

Boletim



SOCIEDADE PORTUGUESA
DE ESTATÍSTICA

Publicação semestral

outono de 2017



O Tema Central da Estatística - um novo olhar

“Estatística, para ver longe”

A Estatística, a indústria e a inovação

Data Science, Big Data e um novo olhar sobre a Estatística

Estatística - “Espelho meu, espelho meu, que futuro terei eu?”

A Revolução dos Dados

A tirania dos jargões

Novo olhar sobre a Estatística, imaginar o mundo

Desafios da Estatística para o século XXI

A minha utopia sobre o Tema Central da Estatística

“... com um novo olhar”

Emília Athayde	26
Cláudia Nunes	27
Isabel Fraga Alves	29
Conceição Amado	32
Pedro Campos	35
Regina Bispo	38
Lisete Sousa	42
Tiago A. Marques	45
Nuno Sepúlveda	49
Emília Athayde	56

Editorial	1
Mensagem da Presidente	5
Notícias	6
Enigmística	25
Ciência Estatística	57
Prémios “Estatístico Júnior 2017”	58
Prémio SPE 2017	61
Prémio Carreira SPE 2017	62
Edições SPE – Minicursos	63
Retrospectiva	64

Informação Editorial

Endereço: Sociedade Portuguesa de Estatística.
Campo Grande. Bloco C6. Piso 4.
1749-016 Lisboa. Portugal.

Telefone: +351.217500120

e-mail: spe@fc.ul.pt

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

ISSN: 1646-5903

Depósito Legal: 249102/06

Tiragem: 500 exemplares

Execução Gráfica e Impressão: Gráfica Sobreireense

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

Sociedade Portuguesa de Estatística desde 1980



**DIA EUROPEU
DA
ESTATÍSTICA
20.10.2017**
**MELHORES DADOS.
MELHORES VIDAS.**
Lisboa, PORTUGAL



**“Estatísticas Oficiais,
um Bem Público”**



Editorial

... com continuidade diretiva...

Um dos vetores, uma das bases para o sucesso na SPE, obviamente, está nos órgãos diretivos. A SPE, nesse domínio, está num momento de continuidade. A Direção SPE e que foi reeleita, tem sabido enfrentar os problemas de toda a conjuntura científica atual e que, evidentemente, também se refletem na nossa sociedade. Com a reeleição dos órgãos diretivos, para além das felicitações que a todos são devidas, aos sócios também cabe a iniciativa da manifestação do seu reconhecimento. Este ato solene deve incluir a declaração de uma maior disponibilidade para uma intervenção mais dinâmica de todos, juntando-se em unísono em torno da Direção para construir um melhor futuro da SPE. Esta de todos depende e, nalgum sentido, toda a comunidade científica dos estatísticos também depende da SPE. A Direção, em cada momento, deve sentir a presença dos sócios para também se sentir mais forte e mais estimulada no seu desempenho.

Reconhecendo todo o trabalho desenvolvido em prol da sociedade, felicitemos a Direção reeleita e, na medida das suas possibilidades, cada sócio torne (ainda mais) visível toda a sua disponibilidade – cada vez mais e em torno dela – para construirmos uma excelente SPE.

A SPE, com o enorme esforço de todos os seus sócios, construiu um passado e foi subindo degraus na escada do reconhecimento de toda a comunidade científica nacional e internacional. O recente acontecimento *Dia Europeu da Estatística 2017* e de que damos relato nesta edição, produziu aquela que podemos apelidar de *Declaração de Lisboa* e foi mais um belo marco na história da SPE. Louvores são devidos à Presidente Eduarda Silva pelo sucesso atingido nessa participação de nível internacional com o devido destaque da Sociedade Portuguesa de Estatística.

E, a propósito de continuidade...

... com uma SPE (ainda) mais dinâmica e com (ainda mais) excelência no futuro...

Escutam-se ainda as palavras de desconforto de muitos, perante o futuro da SPE, em particular no campo da intervenção dos sócios. No recente momento eleitoral, foi muito reduzida a participação dos sócios. Não obstante o esforço desenvolvido pelos membros da comissão eleitoral e que foi notório, a participação foi muito inferior ao número de sócios presentes nesse momento nas atividades do congresso. Num congresso com cerca de 200 participantes “estatisticamente” quantos votos se devem esperar? Menos de setenta, será sempre muito pouco! Porquê assim? O que significa? O que mais se pode fazer para ativar uma melhor participação?

A Sociedade Portuguesa de Estatística vive do dinamismo daqueles que a constituem. Os seus sócios são a sua vida e desta depende o sucesso da Ciência Estatística. Felizmente o grande número de participantes no XXIII congresso é uma excelente notícia e é neles e com eles que está a força necessária para um melhor futuro; como também se pode concluir de todos os textos publicados nesta edição.

Honremos os sócios fundadores e o seu pioneirismo!

A SPE está nas mãos dos sócios, mas se estes não a cuidarem em todos os mais pequenos detalhes...

... com a soma das partes que criam um Congresso...

Convictamente, já escrevi: a concretização de um Congresso SPE é o principal objetivo estatutário, como acontece e deve suceder em todas as associações científicas congéneres. É também a grande oportunidade de iniciação para muitos jovens estatísticos. Para além da partilha de ciência essa é uma das tarefas mais meritórias, sempre que nos reunimos em congresso.

O Congresso é assim, a reunião magna dos estatísticos portugueses!

O Congresso SPE é o resultado de uma soma de esforços: da Comissão Organizadora, da Comissão Científica e da Direção, que realizam uma iniciativa Presidencial mandatada pelos sócios e concretizam os grandes objetivos estatutários. Realizar congressos bienais é cumprir (apenas) metade do objetivo pois a máxima divulgação da Ciência Estatística e a reconhecida adesão de inscritos implicaria a periodicidade anual. É sempre oportuno refletir sobre este assunto.

Do saber de experiência feito aprendemos a melhor avaliar o esforço envolvido na concretização de um Congresso. Mas a soma das partes que concorrem torna mais fácil cumprir a missão. Todos os sócios e, muito em especial, os Seniores que sustentam a SPE e o seu sucesso (também em congressos) devem tomar a iniciativa de oferecer os seus préstimos facilitando assim o trabalho de pesquisa da Direção sobre “potenciais voluntários”. Invertemos assim o ónus da procura, em benefício da SPE. Decerto o trabalho da Direção fica ainda mais estimulado.

Este Boletim, detalhadamente, inclui a expressão da satisfação de missão cumprida em todo o trabalho da Comissão Organizadora do XXIII Congresso SPE. Nesta reunião magna sentiu-se renascimento. É um grande estímulo!

... com novos olhares sobre a Estatística; em espaço editorial iniciado há quase 40 anos...

O Boletim SPE, como se sabe, iniciou a sua publicação em 1979, numa primeira série. O formato atual foi começado em 2006. Passou uma década e, ao entrar numa nova década, o Boletim deve usar todo o seu passado para (também) com a ajuda dele melhorar (ainda mais) o seu futuro editorial. Os *Temas Centrais*, em cada assunto com o apoio editorial de um especialista sénior, e sempre numa perspetiva generalista e de divulgação, criaram edições em *Estatística Educacional*, em *Modelação e Testes* (como seria de esperar e nas suas mais recentes vertentes). Produziram reflexão sobre a *Investigação (em) Estatística* bem como *Estatística no Ensino Superior Politécnico*; além de muitas outras áreas temáticas que se encontram listadas na página *Retrospectiva* do Boletim.

Estas linhas, de algum modo, com os dois editoriais anteriores encerram, ou completam, uma trilogia de reflexão sobre os novos tempos da Estatística, em Portugal e no mundo. Em termos de pesquisar novos desafios para o século XXI, “olhámos” para o passado com o intuito de “melhor descobrir” e preparar o futuro.

... com a Estatística no centro. Investigação e Inovação é a alma...

O *Boletim outono de 2016* veiculou textos opinativos de investigadores seniores. E nessa edição ficou em aberto “continuar a temática com novos olhares”.

Feitos os convites, foi muito reconfortante para o Editor sentir a resposta da comunidade para que, em torno do *Tema Central da Estatística*, se pudesse fazer e continuar um ponto de situação para melhor refletir sobre as áreas que mais nos apelam e nos desafiam como investigadores e como comunidade científica de uma disciplina transversal na Ciência.

A cada autor, como sempre fiz e em sintonia com os líderes “mais seniores” de cada tema central, enviei um mesmo texto de convite associado a uma mesma mensagem convocatória. Com total liberdade nas suas opções, obviamente, enquadradas no tema central, os autores seguem os seus caminhos de reflexão e análise. Desta vez, numa perspetiva de futuro e com especial significado, os textos convergem em grandes temas para o século XXI como são *Data Science e Big Data*. É interessante esta convergência. São dois conceitos que, atualmente, surgem amiúde mas que aos mais seniores devem fazer recordar como *Data Analysis*, a Análise de Dados dos idos anos 70 e 80 do século passado, com toda a polémica que levou talvez duas décadas antes de “cientificamente” estabilizar. Aí se discutia a nível mundial a grande dicotomia estabelecida entre “os mais práticos” que “só faziam umas contas com computadores” – muito recentemente surgidos de tal modo que o próprio termo computador, nalguns casos, ainda carecia de significado – e “os mais teóricos” para quem “toda a investigação tem de ser suportada por uma boa teoria”. A Análise Exploratória de Dados era “quase uma ilegalidade”. Na minha memória surge aqui um célebre congresso sobre Análise Multivariada, em 1982, em Barcelona, onde “as duas fações” se reconciliaram e que eu tive a oportunidade de testemunhar – do ponto de vista muito prático, a escola francesa da Análise de Dados e a escola inglesa, dita mais pura, com “alguns Fisherianos” como defensores. As posições estavam extremadas de modo que “a tradição” (quase) não reconhecia valor ao trabalho científico de *Data Analysis*. Muitos textos sobre esta temática, também históricos, são fundamentais para nos ajudar a perceber “como se faz Ciência ao longo do tempo”. É assim que podemos integrar os novos desafios que agora surgem e que esta edição do Boletim nos apresenta, mas agora já quase 50 anos depois e onde a Ciência Estatística estabilizou em torno daquela polémica. Quais os novos grandes desafios de continuidade?

A comunidade, na sequência dos nossos autores, tem a palavra!

Investigação e inovação é a alma da Estatística e esse pode ser um bom lema para esta atividade editorial. Devo pois agradecer, devemos agradecer, a todos os autores do tema central pois, sempre são eles e só eles que, com toda a generosidade, “geram” cada uma das edições do boletim SPE. Mas, desta vez, muito em especial aquela atitude merece ser reforçada com um agradecimento pela partilha que nos apresentam das suas reflexões pessoais e que a todos nos envolvem. E essa partilha envolvente solicita e desafia a nossa opinião.

Os leitores e os Estatísticos Portugueses, de algum modo também em termos de “olhar para o futuro” devem sentir-se convocados para dar continuidade a esta edição.

No entanto, a propósito de inovação, os custos da edição no formato em papel bem como os modernos meios de comunicação solicitam e aconselham uma análise profunda sobre custos e benefícios. Assim, para consolidar a atividade editorial da SPE e dar-lhe ainda mais força é sempre o momento de reflexão e decisão – é este também “um novo olhar”.

O Tema Central do Boletim outono de 2017 engloba uma perspetiva de futuro.

O momento eleitoral decorrido neste mês de outubro, em particular o programa de candidatura renovada, é também um ponto de partida para novos projetos.

Data Science, *Big Data* e todas as reflexões que os autores nos apresentaram nesta edição são convergentes nos modernos tempos informáticos que se vivem em todos os ambientes; em particular nos meios de comunicação.

Estas linhas podem ser, também elas e de novo, um ponto de partida para um debate “mesmo dentro das páginas do Boletim” sobre a temática nelas envolvida e a necessária reflexão, na qual, já em alguns Editoriais anteriores, deixei palavras de abertura.

Qual o futuro da comunicação digital?

Quais as vantagens e as desvantagens?

Data Science – A Ciência dos Dados é um excelente tema central que se pode criar a partir das conclusões e reflexões apresentadas pelos autores nesta edição.

... com Boletim e / ou eBoletim...

O Boletim SPE, com toda a sua história, é um património dos sócios. É pois, muito importante para os sócios mas ainda mais para toda a comunidade científica, a divulgação do Boletim também para Bibliotecas, em especial as departamentais – que são os centros de divulgação que se encontram mais perto da comunidade e onde “a história científica vive e gera vida”. É nas Bibliotecas que o Boletim convida e se mostra à comunidade. Com um grande esforço a SPE tem levado a cabo essa realização estatutária em favor da Literacia Estatística.

Esta edição de outono integra notícias e temas que a tornam muito especial. Para além do tema central, o registo de prémios e a secção Notícias, ao cumprirem a sua função editorial de registo de atividades e acontecimentos está bem recheada e revela o dinamismo que se vive na comunidade – bem visível, como vimos, no enorme número de jovens estatísticos participantes no XXIII Congresso SPE. É um minimemorial.

A leitura dos registos temáticos e das reflexões que os “futuros seniores” nos proporcionam e que esta edição memoriza gera, de novo, a necessidade de reflexão sobre os modernos meios de divulgação da informação e em particular a dicotomia “papel versus digital” como já por vezes se salientou.

Os meios já intitulados de digitais, em contraposição com os tradicionais em papel, completam-se. Assim tem acontecido com a evolução na comunicação nos *media* em geral. Alguns, iniciando e experimentando uma componente digital mais forte já voltaram a reforçar as respetivas versões não digitais. A comunicação digital mostra-se “mais rápida” e quiçá “mais simples” mas (talvez por isso mesmo) “menos profunda” ou, se quisermos “mais efémera”.

Os *eBoletins*, sem dúvida alguma, têm o seu lugar como uma edição digital em qualquer meio de comunicação e são bem vindos. É um bem acrescido no aumento da literacia. Mas a sua unicidade obscura e desprotege um grande instrumento de divulgação em biblioteca; e esta é “onde está o património”. Além disso, como já se vê, a velocidade do digital e da comunicação vai, mais tarde ou mais cedo, transformar essa edição eletrónica numa “*página eletrónica*” o que em si não é um mal. No entanto, essa decerto não é a vocação de um meio de comunicação científica que se deseja de excelência. Tudo, bem o sabemos, deve ser feito para a promoção da Estatística, como nos diz o Estatuto da SPE que nos rege. Bem, muito bem, esta problemática, também é proposta no Programa de

ação da Direção recentemente eleita – e para uma melhor e mais dinâmica atividade de divulgação científica da SPE: É um novo desafio exigente à comunidade mas que vai dar mais valor à SPE; assim estejamos à altura de a ele responder e para tal tenhamos “engenho e arte”.

A grande vantagem de uma estratégia comunicacional bem construída (também) é a afirmação do ser Estatístico! Este é um grande desafio incluído na era digital em que vivemos e que se pode determinar nos textos do presente tema central bem como no centro das atividades do Dia Europeu.

O digital esteve sempre presente.

E isso é um bem. No entanto:

O “digital” conduz ao “faça você mesmo”?!

Diz-se que antes da ida ao médico, cada vez mais, os sintomas do paciente são “pesquisados e analisados exaustivamente em motores de busca”, cada um à sua maneira e de modo que melhor se adaptem à sua doença para melhor descobrir a cura.

E, quando vai ao médico, porque se de facto está doente – o problema – e se quer tratar, lá terá de ir, já sabe de que sofre e qual a medicina para a sua doença – a solução do problema. No entanto, perante a doença é o médico que sabe encontrar a cura.

Quem é o especialista perante a doença? É o médico!

E, num problema de Estatística, quem é o especialista?

Uma leitura cuidada dos autores nesta edição do Boletim permite fazer uma síntese de “olhar o futuro” e com as principais dificuldades dos (jovens) investigadores e estatísticos portugueses além de uma palavra de confiança na profissão que também envolve a confiança e a segurança nos seus estudos e resultados.

Os utilizadores “não estatísticos”, aqueles que “pesquisam e analisam exaustivamente em motores de busca” não sabem responder sobre a confiança dos seus estudos, dos seus resultados e das suas soluções.

O que são “bons resultados”? Aqueles que “satisfazem” quem os obtém? Ou aqueles outros que, com total apoio e baseados na Ciência Estatística lhes dão credibilidade? Aquela Estatística que a investigação fundamental faz nascer, crescer e desenvolver. É nela que os verdadeiros estudos se devem apoiar. Obviamente que sim!

Quem é o especialista em Estatística?

É o Estatístico! Aquele que a Ciência Estatística forma e que a SPE, de algum modo, inequivocamente e com autoridade, avaliza.

O Tema Central do próximo Boletim SPE será *Estatística Multivariada – perspectiva no século XXI*



Mensagem da Presidente

Caros sócios da SPE,

A acção da ainda Direcção da SPE tem-se pautado pela prossecução das linhas programáticas da nossa candidatura, enunciadas em 2014, a saber: contribuir para os objetivos gerais da SPE de “... *promover, cultivar e desenvolver, em Portugal, o estudo da Estatística, suas aplicações e ciências afins. ... unir todos os estatísticos e juntar pessoas trabalhando em diferentes áreas da estatística nas universidades, no setor privado e na administração pública*”, com ênfase na sensibilização pública para o poder e impacto da Estatística em todos os aspectos da sociedade e do conhecimento; na visibilidade da Sociedade através de acções com impacto na sociedade e no recrutamento de novos sócios quer na comunidade de estatísticos quer na comunidade mais alargada de profissionais que usam ou estão interessados em análise de dados e estatística.

Neste triénio, a SPE reforçou a sua presença junto dos mais jovens com as seguintes iniciativas: palestras em diversas escolas no âmbito da AEVAE; apoio incondicional à Explorística que tem continuado a percorrer o País; sessões públicas na FNAC em 2015 e 2016, na entrega de Prémios Estatístico Júnior; apoio (em 2015, 2016 e 2017) ao ENMath-Encontro Nacional de Estudantes de Matemática; atribuição de bolsas a jovens para participação nos Congressos de 2015 e 2017; instituição de Prémio Iniciação à Investigação (2016); atribuição do Prémio SPE em 2015, 2016 e 2017. A SPE esteve ainda envolvida em outras acções das quais destaco: a participação activa na organização das comemorações do European Statistics Day 2017 que teve como objectivo principal tentar sensibilizar o público em geral para o poder e impacto da Estatística e a presença em sessões sobre ensino e divulgação de matemática no Encontro Nacional da SPM 2016. Deu ainda apoio a diversas iniciativas da comunidade científica para além da organização de 2 Congressos nacionais e 1 Encontro Luso-Galaico. Está em boa marcha a organização de mais um Encontro Luso-Galaico já em 2018 e a organização de um curso no âmbito do European Courses in Advanced Statistics (ECAS) em 2019. A organização do XXIV Congresso da SPE será iniciada brevemente. Estabeleceu-se um acordo bilateral com a nossa congénere Alemã e reforçaram-se os laços com a CLAD. Num país da dimensão de Portugal e numa Sociedade da dimensão da nossa SPE, esta actividade só é possível pelo esforço e espírito de auto-sacrifício de muitos sócios.

Valeu a pena? Como diz o nosso maior poeta, “Tudo vale a pena quando a alma não é pequena”. Continuaremos pois a trabalhar e lutar pela Estatística em Portugal e pela afirmação da SPE.

A Sociedade é dos sócios e para os sócios e é, essencialmente, o que os sócios fizerem dela. Assim, a Direcção continua aberta a apoiar todas as iniciativas em prol da Estatística em Portugal.

Porto, 25 de Outubro de 2015

Cordiais saudações

Maria Eduarda Silva

Notícias

• XXIII Congresso SPE



Lisboa está na moda! Muitas são as pessoas que escolhem viver nesta cidade ou que simplesmente a escolhem para passear. Tem também sido o palco de muitos eventos, e de 18 a 21 de outubro foi anfitriã do XXIII Congresso da Sociedade Portuguesa de Estatística que decorreu nas instalações do Instituto Universitário de Lisboa – ISCTE-IUL.

Estes quatro dias de intenso trabalho contaram com 187 participantes que vieram de norte a sul de Portugal e ilhas, mas também de Espanha, do Brasil, de Itália, de Inglaterra, de França e da Alemanha. O programa científico do congresso foi composto por 4 Sessões Plenárias, 8 Sessões Temáticas com 25 apresentações, 76 comunicações orais e 38 *posters*.

