas as áreas do saber. Peter Diggle (2015) refere que “*Statistics is a data science for the 21st century*”, mas para mim a *Estatística* sempre foi muitas vezes intitulada, mesmo em aulas dos anos 80, como sendo a *Ciência dos Dados*. E a ascensão recente da chamada *Ciência dos Dados*, muitas vezes vista como uma ameaça potencial à *Estatística*, deve ser usada como uma oportunidade de enfatizar a relevância universal da *Estatística* para a interpretação correta dos dados. A importância da Estatística está pois assegurada nos anos vindouros, e a recente pandemia veio mais uma vez mostrar a importância da *Estatística* e da *Literacia Estatística*. Mas temos de estar à altura dos desafios que se avizinham. E reconheço ser necessária uma liderança estatística muito mais forte do que era há uns anos atrás para orientar equipas multidisciplinares na conceção e planeamento da investigação científica e na tomada de decisões robustas e fiáveis com base em dados observados.

Referências

1. Bell, E.T. (1931). *The Queen of the Sciences*, Stechert.
2. Bell, E.T. (1937). *The Handmaiden of the Sciences*, Williams & Wilkins.
3. Bell, E.T. (1988). *Mathematics, Queen and Servant of Sciences*, Reprint, ISBN-13:0883854471.
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, 793-813.
5. Hald, A. (1998). *A History of Mathematical Statistics from 1750 to 1930*. Wiley Series in Probability and Statistics, 1st Edition, ISBN-13: 978-0471179122.
6. Hald, A. (2003). *A History of Probability and Statistics and Their Applications before 1750*. Wiley-Inter-science, ISBN-13: 978-0471471295.
7. Kiaer, A.N. (1895-1896), Observations et expériences concernant des dénombrements représentatives. *Bulletin of the International Statistical Institute*, IX, Book 2, pp. 176-183.
8. Pearson, K. (1982). *The Grammar of Science*. Adam and Charles Black, London.
9. Watson, J.D., and Crick, F.H.C. (1952-1953). Molecular Structure of Nucleic Acids. *Nature* **171**, 737-738.



Ser ou não ser um líder (estatístico)?

Nuno Sepúlveda, *N.Sepulveda@mini.pw.edu.pl*

*Faculty of Mathematics & Information Science, Warsaw University of Technology, Varsóvia, Polónia
& CEAUL – Centro de Estatística e Aplicações da Universidade de Lisboa, Lisboa, Portugal*

Introdução

De tempos a tempos é assim: recebo um correio eletrónico do Prof. Fernando Rosado acerca do Boletim da SPE. No passado, o propósito desse contato foi um convite para contribuir com um artigo em temas tão variados como Investigação Estatística, Estatística Bayesiana, Estatística em Genética ou Tema Central da Estatística. Desta vez, o convite tinha outro cariz. Lançava-me um desafio inédito: que tal eu propor um tema interessante ou, melhor ainda, diferente para o próximo boletim?

Não hesitei por um momento: liderança estatística. É um tema que me tem ocupado nos últimos 5 anos por razões que serão mais evidentes mais à frente. É um tópico que corre subliminarmente ao longo da carreira de todos nós estatísticos. E as implicações de não pensarmos nesse tópico podem ser grandes, enormes ou catastróficas, se quisermos colocar algum dramatismo a este assunto.

Mas afinal de contas, o que é liderança estatística, nomeadamente, num contexto multidisciplinar, seja este empresarial ou académico? Eric Gibson [1] define-a como a capacidade que um estatístico tem em influenciar (sem autoridade própria) o curso, a estratégia e as decisões de uma equipa multidisciplinar. O mesmo autor propõe que esta liderança estatística é resultante de três capacidades ligadas entre si: (i) a de ouvir ativamente (ou atentamente); (ii) a de formar uma rede coesa de colaboradores; e (iii) a de comunicar eficazmente com os seus pares. Já eu aproveitei para discutir com algum detalhe estas capacidades num outro artigo de opinião [2].

A própria liderança estatística pode também ela ser individual, institucional ou política. A liderança individual é exercida por estatísticos ao nível de equipas de investigação. Por seu lado, a liderança institucional é seguida por indivíduos que pretendem a adoção de determinados procedimentos estatísticos ao nível de uma organização ou instituição. Por fim, a liderança política é adotada por estatísticos que, trabalhando em instituições governamentais, visam aconselhar mudanças de políticas baseadas na evidência existente. Neste artigo irei dedicar-me à liderança estatística ao nível individual. É o nível de liderança que conheço melhor e que me sinto mais habilitado a discutir e debater. É também o nível em que a maior parte de nós estatísticos atua. Para uma discussão mais detalhada sobre os outros níveis de liderança estatística, vide o artigo de Gibson [1].

Quando penso em liderança estatística, há uma questão primordial subjacente muito ao estilo de Shakespeare: ser ou não ser um líder estatístico? Será uma escolha por necessidade? Para responder a esta questão, trago a minha experiência pessoal como suporte da discussão (vide o sumário na Figura 1).

No Instituto Gulbenkian de Ciência: ser estatístico ou cientista?

Terminei a Licenciatura em Matemática e Computação, ramo de Probabilidades e Estatística, no Instituto Superior Técnico faz já mais de duas décadas. Tive o prazer de ser orientado pelo Prof. Daniel Paulino no meu trabalho final de curso cujo tema foi “Métodos Inferenciais Exactos em Dados Categorizados”. Por essa altura, intitulava-me estatístico com todo o orgulho e brilho que essa palavra poderia ter, o que contrasta com o panorama atual em que os mais jovens preferem ser chamados de cientista de dados. Assim intitulado, candidatei-me a uma bolsa de investigação científica – as infames BIC que ainda hoje subsistem – no grupo de Dr. Jorge Carneiro sediado no Instituto Gulbenkian de Ciência (IGC) em Oeiras. Não queria ser apenas um estatístico, queria ser um bioestatístico com “B” maiúsculo! Quando ingressei

no instituto, sentia-me sedento de adquirir conhecimento biológico de base que julgava que me faltava para melhor selecionar as técnicas estatísticas a usar nos meus trabalhos. As primeiras dificuldades surgiram: jargão biológico ininteligível, inglês como língua oficial do instituto, e a integração num grupo multidisciplinar e instituto onde toda a gente parecia saber mais do que eu. Prevalcia em mim um profundo sentimento de burrice que equiparo ao mesmo sentimento quando se entra na Universidade vindo do Ensino Secundário. Lembro-me que, por esses tempos, era o único com formação em Estatística a habitar o IGC. Era um verdadeiramente estranho no campus maravilhoso de Oeiras, uma experiência que contei num outro boletim da SPE [3]; como o tempo passa rápido!

Essa experiência inicial foi tornando-me paulatinamente mais forte em Estatística Aplicada através de solicitações de curta duração sobre o teste estatístico a usar num determinado conjunto de dados, ou quantos indivíduos deviam ser incluídos num estudo. Estas solicitações tinham o condão de serem facilmente integradas na minha rotina diária como bolsheiro. E assim sendo, chegaram os primeiros artigos publicados, que coloquei religiosamente no meu currículo, tal como se fossem os troféus dos *Grand Slams* que o Roger Federer, Rafael Nadal e Novak Djokovic tanto gostam de colecionar. Inevitavelmente trouxe-me também a estranha sensação de muitas vezes não ser ouvido por quem me vinha procurar. Passado este tempo todo, consigo entender melhor essa sensação e julgo que resultou da minha falta de experiência nas tais capacidades que Gibson [1] escreve no seu artigo de serem tão essenciais em liderança estatística. Tinha uma autoridade técnica (em Estatística) muito acima dos meus colaboradores que eram muitas vezes investigadores principais, mas ao mesmo tempo, tinha pouca autoridade científica. As minhas artes em escutar atentamente e de comunicar efetivamente eram limitadas para não dizer reduzidas. Em particular, a forma com que me expressava era muito próxima aos ensinamentos que recebi em Estatística e Matemática, o que tornava uma conversa com os meus colegas e colaboradores biólogos, bioquímicos ou físicos, uma tarefa extremamente desgastante para ambas as partes. Nesse aspeto, tive a sorte que o Dr. Jorge Carneiro acreditou sempre no meu potencial como cientista e puxou por mim para que eu não fosse apenas um *estatístico-consultor*, mas sim um *estatístico-cientista*; na visão de Irwin Bross [4], um estatístico-consultor visa apenas satisfazer as necessidades do cliente o que tem as suas vantagens, tais como pouca responsabilidade científica e a neutralidade em situações de conflito, mas desvantagens a longo prazo, tais como a falha em desenvolver uma identidade própria e não seguir um propósito científico específico. O conselho que ele sempre me deu foi de aproveitar as minhas colaborações estatísticas para aprender Biologia e Medicina por um processo que eu agora chamo de “osmose científica”. Aconselhou-me em aceitar qualquer oportunidade para apresentar os meus resultados, não só em conferências ou congressos, mas também em seminários internos ou em outras instituições. A ideia dele era a de melhorar o meu inglês falado e a minha capacidade de adaptar a minha forma de comunicar independentemente do público em questão, seja este mais virado para a Medicina, Biologia ou Estatística.

Ao mesmo tempo, o Prof. Daniel Paulino continuou a ligar-me à área da Estatística, sugerindo-me participações nos congressos anuais na Sociedade Portuguesa de Estatística e em conferências internacionais tais como o Joint Statistical Meetings em Minneapolis no ano de 2005 ou a reunião magna do International Statistical Institutes em Lisboa durante o verão de 2007. E assim, fui-me tornando cada vez mais fluente tanto em inglês como na linguagem científica (estatística ou biológica) e, com esta maior fluência, fui também criando a minha própria rede de colaboradores.

Após um período de quase 10 anos no IGC, tenho o orgulho de escrever que terminei uma tese de mestrado em Matemática Aplicada com um tema em genética humana sob a orientação do Prof. Daniel Paulino e do Dr. Carlos Penha-Gonçalves e uma tese de doutoramento em Ciências Biomédicas sob a orientação do Dr. Jorge Carneiro, tese essa em que tentei compreender como o sistema imunitário evita respostas nefastas ao próprio organismo. Acima de tudo, ganhei experiências interpessoais tão importantes como saber falar com investigadores principais e convencê-los a mudar de teste estatístico e até dar conselhos fora da minha competência técnica. Contudo, estava ainda longe de pensar em liderança estatística ou mais precisamente, de criar um grupo de investigação onde eu pudesse liderar o seu destino. Talvez a razão para tal ausência de pensamento estaria relacionado com um facto que acho dramático na vida científica em Portugal: a precaridade e instabilidade do trabalho. Uma vida de saltapocinhas de bolsa em bolsa não era muito atrativa para mim, principalmente, quando a componente monetária parecia garantir apenas a sobrevivência. Infelizmente esta situação continua a verificar-se no nosso país com a atualização do valor das bolsas praticamente inexistente. Por isso, acho que, no meu

subconsciente, evitei pensar em liderança de qualquer espécie visto a situação profissional dos investigadores principais do IGC não me agradar em nada.

Na London School of Hygiene & Tropical Medicine: ser ou não ser líder?

Exatamente pela razão da precariedade e instabilidade laboral, houve um momento após o fim do doutoramento em que hesitei se devia ou não continuar a trabalhar na academia ou mesmo em Ciência. Por acontecimentos totalmente aleatórios da vida, fui contratado pelo Dr. Jorge Félix da Exigo Consultores, após ter rejeitado um pós-doutoramento em Genética Humana com a Prof. Heather Cordell na Universidade de Newcastle no Reino Unido. A minha ideia na altura foi testar a minha adaptação a uma vida empresarial. A Exigo Consultores parecia uma boa opção para esse teste: uma empresa virada para consultadoria e investigação em Economia da Saúde e com uma forte componente e cultura estatística entre os seus membros. Fui acarinhado por todos desde o primeiro dia. Contudo, o apelo à investigação e à vida académica foi mais forte quando recebi um correio eletrónico do Prof. Taane Clark da London School of Hygiene & Tropical Medicine (LSHTM) a convidar-me para um pós-doutoramento em Genética Estatística e Epidemiologia Genética na área de Malária. E assim, a minha experiência na Exigo durou não mais do que 6 meses, tendo rumado de Lisboa para a grande metrópole londrina no final de Outubro de 2010.

E tanta coisa aconteceu desde esse primeiro dia no escritório em Keppel Street junto ao Museu Britânico (de Arqueologia). A minha primeira impressão foi uma tremenda atividade da Escola que parecia estar em consonância com o ritmo frenético da cidade de Londres, principalmente, na sua zona central. Aprendi também o que era o *Central Point* (um edifício altíssimo junto à estação de metro de Tottenham Court Road), os *West and East Ends* onde os códigos postais passavam de WC (West Central) para EC (East Central), a *Windham Street* onde os célebres Led Zeppelin tiraram umas fotos e as sexta-feiras com *fish-and-chips* e ida ao *pub* depois do trabalho. Sentia-me deslumbrado com este mundo em meu redor e em que todos pareciam ser de longe e estar de passagem.

Em relação ao trabalho em si mesmo, tinha-me juntado à LSHTM mas oficialmente era membro de um consórcio chamado MalariaGEN, que também já falei um pouco neste boletim [5]. Este consórcio estava sediado no *Wellcome Centre For Human Genetics* da Universidade de Oxford e integrava inúmeros parceiros espalhados pelos Trópicos. A minha responsabilidade principal (como pós-doutorado) era a de apoiar estes parceiros em termos de análise estatística, nomeadamente na realização de estudos de associação genética em Malária. Tal como discutido por Irwin Bross [4], a minha responsabilidade era semelhante à de um estatístico-consultor que visava satisfazer as necessidades estatísticas de alguns parceiros. Foi assim que conheci e trabalhei em colaboração estreita com a Dra. Valentina Mangano da Sapienza Università di Roma que tinha um estudo genético nas populações no Burkina Faso, do Alphaxard Manjurano da Kilimanjaro Christian Medical Centre na Tanzânia, do Dr. Boubacar Maiga da Université des Sciences, des Techniques et des Technologies de Bamako no Mali, para nomear alguns. Percebi então que a minha rede de colaboradores aumentava exponencialmente devido à minha integração na MalariaGEN. Percebi também que os investigadores britânicos davam grande importância à criação e manutenção destas redes internacionais, idealmente, de cariz global. No contexto de malária propriamente dito, a criação e manutenção dessas redes advém de uma necessidade essencial dos investigadores britânicos em garantirem acesso aos dados recolhidos no terreno, uma vez que não há transmissão autóctone desta doença no Reino Unido. Num contexto mais general, a existência dessas redes pode ser vista como uma reflexão de políticas fortes do governo britânico no âmbito de ajuda externa aos países pertencentes à *Commonwealth*, para assim manter ou até aumentar a influência política, económica e social nesses mesmos países. Deixando a política de lado, a minha experiência na MalariaGEN trouxe-me imensas oportunidades para desenvolver as minhas competências colaborativas com investigadores de diferentes personalidades e de origens geográficas distintas. Nesse aspeto, concluí que os investigadores europeus tentam muitas vezes aprender e perceber os aspetos essenciais do exercício estatístico em questão, talvez para ganhar mais autonomia no futuro. Nalguns casos extremos, estes investigadores tendem a ter uma opinião prévia bem vincada sobre a metodologia estatística a usar, o que pode ser problemático se a opinião não é partilhada pelo estatístico de serviço, neste caso, eu. Nessas situações, há que saber negociar previamente as análises a realizar de forma a evitar conflitos no futuro (ou seja, quando for a fase de publicação). Por seu lado, os investigadores africanos tendem a encarar a colaboração ao estilo clássico de consultoria estatística onde eles são o cliente e nós os

estatísticos somos o vendedor, mas sob o jugo de uma grande confiança no trabalho estatístico realizado. O que também pode ser mau se isso é indicativo de uma ausência de espírito crítico. Em termos práticos, uma maior capacidade colaborativa conduziu-me a uma explosão de artigos publicados que enriqueceu o meu currículo substancialmente. Contudo, a minha própria investigação foi secundarizada, limitando-se a um artigo metodológico na área da Genómica [6].

Ditou o destino que a minha ligação à MalariaGEN cessasse em finais de 2012. Ainda como pós-doutorado, integrei seguidamente o grupo do Prof. Chris Drakeley na LSHTM. Agora o foco da investigação era Epidemiologia em Malária, nomeadamente, num contexto de eliminação da doença e usando dados serológicos. Mais uma vez, fiquei estupefacto com a quantidade astronómica de colaboradores que o grupo tinha espalhado pelo mundo inteiro. Tornei-me no bioestatístico de serviço e essa responsabilidade trouxe-me mais um grande número de publicações em colaboração nos anos que se seguiram.

Em 2014, achei que era tempo de pensar em voos mais altos e, por isso, pedi uma avaliação interna do meu currículo para uma eventual promoção de Pós-Doutorado a Professor Auxiliar (naquela altura, chamava-se *Lecturer*, agora chama-se *Assistant Professor*). Alguns meses depois, ouvi do Reitor da minha Faculdade, o Prof. Simon Croft, que a avaliação do currículo tinha sido positiva. Fiquei muito contente com a notícia, pois tinha escalado mais um degrau da minha carreira académica. Obtive também um aumento de salário que me trouxe alguma folga orçamental na vida extremamente dispendiosa da cidade de Londres. Neste ponto, introduzo um parêntesis para explicar que a grande maioria dos investigadores da LSHTM é contratada em regime de *soft-funding*. Ou seja, os investigadores são contratados por períodos de curta duração no contexto de um determinado projeto. Os contratos vão sendo renovados sucessivamente, de projeto em projeto, até a própria escola contratar o investigador. Se por algum motivo o financiamento dos projetos termina, o investigador é forçado a *ir pregar para outra freguesia* (ou em momento de inspiração poética inglesa, *to leave for greener pastures*). Nessa altura, não me apercebi das implicações a médio e a longo-prazo deste tipo de gestão da LSHTM para Professores Auxiliares como eu. Estava tão contente com a notícia da promoção e tinha ainda todo esse otimismo que sempre me caracterizou tal como os artigos que tinha publicado até então para continuar a pensar no contínuo progresso da minha carreira.