As quatro Sessões Plenárias foram proferidas por professores de reconhecido mérito científico. A primeira foi proferida pelo Professor Kamil Feridun Turkman, da Faculdade de Ciências da Universidade de Lisboa, intitulada “*Point referenced methods and models for unemployment estimation*”. A Professora Paula Brito, da Faculdade de Economia da Universidade do Porto, proferiu a segunda plenária intitulada “*Multivariate Analysis of Histogram Data*”. Na terceira contámos com a presença do Professor Peter Smith, da Universidade de Southampton, que nos falou sobre “*The Administrative Data Research Centre for England: Making a Methodological Impact*”. Por último mas de todo em último, a quarta e última Sessão Plenária foi proferida pelo Professor Maurizio Vichi, do Department of Statistical Sciences da Sapienza University of Rome, sob o título “*Classification and Dimensionality Reduction in the Era of Big Data*”.



As oito Sessões Temáticas abordaram diversas áreas, *Data Mining*, Biometria, Estatísticas Oficiais, Séries Temporais, Aplicações em Econometria e Inferência Não-Paramétrica. Uma destas sessões foi organizada pela CLAD, com enfoque em *Inteligência Artificial* e em *Data Science*.

No que respeita às 76 comunicações orais, enquadraram-se no âmbito das diversas áreas da Estatística desde Classificação, Multivariada, Modelação, Análise de Sobrevivência, Controlo de Qualidade, Séries Temporais, Extremos, Bayesiana e Espacial, e abrangeram várias áreas de aplicação, tais como Saúde, Biologia, Ecologia e Ambiente, Ciências Sociais, Finanças e Econometria e Gestão.

Para além de todas as sessões orais referidas, foram também apresentados 38 posters, divididos em 2 Sessões de Posters.



Quarta-feira, 18 de outubro de 2017

O XXIII Congresso da Sociedade Portuguesa de Estatística iniciou-se no dia 18 de outubro de 2017, quarta-feira, com o habitual minicurso. Este ano, sob o título “Uma introdução à Meta-Análise” e proferido pela Professora Maria de Fátima Brilhante. Ao longo do minicurso, foi feita uma exposição global, sucinta e completa dos métodos mais usados e dos modelos a usar. Paralelamente, foram também apresentados os principais problemas, diversos exemplos e *software*. Toda a apresentação foi sempre acompanhada por imensas notas úteis e recomendações práticas.



Durante a tarde de quarta-feira, deu-se início às eleições para a nova direção da SPE.



A **Sessão de Abertura do Congresso** decorreu às 16:30 com a presença e as palavras de boas-vindas, da Presidente da Comissão Organizadora, Maria Fátima Salgueiro, da Presidente da Sociedade Portuguesa de Estatística, Maria Eduarda Silva e do Vice-Reitor do Instituto Universitário de Lisboa, Fernando Machado.



Ainda houve tempo para a primeira sessão plenária e o primeiro conjunto de comunicações orais que decorram em paralelo, em 5 salas. Este primeiro e intenso dia, terminou com um cocktail de boas vindas. Este, foi mais um alegre momento de confraternização onde pudemos encontrar ex-colegas, ex-professores, ex e novos participantes, ou seja, amigos que apenas diferem no grau de intensidade!

Quinta-feira, 19 de outubro de 2017

O segundo dia de trabalhos decorreu de forma muito intensa, tendo iniciado por um conjunto de comunicações orais em paralelo, em 5 salas, seguido pela segunda sessão plenária, depois pela primeira sessão de *posters*, pelo primeiro conjunto de sessões temáticas em paralelo, em 3 salas e finalizado por um novo conjunto de comunicações orais em paralelo, em 5 salas. No final da ordem de trabalhos, chegou o tão desejado momento de descontração, fomos visitar o Museu da Água da EPAL, mais concretamente o Reservatório da Mãe D'Água das Amoreiras. Para chegar ao reservatório fomos pelo interior do aqueduto das Águas Livres, subimos ao terraço onde pudemos avistar o castelo de São Jorge, contemplar a magnífica vista da cidade e perceber a importância desta obra de engenharia para o desenvolvimento de Lisboa.



A visita guiada culminou no edifício do reservatório. Mas antes de irmos embora enchemos as garrafas oferecidas pela comissão organizadora e provámos a água, não da fonte onde em tempos correu água, mas do bebedouro que se encontra perto da fonte.



Sexta-feira, 20 de outubro de 2017 - Dia Europeu da Estatística

No terceiro dia, o congresso da SPE associou-se às comemorações do *European Statistics Day*, cujo lema é “*Better Statistics. Better Lives.*”. Durante a manhã estivemos no Ministério das Finanças onde assistimos à conferência sobre “Estatísticas Oficiais, um Bem Público” que foi organizada pelo Comité Consultivo Europeu da Estatística (ESAC), a Sociedade Portuguesa de Estatística (SPE) e o Instituto Nacional de Estatística (INE), e também contou com o apoio do Eurostat.

A sessão de abertura decorreu com os discursos proferidos pela Presidente do Conselho Diretivo do Instituto Nacional de Estatística, Alda Carvalho, pela Presidente da SPE, Maria Eduarda Silva, pela Presidente do Comité Consultivo Europeu da Estatística, Ineke Stoop, pela Diretora-Geral Interina do Eurostat, Mariana Kotzeva e pela Ministra da Presidência e da Modernização, Maria Manuel Leitão Marques.



De seguida ouvimos David Hand, do Imperial College de Londres e deu-se início ao painel de discussão “*As Estatísticas após 2020: desafios e riscos*”, constituído pelo Presidente da Federação Europeia das Sociedades Nacionais de Estatística, Maurizio Vichi, pelo Diretor Geral do Departamento de Estatística do Banco Central Europeu, Aurel Schubert, pelo Membro do Conselho Diretivo do Instituto Nacional de Estatística, Carlos Coimbra, pela Diretora-Geral Interina, Mariana Kotzeva e pela estudante do Mestrado Europeu em Estatísticas Oficiais (EMOS), Universidade de Roma, Giorgia Zaccaria.



Durante esta manhã ouvimos dirigentes de instituições nacionais e internacionais falar sobre a importância das estatísticas oficiais, da constante preocupação em produzir estatísticas de elevada qualidade, da necessidade de cooperação entre as várias instituições para que se possa ter resultados comparáveis, e também do papel do jornalismo, dos jornalistas e dos políticos, bem como da necessidade de promover a literacia estatística.

Também foi uma oportunidade para ver uma das zonas mais emblemáticas da cidade de Lisboa - o Terreiro do Paço - e de apreciar o interior de um dos majestosos edifícios da cidade.

De volta ao ISCTE-IUL, e depois de uma sessão plenária e de sessões temáticas, chegou a altura da atribuição do Prémio SPE 2017 e Prémio Carreira. Estes são sempre momentos de grande expectativa. O prémio SPE 2017 foi atribuído ao jovem investigador Carlos Oliveira que nos apresentou de forma apaixonada o seu trabalho complexo intitulado “Tempo óptimo de paragem em processos com alteração em tempo aleatório”.

Nazaré Mendes Lopes (Maria de Nazaré Simões Quadros Mendes Lopes) foi distinguida com o Prémio Carreira SPE como reconhecimento do papel que desempenhou, no ensino, na investigação e na divulgação da Estatística. Nazaré partilhou, de forma emotiva, o seu merecido prémio com todos os que com ela trabalharam ou que de alguma forma a ajudaram a atingir os seus objectivos.



Depois da Assembleia Geral da SPE, seguiu-se mais um dos pontos altos, o Jantar do Congresso.

O Museu do Oriente foi o local escolhido para este momento de confraternização. Antes do jantar pudemos fazer uma breve visita à exposição do museu e apreciar a riqueza e pluralidade das tradições culturais do Ocidente e do Oriente. O mote perfeito para o momento de convívio que se seguiu, onde houve oportunidade para a troca de experiências profissionais e pessoais, acompanhados por um bom prato e num local descontraído com vista para o rio Tejo e para a ponte 25 de Abril.



Sábado, 21 de outubro de 2017

O último dia do congresso também foi repleto de atividades, com momentos de comunicações orais a decorrer em 4 salas em paralelo, seguidos da segunda sessão de *posters*, depois decorreu a última sessão plenária. A seguir ainda houve lugar para 2 sessões temáticas e um novo conjunto de comunicações orais.

Mesmo a terminar e, durante a atribuição do **Prémio Estatístico Júnior**, decorreu a apresentação intitulada “Paradoxo com grandes números”, proferida por Pedro Freitas. Foram entregues vários prémios a alunos do ensino básico, que se fizeram acompanhar pelos seus professores, familiares e amigos.



O XXIII Congresso terminou com a habitual **Sessão de Encerramento** presidida pela Presidente da Comissão Organizadora, Maria Fátima Salgueiro e pela Presidente da SPE, Maria Eduarda Silva.

Apesar do enorme cansaço, ainda nos brindaram com votos de um bom regresso e um profundo agradecimento a todos os que, de alguma forma, contribuíram para o enorme sucesso deste congresso.

No entanto, não acabou de imediato pois, para os resistentes, ainda houve um alegre lanche de convívio antes do regresso a casa!

Queremos agradecer à Comissão Organizadora do XXIII Congresso da SPE e a todos os voluntários o excelente acolhimento durante estes quatro dias!

Agradecemos a Soraia Pereira, Délia Gouveia-Reis, Anabela Rocha e Ernestina Baptista as fotografias que nos enviaram.

Conceição Ribeiro e Cláudia Marisa Silvestre

• Dia Europeu da Estatística 2017

O dia 20 de outubro, com celebração quinquenal, foi designado o Dia Mundial da Estatística pela Assembleia Geral das Nações Unidas em 2010.

O Dia Europeu da Estatística é comemorado anualmente, nos anos intercalares ao Dia Mundial da Estatística. O primeiro Dia Europeu da Estatística foi celebrado em 2016.

Em 2017, durante o XXIII Congresso da SPE, celebrou-se o Dia Europeu da Estatística com a realização da Conferência internacional “*Estatísticas Oficiais, Um Bem Público*”. Este acontecimento teve lugar na manhã do dia 20 de outubro, no salão nobre do Ministério das Finanças.

Essa Conferência, foi uma iniciativa do Comité Consultivo Europeu da Estatística e foi promovida pela Sociedade Portuguesa de Estatística, pelo Instituto Nacional de Estatística, pela Federação Europeia das Sociedades Nacionais de Estatística, pelo Sistema Europeu de Bancos Centrais e pelo Eurostat. Reuniu produtores e utilizadores das estatísticas oficiais, com palestras e um debate que contribuiu para reforçar a consciência de todos para a importância de dispor de estatísticas oficiais relevantes, fiáveis e facilmente acessíveis por todos os cidadãos. Houve ainda um painel de discussão, composto por produtores e utilizadores, que debateram os desafios e riscos para as estatísticas oficiais. Abaixo, apresenta-se o Programa e a Declaração Final.

“ESTATÍSTICAS OFICIAIS, UM BEM PÚBLICO”

PROGRAMA

08:30 – 09:00 | Receção dos participantes

09:00 – 09:45 | Sessão de abertura

Alda de Caetano Carvalho, Presidente do Conselho Diretivo do Instituto Nacional de Estatística

Maria Eduarda Silva, Presidente da Sociedade Portuguesa de Estatística

Ineke Stoop, Presidente do Comité Consultivo Europeu da Estatística

Helena Figueira, Diretora, Estatísticas Globais das Empresas, Eurostat

Maria Manuel Leitão Marques, Ministra da Presidência e da Modernização Administrativa

09:45 – 10:45 | Keynote Speech

David Hand, Investigador Sénior e Professor Emérito de Matemática, Imperial College, Londres

Debate

10:45 – 11:00 | Coffee break

11:00 – 12:30 | Painel de Discussão: “As Estatísticas após 2020: desafios e riscos”

Moderador: George Arnett, Jornalista *freelance* especialista em Estatística

Maurizio Vichi, Presidente da Federação Europeia das Sociedades Nacionais de Estatística

Aurel Schubert, Diretor Geral do Departamento de Estatística do Banco Central Europeu

Carlos Coimbra, Membro do Conselho Diretivo do Instituto Nacional de Estatística

Helena Figueira, Diretora, Estatísticas Globais das Empresas, Eurostat

Giorgia Zaccaria, Estudante do Mestrado Europeu em Estatísticas Oficiais (EMOS) Universidade de Roma

12:30 – 13:00 | Sessão de Encerramento – Declaração do ESAC

Maria João Valente Rosa, Membro do Conselho Executivo do Comité Consultivo Europeu da Estatística

European Statistics Day

Quality Data and Reliable Statistics: benefits for all Users and Citizens

Quality data and reliable statistics are a public good for everyone. In fact, official statistics:

- are vital drivers to strengthen democracy and to measure the progress of nations and regions;
- shape and are shaped by many individual and social activities in a citizen's daily life;
- play an essential role in the modern development of scientific knowledge;
- provide relevant evidence to support policy formulation and evaluation.

If quality data and reliable statistics are to become successful drivers for democracy, development, innovation, competition, prosperity and well-being, they need to be complete, relevant, coherent, accessible and timely. Official statistics already meet this high-quality standard. However, this is not enough and specifically we must also have:

Trust: all citizens need to trust the reliability of official statistics. Otherwise, there is a danger that demagogues, populists, parties and powerful groups with vested interests or simply illiterate statistical users with their own unchecked data will provide their false "truth" of what is going on across society. That is why strengthening trust in statistics and their appropriate analysis and use is essential for society to prosper.

Such trust must be firmly built on the independence of NSIs and all producers of official statistics from all external political or interest-group interference in the production of their data. This independence requires: clear, rigorous and harmonized statistical concepts and definitions; statistical methodologies based on scientific criteria only; professional independence of statisticians from third parties who commission statistical analysis; capacity to produce all the relevant statistics required to describe what is going on across society; adequacy of human and financial resources and autonomy of NSIs and other statistical agencies from public or private sector operators. Data production processes must be sensitive and responsive to society's needs, and therefore must be transparent and accountable, with confidentiality protected.

Skills: in the 21st century society, citizens need skills to understand and analyse statistics, including appropriate skills for interpreting reality, for decision-making based on empirical evidence and for developing well-founded opinions.

All citizens must acquire these competences at an appropriate level to ensure their inclusion in the Information Society. To achieve this, a stronger engagement is required between statisticians and citizens in relation to data access, communication and training. This is especially relevant for the new potential users at all ages and stages of life, with particular emphasis on fostering statistical literacy in schools. The networked knowledge society of the 21st century, as well as big data, mobile devices and remote sensing, also offer huge new challenges and opportunities for citizens in the decades to come to engage with official statistics.

In summary, the heart of our message for the European Statistics Day lies in the need for a commitment across Europe to increase the culture of statistics as a public good by combining five main statistical goals: independence, quality, trust, accessibility and literacy.

Key words: Quality Data and Reliable Statistics; Official Statistics; Statistical Science; Data Producers Independence and Professionalism; Data Trust and Literacy; Users of Statistics and Citizens

• O caso do Estatístico Grego: uma carta ao Primeiro-Ministro Alexis Tsipras

A Federação Europeia das Sociedades Nacionais de Estatística (FENStatS), em carta enviada ao Primeiro-Ministro Grego manifestou profunda preocupação pelos acontecimentos e procedimentos legais contra Andreas Georgiou, antigo Presidente da Autoridade Estatística Grega e outros altos funcionários dessa instituição.

Abaixo transcrevemos cópia da carta em notícia.

FR



<http://www.fenstats.eu/>

20 October 2017

To

His Excellency Mr. Alexis Tsipras, Prime Minister of the Hellenic Republic

Copies to

His Excellency Mr. Prokópis Pavlóπουλος, President of the Hellenic Republic,
His Excellency Mr Nikolaos Voutsis, President of the Hellenic Parliament.

Your Excellency Prime Minister Alexis Tsipras,

the Federation of European Statistical Societies (FENStatS), whose members are the National Statistical Societies from 23 European countries, wishes to express deep concern over the legal proceedings against Andreas Georgiou, the former president of the Hellenic Statistical Authority (ELSTAT), and other senior ELSTAT officials.

Andreas Georgiou returned to Greece in 2010 to serve your country as President of the then created ELSTAT. Significantly, in the five years preceding Mr. Georgiou's tenure, Eurostat had put reservations on Greek deficit and debt statistics six times in biannual reviews (Excessive Deficit Procedure), meaning Eurostat could not validate and certify the accuracy and overall quality of these statistics. In contrast, during Andreas Georgiou's 5-year term, Greek data did not receive any Eurostat reservation. The figures and the methodology of Andreas Georgiou are still used today by Greece in its official actions vis-à-vis the EU, and have been validated by Eurostat on 14 consecutive semi-annual Excessive Deficit Procedures.

However, this commitment of Andreas Georgiou to accurate statistics and adherence to international statistical principles and European statistical governance concerning all procedures of production and dissemination of official statistics led to a backlash in the form of a broadside of criminal and civil charges. On at least six separate occasions, various Greek judicial officials and panels have proposed or ruled that charges should be dismissed, only to have them resurrected in the wake of pressures. Indeed, in one recent trial, the court unanimously acquitted Andreas Georgiou of all charges, only to have another

prosecutor 10 days later annul the verdict and reorder a new trial on the same charges – a blatant violation of “double jeopardy” principles.

While it is not our aim to comment in detail on those legal proceedings, we wish to express our deep concern about the impact of their outcomes. Andreas Georgiou and his ELSTAT team have successfully established a well-functioning statistical infrastructure according to international standards. The legal convictions against these modern Greek statisticians are unjustified, they destroy their professional and private lives, they undermine international credibility that had only just been rebuilt and, last but not least, they support the incorrect belief that independent and impartial statistical information cannot and do not exist. These prosecutions also create disincentives for Greek statisticians to produce accurate statistics adhering to international statistical principles and European statistical law. Thus, after a longer period of progress and prosperous development, we fear that ‘Greek Statistics’ could return to the situation before 2010, regaining their previous reputation of scandals and misreporting.

Against this background, it should be underlined that we are disturbed by the fact that no legal action has been taken by Greek authorities to bring to account those responsible for the misreporting until 2009.

The goal of an official statistical agency is to describe reality, not create it and it must function independently from any kind of influence. The prosecutions against Andreas Georgiou and his colleagues raised alarms about the continuing politicization of statistics in Greece. We again respectfully urge that the Greek authorities halt any further prosecutions of Andreas Georgiou. We also urge that he be promptly reimbursed for all legal costs and that Greece publically apologize to him.

Confidence in the professional independence and competence of statistical authorities is of utmost importance for the functioning of democratic societies. We, the European national statistical societies, urge the Government of Greece to unambiguously take all actions necessary to publicly defend against any efforts to undermine the credibility of the Greek statistics produced by Andreas Georgiou and his former colleagues.

Sincerely yours;



Prof. Maurizio Vichi

President of the Federation of European National Statistical Societies

European National Statistical Societies that co-signed this letter

Prof. Marek Malý, President

Česká statistická společnost (Czech Statistical Society, ČStS)

Prof. Wolfgang Schmid, President

Deutsche Statistische Gesellschaft (German Statistical Society, DStatG)

Dr. Konrad Pesendorfer, President

Österreichische Statistische Gesellschaft (Austrian Statistical Society, ÖSG)

Prof. David Spiegelhalter, President

Royal Statistical Society (RSS)

Prof. Emilio Carrizosa, President
Sociedad de Estadística e Investigación Operativa (Spanish Society for Statistics and Operation Research, SEIO)

Prof. Maria Eduarda Silva, President
Sociedade Portuguesa de Estatística (Portuguese Statistical Society, SPE)

Prof. Monica Pratesi, President
Società Italiana di Statistica (Italian Statistical Society, SIS)

Prof. Gerard Biau, President
Société Française de Statistique (French Statistical Society, SFdS)

Prof. Matevž Bren, President
Statistično društvo Slovenije (Statistical Society of Slovenia, SSS)

Prof. Fred van Eeuwijk, President
Vereniging voor Statistiek en Operations Research (Netherlands Society for Statistics and Operations Research, VvS+OR)

Dr Eva Laczka, President
Magyar Statisztikai Társaság (Hungarian Statistical Association, HSA)

Prof. Czesław Domański, President
Polskie Towarzystwo Statystyczne (Polish Statistical Association, PSA)

Prof. Marcel Baumgartner, President
Schweizerische Gesellschaft für Statistik / Société Suisse de Statistique / Società Svizzera di Statistica//Societad Svizra da Statistica (Swiss Statistical Society, SGS-SSS)

Prof. Iveta STANKOVIČOVÁ, President
Slovenská štatistická a demografická spoločnosť (Slovak Statistical and Demographical Society, SŠDS)

Prof. F. Thomas Bruss, President
Société Belge de Statistique / Belgische Vereniging voor Statistiek (Belgian Statistical Society, SBS-BVS)

Prof. Jyrki Mottonen, President
Suomen Tilastoseura (Finnish Statistical Society, FSS)

Prof. Gabrielle Kelly, President
Irish Statistical Association (ISA)

Prof. Constantin Mitrut, President
Societatea Romana de Statistica (Romanian Society of Statistics, RSS)

Prof. John Öhrvik, President
Svenska statistikfrämjandet (Swedish Statistical Society, SSFr)

Prof. Claus Thorn Ekstrøm, President
Dansk Selskab for Teoretisk Statistik (Danish Society for Theoretical Statistics, DSTS)

Prof. Ksenija Dumičić, President
Hrvatsko Statističko Društvo (Croatian Statistical Association, CSA)

Prof. Biruta Sloka, President
Latvijas Statistiķu Asociācija the Association of Latvia's Statisticians (ALS)

Dr. Marie Lilleborge, President
Norwegian Statistical Association

• **Novos Órgãos Sociais na SPE**

O Processo Eleitoral para o triénio 2018-2020, para a Sociedade Portuguesa de Estatística, decorreu, durante as atividades do XXIII Congresso SPE, nos dias 18, 19 e 20 de outubro.