A boa maré continuou nos anos seguintes: mais artigos publicados em colaboração ou como autor de correspondência; convites anuais para dar aulas de Bioestatística no IGC e em Cabo Verde no âmbito da Programa Gulbenkian para a Cooperação e Desenvolvimento; convites para dar apresentações em reuniões de trabalho sobre a aplicação de serologia em Malária. Numa das ocasiões tive o privilégio de visitar o *Center for Diseases Control and Prevention* (CDC) nos arredores de Atlanta nos Estados Unidos. Nessa altura (no ano de 2014) pairava no ar uma terrível ameaça de uma pandemia pelo vírus Ébola, que felizmente não se confirmou devido aos esforços tremendos da Organização Mundial de Saúde e dos governos dos países africanos que estavam sob essa ameaça. Quando declarei no controlo de passaportes que vinha ao CDC em trabalho, o funcionário americano disse-me entusiasmado: “*Oh, you’re one of the Ebola guys, right?!?*”. Ao que respondi: “*I am just a malaria guy*”. A conversa prosseguiu por mais uns minutos sob a batuta do funcionário que deseja saber mais sobre os novos avanços em Malária.

Apesar destes momentos de boa memória havia uma pequena pedra no meu sapato que inquietava o meu subconsciente. Todos os anos submetia um trabalho à reunião magna em Malária organizada pela Sociedade Americana em Higiene e Medicina Tropical e era sempre notificado que o meu trabalho tinha sido selecionado para poster. Ao ouvir esta notícia, o Prof. Chris Drakeley torcia inevitavelmente o nariz e dizia que não havia financiamento disponível para um Professor Auxiliar ir a uma conferência para apresentar um mero poster. Eu concordava com a justificação e voltava à minha rotina diária sem pensar que a seleção dos meus trabalhos para apresentação em formato poster simbolizava que, para os organizadores da conferência que conheciam bem o Prof. Chris Drakeley, eu era apenas mais um dos muitos membros do seu grupo de investigação. Em suma, a minha autoridade científica não era minha, mas sim do líder do grupo de investigação e para isso, bastava selecioná-lo (ou invariavelmente convidá-lo) para que a minha investigação fosse comunicada ao mundo. Ou seja, eu era um elemento invisível em termos de investigação em Malária propriamente dita. Embora os meus colegas e o próprio Prof. Chris Drakeley me considerassem um líder estatístico dentro do grupo, a mesma liderança não transpirava para fora das quatro paredes da LSHTM e tal trouxe consequências desagradáveis para a minha carreira.

Comecei a sentir uma insatisfação com a situação precária de contratos de pequena duração. Testemunhei o mesmo sentimento nos meus colegas que vinham e iam num abrir e de fechar de olhos sem possibilidade de renovarem os seus contratos de trabalho. Perguntava-me em como conseguir planear a minha vida futura se os meus contratos de trabalho não excediam um ano e meio. Nem adquirir uma casa em Londres ou nos seus arredores tinha a possibilidade de sonhar. Foi assim que regozizei com a abertura do primeiro Concurso de Estímulo ao Emprego Científico no início de 2018 por parte da Fundação para a Ciência e Tecnologia (FCT). O concurso era apelativo e assegurava financiamento por 5 anos, um horizonte temporal que permitia alguma margem de manobra para pensar no passo seguinte. Ou seja, era um bom ponto de partida para uma vida mais estável. E assim contatei a Prof. Lisete Sousa para que o Centro de Estatística e Aplicações da Universidade de Lisboa, o nosso querido CEAUL, apadrinhasse a minha candidatura como minha instituição de acolhimento. Ela acolheu a ideia com grande interesse e entusiasmo e assegurou o apoio oficial do Centro na realização do projeto.

Preparei então uma candidatura ao concurso com um projeto no campo da serologia de Malária que era uma mescla entre desenvolvimento metodológico em Análise Multivariada e análise de dados reais em colaboração com o Prof. Chris Drakeley da LSHTM. Pedia financiamento para um cargo de Investigador Auxiliar. Tinha optado por submeter a minha candidatura ao painel de avaliação em Ciências da Saúde ou Outras Ciências Biomédicas, deixando de lado o painel destinado à Matemática que seria talvez a opção mais lógica para a maioria de nós com formação de base em Estatística. O meu raciocínio foi que o financiamento para a área da Matemática tende a ser limitado e assim sendo, o respetivo painel tipicamente composto por matemáticos tende a “proteger” os seus. Para além disso, sentia que o projeto tinha mais peso numa área mais aplicada uma vez que: (i) o projeto focava numa doença que continua a afetar milhões de pessoas pelo mundo fora; (ii) o meu currículo académico era suficientemente rico em diferentes aplicações estatísticas (Malária, Genética, Epidemiologia, etc) para ser competitivo; (iii) o projeto proporcionava uma colaboração estreita entre o CEAUL que é um centro de referência nacional e internacional na área de Estatística e a LSHTM que é uma instituição de renome internacional e líder na área de Saúde Pública, nomeadamente, em Malária. Mostrava-me assim otimista para o resultado do concurso.

Foi então no dia 17 de Setembro de 2018 que recebi na minha caixa de correios eletrónicos a mensagem oficial de que resultados do concurso já estavam disponíveis no sítio da FCT. O meu coração aumentou o seu ritmo de batida enquanto abria a ligação à página onde se encontravam os resultados. Descarreguei o ficheiro Excel com as avaliações e abri-o no meu computador. Tanto tempo para abrir, uma eternidade na minha noção de tempo distorcida pela ansiedade de saber o resultado rapidamente. Havia quatro planilhas referentes aos diferentes níveis académicos: Investigador Júnior, Investigador Auxiliar, Investigador Principal e Investigador Coordenador. Abri a dos Investigadores Auxiliares. Bolas, tantos candidatos (1274 para ser mais preciso)! Fui descendo no ficheiro, de uma linha para outra. O meu nome não aparecia no topo do ficheiro que visava aos felizardos que teriam recebido luz verde para o financiamento.

Encontrei o meu nome na 250^a linha que equivalia à 246^a posição da classificação dos Investigadores Auxiliares. A minha pontuação consolidada final foi de 8.9 valores que, não tendo sido má, estava longe de 9.2 que foi o ponto de corte no financiamento a atribuir. Fiquei desapontado com o resultado. Mas o que aconteceu?

Acedi então à avaliação da minha candidatura e fiquei surpreendido com o que estava lá escrito em relação ao meu currículo:

“His scientific output is interesting and of high quality. It is however remarkable that he has not published any paper as senior author considering that he completed his PhD studies eight years ago.”

Importa escrever que *senior author* nas áreas mais viradas para a Biomedicina significa ser o último autor, ou seja, o autor que é visto como líder da investigação. E nesse aspeto o meu currículo estava totalmente carente. Não tinha pensado nisso. Tentei um recurso com o argumento (fraco) que tinha muitas publicações como primeiro autor e de correspondência (que era no meu ponto de vista quem tinha liderado a investigação). A avaliação do painel foi soberana e como diz o povo, o meu recurso “levou sopa”!

Foi então que fiz verdadeiramente uma reflexão sobre a minha carreira em Ciência. E cheguei à conclusão de que queria ser de facto um investigador independente que pudesse seguir as suas próprias

ideias e não ser apenas um estatístico-consultor. Para isso, tinha de mudar radicalmente de estratégia para minha carreira. Percebi tardiamente que uma carreira em Ciência é mais do que uma coleção de artigos e de projetos onde se participou. Há que ter um fio condutor para além de um conhecimento mais aprofundado de métodos estatísticos, genéticos ou epidemiológicos.

Tive então a sorte (a tal aleatoriedade inerente à vida) do meu lado. Por essa altura tinha-me integrado como o representante português na rede europeia em Síndrome de Fadiga Crónica (SFC), uma doença debilitante que afeta múltiplos sistemas do organismo, nomeadamente, o sistema imunitário. Era também membro da comissão executiva. Quase simultaneamente, o Tiago Domingues do Departamento de Estatística e Investigação Operacional tinha começado a ser orientado por mim e pela Prof. Helena Mouriño num projeto de doutoramento em dados serológicos relacionados com SFC. Parecia um bom pronúncio do que aí vinha.