Eleitos

Direcção

Presidente: Maria Eduarda Silva, Universidade Porto

Vice-Presidente: Isabel Pereira, Universidade de Aveiro

Primeiro Vogal: Cláudia Philippart, Universidade de Lisboa

Segundo Vogal: Maria Esmeralda Gonçalves, Universidade de Coimbra

Tesoureiro: Conceição Amado, Universidade de Lisboa

Conselho Fiscal

Presidente: Graça Themido, Universidade de Coimbra

Primeiro Vogal: Carla Henriques, Instituto Politécnico de Viseu

Segundo Vogal: Maria João Polidoro, Instituto Politécnico do Porto

Mesa Assembleia Geral

Presidente: Maria Antónia Turkmann, Universidade de Lisboa

Primeiro Vogal: Carlos Macedo, Instituto Nacional de Estatística

Segundo Vogal: Russell Alpizar-Jara, Universidade de Évora

Programa de Candidatura

As linhas programáticas da nossa candidatura orientam-se segundo os objetivos gerais da SPE de “... promover, cultivar e desenvolver, em Portugal, o estudo da Estatística, suas aplicações e ciências afins. ... unir todos os estatísticos e juntar pessoas trabalhando em diferentes áreas da estatística nas universidades, no setor privado e na administração pública”, almejando contribuir para a prossecução de tais objetivos por parte de Sociedade.

Assim, pretendemos dar continuidade ao trabalho efetuado anteriormente no sentido de:

- aumentar a sensibilização pública para o poder e impacto da Estatística em todos os aspetos da sociedade e do conhecimento;
- aumentar a visibilidade da Sociedade através de ações com impacto na sociedade;
- aumentar o número de sócios através de uma estratégia de recrutamento quer na comunidade de estatísticos quer na comunidade mais alargada de profissionais que usam ou estão interessados em análise de dados e estatística;
- dar continuidade aos acordos, protocolos e iniciativas de colaboração com outras organizações nacionais ou internacionais (científicas ou não);
- assegurar o pleno funcionamento das comissões específicas e grupos de trabalho já constituídos ou a constituir, apoiando a divulgação do produto da sua atividade;
- fomentar a atividade editorial da SPE em língua portuguesa e promover a divulgação das suas publicações na comunidade lusófona.

Para a prossecução destes objetivos pretendemos:

- promover debates sobre temas de âmbito mais ou menos alargado e de inegável interesse para os sócios e para a sociedade;
- melhorar os meios tecnológicos de comunicação e interação entre os membros da SPE;
- promover ligações com outras sociedades, instituições e organizações que recolhem, analisam e usam dados, nomeadamente através da edição de um eboletim para membros e não membros; • prosseguir com a programação de encontros científicos.

• Secretaria da Sociedade Portuguesa de Estatística

Desde o passado dia 2 de outubro, na Secretaria da SPE desempenha funções a Dra Isabel Maurício. Com esta colaboradora foram iniciados novos moldes na Secretaria para atendimento: presencial e à distância.

O Horário de atendimento presencial é o seguinte:
Segundas e quartas, das 10 às 13 horas.

O Horário de atendimento à distância, para esclarecimento de dúvidas via email, spe@spestatistica.pt é o seguinte:
Terças e quintas: das 10 às 13 horas.

FR

• Avaliação de Unidades de I&D - Posição da Sociedade Portuguesa de Estatística

No âmbito de, e a propósito de Consulta Pública à Comunidade Científica sobre os Painéis de Avaliação de Unidades de I & D, e que decorreu no passado mês de Junho de 2017, a Direção da SPE enviou à FCT o seguinte documento:

À FCT

Assunto:

Consultas Públicas à Comunidade Científica: Projeto de Regulamento de Avaliação e Financiamento Plurianual de Unidades de I&D

Posição da Sociedade Portuguesa de Estatística

A Sociedade Portuguesa de Estatística, tendo por objectivos promover, cultivar e desenvolver, em Portugal, o estudo da Estatística, suas aplicações e ciências afins e tendo muitos dos seus associados como investigadores em Unidades de Investigação abrangidas pelo Projecto de Regulamento em epígrafe, considera da maior pertinência comentar um aspecto específico da proposta de regulamento: a constituição dos painéis de avaliação.

A estatística tem, no século XXI, um papel incontestável no desenvolvimento de I&D em todas as áreas científicas tradicionais e emergentes: do Data Science e Big Data, à indústria 4.0, passando pela medicina personalizada e investigação em genómica, é difícil imaginar um projecto de investigação que não dependa da recolha e análise de dados e onde, por isso, a estatística ocupa um lugar de destaque.

Assim, sendo a estatística a área mais transversal a todas as áreas temáticas de investigação reconhecidas pela FCT, o seu papel fundamental em I&D não tem sido reconhecido pela FCT, uma vez que aparece sistematicamente apenas como uma sub-área da área científica Matemática, resultando num menor financiamento de investigação em estatística: menos projectos, menos formação avançada, menor contribuição para o desenvolvimento de métodos avançados fundamentais à excelência na investigação científica.

A SPE vem assim pugnar junto da FCT pela inclusão de estatísticos reconhecidos nos seus painéis de avaliação (Unidades I&D, projectos e bolsas). A presença de estatísticos reconhecidos nos painéis de avaliação na área científica Matemática é crucial no reconhecimento da necessidade de desenvolver a área em Portugal. Por outro lado, é recomendável que sejam incluídos estatísticos especializados nos painéis de avaliação de áreas científicas com forte teor de experimental, como será o caso da "Biologia Experimental" ou de "Ciências Biológicas".

Certos da boa atenção que estas recomendações possam merecer junto da FCT,
Subscrevemo-nos

P' Direcção da SPE

• Ivette Gomes – Insígnias de Professor Emérito da UL

No passado dia 25 de setembro, na Reitoria da Universidade de Lisboa, decorreu a Cerimónia de entrega das insígnias de Professores Eméritos da Universidade de Lisboa.



Ivette Gomes foi um dos agraciados.

A cerimónia contou com as contribuições de José Artur Martinho Simões, Diretor da Faculdade de Ciências e da Professora Maria Isabel Fraga Alves.

Por mais esta distinção a juntar ao seu vasto e insigne CV, Ivette Gomes merece as nossas felicitações.

Uma vez mais e através da Professora Ivette Gomes - uma líder científica nacional e internacional, de todos, muito bem conhecida - a Estatística foi conduzida a lugar cimeiro, também pelo pioneirismo, na lista dos Professores Eméritos da Universidade de Lisboa.

FR

• FENStatS – A Federação Europeia das Sociedades Nacionais de Estatística

FENStatS é uma associação fundada em 2011.

A Federação promove a comunicação mútua, a cooperação e o intercâmbio de objetivos entre os seus membros, em termos de princípios científicos e em sentido geral.

A Federação serve como uma agência de divulgação de informação científica e técnica entre as Sociedades de Estatística que a compõem.

Esta Federação é uma associação que integra 21 Sociedades Europeias de Estatística:

As seguintes sociedades estatísticas são membros fundadores:

- **Česká statistická společnost** (República Checa, ČStS)
- **Deutsche Statistische Gesellschaft** (Alemanha, DStatG)
- **Österreichische Statistische Gesellschaft** (Áustria, ÖSG)
- **Royal Statistical Society** (Reino Unido, RSS)
- **Sociedad de Estadística e Investigación Operativa** (Espanha, SEIO)
- **Sociedade Portuguesa de Estatística** (Portugal, SPE)
- **Società Italiana di Statistica** (Itália, SIS)
- **Société Française de Statistique** (França, SFdS)
- **Statistično društvo Slovenije** (Eslovénia, SSS)
- **Vereniging voor Statistiek en Operations Research** (Holanda, VvS+OR)

Em 2013, aderiram:

- **Magyar Statisztikai Társaság** (Hungria, HSA)
- **Polskie Towarzystwo Statystyczne** (Polónia, PSA)
- **Schweizerische Gesellschaft für Statistik / Société Suisse de Statistique / Società Svizzera di Statistica//Societad Svizra da Statistica** (Suiça, SGS-SSS)
- **Slovenská štatistická a demografická spoločnosť** (Eslováquia, SŠDS)
- **Société Belge de Statistique / Belgische Vereniging voor Statistiek** (Bélgica, BVS)
- **Suomen Tilastoseura** (Finlândia, FSS)
- **Irish Statistical Association** (Irlanda, ISA)
- **Societatea Romana de Statistica** (Roménia, RSS)
- **Svenska statistikfrämjandet** (Suécia, SSFr)

E, em 2016, juntaram-se à FENStatS:

- **Dansk Selskab for Teoretisk Statistik** (Dinamarca, DSTS)
- **Hrvatsko Statističko Društvo** (Croácia, CSA)

FR

Presidente da SPE - Secretária Geral da FENStatS

Após eleições neste mês de outubro, a Prof. Maria Eduarda Silva, Presidente da SPE integra os órgãos dirigentes da Federação Europeia das Sociedades Nacionais de Estatística.

Para o Comité Executivo da FENStatS e para os próximos 3 anos, foram eleitos:

Presidente – Walter Radermacher, da Sociedade Alemã de Estatística

Vice-Presidente – Maurizio Vichi, da Sociedade Italiana de Estatística

Secretária Geral – Maria Eduarda Silva, Presidente da Sociedade Portuguesa de Estatística

Tesoureira – Maria Dolores (Lola) Ugarte, da Sociedade Espanhola de Estatística

Webmaster– Gejza Dohnal, da Sociedade Checa de Estatística

Assim, durante os próximos 3 anos, a Sociedade Portuguesa de Estatística tem responsabilidade acrescida e um maior protagonismo no âmbito da FENStatS.

FR



A **Secção de Biometria da SPE (SBioSPE)** foi novamente convidada a participar numa sessão temática do congresso bienal da *Sociedad Española de Biometría (SEB)*. Isto é, a *XVI Conferencia Española de Biometría (CEB 2017)*, que decorreu de 13 a 15 de Setembro de 2017 na cidade de Sevilha, na Faculdade e no Instituto (IMUS) de Matemáticas da Universidade de Sevilha.

A SBioSPE viu-se representada numa sessão convidada conjunta com as Sociedades de Biometria congéneres: *Eastern Mediterranean Region (EMR)* e *Società Italiana di Biometria (SIB)*. Esta sessão foi aberta pelo presidente da sessão e vice-presidente da *SEB*, Klaus Langohr (*Universitat Politècnica de Catalunya*), depois dos agradecimentos acolheu Euloge Kenne Pagui (*Università di Padova*) que falou sobre um método de redução do viés da mediana, Isaac Meilijson (*Tel Aviv University*) que revisitou o algoritmo EM e a estimação dos erros padrão envolvidos no caso de dados incompletos e Clelia Di Serio (*Università Vita-Salute San Raffaele, Milano*) que apontou vários desafios para a amostragem e modelação de sistemas complexos na biomedicina moderna. Por último, **Isabel Natário** (Universidade NOVA de Lisboa), em representação da SBioSPE, apresentou os desafios atuais na modelação de prevalência para dados provenientes de delineamentos amostrais complexos. Para registo futuro tirou-se uma foto dos participantes desta sessão!

O programa científico foi abrilhantado pela participação dos oradores Inmaculada Arostegui (*Euskal Herriko Unibertsitatea*), que discorreu do ponto de vista estatístico sobre regras de predição clínicas, de Sonjia Greven (*Ludwig-Maximilians-Universität München*) que apresentou uma moldura genérica para a modelação por regressão funcional e de Dankmar Böhning (*Southampton Statistical Sciences Research Institute, UK*) que revelou uma modificação do estimador do limite inferior de Chao para distribuições de contagem truncadas em zero e inflacionadas em um.

Foram ainda apresentados inúmeros trabalhos propostos refletindo desenvolvimentos atuais e aplicações da Biometria a problemas da vida real. Houve lugar a uma sessão especial para estudantes, onde foram apresentados trabalhos sobre AUC em modelos de predição logísticos, ensaios clínicos, sobrevivência, mapeamento de doenças e estudos longitudinais.

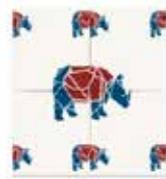
Foi assim muito enriquecedor o CEB 2017, com participantes de várias nacionalidades, ainda que a grande maioria fossem espanhóis, em que foi possível trocaram-se ideias e discutiram-se resultados vários de investigação na área da Biometria.



Isabel Natário e Giovanni Silva

• 11th European Conference on Mathematical and Theoretical Biology

A Conferência **ECMTB 2018** realiza-se em Lisboa, de **23 a 27 de Julho de 2018** na Faculdade de Ciências e Tecnologia da Universidade de Lisboa, sendo anfitrião o centro de investigação CMAF-CIO



ECMTB 2018
LISBON

Será um acontecimento principal do **YEAR OF MATHEMATICAL BIOLOGY (YMB; <http://euro-math-soc.eu/year-mathematical-biology-2018>)**, promovido pela European Society for Mathematical and Theoretical Biology (ESMTB) e pela European Mathematical Society (EMS). Por essa razão, a ECMTB 2018 será, pela primeira vez, uma conferência conjunta ESMTB-EMS e será coorganizada pela Sociedade Portuguesa de Matemática. A Sociedade Portuguesa de Estatística é uma das instituições patrocinadoras.

Convidamos todos os investigadores e estudantes interessados em Biologia Matemática e suas aplicações a se juntarem a esta excitante conferência! As inscrições estão abertas na página web da conferência <http://www.ecmtb2018.org>. Também estão abertas as candidaturas a Mini-Simpósios e a submissão de comunicações orais e em poster, estando os formulários para os respetivos Resumos também disponíveis na página web.

INFORMAÇÃO IMPORTANTE e PRAZOS:

Mini-Simpósios: candidaturas abertas!

Data limite de candidatura: 15 de novembro de 2017

Notificação de aceitação: 15 de janeiro de 2018

Comunicações orais e em poster: submissão de Resumos aberta!

Data limite de submissão: 20 de fevereiro de 2018

Notificação de aceitação: 2 de abril de 2018

Pagamento da taxa de inscrição:

Para beneficiar de tarifa reduzida, deverá inscrever-se antes de 15 de abril de 2018. Para se manter a par das últimas novidades sobre a ECMTB 2018, pode seguir-nos em <https://www.facebook.com/ecmtb2018/>

Esperamos encontrá-los em Lisboa para a ECMTB 2018

Maira Aguiar, Carlos Braumann, Nico Stollenwerk
(Conference Chairs)

• Sobre a sessão de entrega dos Prémios “Estatístico Júnior 2017”

A Sociedade Portuguesa de Estatística promove anualmente o Prémio Estatístico Júnior, com o patrocínio da Porto Editora.

Com esta iniciativa pretende-se incentivar o interesse pelas áreas de Probabilidades e Estatística dos estudantes dos Ensinos Básico e Secundário, e dos Cursos de Educação e Formação (CEF) e de Educação e Formação de Adultos (CEFA).

O Prémio Estatístico Júnior (PEJ) distingue anualmente sete trabalhos e é atribuído aos estudantes que os realizaram e a alguns dos professores orientadores, sendo a sua entrega formal realizada numa sessão que lhe é expressamente consagrada.

A entrega dos PEJ 2017 decorreu numa sessão especial, integrada nos trabalhos do XXIII Congresso SPE e da qual, mais adiante neste Boletim, damos o devido destaque.

O Júri dos Prémios Estatístico Júnior 2017 integrou a Doutora Maria Eugénia Graça Martins e a Doutora Manuela Neves.

• Prémio SPE 2017

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.

Júri do Premio SPE 2017:

- Prof. Paula Manuela Lemos Pereira Milheiro de Oliveira (Presidente), da Faculdade de Engenharia da Universidade do Porto.
- Prof. Marco André da Silva Costa, da Escola Superior de Tecnologia e Gestão de Águeda, Universidade de Aveiro.
- Prof. Patrícia Andreia da Silva Filipe, da Escola de Ciências e Tecnologia da Universidade de Évora.

O Prémio SPE 2017 foi atribuído a Carlos Miguel dos Santos Oliveira, estudante de Doutoramento em Matemática no Instituto Superior Técnico da Universidade de Lisboa.

No final desta edição do Boletim damos o devido destaque desta notícia.

• Sessão Prémio Carreira da SPE

O Prémio Carreira – SPE foi instituído, em 2013, pela Sociedade Portuguesa de Estatística (SPE) e propõe-se reconhecer a atividade de estatísticos portugueses com papel de relevância no desenvolvimento científico, pedagógico e de divulgação da Estatística em Portugal.

Em 2017, o Prémio Carreira foi atribuído à Doutora Maria de Nazaré Mendes Lopes numa sessão integrada no programa do XXIII Congresso da SPE, na tarde de 20 de Outubro de 2017.

A sessão iniciou-se com algumas palavras proferidas pela Presidente da SPE sobre o Prémio e a sua atribuição, seguida de uma breve intervenção do Professor Dinis Pestana, membro do júri, sobre a relevância da atividade científica da Doutora Nazaré Mendes Lopes, destacando em particular a sua enorme influência na criação e desenvolvimento do Grupo de Probabilidades e Estatística da Universidade de Coimbra. Em seguida, foi feita uma apresentação sucinta dos principais elementos biográficos e profissionais da Doutora Nazaré Mendes Lopes.

A Doutora Nazaré Mendes Lopes tomou então a palavra e manifestou a sua gratidão e reconhecimento por esta distinção, partilhando este Prémio com Professores, Colegas e Familiares que sempre a apoiaram e nela confiaram ao longo do seu percurso profissional. Antigos alunos e instituições, como o DMUC e o DEIO, foram também mencionados.

A Presidente da SPE entregou então o diploma do Prémio Carreira – SPE e uma lembrança comemorativa à Doutora Nazaré Mendes Lopes que é, a partir de agora, Sócia Honorária da SPE.

No final desta edição é apresentado um testemunho.

Enigmística de mefqa

$$e^Y e^X e^a e^b e^c$$

AMOSTRA

No Boletim SPE primavera de 2017 (p.12):

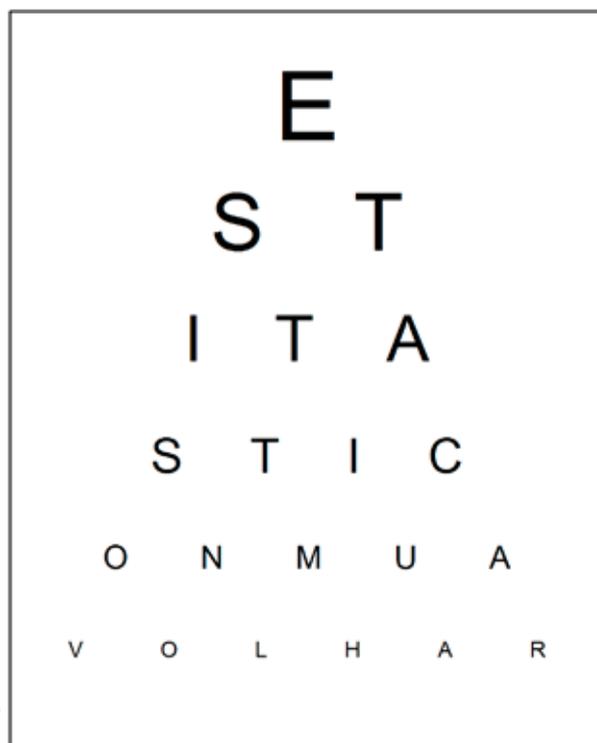
ESTIMADOR

MÉDIA

Estimador de mínimos quadrados

Média aparada

Tema Central da Estatística - um novo olhar



A Estatística, a indústria e a inovação

Cláudia Nunes, cnunes@math.tecnico.ulisboa.pt

IST - CEMAT

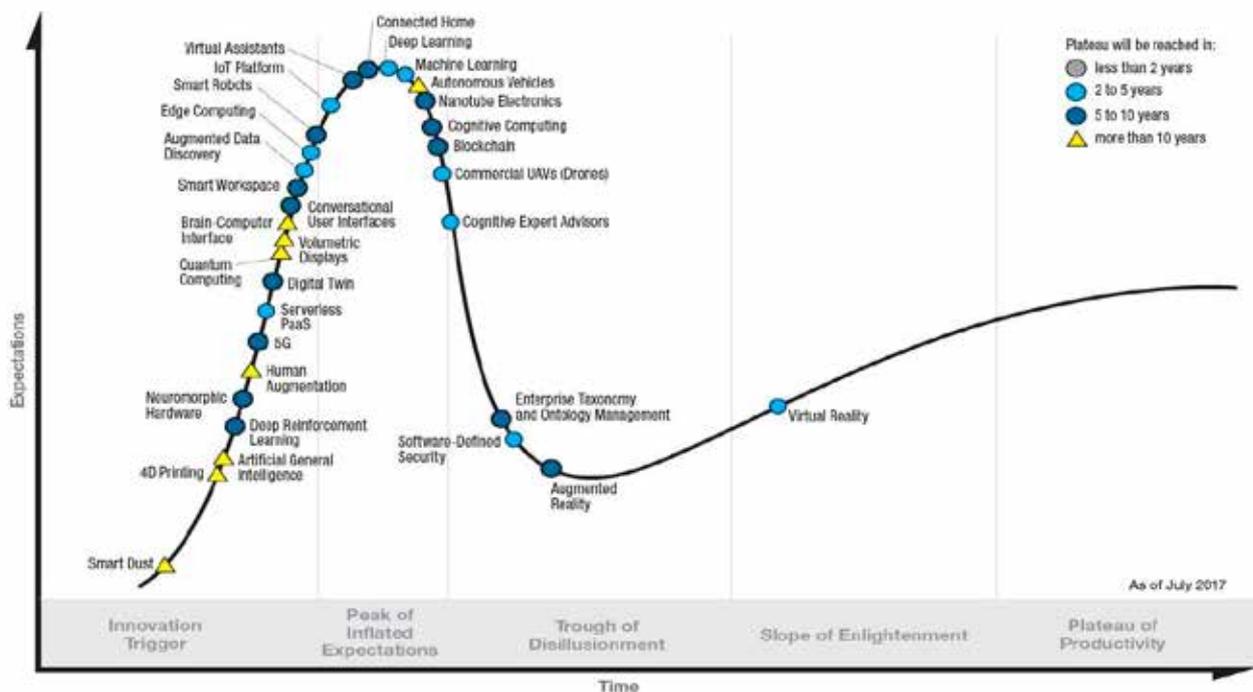
O desafio colocado pelo editor deste boletim era claro: mostrar um novo olhar sobre a Estatística, no seu significado mais amplo. Mas a resposta não é tão clara...

Antes de sermos matemáticos, probabilistas/estatísticos, somos cidadãos e elementos de uma sociedade global, sempre em movimento. Mas esta sociedade global evoluiu, altera-se, contradiz-se, apoia, sempre num ritmo frenético, no qual já todos temos algumas dificuldades em seguir. Por isso o olhar de hoje pode mudar radicalmente amanhã ou depois. E isto é um desafio, quer a nível profissional quer a nível pessoal.

Tempos houve em que era possível planear a nossa pesquisa, o nosso estudo, o nosso trabalho tendo (quase) a certeza que o olhar da altura estaria ainda actual. Mas será que agora podemos dizer o mesmo? Parece que o relógio do tempo não tem uma velocidade constante... parece que está sempre a passar cada vez mais depressa!

A esse propósito, recomendo uma visita ao site <http://www.gartner.com/smarterwithgartner>, onde poderemos encontrar uma previsão para quais os tópicos mais relevantes a nível de ciência e tecnologia para os próximos tempos. E todos os anos há actualizações! Por exemplo, a previsão divulgada em 2017 continua a apontar “*machine learning*” como um dos tópicos em expansão. E a comunidade estatística tem de reclamar a sua posição neste (e noutros temas).

Gartner **Hype Cycle** for Emerging Technologies, 2017



Mas entre toda esta incerteza em que vivemos, teremos sempre alguns pilares, verdadeiras tábuas de salvação para os nossos planos. Por isso eu reformulo a questão do nosso editor, colocando-a da seguinte forma: qual (ais) são os desafios e os pilares da Estatística no futuro?

Pessoalmente acredito que a resposta não é única, mas um dos pilares que eu rapidamente identifico é a Indústria. E é sobre isto que eu vou partilhar com os leitores a minha visão.

Nos dias de hoje a indústria apresenta inúmeros desafios à matemática e à estatística. Cabe-nos termos atitudes pró-activas e, simultaneamente, pró-activas.

E estaremos preparados para tal? Claro que sim, mas receio que as nossas instituições académicas não estejam ainda, nomeadamente os nossos pares que nos avaliam. No seio da comunidade matemática ainda prevalece um sistema de castas, em que a casta superior é a chamada “matemática pura” e as aplicações pertencem a uma casta inferior.

Esta situação é comum na maior parte das universidades europeias, com quem muitas vezes nos comparamos. Mas há cada vez maior percepção que esta é a atitude errada. E a prova disso reside inclusivamente nas políticas de financiamento de projectos a nível europeu que, infelizmente, não temos aproveitado na sua globalidade.

No Horizon 2020, o programa FET (Future and Emerging Technologies) identifica várias áreas prioritárias, não estando explicitamente a Estatística. Mas as oportunidades estão lá! A estatística está quase em tudo, e cabe a nós mostrar como a nossa intervenção é fundamental.

Se não o fizermos, outros o reivindicarão. É o caso do agora muito na moda *machine learning vs data science*. Em muitas universidades (quer em Portugal quer fora de Portugal) estão a ser criados programas de mestrado liderados por... informáticos! E porquê eles? Certamente por muitas razões, que ultrapassam a minha razão e o âmbito deste pequeno artigo de opinião, mas talvez haja alguma responsabilidade da nossa comunidade. Eles tomaram a dianteira; nós ficámos para trás. Não fomos pró-activos...

A indústria e as nossas ligações, em grupos de especialistas de diversas áreas, quer da estatística, quer das probabilidades, e da matemática, pode levar-nos muito longe. A indústria começa a estar preparada para nós. E nós temos de estar preparados para a indústria. Precisamos ser pró-activos...

Já há em Portugal e na Europa uma sensibilidade mais apurada para este tema. A nível nacional foi recentemente lançada uma rede de centros de matemática designada por “PT-MATHS-IN”, membro integrante de uma organização europeia designada por “EU-MATHS-IN”. E em ambos a palavra chave é IN: IN de “Industry and Innovation”.

Ainda a nível europeu, o ECMI (European Consortium for Mathematics in Industry) tem liderado este processo (lento, é certo...) de alertar a comunidade científica matemática para o seu papel relevante a nível da indústria e da inovação, ao mesmo tempo que mostra à indústria (alguma ainda renitente) os casos de sucesso da matemática. O ECMI tem apostado em “moldar” as novas gerações, promovendo actividades de MSO (modeling, simulation and optimization) em alunos de mestrado (as designadas Modelling Weeks) e em académicos (os designados ESGI-European Study Groups in Industry).

Mas lamentavelmente o envolvimento de estatísticos nestas organizações e programas ainda é reduzido, muito reduzido! E a indústria reconhece, sem qualquer hesitação, a importância da estatística. Paradoxalmente, reconhece até como sendo mais importante a estatística que outras áreas da matemática.

Por isso, do que estamos nós à espera? Continuará a ser necessário pesquisa fundamental, mas precisamos urgentemente de outro olhar: não estamos sozinhos no nosso confortável gabinete. Estamos rodeados de pessoas, com outros interesses, com muitos desafios, e devemos ocupar a nossa posição charneira, que a Estatística, a Indústria e a Inovação merecem!

E é este o meu “novo” olhar da estatística...



Data Science, Big Data e um novo olhar sobre a Estatística

Isabel Fraga Alves, *isabel.alves@fc.ul.pt*

CEAUL, DEIO, FCUL

Recentemente, tem havido muita discussão sobre o papel da Estatística nas áreas de Data Science (DS) e Big Data (BD). A título de nota de rodapé, embora o Glossário Inglês-Português de Estatística SPE-ABE não tenha entrada para estas designações em língua portuguesa, parece consensual a denominação de “Ciência de Dados” para “Data Science”, e em paralelo a não tradução de “Big Data”. Vejam-se a este propósito os cursos de especialidade que vão surgindo aqui e ali, como é o caso da região da Grande Lisboa, com os recém-criados *Mestrado em Ciências de Dados* (FC-UL), na sequência da *Pós-Graduação em Data Science*, o *Mestrado em Engenharia e Ciência de Dados* (IST-UL), ou ainda o *Mestrado Análise e Engenharia de Big Data* (FCT-UNL).

DS não é “apenas” Estatística, contemplando efectivamente áreas que tradicionalmente são consideradas fora do âmbito da investigação estatística.

Recentemente, no ISI2017 (Marrakech) tomei conhecimento com uma posição algo extremada de Hadley Wickham, um cientista-chefe no RStudio e Professor de Estatística das Universidades de Auckland (NZ), Stanford (USA) e Rice (USA). O seu trabalho inclui R Packages para DS (o tidyverse: incluindo ggplot2, dplyr, tidyr, purrr e readr). Conferencista na ISI President’s Invited Keynote Lecture, quando questionado por Pedro Silva, Presidente da ISI, acerca do que considera constituir a maior diferença entre Estatística e Data Science a resposta de Wickham foi curta e no mínimo polémica: “*Data Science is useful Statistics*”.

Dois pontos fulcrais constituem o motor desta área emergente: por um lado, a natureza interdisciplinar para a solução dos problemas em DS, integrando cientistas num leque de campos como a Estatística, Bases de Dados, através de questões relevantes em campos como a biologia molecular, geofísica, astronomia, física, meio empresarial e tantos outros, chama ao palco ramos como a bioestatística, geoestatística e ciências informáticas, desejavelmente actuantes em equipa; por outro lado, e neste cenário, é exigida aos futuros estatísticos uma constante actualização da literacia computacional. Estes pontos centrais devem ser tomados em consideração pelos jovens alunos, investigadores e futuros profissionais, no sentido de que será aconselhável que o seu percurso de formação integre conhecimentos de novas ferramentas computacionais, em constante mudança, e preferencialmente em instituições que ofereçam exactamente esta diversidade de competências.

Na minha óptica, a área emergente de DS/BD pode-se revestir de um binómio quer de ameaça, quer de oportunidade e desafio. Parece-me inquestionável que a área da Estatística constitui uma parte crucial da DS/BD, mas simultaneamente, a realidade portuguesa dos Departamentos de Matemática/Estatística do nosso ensino superior e o seu contínuo envelhecimento, não acompanhado de uma renovação dos seus quadros de jovens professores, pode ter implicações drásticas na relevância destes departamentos neste novo cenário.

As etapas num projecto em DS deveriam passar por aquisição de dados, face a questões que são pertinentes para o problema em estudo, a sua análise envolvendo formas de visualização e modelação e por fim a comunicação dos resultados. Ora na prática nem sempre o curso dos acontecimentos decorre assim; é cada vez mais usual, face a quantidades massivas de dados (BD) que nos inundam no nosso dia-a-dia, extrair conclusões sobre problemas postulados *a posteriori*, deixando para segundo plano exactamente um bom planeamento da experiência. Boas questões são cruciais para uma boa análise de dados, mas quanto a mim não tem havido em investigação o investimento desejável exactamente acerca da condução mais correcta para esta formulação de problemas.

Acresce ainda que qualquer análise de dados reais envolve também uma maior ou menor manipulação de dados, de forma a tornar viável uma conveniente modelação, complementada e desafiada por boas ferramentas de visualização.

Contudo, o grande peso em investigação no ramo da Estatística recai na modelação e metodologia, muito menos em visualização e muito pouco na comunicação dos resultados. A este propósito, relembro Holger Rootzén que no EVA2013 Vimeiro fez exactamente este alerta acerca da importância de uma boa transmissão de conclusões para um público-alvo não especializado.

Em concordância com Wickham (2014), julgo que uma análise não tem sentido, a menos que convença alguém a agir. Em gestão, isso normalmente significa convencer a alta administração que usualmente não está equipada de grande conhecimento estatístico. Em investigação científica, porém, normalmente isso significa convencer os *reviewers* do periódico onde se pretenda publicar.

Retomando a (im)preparação de alguns dos Departamentos/Secções de Estatística nesta área de DS, esta não é alheia a uma política estratégica de (de)crescimento curricular/académico: historicamente a ascensão académica, essencialmente baseada na publicação científica em revistas científicas com forte Factor de Impacto na área da Estatística, não se compadece com este novo paradigma da DS. O contacto com a realidade portuguesa, por exemplo, quer a nível individual de avaliação curricular, quer mesmo de avaliação de Centros de I&D, “condenou” à depreciação durante demasiado tempo a chamada Estatística Aplicada, fortemente apoiada em métodos computacionais. E isso em favor dos chamados desenvolvimentos teóricos e metodológicos, com lugar de destaque na cena da publicação científica, contrastando muitas vezes com um desrespeito pelos estatísticos aplicados (veja-se Jeff Leek, 2014), por parte dos chamados peritos em Big Data.

Mas, como disse acima, estamos submersos em quantidades massivas de dados (BD), armazenados por organizações em todo o mundo ao longo de décadas e a capacidade de aceder e analisar estas bases de dados revela-se cada vez mais importante. Mark Mulcahy (2017) refere alguns:

- 2.7 Zetabytes de dados existem hoje no mundo digital.
- 235 Terabytes de dados foram coligidos pela U.S. Library of Congress em April 2011.
- A administração Obama investiu is 200 milhões\$ em projectos de investigação em Big Data.
- Facebook armazena e analisa 30+ Petabytes de dados gerados por usuários.
- Mais do que 5 biliões de pessoas estão a fazer chamadas, digitar texto ou tweeting em telemóveis em todom mundo.
- Em 2008, o Google processava 20,000 terabytesde dados (20 petabytes) por dia.
- 571 novos websites são criados em cada minuto.
- Marcas e Organizações no Facebook recebem 34,722 Likes por minuto.

Devemos nós depositar uma fé cega nestas bases de BD? Estas e outras perguntas são alvo de análise crítica por parte de uma cientista em DS, Cathy O'Neil, que não posso deixar de referir neste artigo de opinião. Cathy O'Neil, matemática pura de formação, e após passar pelo RiskMetrics no sector financeiro, infectiu a sua área de investigação para DS. Recentemente tem vindo a difundir apresentações públicas sob o título “*The era of blind faith in big data must end*”, publicitando o seu livro de *Weapons of Math Destruction*, escrito de forma bastante coloquial.

Convido-vos, a todos que tiveram a persistência de ler este meu apontamento até ao fim, a lerem as referências abaixo, deixando a mensagem final de que é pertinente

Um Novo Olhar Sobre A Estatística.

Referências

- Jeff Leek (2014). Why big data is in trouble: they forgot about applied statistics - <https://simplystatistics.org/2014/05/07/why-big-data-is-in-trouble-they-forgot-about-applied-statistics/>
- Mark Mulcahy (2017). Big Data – Are You In Control? *Waterford Technologies*. <https://www.waterfordtechnologies.com/big-data-interesting-facts/>
- ISI2017 Marrakech - Daily news 20 July 2017, Thursday
<http://payment.isi2017.org/wp-content/images/ISI2017-Marrakech-DailyNews-5-Thu20072017.pdf>
- Piercesare Secchi (2017). On the role of statistics in the era of big data: a call for a debate, *MOX-Report No. 35/2017*.

Cathy O'Neil (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown Publishers.

Cathy O'Neil (2017). The era of blind faith in big data must end, *talk presented at an official TED conference*.

https://www.ted.com/talks/cathy_o_neil_the_era_of_blind_faith_in_big_data_must_end/transcript#t-39368

Hadley Wickham (2014). Data science: how is it different to statistics? *IMS Bulletin online*.



Estatística - “Espelho meu, espelho meu, que futuro terei eu?”

Conceição Amado, conceicao.amado@tecnico.ulisboa.pt

*Departamento de Matemática e CEMAT
Instituto Superior Técnico
Universidade de Lisboa*

A leitura do Boletim de Outono de 2016 da nossa Sociedade revela importantes reflexões sobre a Estatística e o seu futuro. Artigos escritos com grande rigor pelos colegas Paula Brito (2016, pag.40) e João A. Branco (2016, pag.49) evidenciam a atualidade e importância desta área do saber. Por outro lado, dificultam a minha tarefa de retornar a este assunto.

Não há dúvida que uma das características inerentes ao ser humano é a de querer prever o futuro. Desde a antiguidade que o Homem tentou antecipar o que o futuro lhe reservou. Os oráculos Chineses ou os adivinhos da Mesopotâmia antiga tentavam prever o efeito incerto das ações humanas, desde a sorte no casamento ao resultado de uma batalha. Em relação à Estatística, a questão do seu futuro não é atual, pois sempre que há mudança de paradigma esta discussão imerge. Tomemos, como exemplo, as reflexões de vários estatísticos, Hotelling (1940) Kendall (1942, 1968), Tukey (1962) e Yates (1966), durante a segunda guerra mundial ou o início do poder computacional. Em 1942, Kendall no início de um dos seus primeiros artigos sobre o tema do futuro da Estatística escreve: “*We statisticians, I suppose, are naturally diffident about extrapolating from the experience of the past into the future*”.

Na denominada Era da Informação, qual será o futuro e papel da Estatística? Muitos, provavelmente, assustados com os termos travestidos – *Big Data, Data Science, Machine Learning* - prevêm-lhe um futuro negro. Mas será que é realmente assim?

Na discussão do artigo de Kendall (1942), George Yule menciona o seguinte:

“I propose accordingly to traverse Mr. Kendall's proposition that "the estimation of properties of a population from a sample is the most important practical problem in statistics and is long likely to continue so." It never was, in my opinion, the most important practical problem in statistics, and so cannot continue to be that which it never was. I put aside the point that the most important practical problem for the statistician is really how to get his sample, for the sample is evidently assumed to have been obtained. The initial problem of the statistician, or of statistical method, if we like to personify it, is then simply the description of the data presented; to tell us what the data themselves show.”

Atualmente, não se pode colocar de lado a questão da importância da seleção da amostra. Já que a maioria dos dados disponíveis não são selecionados com o método estatístico. Os dados perseguem-nos quando ligamos um computador, transportamos um telemóvel ou um *smartwatch*, e, à primeira vista, parece que não precisamos de os selecionar!

Esta questão é, na minha opinião, uma das fundamentais e atuais com que a Estatística se tem de preocupar. Há, hoje em dia, um perigo de mudança das ciências dedutiva e indutiva para a extrapolativa. Com a quantidade de dados massivos disponível corre-se o risco de, sem a intervenção da probabilidade na sua análise, tomar decisões, criar leis apenas com base na descrição e na

extrapolação. Um dos exemplos mediáticos desta análise extrapolativa é a devida a Marissa Mayer¹, a diretora executiva da Yahoo, que em 2013, através de uma análise dos dados das ligações em rede dos seus colaboradores em teletrabalho, chegou à conclusão que as pessoas só trabalhavam metade do tempo. Esta análise de “*Big Data*” conduziu à proibição do teletrabalho na Yahoo e a uma grande polémica sobre os direitos dos trabalhadores.

O impacto nas diversas áreas do trabalho e da ciência que o desenvolvimento do mundo digital opera motivado pelo tratamento de um enorme volume de dados, onde a importância da maioria dos dados nem sempre é confirmada, é uma das questões importantes para a análise Estatística. O desenvolvimento de metodologias de análise da fiabilidade da informação será fundamental.

Para além desta questão, e como tem sido habitual na investigação estatística, mas atualmente com maior relevo, é a da necessidade de a Estatística se fortalecer com novas ferramentas teóricas quando a teoria atual se baseia em pressupostos que não são concretizados na prática.

Numa das comunicações mais importantes dos últimos tempos, David Donoho (2015), no centenário de John Tukey, reflete sobre as questões de *Data Science* e da Estatística. Numa nota de rodapé, David Donoho refere:

“I personally believe that mathematics offers the only way to create true breakthroughs. The empirical method is simply a method to avoid self-deception and appeals to glamor.”