Em Fevereiro de 2019 peguei na revista *Nature* enquanto voava de Londres para Bucareste, o local de uma das reuniões da rede em SFC. Comecei a folhear os artigos publicados nessa edição e deparei-me com um que me chamou à atenção. Estudava as células T reguladoras e o impacto da alteração da sua mitocôndria na regulação do sistema imunitário [7]. Neste ponto, preciso de mencionar o facto de que estas células tinham sido um dos tópicos do meu doutoramento e por isso fiquei extremamente curioso em saber qual era a grande novidade do artigo. Foi nesse momento que tive um momento *eureka* para a minha carreira. Que tal investigar o papel dessas células em SFC uma vez que havia múltiplos estudos indiciando perturbações graves no sistema imunitário dos pacientes com esta doença?

Depois da reunião, voltei a Londres todo entusiasmado e fiz uma revisão bibliográfica onde percebi que a minha tese de doutoramento poderia servir de base para construir uma nova teoria sobre a patogénese da doença. A ideia era simples: SFC de alguns pacientes pode ser o resultado de uma hiper-regulação do sistema imunitário sob a ação das células T reguladoras que não permitem livrar de infeções latentes por vírus da família herpes. Arregacei as mangas e, com energia renovada, trabalhei afincadamente até conseguir publicar o meu primeiro artigo (ainda como primeiro autor) em SFC em Novembro de 2019 [8]. O artigo teve muito boa aceitação entre a comunidade de investigadores e também pacientes, tendo recebido mais de 13,000 visualizações e 1,300 descargas até ao momento. Resolvi também que queria montar um pequeno grupo para investigar essa ou outra doença complexa com uma forte componente do sistema imunitário. Chamei-o de Immune-Stats Group cuja missão consiste em compreender o funcionamento do sistema imunitário na saúde e na doença através de análises estatísticas avançadas.

Avançando a minha investigação em SFC, a minha participação na rede europeia sobre esta doença possibilitou-me aumentar rapidamente o meu leque de colaboradores, tais como o Dr. Francisco Westermeier da FH Joanneum University of Applied Sciences em Graz na Áustria, a Prof. Carmen Scheibenbogen da Charité - Universitätsmedizin Berlin na Alemanha, e Luis Nacul e Eliana Lacerda da LSHTM no Reino Unido. Com os meus primeiros trabalhos na área, constatei que alguns estudos careciam de análises estatísticas mais sofisticadas e assim optei por reanalisar dados disponíveis na literatura. Numa dessas análises, os resultados foram tão promissores que deram origem ao meu primeiro projeto financiado como investigador principal. Por sorte, fui também contactado pelo João Malato e André Fonseca que tencionavam realizar projetos de doutoramento em Biomedicina. Propus um projeto em SFC ao João e outro em Malária ao André, projetos esses que foram submetidos à FCT no concurso para bolsas de doutoramento de 2019. Ambos os projetos tinham um grande cariz estatístico, mas sugeri ao João e ao André a submetê-los ao painel de Investigação Clínica (e não de Estatística). Foi um passo arriscado em virtude da minha experiência prévia com a FCT, mas desta vez os projetos foram bem recebidos pelo painel da avaliação, tendo sido aprovados para financiamento. Deste modo, o pequeno Immune-Stats Group passou a contar com mais dois elementos. Passei também a contar com mais dois colaboradores, o Prof. Luís Graça da Faculdade de Medicina da Universidade de Lisboa e Prof. Clara Cordeiro da Faculdade de Ciências e Tecnologia da Universidade do Algarve, que concordaram em assegurar a co-orientação do João e André, respetivamente. Várias peças estavam a alinhar-se à minha volta e assim, deparei-me com outros desafios relacionados com liderança.

Em Varsóvia: como ser líder?

Em final de Setembro de 2019, mudei-me de Londres para Varsóvia por motivos familiares – a minha esposa, sendo polaca, tinha recebido uma proposta de emprego na sua cidade natal e para além disso, estávamos à espera do nascimento do nosso filho, Artur. Contudo, continuava ligado contratualmente à

LSHTM e por esse motivo, passei 3 meses a trabalhar à distância muito antes do teletrabalho passar a ser moda durante a pandemia da COVID19. Continuei a trabalhar desse modo a partir de Janeiro de 2020 quando o meu vínculo profissional passou a ser com a Charité - Universitätsmedizin Berlin a convite da Prof. Carmen Scheibenbogen, que pretendia um colaborador (ou consultor) na área de Bioestatística e Bioinformática. Como consequência da minha residência na capital polaca e do advento da pandemia da COVID19 que tinha, entretanto, virado o mundo de pernas para o ar, o primeiro grande desafio que me deparei foi como gerir a orientação de três estudantes de doutoramento (Tiago, João e André) à distância e como criar alguma coerência e espírito de grupo. Uma solução para esse desafio foi “obrigar” a todos a usarem a plataforma digital *Slack* como forma de definir rotinas claras de comunicação e de trabalho; o envio dos tradicionais e-mails passou a ser secundário. Contudo, importa ressaltar que a gestão desta plataforma pode ser bastante problemática com o aumento do número de membros de um grupo; para isso, é preciso estabelecer regras bem claras que garantam uma comunicação eficiente entre todos. Uma outra solução foi agendar uma reunião (*online*) semanal entre todos, excluindo o Tiago que começava a entrar numa fase crítica do seu doutoramento e precisava de estar focado na sua investigação. A minha ideia era que, de um modo rotativo, um de nós estivesse incumbido de preparar uma apresentação longa sobre o seu projeto enquanto os restantes faziam uma pequena apresentação sobre o progresso realizado na última semana. Depois seguia-se uma semana onde a apresentação longa era substituída por uma discussão detalhada de um artigo publicado que um de nós tinha escolhido em apresentar. Para finalizar o ciclo, seguia-se uma semana onde um de nós convidaria um investigador amigo, antigo colega, ou colaborador a apresentar a sua investigação em qualquer tema científico na área de Biomedicina ou Estatística. No momento de escrita deste artigo, tivemos cerca de 7 seminários convidados onde se falou de temas tão variados como deteção de múltiplas modas em distribuições (Dr. José Ameijeiras-Alonso da Universidade de Santiago Compostela) e problemas atuais dos testes de diagnósticos rápidos em Malária (Dr. Khalid Beshir da LSHTM). Finalizado o ciclo, retomávamos às apresentações longas de cada um de nós. Com o relaxamento das políticas de controle da COVID19, as reuniões passaram a ser realizadas de duas em duas semanas, nomeadamente, quando o João e o André vieram a Varsóvia em visitas oficiais no âmbito dos seus projetos.

Um outro desafio foi operar uma mudança no estado de espírito relacionado com as minhas colaborações. Agora, tinha de tentar equilibrar os interesses a curto e a médio prazo do Immune-Stats Group com o grau de relevância e de prioridade de aceitar um determinado trabalho ao estilo de consultoria estatística. Também queria seguir a ideia de que, no final dos seus doutoramentos, Tiago, André, e João tivessem um currículo competitivo no mercado de trabalho, seja este académico ou fora dele. Para isso, a decisão de aceitar determinados projetos em colaboração teria de obedecer uma estratégia global e não apenas no resultado final que, em muitos casos, seria a publicação de um artigo em autoria conjunta. O primeiro critério foi que as colaborações teriam de ser em áreas de investigação relacionadas com as questões científicas seguidas pelo grupo, ou seja, Malária, doenças complexas (SFC, esclerose múltipla ou outra doença autoimune), estratificação de pacientes, uso de técnicas de aprendizagem de máquina, entre outros. Deste modo, eu tentava garantir que as colaborações iriam enriquecer o grupo de algum modo. O segundo critério foi um pouco mais de cariz pessoal e estabelecia uma maior prioridade aos colaboradores que já conhecia e que sabia de antemão serem pessoas de trato fácil com quem partilhava os mesmos ideais, princípios e valores. Assim, minimizava a probabilidade de problemas interpessoais no futuro. O terceiro e último critério foi dizer “não” a projetos em que a exequibilidade do trabalho envolvido era irrealista em termos de volume ou de prazo, nomeadamente, quando a colaboração era realizada num regime *pro-bono*. Por outras palavras, passei a rechaçar qualquer colaboração em que o trabalho a realizar tinha o selo de “era para ontem”; esta estratégia evita entrar em ciclos de stress totalmente desnecessários! Contudo, sobre este último critério, importa notar que, segundo Daniel Kahneman [9], nós os humanos sofremos de um viés cognitivo crónico que tende a subestimar o tempo necessário para realizarmos uma determinada tarefa ou projeto. No seu livro intitulado *Thinking, Fast and Slow*, esse autor conta que tinha planeado escrever esse mesmo livro em cerca de um ano. Acabou por terminá-lo uma década depois. Esta história de puro otimismo utópico ficou gravada na minha mente e agora tento adotar uma perspetiva oposta, muito ao estilo da famosa “Lei de Murphy”: quanto tempo preciso para realizar o que me pedem se o tempo é escasso? Vale mais ser pessimista e entregar antes do planeado do que ser um otimista irracional que não cumpre prazos ou promessas. De facto, a gestão das expectativas é um ponto fulcral e muito sensível entre colaboradores

e, na minha opinião, o não-cumprimento de um prazo acordado pode erodir imediatamente a confiança entre dois colaboradores.