E, sobre o futuro refere:

“In the future, there may be great industrial demand for the skills inculcated by GDS²; however the core questions which drive the field are scientific, not industrial. GDS proposes that Data Science is the science of learning from data; it studies the methods involved in the analysis and processing of data and proposes technology to improve methods in an evidence-based manner.”

Não há dúvida que a Estatística, para além de ser uma das áreas conducente à “melhor” profissão do mundo³, continuará a ter um papel fulcral na análise de dados (*Small* ou *Big*) e na geração de conhecimento científico. Os oito pontos referidos em Branco (2016) constituem uma excelente síntese do caminho que a Estatística terá de prosseguir.

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

“Poderia dizer-me, por favor, que caminho devo tomar para ir embora daqui?”

“Depende bastante de para onde quer ir”, respondeu o Gato.

“Não me importa muito para onde”, disse Alice.

“Então não importa que caminho tome”, disse o Gato.

Alice no País das Maravilhas, Lewis Carrol

¹ <http://www.businessinsider.com/how-marissa-mayer-figured-out-work-at-home-yahoos-were-slacking-off-2013-3>

² *Greater Data Science*

³ <http://www.careercast.com/jobs-rated/best-jobs-2017>

Referências

Branco, J.A. - *Boletim SPE outono de 2016*, 2, 49-58, 2016.

Brito, P. - *Boletim SPE outono de 2016*, 2, 40-42, 2016.

Donoho, D. - "50 Years of Data Science" [online]

Available:<http://courses.csail.mit.edu/18.337/2015/docs/50YearsDataScience.pdf>, 2015.

Hand D.J. *et al.* - Classifier technology and the illusion of progress. *Statistical science*, 21(1):1-14, 2006.

Hotelling, H. - The teaching of statistics. *The Annals of Mathematical Statistics*, 11(4):457-470, 1940.

Kendall M.G. - On the future of statistics. *Journal of the Royal Statistical Society*, 105: 69-91, 1942.

Kendall M.G. - On the future of statistics - a second look. *Journal of the Royal Statistical Society Series A*, 131: 182-204, 1968.

Tukey, J.W. - The future of data analysis. *The Annals of Mathematical Statistics*, 33(1):1-67, 1962.

Yates, F. - Computers, the second revolution in statistics. *Biometrics*, 22 (2): 233-251, 1966.



A Revolução dos Dados

Pedro Campos, pcampos@fep.up.pt

Faculdade de Economia do Porto, INESC TEC e Instituto Nacional de Estatística

Introdução

Os dados são a força vital da tomada de decisões e a matéria-prima para a responsabilização. Sem dados de qualidade que forneçam as informações certas sobre as coisas certas no tempo certo, conceber, monitorizar e avaliar políticas eficazes torna-se uma tarefa impossível. Este texto resulta de uma iniciativa no âmbito do novo olhar sobre o Tema Central da Estatística, que o Prof. Fernando Rosado promoveu nos últimos números do Boletim da SPE. A “Revolução dos Dados” foi o título que usei para o texto, nome igual ao que foi dado ao estudo das Nações Unidas com vista a fazer recomendações concretas sobre como trazer a “data revolution” para desenvolvimento sustentável (Data Revolution Group, 2014). Organizei as ideias em quatro partes: os dados, o ensino, a indústria e o futuro, que desenvolverei de seguida.

Dados

Por ocasião de um ICOTS (International Conference on Teaching Statistics), tive a oportunidade de jantar ao lado de um professor emérito do departamento de matemática da Universidade de Chicago. Durante o jantar, comunicou-nos a forma como foi visto pelo departamento ao lecionar processos estocásticos. Foi algo “impuro”, por ser demasiado aplicado e logo começou a sentir-se fora da elite dos matemáticos do departamento. Estávamos entre estatísticos, a história foi engraçada, pois falou-se em “sujar” as mãos com dados, mas este exemplo retrata o afastamento real entre a educação de tradição mais pura, baseada em modelos e o ensino com os dados.

Numa pesquisa rápida no Google, não se conseguem encontrar, hoje em dia, métodos para ensinar estatística e análise de dados, sem dados. O mundo está cheio de fontes de dados e de tecnologias para os tratar. Nunca existiram tantas iniciativas para tornar os dados abertos e acessíveis às populações. São disso bons exemplos o OpenDataWatch¹ (um projeto para melhorar o acesso à informação nos países em desenvolvimento e em formatos livres), data2x² (uma iniciativa para obtenção de dados sobre a igualdade de género), a African Maths Initiative³ (uma ONG queniana formada por professores de Matemática e Estatística para promover a cultura matemática e estatística em África). A necessidade de recolher e divulgar dados surge não só para apoio à tomada de decisão, mas também por motivos de índole política, como forma de combater problemas graves, tais como: fome⁴, racismo e desigualdade⁵ e outros (ver ProCivicStat⁶, um projeto com o objetivo de compreender a evidência quantitativa de fenómenos sociais e da vida cívica). Estão envolvidas instituições como a OCDE,

¹ <http://opendatawatch.com/>

² <http://data2x.org/>

³ www.africanmathsinitiative.net/

⁴ <https://data.unicef.org/topic/nutrition/malnutrition/>

⁵ <http://www.enar-eu.org/Equality-data-collection-151>

⁶ <http://community.dur.ac.uk/procivic.stat/>

Nações Unidas, Eurostat, etc. que colaboram em rede, tal como no caso do PARIS 21⁷, iniciativa para promover a produção de informação estatística no mundo em desenvolvimento.

Ensino

A revolução dos dados permite que possamos resolver problemas práticos em sala de aula, fazendo dos alunos os nossos heróis e mostrando-lhes o potencial da ciência estatística. Num artigo recente da AmstatNews (2016), newsletter da American Statistical Association, reflete-se sobre a importância dos dados no ensino da estatística nos Estados Unidos. As Diretrizes Curriculares da ASA para Programas em Ciências Estatísticas afirmam que as instituições precisam de garantir que os alunos que entram no mercado de trabalho ou em pós-graduações tenham a capacidade adequada para "pensar com dados" e para responder e responder a perguntas estatísticas. As diretrizes também observam a crescente importância da ciência dos dados. Em particular, as diretrizes da ASA apontam para:

- "Um foco nos dados, que deveria ser um componente importante dos cursos de estatística introdutória e avançada, incentivando o trabalho do aluno com dados autênticos ao longo do currículo".
- "Uma série de tópicos de ciências de dados precisam ser considerados para inclusão em cursos introdutórios, secundários e avançados em Estatística para garantir que os alunos desenvolvam a capacidade de enquadrar e responder a perguntas estatísticas com dados de suporte ricos no início dos seus programas e avançar em direção ao desenvolvimento e destreza na capacidade de usar com dados em cursos posteriores.

Estes assuntos não constituem novidade, porquanto desde há muito tempo se fala sobre a importância da aprendizagem com dados. Roiter e Petrocz, (1996) consideram existirem vários paradigmas do ensino da Estatística, dos quais se destacam aqueles em que a Estatística surge como planeamento de experiências ou como um assunto baseado em problemas. É precisamente neste paradigma que assenta a minha visão do ensino e aprendizagem da estatística: os métodos estatísticos devem servir para enquadrar um conjunto de dados e não o conjunto de dados para servir os métodos (Oliveira e Campos, 2008). Os alunos deverão ser progressivamente orientados para o trabalho com dados, apoiando-se no conhecimento prévio que já possuem. Este é, aliás, o domínio da denominada aprendizagem significativa. A importância atribuída ao trabalho com os dados exige o envolvimento dos alunos em novas ideias e procedimentos, à medida que avançam nos anos de escolaridade, por oposição à constante repetição dos mesmos temas e actividades (NCTM, 2007, p.52).

Indústria

As inovações no âmbito da indústria 4.0 envolvendo a automação, a Internet of Things, o Big Data e o Cloud Computing estão a gerar dados de forma exponencial. Nos serviços, a utilização de Big Data nas Estatísticas oficiais (a partir de social media, dados de sensores, transacionais, ou de satélite), constitui um bom exemplo sobre a forma como os Institutos Nacionais de Estatística e os estatísticos estão a lidar com esta massificação de dados. Há diversos estudos em que se antecipa com sucesso tendências de crescimento do PIB e de outros indicadores macroeconómicos com base em contadores de tráfico automóvel (Daas, et al., 2015). Existe, no entanto, muita controvérsia na utilização de fontes de Big Data, nomeadamente pelo facto de os dados poderem não ser tratados como amostras aleatórias e também pelo facto dos erros gerados na combinação de diferentes fontes serem muito difíceis de medir, já que as fontes de Big Data são, em geral, proxies vagas de fontes originais, (Shlomo, 2015). A acrescentar a tudo isto, temos ainda os problemas da ética e da privacidade que preocupam instituições como o INE e a Comissão Nacional da Proteção de Dados.

Apesar de todos estas desvantagens, sabemos que o potencial da utilização de fontes alternativas na produção de estatísticas oficiais é inequívoco, pois contribui claramente para a diminuição da carga

⁷ www.paris21.org

estatística sobre os respondentes, para a antecipação de alguns indicadores macroeconómicos e para a óbvia redução dos custos de operação e de recolha de informação.

Futuro

É estranho estarmos sempre a pensar no futuro da Estatística, tal como refere João Branco (Branco, 2016). Estou certo que este passa pelo contributo fundamental que a Estatística pode dar para a Data Science. Mas estamos a perder muito tempo. Assistimos em Portugal a sucessivos avanços e recuos na inclusão de conteúdos sobre Estatística nos programas curriculares do ensino secundário. Precisamos urgentemente de uma revolução dos dados no ensino da Estatística que permita verdadeiramente aos futuros estatísticos ganharem competências para lidarem com grandes quantidades de dados e em fluxo contínuo (streaming).

Mas o salto já deveria ter sido dado. Hoje em dia é mais fácil contratar alguém que domine outras abordagens, computacionais, como o machine learning e o data mining, que recorrem a algoritmos extremamente eficazes e cujos pressupostos são poucos ou inexistentes.

Tal como referia a Paula Brito num artigo do Boletim (Brito, 2016), “Se os estatísticos abraçarem o desafio de se abrirem a novas formas de pensar e de comunicar [e eu acrescento, “com dados”], a Estatística terá um papel relevante e que se nos afigura imprescindível para o desenvolvimento sadio da Data Science”.

Referências

- AmstatNews (2016), newsletter da American Statistical Association, 1 Março 2016, disponível em: <http://magazine.amstat.org/blog/2016/03/01/introstats16/>
- Branco, J., (2016), O Futuro da Estatística, Boletim da SPE, Outono de 2016
- Brito, P., (2016), Data Science – um desafio para os estatísticos? Boletim da SPE, Outono de 2016
- Daas, P., Puts, M. J., Buelens, B., and van den Hurk, Paul A. M., (2015), Big Data as a Source for Official Statistics, Journal of Official Statistics, Vol. 31, No. 2, 2015, pp. 249–262
- Data Revolution Group, (2014), A World That Counts, Mobilising the Data Revolution for sustainable Development, Report Prepared by request of The United Nations Secretary-General by the Independent Expert Advisory Board on a Data Revolution for Sustainable Development, available at: <http://www.undatarevolution.org/wp-content/uploads/2014/12/A-World-That-Counts2.pdf>
- NCTM (2007). Princípios e Normas para a Matemática Escolar. Lisboa: APM.
- Oliveira, E., Campos, P. (2008), Pensar com os dados: o contributo do ALEA, PROFMAT 2008 - Congresso da Associação de Professores de Matemática
- Roiter, K. e Petrocz, P. (1996) "Introductory Statistics Courses — A New Way of Thinking", Journal of Statistics Education, v.4, n.2.
- Shlomo, N., Editorial: Big Data in Social Research (2015), Journal of the Royal Statistical Society A, 178, Part 4, pp. 787-790



A tirania dos jargões

Regina Bispo, rbispo@ispa.pt

ISPA – Instituto Universitário
MARE – Centro de Ciências do Mar e do Ambiente
CEAUL - Centro de Estatística e Aplicações da Universidade de Lisboa

Quem, de entre os leitores do Boletim da SPE, não conhece a citação “*All models are wrong, but some are useful*”? Atrevo-me a dizer que todos conhecerão bem a célebre afirmação publicada, pela primeira vez, por George Box, em 1976, na revista *Journal of the American Statistical Association*. Curioso é (pelo menos eu assim o acho) que 40 anos passados, a frase me pareça mais atual do que nunca, sobretudo a primeira parte da afirmação... *all models are wrong*.

Ao longo da última década, tenho tido o privilégio de poder trabalhar em Estatística Aplicada. Quer isto dizer que estou entre aqueles que usam a Estatística sobretudo ao serviço de outras ciências. No meu caso, o uso da Estatística atravessou ciências tão diversas como, para citar alguns exemplos, a meteorologia, a agronomia, a bioquímica, a química, a ecologia, a psicologia, a medicina e a geologia. Uma das principais batalhas que neste percurso tenho travado é a busca do modelo parente dos dados que me chegam às mãos e que, para meu desalento e mais frequentemente do que eu desejaria, não se ajustam satisfatoriamente aos conhecidos/tradicionais modelos probabilísticos.... e lá penso eu... *all models are wrong!*...

Serão estas dificuldades uma fatalidade intransponível?... Certamente não existirá uma única e simplista resposta a esta questão... Mas, em anos recentes, a crescente consciência que a análise de dados pode representar uma poderosa ferramenta no mundo dos negócios tem impulsionado o desenvolvimento de métodos que não impliquem necessariamente o assumir de um modelo estocástico.

Neste novo contexto, a ênfase é colocada não no desenvolvimento teórico mas no desenvolvimento computacional de algoritmos de previsão apoiados em padrões “aprendidos” nos dados (em regra, de grande volume) – *Machine Learning* (ML) – que permita apoiar e suportar a tomada de decisão nos negócios, bem como, gerir de forma sustentada a performance organizacional, e com isso trazer claros benefícios económicos.

Em Macfee e Brynjolfsson (2012), pode ler-se “...*the more companies characterized themselves as data-driven, the better they performed on objective measures of financial and operational results ... companies in the top third of their industry in the use of data-driven decision making were on average 5% more productive and 6% more profitable than their competitors*”. Parece relevante.

Há não muito tempo, um querido amigo, perguntava-me:

– Estou a analisar um conjunto de dados com técnicas de *Machine Learning*. Tens experiência nesta área?

Não querendo dar o braço a torcer, afinal ele acreditava estar a falar com uma especialista na área, retorqui, na esperança de me sair airoso, com uma pergunta:

– Que dados são esses e quais os objetivos da análise que estás a fazer?

– Estou a trabalhar com um *high-dimensional data set* ou *big data*. *Terabytes* de dados, entendes? O objetivo é criar um algoritmo baseado em *supervised learning* de forma a prever um *label* binário em função de um conjunto de *features*.

Que diabo!... Estaríamos a falar em que língua?... “Jargonês”?... Não me dei por vencida (quem me conhece bem, sabe quão teimosa posso ser...) e lá fui desbravando pouco a pouco aquele discurso, crente no que, pelo menos eu considerava ser, uma sólida formação académica e relevante experiência profissional, consolidada ao longo de duas décadas de trabalho em Estatística. Por fim, concluí:

– Tens portanto um conjunto de dados de grande dimensionalidade (*big data...*), sendo o teu objetivo construir um modelo de regressão (*supervised learning...*) de forma a explicar a variabilidade de uma variável resposta (*label...*) binária em função de um conjunto de variáveis explicativas (*features...*). Entendi bem?

– Sim! – exclamou com um sorriso estampado no rosto, ciente da clareza do seu discurso. A medo, mas com expressão convincente, atrevi-me a prescrever:

– Parece-me, à partida e apesar de não ter “apalpado” os dados, que poderás usar técnicas multivariadas de redução da dimensionalidade dos dados e um modelo de regressão logística...

– Sim – interrompeu – É isso mesmo que estou a fazer! *Machine Learning!* – afirmou, com ar de quem falava com alguém preso num passado distante e obsoleto (e na realidade confesso que foi assim mesmo que me senti...).

Este curto, mas simpático, diálogo despertou em mim a curiosidade de pesquisar sobre “esta coisa” do *Machine Learning*. Pesquisei então (de forma despreocupada) definições para o conceito:

1. “*The aim of machine learning is to uncover hidden patterns, unknown correlations, and find useful information from data*” (Chiu, 2015.)

Nada de novo, pensei... Tudo isto me parecia caber de forma adequada na definição de Estatística (seja lá qual ela for). O mesmo autor continuava:

2. “*In addition to this, through incorporation with data analysis, machine learning can be used to perform predictive analysis*” (Chiu, 2015).

Como se a primeira frase não fosse já análise de dados...

3. “*Machine learning is all about using the right features to build the right models that achieve the right tasks*” (Flach, 2012).

Construir modelos parecia-me, de novo, caber na definição de Estatística. Já a parte do *right* “qualquer coisa”, despertava em mim um sentimento de desconfiança profunda...

4. “*Machine Learning is an algorithm that can learn from data without relying on rules-based programming*” (Shah, 2016).

Talvez nas entrelinhas aqui se possa ler afinal uma primeira diferença – procura de um algoritmo que opere sobre uma matriz de variáveis explicativas de modo a prever uma variável resposta, sem assumir regras operativas explícitas.

5. “*Machine learning is a subfield of computer science and artificial intelligence. It deals with building systems that can learn from data, instead of explicitly programmed instructions*” (Shah, 2016).

Por aqui eu lia também uma clara diferença de parentalidade entre ML e Estatística. A eleger um parente para a Estatística, não seria certamente as Ciências de Computação.

Ao fim de algumas, poucas, leituras sobre ML, confesso que acabei com mais perguntas do que respostas. Seriam as técnicas de ML um subconjunto da Estatística ou seriam alguns métodos estatísticos usados do âmbito do ML? Serão duas áreas complementares ou será que se intercetam? Haverá, em todas as circunstâncias, um paralelismo de termos? Será então apenas e sobretudo um problema de *rebranding* da Estatística? Não tenho respostas claras...

Mas algo que retive, nas conversas que fui tendo, foi a clara reação positiva ao *fancy* termo *Machine Learning* quando comparada com o típico – *Oh, não!...* – despoletado pela referência à Estatística numa qualquer conversa com não estatísticos. Como se a velha Estatística precisasse urgentemente de um novo fato, mais apelativo e moderno, sob pena de perder o comboio que a rápida evolução das ciências computacionais implicaram ao permitir facilmente aceder e armazenar um grande um volume de informação, referido como *big data*, sendo a sua análise e exploração tipicamente associadas às técnicas de ML.

A par da minha pesquisa sobre ML, quase sempre me cruzava também com as expressões *Data Analysis* e *Data Science*, diferenciadas da ciência Estatística. Serão estas apenas mais duas expressões “*sexed-up*”?

Devo confessar que tendo a concordar com a argumentada ambiguidade do termo Estatística que, dependendo do contexto, pode significar a ciência em sim mesmo, uma mulher licenciada em Estatística (ou será isto um Estatístico, por força do irrefutável machismo gramatical?), uma função de uma amostra aleatória ou até mesmo, no senso comum, uma estimativa pontual, podendo assim o nome deixar pouco explícito toda a potencialidade da área científica. Por seu turno, as expressões *Data Analysis* e *Data Science* parecem explicitar a excelência da função da Estatística, usada em combinação com outras áreas do conhecimento, nomeadamente com as ciências de computação. Mas, assim sendo, serei eu um “Cientista de Dados”? ...um “Analisador de Dados”?...

Pesquisando (rapidamente e sem grandes preocupações de rigor científico) pelas expressões “*Statistician jobs*”, “*Data Scientist jobs*” e “*Data Analyst jobs*”, numa plataforma internacional online aparecem os resultados: 1400 vagas para “*Statistician*”, 5600 vagas para “*Data Scientist*” e 7100 vagas para “*Data Analyst*”.

Consultado um (novamente sem preocupações de representatividade científica) dos anúncios para “*Data Analyst*” são múltiplas as competências requeridas:

- Ciências de computação/Engenharia;
- Estatística, com experiência na análise de grandes quantidades de informação;
- Experiência em Hadoop, R, Python e JavaScript,

além das habituais “paixão pelo trabalho”, “criatividade” e “empreendedorismo”, entre outras da mesma natureza (mas estas, todos temos...).