Neste momento, o Immune-Stats Group vai entrar no seu terceiro ano de existência. O Tiago já terminou o seu doutoramento e novas caras já entraram (Frederico Moreira) e saíram do grupo (Fernando Madaíl). As reuniões passaram também a ter uma participação frequente da Prof. Clara Cordeiro e Dr. Francisco Westermeier que, com o seu entusiasmo e curiosidade, elevaram as discussões científicas dos nossos projetos. Continuo a sentir que o exercício de “ser líder” de um grupo é apenas uma aprendizagem motivada, ou melhor ainda, modelada pelas pessoas (e suas personalidades) em nosso redor. Por isso, sinto que o caminho da liderança (estatística ou não) é apenas uma viagem tanto intrapessoal como interpessoal que vai sendo afinada ao longo do tempo. No fim, pretende-se atingir utopicamente a tal sabedoria subliminar dos mais velhos que às vezes sentimos sem dar conta. Esperamos conseguir proferir no momento certo a tal palavra que ilumina, motiva ou apazigua.

Tenho a sorte de ter recebido financiamento da Agência Polaca NAWA (Polish National Agency for Academic Exchange) que garante a minha pertença à família da Politechnika Warszawa por um período de 2 anos. Fui recebido pelo Prof. Przemyslaw Biecek na Faculdade de Matemática e da Ciência da Informação que de braços abertos, e após 8 anos de nos termos conhecido numa conferência, abriu as portas do seu grupo (MI² data lab). Do meu lado vi esta abertura como uma nova oportunidade para aprender novas maneiras de pensar em Estatística tanto para mim como para a pequena família do Immune-Stats Group. Já o escritor inglês John Donne escrevia e posteriormente adaptada pelos roqueiros Bon Jovi na canção *Santa Fe: “no man is an Island”*. E o Immune-Stats Group não é uma ilha. Faz parte de um ecossistema científico rico herdado dos investigadores e líderes do passado e modelado pelas interações pessoais do presente. E o presente é Varsóvia. Neste momento, acho que ser líder é apenas ter a liberdade de esticar o braço e poder tocar em nome próprio os vários ramos da Ciência, seja esta Estatística, Ciência de Dados ou Imunologia. Em última análise, somos todos líderes da nossa vida e já o povo dizia: “cada cabeça sua sentença”.

Últimas palavras

Para concluir faço uma pequena reflexão sobre a questão primordial: será a liderança estatística uma escolha por necessidade? Acho que a resposta a esta pergunta é como um modelo estatístico de misturas onde várias componentes de diferentes pesos estão presentes. Por um lado, não posso negar que houve uma necessidade urgente de mudar o rumo da minha carreira a certo ponto da minha estadia na LSHTM. Essa necessidade levou a uma escolha consciente de ser um investigador independente e de formar um pequeno grupo onde a minha responsabilidade aumentou, mas que trouxe consigo uma maior liberdade de escolha. Essa liberdade, embora não seja uma garantia por si só de uma motivação constante, aumenta a longo prazo a satisfação de estarmos a seguir o nosso caminho e não o dos outros. Contudo, penso que a liderança estatística foi-me surgindo naturalmente ao longo da minha carreira e por isso, o caminho para a liderança já estava traçado antes da necessidade bater à porta. Assim sendo, resta-me dizer que, se dermos oportunidade ao nosso bichinho da liderança para crescer e de se tornar mais marcante no nosso consciente, podemos tornar-nos em algo diferentes daquilo que inicialmente idealizamos para nós próprios. Ser líder principalmente da nossa vida tem a virtude de torná-la mais divertida e preenchida. É como se nos levantássemos todos os dias e o sol nos desse um bom-dia logo pela manhã e nos perguntasse: qual é o próximo passo que vais dar na tua vida?

Agradecimentos

Ao escrever este texto, fiquei surpreendido com os vários artigos que já escrevi para o Boletim da SPE. Por isso, o meu primeiro agradecimento é para o Editor do Boletim o Prof. Fernando Rosado. O segundo agradecimento é para todos os orientadores, colaboradores, autores e estudantes com quem tive o prazer de trabalhar. Agradeço em especial ao Prof. Carlos Daniel Paulino que sempre me apoiou desde os tempos da Licenciatura em Matemática Aplicada e Computação no Instituto Superior Técnico, ao Dr. Jorge Carneiro que orientou o meu doutoramento e muito contribuiu para que eu fosse um Cientista-Estatístico (e não apenas um Estatístico), a Prof. Solange Gil da Faculdade de Medicina Veterinária da Universidade de Lisboa pela longevidade da nossa colaboração que data desde o longínquo ano de 2003, ao CEAUL que é quase como família para mim, e ao Prof. Przemyslaw Biecek que acolheu a mim e ao

Immune-Stats Group em Varsóvia. Agradeço a todos os membros do Immune-Stats Group, nomeadamente, ao João Malato que, com toda a dedicação e amizade, desenhou a Figura 1 e que fez uma revisão editorial deste artigo. Finalmente, este trabalho teve o apoio da NAWA no âmbito do projeto de referência PPN/ULM/2020/1/00069/U/00001.

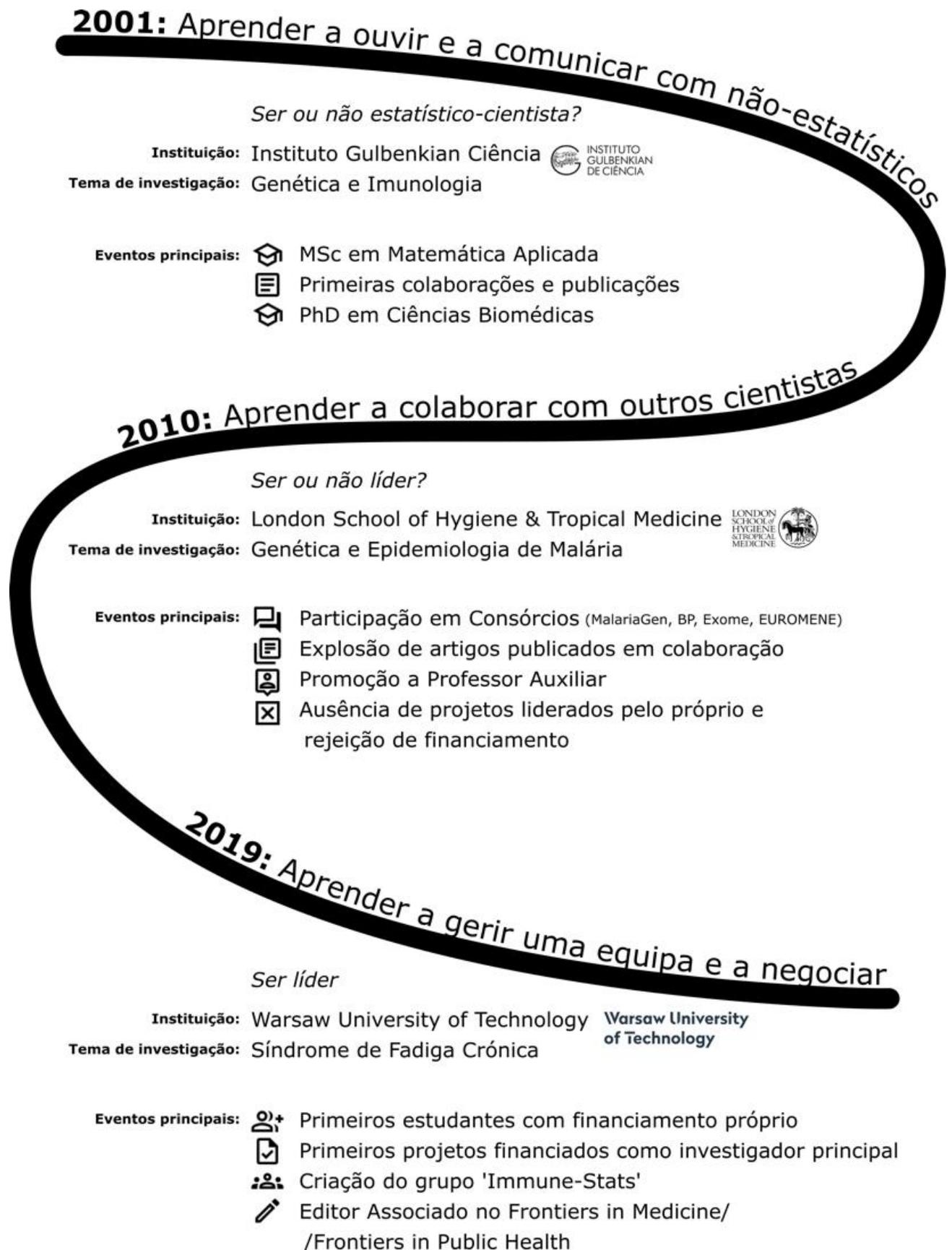


Figura 1: Sumário da minha carreira em termos de eventos principais.