Parece pois que, independentemente da expressão que possa designar melhor ou pior o que se faz na prática, hoje em dia, e certamente no futuro, a análise de dados (nomeadamente atendendo ao seu volume, variedade e/ou velocidade de produção) obriga à dotação de múltiplas competências que vão de facto além das estritamente associadas à formação clássica em Estatística. Assim, tudo (me) indica que estamos perante uma oportunidade de mudar e evoluir, trabalhando em colaboração e complementaridade com outras áreas do saber. Reconhecer esta oportunidade obriga, também e certamente, a uma reflexão sobre as atuais abordagens pedagógicas. O mercado de trabalho é hoje

muito exigente e competitivo e a formação acadêmica oferecida em Estatística não pode, em minha opinião, alhear-se desta nova realidade.

Apesar de em mim encerrar o pecado capital de uma formação base em Engenharia (e não em Matemática...), sempre me dei, se não por Estatística, por alguém que faz ciência em Estatística Aplicada. Hoje e depois de escrever este texto não sei bem o que sou... *Statistician? Data scientist? Data analyst?* Um elemento da sua interseção? Nada disto?... Acabei com uma profunda crise de identidade profissional... e, pelo sim pelo não, a quem me pergunta, vou dizendo que sou Professora do Ensino Superior!

Referências Bibliográficas

Box, G. E. P. (1976) – Science and statistics. *Journal of the American Statistical Association*, 71: 791–799, doi:10.1080/01621459.1976.10480949.

Chiu, D. (2015) – *Machine learning with R cookbook*. Packt Publishing.

Flach, P. (2012) – *Machine Learning. The art and science of algorithms that make sense of data*. Cambridge University Press.

Macfee, A. and Brynjolfsson, E. (2012) – Big data: The Management Revolution. *Big Data*, 1(1): 51 - 59).

Shah, A. (2016). *Machine Learning vs. Statistics*. Retrieved October 2017, from <http://www.kdnuggets.com/2016/11/machine-learning-vs-statistics.html>.



Novo olhar sobre a Estatística, imaginar o mundo

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

Faculdade de Ciências da Universidade de Lisboa, CEAUL

O Tema Central da Estatística - um novo olhar! Foi este o desafio lançado pelo editor do Boletim da SPE. Sem ideias muito concretas sobre o que escrever neste âmbito, o tema em si levou-me a uma reflexão sobre um possível significado para este “novo olhar”; quais as perspetivas deste “novo olhar”. Sem fórmulas simples ou complexas, sem raciocínios matemáticos, optei por refletir um pouco sobre a nova forma de olhar o ensino no presente e no futuro. Assim, vou debruçar-me sobre o primeiro contacto dos mais novos com a Estatística, sobre a importância de os motivar, sobre como isto constitui um desafio para os professores.

É DE PEQUENINO...

Quem contacta com crianças pequenas, como é o meu caso, tem o privilégio de acompanhar o seu crescimento e aprendizagem. Tenho verificado que os programas escolares vão tentando adaptar-se às exigências do mundo moderno e, como não poderia deixar de ser, sendo a Estatística reconhecidamente uma das áreas mais importantes da atualidade, começa a ser explorada desde cedo pelos pequenitos.

Pré-Escolar: Sim, é verdade, os meninos no ensino pré-escolar têm contacto com a estatística! Um livro de matemática da educação pré-escolar, da Porto Editora, tem uma página dedicada à temática “Recolher, organizar e interpretar dados”. Nessa página, os meninos são convidados a completar um diagrama de barras, onde se contabiliza o número de insetos/flores que se encontram num certo jardim (Fig. 1). É um exercício alusivo à temática “Recolher, organizar e interpretar dados” retirado do livro “Alfa no Jardim de Infância – Iniciação à Matemática, 5-6 anos”, da Porto editora.



Figura 1

1.º Ciclo: Entre os 6 e os 10 anos, as crianças são desafiadas a abraçar a Estatística na sua vertente gráfica, desde a construção de diagramas de barras, de diagramas circulares ou de diagramas em caule-e-folhas. A noção de Probabilidade é adquirida já no final do 1.º ciclo, estabelecendo-se o conceito de acontecimentos muito ou pouco prováveis, improváveis ou certos. Estas noções surgem no 4.º ano, como se pode ver no livro de matemática, da Editora Santillana (Fig. 2). É um exercício alusivo à temática “Probabilidade” retirado do caderno de atividades “Projetos e Desafios – Matemática 4.º Ano”, da autoria de Ana Sofia Ribeiro, Editora Santillana.

Situações aleatórias

1. Num saco estão 6 cartões numerados da forma indicada.

9	17	33	27	12	15
---	----	----	----	----	----

a) Ao retirar um cartão do saco, é mais provável sair um número par ou um número ímpar? Justifica. *É mais provável sair um número ímpar, porque em 6 cartões só há um par.*

b) Enuncia um acontecimento pouco provável e outro impossível. *É pouco provável sair o número 12. É impossível sair um número maior que 33.*

c) Podes afirmar que «sair um múltiplo de 3» é um acontecimento certo? Justifica a tua resposta. *Não, não posso afirmar, porque há um cartão que não é múltiplo de 3, por isso, é um acontecimento muito provável.*

Figura 2

A curiosidade destes meninos é imensa e a motivação é indispensável. Não resisto a contar-vos uma história deliciosa que aconteceu no ano passado, quando fui à sala da minha filha fazer uma atividade matemática no âmbito da semana da família. Tratou-se de um jogo matemático, com algumas tabelas construídas com base na tabuada e em que eu supostamente adivinhava o número que eles tinham escolhido e que se encontrava em algumas tabelas. Magia! Diziam alguns. Como consegues? Perguntavam outros. A minha filha, conhecedora do segredo, também pôde fingir ser mágica! O entusiasmo foi tal, que aprenderam a mecânica do jogo rapidamente, reproduziram-no em casa e manteve-se o sucesso da magia! Alguns pais vieram ter comigo dizendo que os filhos fizeram o jogo em casa e estavam muito entusiasmados! Sei que não se tratava de um jogo estatístico, tratava-se de um jogo matemático, no entanto, mostrou que “casar” o ensino com a motivação é meio caminho andado para que as crianças queiram aprender mais e mais. As crianças de hoje vivem num mundo tecnologicamente avançado, dominam vários dispositivos eletrónicos, como *tablets*, computadores e *smartphones*. O ensino tem que ser adaptado fazendo uso deste novo mundo para os cativar. Ou, como na história que contei, pequenos exemplos, simples e mágicos, que despertem o seu interesse para a Matemática e para a Estatística.

ESTATÍSTICOS DO FUTURO...

À semelhança dos alunos do ensino pré-escolar e do 1.º ciclo, os estudantes universitários são um constante desafio para os professores. Mudaram os hábitos de estudo, a forma como estudam, o constante contacto com materiais digitais em grande quantidade, a possibilidade de participarem em aulas dinâmicas, de pesquisarem durante a aula recorrendo a *smartphones* ou *tablets*, preparando breves apresentações sobre o tema da pesquisa. Mais uma vez, o uso da tecnologia para “agarrar” os alunos! Isto implica explorar novas formas de ensino, tal como o *Web-Based Learning Technologies* (Gosper *et al.*, 2008), por exemplo.

Continua a haver alunos muito capacitados, com grande vontade de trabalhar. Alunos que prescindem das suas férias para realizarem estágios de verão, simplesmente pela possibilidade de aprender, de crescer profissional e socialmente, de fazer currículo. Um processo de simbiose! Muitos começam a trabalhar quando estão ainda no 1.º ano do mestrado e fazem um grande esforço, que no final é muito compensador. Partem para o mercado de trabalho ou mesmo para doutoramento, pedem-nos cartas de referência, conselhos. É muito gratificante poder continuar a participar neste processo de crescimento pessoal e profissional. São estes alunos que nos dão ânimo para continuar. Dão significado à palavra professor, este ser que não só transmite conhecimento, mas forma cidadãos. Melhores cidadãos. (Lin *et al.*, 2013)

NOVOS SENIORES, VELHOS JUNIORES

Os novos seniores, como carinhosamente refere o editor do boletim, são, na verdade velhos juniores, no contexto universitário e politécnico. A dificuldade na abertura de concursos, faz com que continuemos a ser os docentes mais novos, ou seja, eternos juniores, mas com a idade de quem está a meio da carreira e que, portanto, adquiriu algum conhecimento; seremos então “novos seniores”?

Cientificamente falando, deparamo-nos com muitos desafios, como é o caso da análise de grandes quantidades de dados, atualmente designados por *Big Data*, mas cujo conceito já era familiar a todos os que trabalham em análise de dados genómicos, por exemplo. Infelizmente, estes desafios são pouco compatíveis com as *Big Burocracias* e com as *Big Horas de Serviço Docente*, mas vamos fazendo o melhor que sabemos e podemos, porque na verdade, gostamos daquilo que fazemos!

REFERÊNCIAS

- Gosper M, Green D, McNell M, Phillips R, Preston G e Woo K (2008). *Web-Based Lecture Technologies: Highlighting the Changing Nature of Teaching and Learning*. Australian Learning and Teaching Council.
- Lin X, Banks D, Genest C, Molenberghs G, Scott D e Wang J-L (2013). *Past, Present and Future of Statistical Science*. Taylor & Francis, Boca Raton, FL.



Desafios da Estatística para o século XXI

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

*Centre for Research into Ecological and Environmental Modelling,
The Observatory, University of St Andrews*

e

Centro de Estatística e Aplicações, Universidade de Lisboa

Quando aceitei o convite para contribuir para o tema “*O Tema Central da Estatística – um novo olhar*” assumo que não pensei desde logo na responsabilidade que isso acarretava. Tudo bem, falta quase um ano, sem stress, pensei na altura. Mas depois de, no início de Setembro, me sentar a olhar para a página em branco, rapidamente me apercebi de que, afinal, não é bem assim! Quando nos pedem para dar uma opinião sobre um determinado tema estão implicitamente a mostrar que é legítimo termos uma opinião sobre ele. E isso não deixa de ser surpreendente, na medida em que seria uma pergunta para fazer a um estatístico, quando, cá dentro, continuava, até agora, a assumir-me como um biólogo. Ou melhor, como um biólogo quando entre estatísticos, como um estatístico quando entre biólogos. Trata-se de uma posição confortável, no sentido em que me retira responsabilidades, mas percebo agora que se tornou insustentável. É necessário definir o que sou e assumi-lo, e penso que vou recorrer ao inglês para o fazer. *I am an Ecological Statistician*. Um estatístico ecológico? Não soa tão bem. (Já diziam os Clã na voz da sua Manuela Azevedo, em “Problemas de Expressão”, *a língua inglesa soa sempre bem...*). Mas sim, sem dúvidas, um estatístico. E a partir de hoje conscientemente assumido como tal, até prova em contrário.

Estando definido que me assumo como um estatístico, surge a pergunta: mas o que faz um estatístico? Bem, de facto faz Estatística. Mas o que é a Estatística? Provavelmente, existem tantas definições como tentativas de a definir. No entanto, parece-me natural que, se pretendo dar uma opinião sobre os desafios que a Estatística enfrenta, seja sensato começar por conseguir definir o que ela é. Nem que seja apenas aqui, para que o texto seja internamente coerente. E assim me lanço nessa tarefa, juntando mais uma definição às muitas existentes. O aspeto fundamental que torna a Estatística única é ser a ciência que lida com a incerteza. Eu diria que a Estatística é a ciência que estabelece as regras para a recolha e análise de dados e sua posterior interpretação e visualização, com ênfase último em inferência e predição. Em particular, a Estatística permite-nos quantificar a incerteza de uma forma rigorosa. A única forma adequada de lidar com a incerteza é através do conceito de probabilidade que, como tal, toma na Estatística um papel de destaque.

Para perceber o futuro, é fundamental conhecer e compreender o passado. O século XX assistiu ao nascimento e consolidação do que entendemos como a Estatística moderna, através de cientistas de exceção, com a axiomatização da probabilidade por Kolmogorov, o teste de qui-quadrado por Pearson, os testes de hipóteses e intervalos de confiança de Fisher e Neyman/Pearson, e, na generalidade, o domínio da Estatística frequentista clássica. Mais do que uma mudança de perspetivas, o final do século XX veio trazer a Estatística Bayesiana para um lugar de destaque, fruto do crescente desenvolvimento das capacidades computacionais que permitem, hoje em dia, fazer em segundos, cálculos impensáveis no início do século. Hoje em dia, assiste-se a um pragmatismo de paradigmas inferenciais, sendo, frequentemente, quer o paradigma bayesiano, quer o clássico, adotados não por considerações filosóficas, mas em função das ferramentas que estão disponíveis para realizar as análises necessárias para responder às questões de interesse. Para dar um exemplo numa área que me é familiar, se no século XX as técnicas de captura-recaptura eram exclusivamente frequentistas, baseadas em máxima verosimilhança, neste século, fruto do desenvolvimento de inúmeros *packages* de

R que implementam metodologias bayesianas, as posições equilibraram-se. No entanto, se perguntarmos à generalidade dos utilizadores se se considera Bayesiano ou Frequencista, a resposta mais comum será “não sei” ou “nunca pensei nisso”. Penso que, de facto, o paradigma mais adequado depende do problema em análise. A noção de que os parâmetros são variáveis aleatórias e que o que observamos é apenas uma realização de um processo estocástico torna-se muitas vezes apelativa e próximo de Bayes. Mas a verdade é que, em muitas circunstâncias, não estamos minimamente interessados nesse processo estocástico não observado e queremos apenas saber, para uma população, qual o valor de um determinado parâmetro que, sendo desconhecido, é fixo. No meu entender, o século XXI vai assistir a uma convergência e convivência pacífica entre estes dois paradigmas inferenciais que coabitam lado a lado sem problemas. Fazendo uma analogia familiar, será mais ou menos como dois irmãos que na juventude andaram constantemente em conflito, mas que quando chegam à idade adulta já não sentem a necessidade de afirmação, para eles ou para os outros, e se admiram mutuamente. Cada um deles é o que é e não sente necessidade de mudar o outro.

A Estatística precisa de parar de ser redescoberta. É preciso fazer uma “limpeza da casa”. É preciso perder algum tempo a sistematizar o conhecimento e a torná-lo acessível a todos. Não podem existir 32 nomes pomposos para o que são apenas diferentes sabores de modelos mistos. É preciso repensar os currículos e apresentar aos futuros Estatísticos uma metodologia transversal, não pequenas caixas cuja utilidade deixa de ser óbvia. A Estatística debate-se atualmente com uma crise de identidade. Um sentimento de estar sob ameaça, que levou por exemplo Les Nicholls a considerar dar o título “The survival of the species?” ao seu *presidential address* de 1999 à *Statistical Society of Australia* (Nicholls, 1999). Alguns anos mais tarde (poucos), o aparecimento de uma competição feroz por parte de informáticos do campo da chamada Ciência de Dados (*Data Science*) veio abalar as fundações de uma casa então estável. Surge o medo de que desapareçam os empregos para os Estatísticos, tomados de assalto por informáticos que dominam as ferramentas necessárias para processar as grandes quantidades de dados que lhes chegam às mãos e para as quais o Estatístico de ontem, e porventura o de hoje também, se limita a admirar como “um burro faz com um palácio” (Não pretendo insultar ninguém, estou a incluir-me na manada conceptual!). Na verdade não me parece que devamos sentir uma ameaça pois, na realidade, o que está à porta é uma oportunidade. E esta, não é uma ideia minha (bem, no fundo no fundo, raramente temos ideias originais, são sempre adaptações, permutações ou transformações de outras, de outros – *we stand on the shoulders of Giants*, uma citação cuja real origem parece difícil de precisar). Peter Diggle, no seu discurso à *Royal Statistical Society* em 2015, faz um paralelismo com o momento vivido a meio do século XX, com a entrada dos computadores em cena (Diggle, 2015). Temia-se que, dado que as máquinas conseguiam fazer cálculos tão rápidos, mais cedo ou mais tarde ninguém iria precisar de Estatísticos para fazer essa tarefa. No entanto, já estamos no “mais tarde” e isso não aconteceu. Muito pelo contrário! O problema desta linha de argumentação é que, infelizmente, ou, provavelmente, muito felizmente, as máquinas ainda não pensam. E como tal, não são minimamente úteis para conseguir definir qual a melhor abordagem quando surge um qualquer problema não banal. Depois de definida a abordagem a tomar, os computadores são claramente muito úteis para fazer os cálculos necessários para implementar a solução. Mas, no final, novamente, não conseguem fazer a interpretação dos resultados. Na realidade, o que as máquinas não têm é o que o meu amigo e mentor, Dinis Pestana, já há 20 anos, me dizia ser fundamental: a Estatística é feita muito à base de sensibilidade e bom senso e, até ver, nenhuma máquina tem essas características – são características humanas. Quem diria, a Estatística é uma ciência fundamentalmente humana! Roubando uma citação comumente atribuída a Einstein, mas que não é dele, sendo efetivamente de autor desconhecido: *Computers are incredibly fast, accurate and stupid; humans are incredibly slow, inaccurate and brilliant; together they are powerful beyond imagination*; a ênfase está aqui, em *together!*

Existem inúmeros chavões usados hoje em dia, como *big data*, *data science*, *business intelligence*, *statistical learning*, *deep learning*, *data mining*, *artificial intelligence*. Aquilo que estes conceitos têm em comum, na minha opinião, são três aspetos fundamentais: (1) ninguém os consegue definir rigorosamente, (2) são usados e abusados na proporcional medida em que impressionam quando são introduzidos no discurso, e (3) irritam em maior ou menor grau os Estatísticos que os vêem apenas como uma ameaça, que representam “papões” não assumidos para a Estatística, associados mais a coisas “estranhas” que Engenheiros e Informáticos fazem do que verdadeira Estatística. Mais uma vez, na minha opinião, estes conceitos representam uma extraordinária oportunidade. Uma equipa com um

programador e um estatístico (ou vários, claro, depende da envergadura de cada projeto concreto!) será tudo o que é suficiente, mas também necessário, para conseguir responder às perguntas que surgem associadas aos chavões acima. Como tal, está na altura da Estatística assumir a sua posição central em todos estes chavões, em vez de se refugiar no seu reduto purista de que essas coisas são apenas, de uma forma ou de outra, um desvirtuar de coisas que na realidade já fazemos.

Dos conceitos que referi acima, há um que talvez mereça um parágrafo. Porquê tanto alarido à volta do *big data*? Como se, por termos *big data*, deixasse de ser preciso modelos, ou pior, deixasse de ser necessário ligar o cérebro. Nada podia ser mais falso. Na realidade, se uma pequena quantidade de maus dados pode ser muito enganadora, uma grande quantidade de maus dados pode ser muito pior! (uma ideia introduzida por exemplo por Cox & Efron, 2017). Penso que a capacidade de interagir com grandes quantidades de dados fará com que se assista a parcerias estratégicas entre a Estatística e a Informática/Computação. Neste momento assistimos a uma situação em que os dados se acumulam a uma velocidade maior do que uma única máquina pode processar. É impossível pensar que uma única máquina vai conseguir processar em tempo real as pesquisas feitas no *google* se elas acontecem a uma média de 2 milhões por minuto! E mais do que multivariadas, são *hipervariadas*¹, no sentido em que podemos fazer um número infinito de perguntas sobre essas pesquisas! Quem as faz? Onde? Que termos são mais utilizados? Qual a relação do número de pesquisas em função do número de *tweets* no minuto anterior? E na hora anterior? São megas e teras e gigas e petas e hexas e infinitos bites de potencial informação, quando pensamos na nova informação derivada que surge ao procuramos coisas tão simples como correlações entre estas quantidades. Assim, parece-me óbvio que a Estatística tem muitos campos férteis onde ir procurar parcerias estratégicas, nas empresas e no governo, encontrando, como no passado fez talvez de forma ímpar nas ciências da vida, novas perguntas e novos limites para serem ultrapassados. Porque se há uma coisa de que não nos podemos esquecer é que o progresso não se pode contrariar, podemos apenas decidir como vamos lidar com ele e que aquilo que hoje nos parece impossível será, dentro de 20 anos, uma realidade. Basta pensarem (bem, se tiverem mais do que, vá... 40 anos) que quando concluíram a licenciatura quase não havia computadores, hoje em dia é possível ver uma peça única em Hong Kong, tirar-lhe uma fotografia e um, sim, um, minuto depois uma cópia estar nas Ilhas Desertas, impressa em 3D. Tudo consequência desse monstro que se tornou a internet, sem a qual a maioria de nós já nem saberia (sobre)viver.

A Estatística tem de sair do escritório. A Estatística tem de sair do departamento. Atrevo-me a dizer, no nosso pequeno universo, que a estatística tem de sair da SPE, temos de passar da Sociedade Portuguesa de Estatística para uma sociedade portuguesa estatística. E essa é também uma missão que deve ser assumida pela SPE. A nível pessoal, a saída do escritório foi natural, até porque o que eu tive de fazer foi conquistar um lugar no escritório, literal e figurativamente. Eu já vinha do “outro lado”, da biologia, onde cada vez há mais dados e menos capacidade de os tratar devidamente. Sintomaticamente, as metodologias que hoje em dia se consideram de ponta na biologia, já não são publicadas em revistas biológicas, mas em revistas na área da Estatística ou em revistas que vivem na interface entre disciplinas, como é o caso da *Methods in Ecology and Evolution*, *Biometrics* ou *JABES*, entre outras. Os Estatísticos precisam de trabalhar em conjunto com cientistas de outras disciplinas na procura das soluções para os problemas nessas outras áreas. Nesse contexto, vejo com particular interesse o comentário de Peter Diggle quando refere que “toda e qualquer Universidade orientada para a investigação deveria ter um instituto de estatística transversal” (Diggle, 2015). Os seus Membros teriam posições mistas, a tempo parcial, num departamento de Estatística e noutra departamento, permitindo, assim, um rápido intercâmbio de problemas e soluções entre quem tem uns e outros. É uma ideia velha dessa sábia Natureza, tantas vezes repetidas: simbiose. Nunca falha!

Um dos desafios que vejo actualmente, consiste no desenvolvimento de estratégias que permitam diminuir o tempo que permeia entre a publicação de uma metodologia e a sua aplicação no campo a que se destina. Isso passa pelo investimento dos investigadores no desenvolvimento de material que sirva esse objectivo, como livros, *software*, cursos ou workshops. Estes são temas que tendemos a deixar para mais tarde pura e simplesmente porque já não há nem tempo, nem energia, para lá chegar.

A Estatística, e aqui creio que a SPE tem o seu papel dentro do nosso país, tem, de uma vez por todas, sanear uma imagem que de facto nunca se preocupou em defender. *Lies, damn lies, and statistics* não

¹ Vejo aqui uma oportunidade para criar um novo termo ameaçador: a "*hipervariate statistics*" – zero resultados no Google hoje, quantos daqui a um ano, se o lançar no *twitter* logo à noite?

pode ser a visão do público (cf. Nicholls 1999). Além disso, a literacia Estatística tem de ser fomentada. Um público informado é um público menos fácil de enganar. A maioria das Estatísticas que hoje vejo reportadas num telejornal são, infelizmente, na melhor das hipóteses enganadoras ou na pior das hipóteses falsas – “38 % dos portugueses acha que o Facebook é uma fabricação da CIA para distrair as massas (amostra: 278 entrevistas), 77% das mulheres acha que os seus cabelos ficaram mais sedosos depois de usar o nosso amaciador (amostra de 21 mulheres das quais 13 eram nossas primas e lhes demos um cabaz de natal), 78.89 % de todas as Estatísticas são inventadas. As 3 anteriores foram... passou para 78.97 (mais uma...!)”. Essa conclusão é óbvia usando um pouco de sensibilidade e bom senso: não consigo imaginar como é que eu as poderia obter, mesmo que quisesse. Vejo problemas na definição das bases de amostragem, tamanhos de amostra reduzidos, vieses na seleção de amostras, inúmeros fatores de confundimento, etc.. A missão da SPE poderia ser essa também, criar mecanismos para promover a literacia Estatística, mecanismos para certificar o trabalho dos Estatísticos. Fica aqui a sugestão!

A força da Estatística, aquilo que a leva a ter o futuro assegurado, é servir hoje de base à generalidade das restantes ciências, com especial ênfase nas ciências da vida. Isto porque a incerteza e variabilidade são transversais ao conhecimento, e dessas sabemos nós! É a Estatística que permite definir uma amostragem; que permite explicitar que dados recolher; e como estes devem ser recolhidos. Como sumariar os dados. Como produzir inferências para a população de onde estes vieram. Como prever novas observações. Em todos os níveis deste processo de descrição, inferência e previsão, é preciso saber como lidar, ou seja, como reduzir, incorporar e quantificar a variabilidade, a incerteza, que servem de combustível à Estatística. Venha esse *big data* ou qualquer novo nome pomposo que lhe queiram dar (... bigdata4.0, LOL), venha o que vier, se implicar variabilidade e incerteza, vai ser necessário um estatístico. E se tempos incertos houve, estes também o são. Paradoxalmente, isso leva-me a uma certeza: não nos vejo sem trabalho num futuro próximo.

Referências

- Cox, D. R. and Efron, B. (2017) Statistical thinking for 21st century scientists. *Science Advances* **3**:e1700768
- Diggle, P. J. 2015 Statistics: a data science for the 21st century. *Journal of the Royal Statistical Society A* **178**:793-813
- Nicholls, D. 1999 Statistics into the 21st century. *Australia & New Zealand Journal of Statistics* **41**:127–139



A minha utopia sobre o Tema Central da Estatística

Nuno Sepúlveda, nuno.sepulveda@lshtm.ac.uk

*London School of Hygiene and Tropical Medicine, Reino Unido
Centro de Estatística e Aplicações da Universidade de Lisboa, Portugal*

Introdução

Quis o destino que o professor Fernando Rosado, que agradeço desde já, me lançasse o desafio de refletir sobre o Tema Central da Estatística. Depois de ler o Boletim de Outono do ano passado, pensei em dar uma visão pessoal sobre um possível futuro da Estatística, tal como tão bem redigiu o professor João Branco [1]. “Mas o que escrever exatamente?”, foi a questão que coloquei a mim mesmo meses a fio.

É certo e sabido que a inspiração nem é fácil de invocar nem de saber quando surge. Mais ainda quando a minha possível qualidade de vidente sofreu um profundo abalo após a leitura do livro ‘*thinking, fast and slow*’ durante a primavera deste ano [2]. Nesse livro, o psicólogo e vencedor do Prémio Nobel da Economia Daniel Kahneman coloca a nu a inépcia de muitos peritos da nossa sociedade em preverem eventos futuros das suas áreas de competência. Portanto, refutei a ideia de me tornar num novo professor Chibanga dos ‘Gato Fedorento’, que só tem poderes para ler o futuro para lá de 2070 [3].

O livro de Daniel Kahneman também me chamou à atenção para um fenómeno psicológico que aparentemente muitos de nós padecemos. Segundo este psicólogo, o ser humano tem uma inclinação de tomar decisões baseadas naquilo que vê e sente no momento (“*what you see is all there is*”). E ao tomar conhecimento deste fenómeno, descobri o caminho que me levou ao conteúdo deste pequeno artigo.

Resolvi, então, consultar a minha memória mais recente e deparei-me com dois grandes temas que, por falta de tempo, nunca consegui aprofundar. O primeiro refere-se à moda do ‘*big data*’ - como se houvesse uma dicotomia entre dados pequenos e grandes?! O segundo tema reflete nalguns factos surpreendentes, mas também preocupantes sobre o atual estado da literatura científica.

Ai, esse admirável mundo novo dos grandes volumes de dados!

No prefácio do ‘*Statistical Methods for Research Workers*’ de 1925, Ronald Fisher escrevia o seguinte: “*often the analysis of large masses of data by statistical methods is necessary, and the biological worker is continually encountering advanced statistical problems the adequate solutions of which are not found in current statistical text-books.*” Será que Fisher previu que esses grandes volumes de dados estariam na ordem do dia passado mais de 90 anos?

Durante o ano de 2017, o campus da *University College of London* onde presentemente trabalho recebeu inúmeras iniciativas sobre o uso de grande volume de dados em ciência e na sociedade. Por exemplo, o Centro para a Metodologia Estatística da *London School of Hygiene & Tropical Medicine* organizou dois ciclos de seminários sobre esta temática e criou uma nova linha de investigação [4]. O 61º Congresso do *International Statistical Institute* (ISI), em que tive o prazer de participar na bonita cidade de Marrakech, não ficou atrás no fervor sobre esta temática, tendo havido 23 apresentações e 12 sessões convidadas que continham o termo ‘*big data*’ no seu título.

Esta moda estimulou a minha curiosidade natural. Afinal de contas, dados são dados, pequenos ou grandes. Será que é mais do que isso mesmo? Em geral, mostro-me sempre relutante em seguir modas que não entendo bem. Arregacei, então, as mangas e resolvi informar-me, lendo alguns artigos da literatura e assistindo a alguns seminários, nomeadamente, o da professora Nancy Reid da

Universidade de Toronto que, na manhã do último dia do congresso do ISI, fez uma excelente síntese de um artigo seu sobre este tema [5].

É um facto inegável de que há atualmente uma grande facilidade na aquisição de dados através da internet, telemóveis, ou outras tecnologias populares nas sociedades mais evoluídas. Esta explosão de dados trouxe também consigo novos formatos de dados passíveis de serem analisados, tais como os de vídeo ou de som, que necessitam de ser harmonizados com outros de carácter mais tradicional. A análise de ‘*big data*’ baseadas no uso destas plataformas de massas centra-se invariavelmente na predição de um comportamento ou de uma performance futura, predição essa que se espera ter uma grande precisão, exatamente por se ter disponível um grande volume de dados.

Grandes volumes de dados surgem também nas novas disciplinas ‘*omics*’ (e.g., genómica, proteómica ou metabolómica) devido a um forte desenvolvimento tecnológico. Em jeito de curiosidade, as atuais técnicas de sequenciação permitem atualmente obter informação detalhada de todo o genoma e da sua expressão genética em diferentes organismos importantes para a saúde humana. Um dos objetivos de análise reside em identificar e catalogar a diversidade genética existente numa determinada espécie biológica, diversidade essa que pode rondar a ordem dos milhões no ser humano. Outro objetivo de análise consiste em compreender a relação entre milhares, ou mesmo milhões de variáveis de natureza genética e uma característica biológica (e.g., presença de um cancro) observada num conjunto relativamente reduzido de indivíduos (da ordem dos milhares, na melhor das hipóteses).

Em termos práticos, a era do ‘*big data*’ conduziu a uma redução drástica do tempo despendido na recolha dos dados e dos seus custos associados. Por sua vez, aumentaram os custos de análise propriamente dita devido a um maior tempo requerido para a sua execução. Por exemplo, na área da Genómica, Sboner et al [6] sugere que, em 2020, a análise dos dados contribuirá em mais de 50% para o tempo total de um projeto. Sem ter estatísticas concretas, parece-me que há mais dados do que pessoal qualificado capaz de lidar com os mesmos e, por isso, sinto que estamos a viver uma era dourada para os estatísticos aplicados.

Contudo, parece que outro tipo de profissionais tais como os bioinformáticos ou os cientistas de dados estão a tomar conta da análise desses grandes volumes de dados. Se um estatístico aplicado é tipicamente talhado para a análise de dados, porque razão este não é chamado para este tipo de aventuras científicas?

Pelo que consegui inferir da minha experiência pessoal em Londres e da pequena pesquisa que fiz sobre este tema, estes projetos requerem conhecimentos computacionais avançados sobre o armazenamento, a manipulação e gestão dessas grandes coleções de dados, conhecimentos esses que não fazem normalmente parte de uma formação tradicional em Estatística, tal como é o meu caso. Para colmatar essas lacunas técnicas, têm surgido inúmeros cursos em ‘*Data Science*’ (Ciência de Dados) com o intuito de habilitarem futuros profissionais com conhecimentos básicos de Estatística e de Computação; vide o artigo bastante esclarecedor sobre o que é *Data Science* da professora Paula Brito [7].

Outro fator de possível exclusão de um estatístico destes projetos está relacionado com as suposições simplificadoras e redutoras sobre o que os dados representam. Por um lado, projetos baseados na internet ou noutras plataformas tecnológicas de massas partem do pressuposto de que os dados recolhidos representam a população no seu todo. Por outro lado, nas disciplinas ‘*omics*’, questões relacionadas com a definição da população ou com o papel da amostragem são colocadas quase em jeito de nota-de-rodapé. O que interessa é adquirir amostras biológicas e aplicar as novas tecnologias de sequenciação para depois compilar e catalogar os resultados que se encontram. Nestes dois cenários, a análise estatística limita-se essencialmente na descrição e na visualização dos dados, duas tarefas que necessitam de ‘menos’ conhecimentos técnicos de Estatística

Contudo, a acumulação de estudos de ‘*big data*’ tornou evidente de que nem tudo são rosas nesta área. Retomemos então à questão central num estudo estatístico: a representatividade dos dados da população em estudo. Ora, é sabido que a frequência do uso de telemóveis ou da internet está inversamente correlacionada com a idade de um indivíduo e com o seu nível socioeconómico [8]. Agora, imagine que se pretende efetuar um estudo de ‘*big data*’ com o objetivo de inferir o estado de saúde da população e identificar potenciais grupos de risco através da recolha massiva de dados pela internet ou por telemóvel. Se esse estudo não tiver uma definição clara e rigorosa da população em estudo, arrisca-se a excluir os indivíduos mais idosos e/ou mais pobres da sociedade que, por sua vez,

constituem os respetivos grupos de risco de muitas doenças comuns. Este exemplo específico mostra bem que os projetos de ‘*big data*’ não estão isentos de produzirem resultados extremamente precisos, mas também enviesados.

Problemas relacionados com a definição da população ou da representatividade dos dados estendem-se para além das investigações baseadas em dados da internet ou telemóvel. Tal como referido acima, é prática comum em estudos ‘*omics*’ de simplesmente referirem que as amostras recolhidas provêm de um determinado país ou de um grupo de pacientes sem a especificação do mecanismo de amostragem ou a localização específica dessas amostras. A ausência de uma definição clara da população em estudo torna difícil a avaliação de possíveis enviesamentos de amostragem, a reprodutibilidade de resultados, ou a comparação direta entre diferentes estudos.

Em projetos de ‘*big data*’, o entendimento causal entre diferentes variáveis aleatórias parece também ter sido colocado para um segundo plano. Interessa apenas identificar correlações ‘úteis’ que assegurem uma boa predição com elevada precisão. Esta abordagem é cientificamente limitada, pois afasta-se da génese de toda a atividade científica, que deveria ser a compreensão do mundo que nos rodeia.

Um exemplo que se tem tornado um clássico entre os opositores à moda do ‘*big data*’ é dado pela famosa *Google Flu Trends* (GTF) que, em 2009, espantou o mundo ao estimar incidência de gripe através da análise dos termos de busca na internet [9]. Em 2014, correu a triste notícia de que a GTF estava sistematicamente a sobrestimar o número oficial de casos de gripe registados nos Estados Unidos da América [10]. Segundo reza a história, os algoritmos da GTF foram essencialmente calibrados com as buscas realizadas entre 2003 e 2008, um período de tempo em que o ciber-espaço estava povoado por previsões apocalípticas de pandemias de gripe, tal como o da estirpe H1N5 (2005-2006), que felizmente acabaram por não acontecer.

Os problemas acima mostram que os projetos de ‘*big data*’ estão ainda longe de produzirem um conhecimento perfeito sobre a realidade. A multidisciplinaridade parece uma chave para o futuro onde diferentes competências e maneiras de olhar o mesmo problema se conjugam. Nesse aspeto, um estatístico de formação mais tradicional mostra a sua mais-valia devido à sua maior preocupação em definir corretamente a população em estudo, em identificar possíveis enviesamentos na aquisição de dados, e em ser mais cauteloso na interpretação e no uso de uma determinada correlação.

Ao encerrar este assunto, parece-me um facto adquirido de que os projetos de ‘*big data*’ estão aí para ficar devido ao constante avanço tecnológico na aquisição de dados. Contudo, devemos redobrar as nossas cautelas: estes projetos oferecem de facto oportunidades únicas para entender o mundo e a humanidade, mas também oferecem a possibilidade de fazermos erros a uma escala nunca antes imaginável.

Publicar, publicar, publicar: um mantra para o futuro?

Depois desta minha pequena dissertação sobre o fenómeno ‘*big data*’, pretendo agora abordar um outro tipo de dados, mas também de grande escala: o das publicações científicas. Este tema foi-me motivado pela leitura de um artigo publicado na versão *online* do jornal britânico *The Guardian* no dia 27 de Junho deste ano [11]. Neste artigo, o jornalista Stephen Buranyi faz uma resenha muito interessante sobre o mercado global das publicações científicas. Resultado final: fiquei completamente atónito após a sua leitura!

Quando ingressei em 1994 na Licenciatura em Matemática Aplicada e Computação do Instituto Superior Técnico, estava longe de pensar que o mercado global das publicações em 2017 pudesse gerar lucros de mais 6 bilhões de euros para os grandes grupos editoriais [12]. Sim, leram bem, lucros de mais de 6 bilhões de euros, que apenas rivalizam com os da empresa *Google*! Estes números são reveladores de uma verdade tão inegável como desconfortável: a atividade científica é também um negócio e, sim, mesmo muito rentável para muitos cofres privados.

Como é possível gerar tanto lucro num mercado em que os seus consumidores são ‘meia-dúzia’ de cabeças-duras que sonham que o conhecimento científico é uma das alavancas para o progresso individual e civilizacional? Vamos a factos.

O atual mercado das publicações obedece a um modelo económico baseado em três pilares. O primeiro consiste no facto de que o dinheiro público está na génese da grande maioria da atividade científica. O segundo é o financiamento público aos cientistas que, para além da sua investigação, ainda asseguram gratuitamente os pareceres técnicos dos artigos submetidos para publicação. O

terceiro assenta no pagamento de taxas de subscrição por parte das instituições públicas para garantirem o acesso ao produto final da investigação, que os seus próprios membros produziram.

Os eventuais custos de publicação não justificam este modelo economicamente estranho para as instituições públicas, uma vez que estes são praticamente zero num panorama de publicação em formato digital. A pequena farpa de obrigar a publicação de artigos em regime de livre acesso, embora uma boa ideia em teoria, veio introduzir alguns abusos do sistema, que enumero de seguida.

O abuso mais inconsistente com os ideais científicos parece ser a ideia tácita de que, quando se paga para publicar, então qualquer manuscrito é publicável, independentemente da sua qualidade. Se não é no primeiro jornal submetido, é no segundo, se não é no segundo, é no terceiro e por aí fora, até chegar a um jornal menos exigente que aceite a sua publicação. Um outro abuso é a rápida proliferação de novos jornais digitais de qualidade duvidosa, cujas práticas editoriais visam pura e simplesmente a cobrança das respetivas taxas de publicação em regime de livre acesso. Um último abuso é o constante assédio de algumas revistas a possíveis autores, ora convidando-os a submeter um artigo que terá um parecer no espaço de duas semanas, ora a serem editores convidados para um número especial num tema à sua escolha – sim, já recebi convites destes dois tipos no mesmo correio eletrónico!

Porque razão estes abusos estão relacionados com a Estatística e o seu tema central? Fiz uma pequena pesquisa na base de dados *Web of Science* sobre o número total de artigos publicados por ano em geral e na área da Estatística em particular, entre 1990 e 2016 (Figura 1A). Estes números surpreendem-me pela sua magnitude. Durante esse período, o número total de artigos publicados cresceu de 579,089 para 1,681,205. A área da Estatística também abraçou este crescimento na produção científica, passando de 2,908 artigos publicados em 1990 para 10,338 em 2016. Curiosamente, o crescimento do número de artigos publicados parece ter sido mais rápido na nossa disciplina do que em geral até 2009, estando a abrandar nos últimos 7 anos (Figura 1B).