Referências

- [1] Gibson, E. W. (2019). Leadership in Statistics: Increasing Our Value And Visibility. *The American Statistician*, 73:109-116.
- [2] Sepúlveda, N. (2022). How to increase the visibility of statisticians in the modern world of Dataism?. Aceite para publicação nas Actas do XXV Congresso da SPE.
- [3] Sepúlveda, N. (2009). Surfando pelos mares da Biologia: peripécias de um jovem estatístico aspirante a biólogo. *Boletim SPE (Primavera)*, 75-81.
- [4] Bross, I. D. J. (1974). The role of the Statistician: scientist or shoe clerk. *The American Statistician* 28, 126-127.
- [5] Sepúlveda, N. (2015). Tudo sobre Malária, Genética, e Estatística ou talvez não!. *Boletim SPE (Outono)*, 41-48.
- [6] Sepúlveda, N., Campino, S. G., Assefa, S. A., Sutherland, C. J., Pain, A., & Clark, T. G. (2013). A Poisson hierarchical modelling approach to detecting copy number variation in sequence coverage data. *BMC Genomics*, 14, 128.
- [7] Weinberg, S. E., Singer, B. D., Steinert, E. M., Martinez, C. A., Mehta, M. M., Martínez-Reyes, I., Gao, P., Helmin, K. A., Abdala-Valencia, H., Sena, L. A., Schumacker, P. T., Turka, L. A., & Chandel, N. S. (2019). Mitochondrial complex III is essential for suppressive function of regulatory T cells. *Nature*, 565, 495–499.
- [8] Sepúlveda, N., Carneiro, J., Lacerda, E., & Nacul, L. (2019). Myalgic Encephalomyelitis/Chronic Fatigue Syndrome as a Hyper-Regulated Immune System Driven by an Interplay Between Regulatory T Cells and Chronic Human Herpesvirus Infections. *Frontiers in Immunology*, 10, 2684.
- [9] Kahneman, D (2011). Thinking, Fast and Slow. Penguin books, Londres.



How to be an effective statistician! With a focus on your leadership skills

Alexander Schacht, alexander.schacht@veramed.co.uk

*Executive Vice President Launch and Commercialisation Data Sciences
Veramed*

Decades ago management guru Peter Drucker wrote his impactful book "[The Effective Executive](#)". In this book, he explains, what it means to be effective as an organizational leader. Further, he dives into the different tools that executives mainly use to achieve results in an efficient way. Like an executive, you as a statistician work effectively if you get the right things done in the right way at the right time. My experience tells me that you need to focus on four dimensions and excel in all four to be an effective and impactful statistician over time. These dimensions were first introduced to me about 10 years as the framework for the functional strategy for the statistical sciences group while I was at Eli Lilly and Company. Although they were established as the strategy and direction for the entire function (>400 statisticians) to positively impact the business - and proved to be very effective - I realized that they also served as an excellent framework for individual development and improvement. I have adopted and used them as my guidepost since that time. The four areas, which the statistical leadership at Lilly referred to as "pillars", are:

- Leadership
- Innovation
- Knowledge
- Excellence

Credit as well to the Lilly statistical leadership for the easy to remember acronym: LIKE

In my podcast – specifically episode 199 with Olaf Kapinski - I cover each of these areas and explain how they can be leveraged as a guide for individual development and, ultimately, how this will result in greater impact on your organization as well as personal fulfilment & success. In this article, I will dive only into leadership.

Let me begin by sharing my background and experience to give you an idea of my journey and the impact these areas have had on the direction my career.

I have worked as a professional statistician for over 20 years, most of that time working for major pharmaceutical companies. I have several years of experience supervising other statisticians and have been recognized in my career with several technical promotions. I have authored or co-authored over 75 peer reviewed manuscripts, more than 100 presentations and posters, and present regularly at conferences around the world.

In 2018, I founded "The Effective Statistician" - a podcast dedicated to helping statisticians in the health care sector achieve their maximum potential to serve patients. This podcast has already more than 200 episodes and serves a global audience with more than 130,000 downloads. Some years, I launched a first of its kind on-line leadership course for pharmaceutical statisticians with my business partner Gary Sullivan, a professional statistician with almost 30 years of experience, including 10 years as of leadership trainer for statisticians. We have now trained nearly 300 statisticians, data scientists and programmer in this in-depth 5 months program and seen many success stories. Several large pharma companies started to train large cohorts of statisticians via this program.

My experience, success and personal growth is largely a result of focusing on the four areas.

Now, we're all standing on the shoulders of giants. Many topics I speak about come from a variety of sources and being a life-long learner in these areas is key to becoming successful from my point of view. This goes hand in hand with being gritty in the sense as Angela Duckworth talks about it in her book *Grit: The Power of Passion and Perseverance*.

Since I learned about these 4 pillars at Eli Lilly, I'm constantly watching statisticians around me to assess how capable they are with respect to the LIKE pillars. It's amazing how clearly those that are successful separate from those that have average careers and those that fail to move forward at all. Those that are moving fast in their career and that are respected by many other statisticians as well as other business partners, are the statisticians that excel in each of these 4 areas. All those that have more average careers miss at least 1 of the four pillars and those that fail to progress usually miss out on at least 2 pillars.

You need to be strong in 4 LIKE pillars!

Throughout my podcast, I give you guidance on the four different pillars and what actions you can take, to improve in these areas. It's impossible to cover everything in here – I have 200+ episodes. Thus, I will focus on only some practical tips as well as explaining some key concepts.

Also, the four pillars are not mutually exclusive. They overlap and if you are strong on all of them, they work synergistically for you. This will be self-evident from the description of the pillars.

If you want to learn more, you may listen to my [podcast](#) “The Effective Statistician”. The podcast episodes are all dedicated to this topic and each episode is tagged according to the four different LIKE pillars. Here, I will concentrate on the first pillar – leadership.

LEADERSHIP

Leadership consists of two concepts: self-leadership as well as leading or in other words influencing others.

SELF-LEADERSHIP

To be a great leader, you first need to lead yourself. This includes having good personal systems, habits and behaviors to use your time and energy in the best way. Many tips around this are published under the terms of time management, employing good habits, creating a clear vision and having goals and action plans.

Statisticians are brain workers. Many of the different products we create need deep and uninterrupted working time to produce meaningful results. However, our modern style of working with the constant meeting culture, social media interruptions, regular email checking, and open plan office set ups, works against this so-called deep work. We need to actively and consciously create boundaries in which we as statisticians can work on larger deliverables. These may include study protocols, analysis plans, programs, sets of results, or other kind of reports.

Cal Newport wrote the book [“Deep Work”](#) exactly about this dilemma. Being an analytical thinker himself, he noted different ways of how people can organize their life to make deep work possible. He provides lots of practical advice on the different topics, that I don't need to repeat here. I highly recommend this fundamentally important book as it will help you to organize your work life better and to see, what kinds of habits are good to adopt and which you should avoid or stop.

Another part about self-leadership consists of developing a clear vision for yourself. Not something, that is told to you by senior management, but a vision for how you want your life or at least your career to look like in the future. Having a 5-year vision, will help you to determine your yearly goals and these in turn will help you to create weekly and daily action plans. Michael Hyatt has published extensively about these concepts and his [podcasts](#) are highly recommended to learn about these topics.

LEADING OTHERS

The concept of leadership is often misunderstood as management or assigned leadership with having a title and being head of some department or group. The relevant leadership definition, that Gary Sullivan also adopted for the leadership courses for the American Statistical Association, reads as follows:

The ability to consistently deliver value to an organization or cause by inspiring others to take a specific direction when they truly have the freedom or choice to do otherwise.

It's worthwhile to investigate this definition closely. First, leadership is an ability. It's not a title or a job description. You don't get it from an impressive title on your business card. It's an ability, that can be learned and which like all skills needs training.

Vince Lombardi worded it quite nicely by saying: “Leaders are made, they are not born. They are made by hard effort, which is the price which all of us must pay to achieve any goal that is worthwhile.”

Second, this ability is characterized by the results: consistently delivering value. And these will mostly be towards your department or company. But if you work together on a cross-companies project, like creating a better SAS macro, this becomes the common cause.

Third, you’re not commanding others, you’re not telling others, you inspire them. You help them understand, why a specific action makes sense. You show them, how they win in it. You create win-win situations, where others want to follow your guidance and advice. Thus, this is not about order and command - something that rarely works in companies and even less in voluntary groups. Leading in such a way is called emergent leadership in contrast to the assigned leadership with huge titles and large offices.

Unfortunately, this leadership is the pillar where most statisticians fail. And thus, I’m focusing most of this manuscript about leading others. But why do statisticians fail to lead?

From a webinar, which Gary Sullivan and myself delivered in early 2019, I took the 4 big reasons, why based on our experience, statisticians fail to lead. These are

- Mindset
- Bias toward technical skills
- Lack of leadership understanding
- Lack of “big picture” thinking

In terms of mindset, as a statistician, we are predominantly trained to help and support others in their fields. They come with their problems to us and we consult and teach them to solve their problems. The emphasize is on *their* problems. We don’t tend to think of these as our projects or our problems. We’re just the statistical researchers and they can follow our advice or choose not to. Of course, we’re a bit disappointed, if they don’t but in the end, we like to feel - it’s their project not ours.

Also, we become busy in what we do and find success in moving things around from collecting data to analysing data, to communication of data. We get so busy, that we fall into this routine of just producing analyses and study designs and many organizations reward such behaviour - so why change?

Similarly, we often think about leadership as something that is not our responsibility. It’s not on our job description or in our email signature. But who’s responsibility is it, if we think about leadership in the definition mentioned above. We might never ask ourselves, whether it’s up to us to lead. And thus, we never see the lost opportunities. Or even, whenever someone else takes the lead and we’re not convinced from their decisions, we get back into some victim mentality, in which we lose, because the other has more power.

Naturally, our function has a strong bias toward technical skills. This is what makes us be statisticians in the end. This is what we’re passionate about and the amount of technical training, that many statisticians pick up in relationship to non-technical training is the best proof for this. We don’t allow a debate to happen, that our core strength and foundation are *only* technical skills. This make us believe, that more technical expertise is always better and nearly all our self-development time should be geared towards this goal.

In this group thinking, we all find ourselves in a nice cosy comfort zone, where it’s all a good.

It’s all good until inevitably in your career your technical skills aren’t helping you. Where you realize, that no further technical argument gets you closer to your desired direction. Or where you may even see the opportunity, but you have no idea how to seize it.

Another large area of failure for statisticians is the lack of leadership understanding. As a statistician you may think, that you need to

- Communicate more effectively,
- Network,
- Influence,
- Negotiate,
- Drive change, or
- Lead.

But you may realize, that you have no clue how. Our training doesn't focus on these skills and more technical training doesn't help either. Colleagues in other areas spend much more time on this and naturally get better at it. They even enjoy doing these things. A situation, that puts statisticians at risk of being the nerd, that doesn't get it.

Furthermore, based on the above-mentioned habits of statisticians, there's often a lack of "big picture" thinking. The day to day activities of statistical consulting provides lots of tactical challenges, which statisticians generally love to dive into. The passion for detail of statisticians, which in many areas is a strength, turns out to be a disadvantage with respect to understanding how the tactical challenges fit into the bigger picture. E.g. is it worth investing a lot of time in fine-tuning the methodological approach for a challenge? Such questions are either not asked in the first place by statisticians or they struggle to answer such a question due to a lack of the broader impact on the business.

This lack of big picture thinking also excludes us as statisticians from the decision-making process, where this capability is essential. Now some might say, that this is not the role of the statistician, but as said above. This leads to average careers at best.

Now what should you as a statistician do about these topics? Reading this article shows, that you at least caring about this topic and changing the mindset on how you see yourself moves you a great step forward. This mindset will make you aware of opportunities, that you may have missed before and just the curiosity to change will naturally drive learnings in terms of leadership.

Although it's never too late to invest in leadership, I'm recommending starting as early as possible in your career. The longer you wait, the more time you lose to have the full impact, that your overall capabilities allow for. Become a student of leadership and embrace a life-long journey of improving your skills in the different parts of leadership. There's always an area where you can further improve your skills and usually you get better, while doing your day-to-day work just more consciously, open-minded, and reflective.

In terms of specific actions to take, I recommend taking on challenging roles and opportunities. Many organizations establish teams to drive forward e.g. process improvements. Such cross-functional groups lend them self perfectly to learn about emergent leadership. Of course, it's great to have them within your business, but something outside of work might help you as well. As mentioned above, voluntary groups and organizations depend nearly exclusively on emergent leadership. Driving forward such initiatives in your local community offers lots of areas for learning.

Furthermore, sign up for non-technical trainings. Your annual performance plan, which you are measured against should optimally also contain a training and development part. Check, which non-technical trainings are offered by your organization or which of such trainings are taken by colleagues outside of your statistical department. Such trainings include but are not limited to:

- Cultural training,
- Communication,
- Leadership development,
- Change management,
- Presentation training, or
- Personality training.

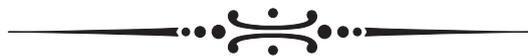
If you're a reader, I recommend reading books about exceptional leaders (e.g. biographies) that you're interested in. These leaders may come from business, politics, sports, art, science, or other areas, that you're interested in. I personally loved the books about Steve Jobs or Pep Guardiola. Many of such books are also published as audio books allowing you to listen to them, e.g. while you're commuting. Also, there are many great podcasts about leadership and the one by Michael Hyatt – even if it's mainly targeted towards supervisors - helps in this aspect a lot.

Finally, you need to get out of your comfort zone to grow these leadership skills. Speak with your supervisor, if there are work assignments that helps you to get there. These might be larger projects or projects including other people, that you have not worked with before or in areas, that you're not familiar with. Such new assignments provide you with the opportunity of a fresh start in terms of building relationships and trust. Test different concepts of becoming a leader in these projects.

Furthermore, I strongly recommend finding an experienced leadership coach. Of course, this could be an officially assigned coach, which is paid for helping you, but mostly these are unofficial agreements. Look out for great leaders in your organization and approach them. In addition, you may want to speak with your supervisor in terms of the suggestions he might have.

A very important way to improve for you is establishing a daily to routing of committing to driving action. Create a morning routine at work during which you review the opportunities of the day to drive actions. Think proactively about the different meetings and situations that will arise over the day and how you can create win-win situations for the different participants to make sure, that the projects move forward into the right direction. Make it a routine to review this list in the evening before you quit to assess how successful you were. Then write down what worked and what needs further improvement. One step each day in the right direction will build lots of momentum over time.

Leadership and delegation go hand in hand. As an emergent leader, of course, you need to approach delegation from a different attitude. You can't just tell people what to do, but this is a benefit of emergent leaders. Anyhow, you need to understand the basics of delegation as not only supervisors delegate. If you're in a project meeting and you note down a follow-up action – who is doing what by when, this is essentially delegating. If you're working with a vendor, which does the programming for you, you're delegating. If you're asking a colleague for help, this is delegation. Even if you're asking your supervisor to do something for you, this is delegation. Get clear on how to delegate effectively. I've recorded a podcast episode on this topic, and you'll find further references there. [Just click here.](#)



Mensagem do Presidente

Caros Sócios da Sociedade Portuguesa de Estatística,

O Boletim da Sociedade Portuguesa de Estatística é desta vez dedicado ao tema da *Liderança Estatística*.

O que é a Liderança Estatística?

Muito provavelmente se perguntássemos a 5 estatísticos profissionais ao acaso obteríamos 5 respostas diferentes—o que demonstra apenas a riqueza e escopo do tema.

Uma referência moderna acerca do tema é a seguinte:

Gibson, Eric W. "Leadership in statistics: Increasing our value and visibility," *The American Statistician* (2018).

Gibson argumenta que houve nos últimos anos um *choque na oferta* de “dados”, o qual gerou um *choque na procura* de “estatísticos”. E paradoxalmente a Estatística tem um papel menos visível do que seria de esperar – dada a magnitude do tal *choque na oferta*. Gibson argumenta assim que não basta que a profissão se **adapte** a este “choque”, mas é necessário que **lidere**; e para isso necessitamos desenvolver novas habilidades que nos permitam ocupar uma posição de destaque, guiar equipas multidisciplinares e comunicar de forma mais eficiente com profissionais de outras ciências.

Atrevo-me a acrescentar o óbvio: Aumentarmos o valor da nossa profissão e a nossa visibilidade, começa na sala de aula. Essa é a nossa primeira oportunidade para influenciarmos a próxima geração acerca do propósito da nossa profissão, da sua importância e significância para o Mundo em que vivemos. Se adquirirmos novas habilidades, mas não adaptarmos a nossa forma de ensinar – semestre após semestre – perdemos a oportunidade de liderar e de assim aumentar o nosso valor e a nossa visibilidade. E não devemos parar na sala de aula, temos os nossos colaboradores de outras áreas a quem podemos ensinar e com quem podemos ao mesmo tempo aprender enquanto colaboramos. Até porque para liderarmos necessitamos de entender as razões e as necessidades de quem necessita dos nossos serviços. Isto recorda-me o seguinte. Escrevi recentemente um artigo para a Gazeta da Sociedade Portuguesa de Matemática, intitulado:

“Espelho meu, espelho meu: Há alguém que gaste mais por mês do que eu?”

O artigo é relacionado com um problema de consultoria no qual estive envolvido recentemente. Uma das primeiras etapas em qualquer problema de consultoria é entender exatamente qual é o problema com que se confronta o cliente: Quais são as suas

necessidades, quais são os seus objetivos? E claramente o cliente não costuma chegar à reunião com o problema formulado numa forma de teste de hipóteses ou com um modelo de regressão. O cliente no problema de consultoria em questão chegou à reunião com uma base de dados enorme, recorria a jargão da área da *fintech*,¹ que nem eu nem os meus colaboradores dominávamos, e o meu primeiro objetivo foi **converter as suas necessidades numa questão simples** que descrevesse aproximadamente o objetivo; a primeira versão dessa questão foi a seguinte:

“Como é que comparo estatisticamente as despesas de um dos utilizadores da nossa plataforma com as de outros que ganham aproximadamente o mesmo?”

Esse foi o ponto de partida. Acontece que afinal uma versão mais elaborada da questão do cliente tinha alguns aspetos interessantes do ponto de vista conceptual e a nossa solução acabou por ser publicada no *Journal of the Operational Research Society*:

<https://www.tandfonline.com/doi/full/10.1080/01605682.2021.1952115>

Não há liderança sem comunicação eficiente. Mas nem sempre é fácil entender quem nos solicita e nem sempre é fácil comunicar de forma clara o mérito da nossa análise.

A liderança requiere o empenho de todos nós, em todas as facetas da nossa profissão.

Até breve,

Edimburgo, 3 de Março de 2022

Cordais saudações

Miguel de Carvalho

¹ De modo grosseiro, a *fintech* é uma área emergente que recorre ao uso de tecnologias e técnicas inteligentes de automação para desenvolver novas ferramentas e serviços financeiros.

Ciência Estatística

• Artigos em Revistas

- Brites, N.M.; Braumann, C.A. (2021). Harvesting optimization with stochastic differential equations models: is the optimal enemy of the good? *Stochastic Models*. Publicado online em 2021.12.19. <https://doi.org/10.1080/15326349.2021.2006066>
- Brites, N.M.; Braumann, C.A. (2022). Profit optimization of stochastically fluctuating populations under harvesting: the effects of Allee effects. *Optimization*. Publicado online em 2022.01.31. <https://doi.org/10.1080/02331934.2022.2031191>
- Jacinto, G.; Filipe, P. A.; Braumann, C. A. (2021). Profit optimization of cattle growth with variable prices. *Methodology and Computing in Applied Probability*. Publicado online 2021.10.08. <https://doi.org/10.1007/s11009-021-09889-z>
- Jamba, N.T.; Jacinto, G.; Filipe, P.A.; Braumann, C.A. (2022). Likelihood function through the delta approximation in mixed SDE models. *Mathematics*, 10, 385. <https://doi.org/10.3390/math10030385>
- Larguinho, M.; Dias, J.C.; Braumann, C.A. (2022). Pricing and hedging bond options and sinking-fund bonds under the CIR mode[J]. *Quantitative Finance and Economics*, 6(1): 1-34. <https://doi.org/10.3934/QFE.2022001>
- Papança, F. (2021). A Influência do Ensino Militar na Evolução da Matemática e da Estatística. *Revista Militar, II Século -73º Volume - nº 10. Nº 2637. p. 7955-812.*

• Revista

Título: *REVSTAT* – Statistical Journal

Ano: 2021. Volume 19 – Números 1 a 4, (<https://revstat.ine.pt/index.php/REVSTAT/issue/archive>).
Listagem dos artigos na página 6 deste Boletim SPE.



Retrospectiva do Boletim SPE - Tema Central

Primavera de 2022 – Liderança Estatística

Outono de 2021 – *Machine Learning* e Inteligência Artificial

Primavera de 2021 – Especial Covid: a Estatística ao serviço da sociedade

Outono de 2020 – 40 anos SPE: De onde viemos? Onde estamos? Para onde vamos?

Primavera de 2020 – INE-85 anos de estatísticas a servir o país

Outono de 2019 – Estatística nas Ciências da Saúde

Primavera de 2019 – Séries Temporais de Valor Inteiro

Outono de 2018 – Equações diferenciais estocásticas e algumas aplicações

Primavera de 2018 – Estatística Multivariada – perspectiva no século XXI

Outono de 2017 – O Tema Central da Estatística - um novo olhar

Primavera de 2017 – Incerteza em Engenharia

Outono de 2016 – O Tema Central da Estatística

Primavera de 2016 – Séries Temporais e suas aplicações

Outono de 2015 – Estatística em Genética

Primavera de 2015 – Estatística no Desporto

Outono de 2014 – Estatística no Ensino Básico e Secundário

Primavera de 2014 – (Um) Ano Internacional da Estatística

Outono de 2013 – A “Escola Bayesiana” em Portugal

Primavera de 2013 – Estatística não - paramétrica

Outono de 2012 – Métodos Estatísticos em Medicina

Primavera de 2012 – Estatística no Ensino Superior Politécnico

Outono de 2011 – Análise de Sobrevivência

Primavera de 2011 – Sondagens e Censos

Outono de 2010 – Estatística Espacial

Primavera de 2010 – *Data Mining* - Prospecção (Estatística) de Dados?

Outono de 2009 – Modelos Económétricos

Primavera de 2009 – Investigação (em) Estatística

Outono de 2008 – Processos Estocásticos

Primavera de 2008 – ALEA - Um sítio do nosso mundo

Outono de 2007 – Bioestatística

Primavera de 2007 - A “Escola de Extremos” em Portugal

Outono de 2006 – Ensino e Aprendizagem da Estatística

também disponíveis em <https://www.spestatistica.pt/publicacoes/categoria/boletim-da-spe>

Edições SPE - Minicursos

Título: *Uma introdução à Meta-Análise*

Autora: Maria de Fátima Brilhante

Ano: 2017.

Título: *Estatística Bayesiana*

Computacional – uma introdução

Autores: M. Antónia Amaral Turkman e

Carlos Daniel Paulino

Ano: 2015.

Título: *Análise de Valores Extremos: Uma Introdução*

Autoras: M. Ivette Gomes, M. Isabel Fraga Alves e Claudia Neves

Ano: 2013.

Título: *Modelos com Equações*

Estruturais

Autora: Maria de Fátima Salgueiro

Ano: 2012.

Título: *Análise de Dados Longitudinais*

Autoras: Maria Salomé Cabral e

Maria Helena Gonçalves

Ano: 2011

Título: *Uma Introdução à Estimação Não-Paramétrica da Densidade*

Autor: Carlos Tenreiro

Ano: 2010

Título: *Análise de Sobrevivência*

Autoras: Cristina Rocha e

Ana Luísa Papoila

Ano: 2009

Título: *Análise de Dados Espaciais*

Autoras: M. Lucília de Carvalho e

Isabel C. Natário

Ano: 2008

Título: *Introdução aos Métodos*

Estatísticos Robustos

Autores: Ana M. Pires e

João A. Branco

Ano: 2007

Título: *Outliers em Dados Estatísticos*

Autor: Fernando Rosado

Ano: 2006

Título: *Introdução às Equações*

Diferenciais Estocásticas e

Aplicações

Autor: Carlos Braumann

Ano: 2005

Título: *Uma Introdução à Análise de Clusters*

Autor: João A. Branco

Ano: 2004

Título: *Séries Temporais – Modelações lineares e não lineares*

Autoras: Esmeralda Gonçalves e

Nazaré Mendes Lopes

Ano: 2003 (2ª Edição em 2008)

Título: *Modelos Heterocedásticos.*

Aplicações com o software Eviews

Autor: Daniel Muller

Ano: 2002

Título: *Inferência sobre Localização e Escala*

Autores: Fátima Brilhante, Dinis

Pestana, José Rocha e

Sílvio Velosa

Ano: 2001

Título: *Modelos Lineares*

Generalizados – da teoria à prática

Autores: M. Antónia Amaral

Turkman e Giovani Silva

Ano: 2000

Título: *Controlo Estatístico de Qualidade*

Autoras: M. Ivette Gomes e

M. Isabel Barão

Ano: 1999

Título: *Tópicos de Sondagens*

Autor: Paulo Gomes

Ano: 1998



O MUNDO DA ESTATÍSTICA

ORGANIZAÇÃO PARTICIPANTE



Federation of European National
Statistical Societies