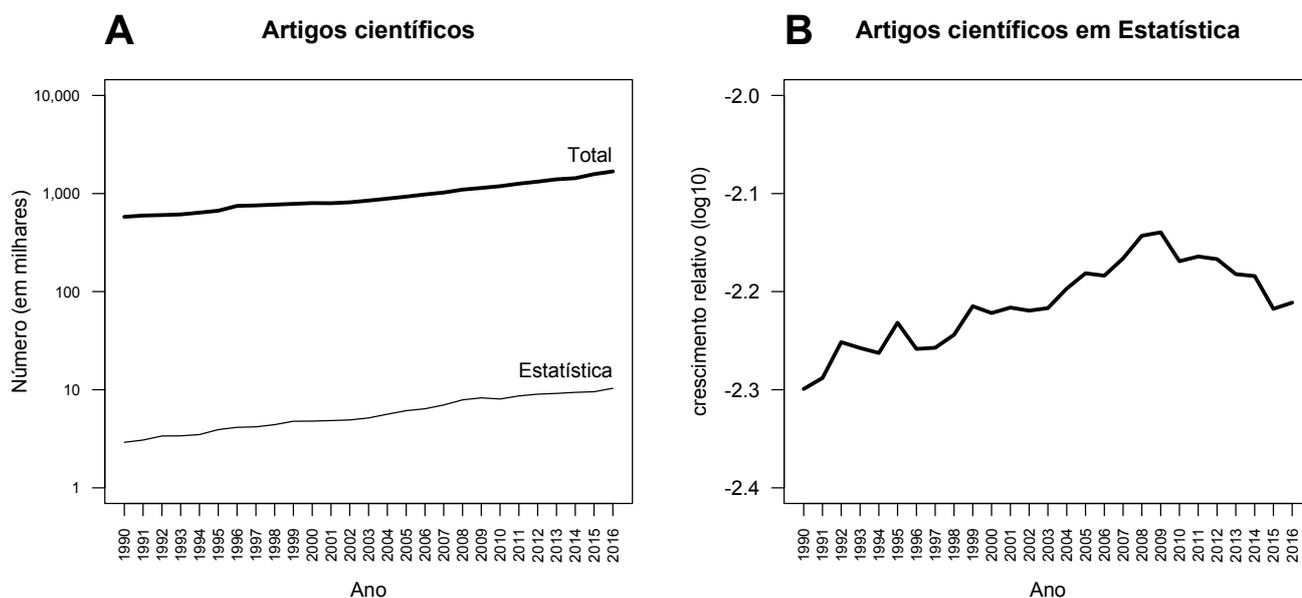


Figura 1: A. Número total de artigos científicos publicados em geral e em Estatística entre 1990 e 2016.

B. Crescimento em escala \log_{10} do número de artigos científicos publicados em Estatística relativamente ao número total de artigos publicados em cada ano. Fonte: *Web of Science* [13].

Para este artigo, a ideia não é tanto aprofundar as razões para o crescimento global da produção científica, mas sim entender as suas implicações para a credibilidade da Ciência e da Estatística. É preciso não esquecer que o método científico assenta na recolha e análise de dados como veículo de verificação de uma hipótese. Por isso, falar de Ciência é, como quem diz, falar um pouco de Estatística.

Com o número de artigos publicados a aumentar, há um sentimento latente de crise na reprodutibilidade científica, com alguns ecos para a sociedade civil [14,15]. Para saber a extensão desse sentimento de crise, uma sondagem foi recentemente realizada a 1,576 cientistas de diferentes áreas dentro do universo de leitores da revista *Nature* [16]. Embora possa haver um possível

enviesamento no universo inquirido, os resultados parecem ser inequívocos e merecem uma reflexão séria por parte da comunidade científica.

Primeiro, cerca de 90% dos inquiridos acha que há de facto uma crise ligeira ou significativa. Segundo, mais de 70% dos participantes confessa que não conseguiu reproduzir algum dos resultados publicados (do próprio ou de outrem). Terceiro, vários fatores foram apontados como redutores da reprodutibilidade científica, onde se destaca a publicação de resultados selecionados ‘a dedo’, ou a pressão para publicar (com uma frequência de respostas superior a 90%). Eventuais erros relacionados com a aplicação de métodos estatísticos ocupam o terceiro lugar com uma frequência de resposta rondando os 80% dos inquiridos.

Os resultados desta sondagem são preocupantes e mostram que o mantra de ‘publicar, publicar, publicar’ tem repercussões para a credibilidade da Estatística como disciplina de suporte a outras áreas científicas. Nas ciências biomédicas que conheço melhor, a crise na reprodutibilidade científica pode advir de muitos fatores, tais como imprecisões na definição da população em estudo, uso de desenhos experimentais incorretos ou enviesados, ou ainda, a especificação de tamanhos amostrais insuficientes. A incorreta especificação do modelo amostral ou o uso de testes de hipóteses inadequados também podem estar a contribuir para esta crise.

Uma razão mais sofisticada para o problema é sugerida pela prática generalizada de aplicar-se testes múltiplos ao mesmo conjunto de dados (até encontrar-se um resultado estatisticamente significativo). Em rigor estatístico, tal prática deveria ser acompanhada por uma correção do nível de significância de cada teste efetuado de forma a garantir um determinado nível de significância global (associado ao artigo no seu todo). Essa correção, sendo aplicada corretamente, teria o condão de pelo menos reduzir a chance de obter resultados espúrios sob a hipótese nula. É escusado dizer que essa correção é raramente aplicada na prática, com algumas áreas científicas de exceção, tal como é o caso da Genética, uma disciplina tradicionalmente ligada a alguma sofisticação matemática e estatística.

A ignorância das consequências associadas à prática acima referida parece ter motivado uma decisão editorial inédita em não publicar artigos que baseassem as suas conclusões no infame valor p [17]; não quero pensar que esta decisão tenha sido tomada por meras questões de *marketing*. Felizmente, esta sanção ao estilo medieval – tudo que o que não se compreende (bem), bane-se ou proíbe-se – conduziu a uma reação pedagógica por parte da comunidade estatística internacional [18], talvez não tão célere quanto desejada [1]. Neste clima de suspeição sobre o valor p , lanço a pergunta: como garantir que os artigos tenham a qualidade estatística desejável e que ao mesmo tempo sejam reprodutíveis?

A resposta a esta pergunta é evidentemente complexa. Da minha parte, parece-me óbvio que algo deve mudar na maneira de como se faz ciência atualmente. A sondagem da *Nature* sugere que a pressão para publicar deve ser eliminada, ou pelo menos reduzida, uma vez que o produto científico onde se inclui a análise estatística é como o vinho: precisa de tempo para ganhar corpo e alma. Contudo, penso que o paradigma baseado na publicação (rápida) será dificilmente mudado a curto-prazo, pois requer uma reforma estrutural de toda a atividade científica e a uma mudança cultural por parte dos cientistas.

Uma forma mais imediata de operar mudanças parece então passar por exigir uma maior responsabilidade dos cientistas e das comissões editoriais das revistas no entendimento dos princípios básico que regem uma análise estatística de qualidade.

Algumas revistas, tais como The Lancet ou PLoS One, já começaram a dar os primeiros passos nesse sentido, tendo painéis específicos de avaliadores pagos para a componente estatística dos artigos submetidos. Estes avaliadores não fazem normalmente parte das equipas editoriais das revistas, sendo pagos por manuscrito revisto (entre 70-80€). Pela minha experiência pessoal, estes avaliadores são jovens investigadores que usam essas remunerações extras para complementar o seu (magro) salário. Contudo, prevejo que, num futuro próximo, muitos desses avaliadores estatísticos se tornem verdadeiros profissionais dessa atividade devido à elevada competitividade no financiamento científico e ao estrangulamento das vagas relacionadas com a docência universitária em muitos países europeus.

Uma maior exigência estatística dos artigos submetidos levará a uma resposta da comunidade científica, que se espera positiva. Em primeiro lugar, haverá uma motivação extra de incluir um estatístico desde a conceção inicial de um projeto até à elaboração dos respetivos artigos. Em segundo lugar, a comunidade científica no seu todo tentará realizar esforços para elevar o seu nível de literacia estatística.

Já dizia o povo com razão que “é de pequenino que se torce o pepino” e tal provérbio parece ser bem adequado para o desenvolvimento de uma literacia estatística consciente e responsável. Deste ponto de vista, iniciativas como os projetos *Explorística, Radical Estatística ou Estatística Vai à Escola* para o ensino pré-universitário são apostas garantidas para o futuro da Estatística na sociedade portuguesa [19-21].

Ao nível global, a grande acessibilidade da internet tem proporcionado novos meios de divulgação e partilha do conhecimento estatístico. A enciclopédia online *Wikipedia* é um dos expoentes máximos desse fenómeno, permitindo a qualquer indivíduo aceder gratuitamente a conteúdos estatísticos de qualidade e amplamente diversos [22].

Os eventos *Technology, Entertainment & Design* (TED) podem também contribuir para uma maior disseminação de como pensar e fazer estatística. Estes eventos estão baseados na partilha de ideias que possam inspirar o mundo (*‘ideas worth spreading’*) através de apresentações cativantes de cerca de 15 minutos [23]. Não há restrições quanto aos temas em discussão, que podem abranger questões sobre alterações climáticas até aos fundamentos da felicidade individual. Estas apresentações são normalmente gravadas ao vivo em eventos próprios, sendo posteriormente partilhadas *online*.

A Estatística já conta com várias apresentações na comunidade TED. Destaco a palestra *‘The best stats you’ve ever seen’* do falecido professor Hans Rosling do Karolinska Institute que, desde 2006, tem vindo a colecionar número de visualizações, que já chega às 12 milhões neste momento. Num contexto nacional, encontrei cerca de 175 eventos com a assinatura TED realizados em Portugal, mas infelizmente não encontrei nenhum sobre a Estatística e o seu papel no mundo. Acho que está na altura de lançar esse repto aos membros mais jovens da nossa querida Sociedade Portuguesa de Estatística. Será que há alguém por aí com a coragem e bravura de aceitá-lo?

Conclusões

Ao finalizar este artigo, retomo o Tema Central da Estatística proposto pelo professor Fernando Rosado. Afinal, qual é de facto esse tema? Neste momento, há um admirável mundo novo de grandes volumes de dados com as suas promessas de revolucionar a forma de como fazemos ciência. Por outro lado, existe um complexo mundo de publicações em que objetivos comerciais se cruzam com uma crise latente na reprodutibilidade científica, onde a Estatística se encontra intimamente envolvida. O fio condutor entre as promessas do primeiro mundo e os problemas do segundo parece ser a forma, talvez errante, em como podemos ler os mesmos dados, pequenos ou grandes, publicados ou por publicar. A objetividade científica deveria imperar acima de tudo, não é assim?

Deste ponto de vista, o Tema Central da Estatística não assenta então, para mim, nos desenvolvimentos estatísticos do passado, do presente e do futuro, embora estes sejam peças fundamentais para a existência e sobrevivência da nossa disciplina científica. Assenta, sim, numa aspiração quase utópica de que, num futuro próximo, a sociedade que nos rodeia partilhe uma linguagem, uma cultura e uma ética estatística que nos una numa forma objetiva de encararmos o mundo. Afinal de contas, e parafraseando o estatístico Alan Smith [24], a Estatística não é mais nem menos do que a ciência de todos nós e, assim sendo, todos nós temos um pouco de estatísticos.

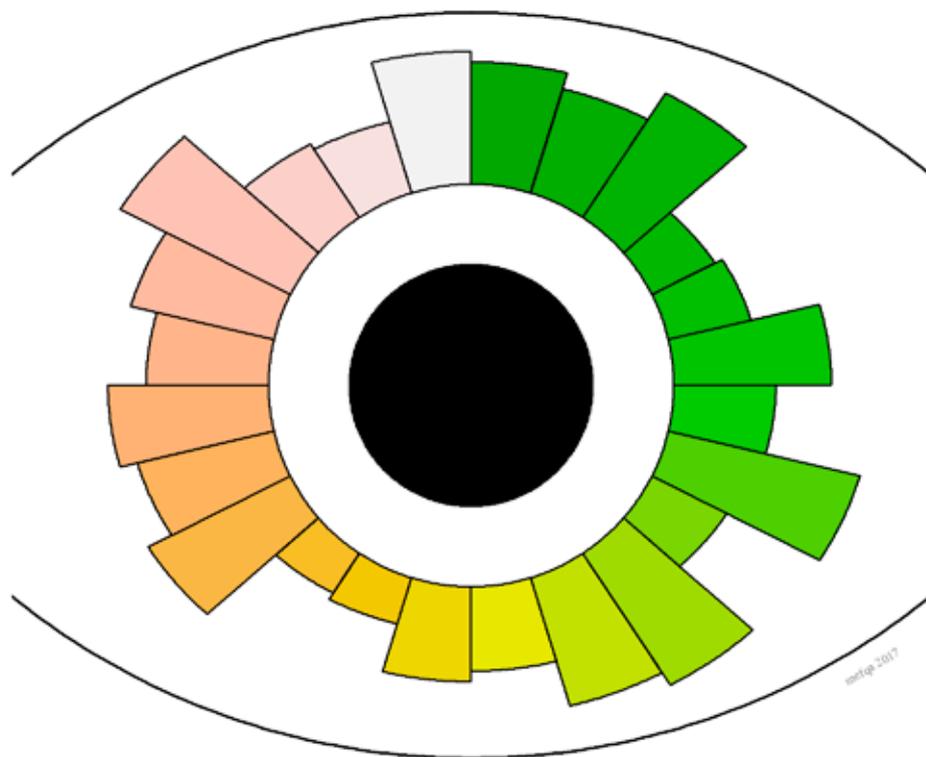
Referências

- [1] Branco, J. (2016). O Futuro da Estatística. *Boletim de Outono da SPE*, p. 49-58.
- [2] Kahneman, D (2011). *Thinking, Fast and Slow*. Penguin books, Londres.
- [3] Gato Fedorento - Magazine - Professor Chibanga.
Disponível em <https://www.youtube.com/watch?v=8fshrTYMwL4>.
- [4] “Statistical Methods for Big Data Symposium”, Centre for Statistical Methodology of the London School of Hygiene and Tropical Medicine. Disponível em <http://csm.lshtm.ac.uk/statistical-methods-big-data-symposium/>.
- [5] Franke, B., Plante, J.-F., Roscher, R., Lee, E.-S., Smyth, C., Hatefi, A., Chen, F., Gil, E., Schwing, A., Selvitella, A., Hoffman, M. M., Grosse, R., Hendricks, D., Reid, N. (2016). Statistical Inference, Learning and Models in Big Data. *International Statistical Institute* 84:371-389.
- [6] Sboner, A., Mu, X.-J., Greenbaum, D., Auerbach, R. K., Gerstein, M. B. (2011). The real cost of sequencing: higher than you think! *Genome Biology* 12:125.
- [7] Brito, P. (2016). Data Science – um desafio para os estatísticos?
Boletim de Outono da SPE, p. 40-42.

- [8] Blumenstock, J., Cadamuro, G., On, R. (2015). Predicting poverty and wealth from mobile phone metadata. *Science* 350, 1073-1076.
- [9] Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature* 457:1012-4.
- [10] Lazer, D., Kennedy, R., King, G., Vespignani, A. (2014). Big data. The parable of Google Flu: traps in big data analysis. *Science* 343:1203-5.
- [11] Buranyi, S. (2017). Is the staggeringly profitable business of scientific publishing bad for science? *The Guardian*. Disponível em <http://goo.gl/n2nCms>.
- [12] Ware, M., Mabe, M. (2015). The STM Report – An overview of scientific and scholarly journal publishing. Disponível em: <http://www.stm-assoc.org/>.
- [13] Web of Knowledge. Disponível em: <https://www.webofknowledge.com/>.
- [14] Jump, P. (2015). Reproducing results: how big is the problem? *Times Higher Education*. Disponível em <https://www.timeshighereducation.com/features/reproducing-results-how-big-is-the-problem>
- [15] Feilden, T. (2017). Most scientists ‘can’t replicate studies by their peers. *BBC News*. Disponível em <http://www.bbc.co.uk/news/science-environment-39054778>.
- [16] Baker, M. and Penny, D (2017). Is there reproducibility crisis? *Nature* 533:452-454
- [17] Trafimow D. and Marks, M. (2015). Editorial. *Basic and Applied Social Psychology* 37:1-2.
- [18] Wasserstein, R. L., Lazar, N. A. (2016) The ASA's Statement on p-Values: Context, Process, and Purpose. *The American Statistician*, 70:129-133.
- [19] Exploristica – Aventuras na Estatística. Disponível em <http://www.exploristica.com/>.
- [20] Radical Estatística. Disponível em <http://www.spestatistica.com/>.
- [21] A Estatística vai à Escola. Disponível em <http://aevae-aie2013.weebly.com/>.
- [22] Wikipedia. Disponível em <http://pt.wikipedia.org>.
- [23] TED Ideas worth spreading. Disponível em <https://www.ted.com>.
- [24] Smith, A. (2016). Why we are so bad at statistics. TEDxExeter. Disponível em: https://www.ted.com/talks/alan_smith_why_we_re_so_bad_at_statistics.



ESTATISTICA



UM NOVO OLHAR

Ciência Estatística

• Teses de Mestrado

Título: Correlação de eventos ocorridos em sistemas de *metering* de energia, instalados em redes de distribuição de baixa tensão

Autora: Sara Raquel da Silva Almeida, *junior_almeida_6@hotmail.com*

Orientadoras: Maria do Carmo Vaz de Miranda Guedes e Rute Alexandra Borges de Almeida.

Título: Formulação de aflúências sintéticas aos aproveitamentos hidroelétricos existentes e futuros

Autor: André Gustavo Pereira Moutinho, *andregmoutinho@gmail.com*

Orientadoras: Maria do Carmo Vaz de Miranda Guedes e Rute Alexandra Borges de Almeida.

• Tese de Doutoramento

Título: Análise Estatística de Séries de Contagem com Estrutura Periódica

Autora: Cláudia Susana Pereira dos Santos, *csps@ua.pt*

Orientadores: Isabel Maria Simões Pereira e Manuel González Scotto

Na minha tese os modelos autoregressivos de valores inteiros multivariados (MINAR) foram objeto de estudo. Estes desempenham um papel central na análise estatística de séries temporais de contagem. Dentro do razoavelmente grande espectro de modelos MINAR propostos na literatura, muito poucos focam a análise de séries de contagem com estrutura periódica. A análise dos processos de contagem multivariados apresenta muitos desafios que vão desde a especificação do modelo até à estimação de parâmetros. A minha tese teve como objetivo dar uma contribuição nessa direção. Especificamente, o objetivo deste trabalho foi duplo: primeiro, introduzimos o processo multivariado periódico de ordem um, PMINAR(1). As propriedades probabilísticas e estatísticas do modelo foram estudadas em detalhe. Para superar as dificuldades computacionais decorrentes da utilização do método da máxima verosimilhança introduzimos uma abordagem baseada na verosimilhança composta. O desempenho do método proposto e outros métodos concorrentes na estimação dos parâmetros foi comparado através de um estudo de simulação. A previsão também foi abordada. Uma aplicação de dados reais relacionados com a análise de fogos foi apresentada.

Em segundo lugar, propusemos dois modelos INAR (univariado e bivariado) com estrutura periódica, S-PINAR(1) e BS-PINAR(1), respetivamente. Ambos os modelos são baseados no operador *signed thinning* permitindo contagens de valores positivos e negativos. Apresentamos as propriedades probabilísticas básicas e estatísticas dos modelos periódicos. As inovações foram modeladas através das distribuições *Skellam* univariada e bivariada, respetivamente. Para avaliar o desempenho dos estimadores dos mínimos quadrados condicionais e da máxima verosimilhança condicional, foi realizado um estudo de simulação para o modelo S-PINAR(1).

Cláudia Santos

• Livros

Título: *SPE 2017 – Programa e Resumos.*

Autoras: Maria de Fátima Salgueiro e Paula Vicente.

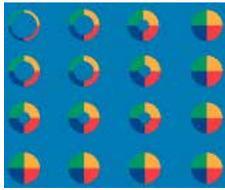
Ano: 2017. Editora: Edições SPE. ISBN: 978-972-8890-41-4. Depósito Legal: 432643/17.

Título: *Uma Introdução à Meta-Análise*

Autora: Maria de Fátima Brilhante

Ano: 2017. Editora: Edições SPE. ISBN: 978-972-8890-40-7. Depósito Legal: 432642/17.

Sessão de Entrega dos Prémios Estatístico Júnior 2017



Sociedade
Portuguesa de
Estatística

Porto
Editora

21 de Outubro 2017, 15:45
Grande Auditório
ISCTE, Lisboa

O Prémio Estatístico Júnior 2017, promovido pela Sociedade Portuguesa de Estatística e patrocinado pela Porto Editora, pretende incentivar o interesse pelas áreas de Probabilidades e Estatística dos estudantes dos Ensinos Básico e Secundário, e dos Cursos de Educação e Formação (CEF) e de Educação e Formação de Adultos (CEFA). Nesta sessão que teve lugar durante o XXIII Congresso da Sociedade Portuguesa de Estatística, foram entregues os Prémios atribuídos aos trabalhos distinguidos este ano, que versam estudos sobre temas relevantes para a sociedade e o indivíduo.

15:45 - Abertura

15:55 - Orador Convidado

Professor Pedro Freitas (FC- UL) - *Paradoxos com grandes números!*

16:30 - Entrega dos Prémios

17:15 – Lanche Convívio





Prémios Estatístico Júnior 2017

Trabalho classificado em 1º lugar *ex-aequo* (3º ciclo do Ensino Básico)

Título: *Hábitos de Vida*

Autores: Emanuel Sousa Roque, João Filipe Cruz Caçador e João Francisco Caçador Dinis

Professora orientadora: Eneida Edite Spencer Monteiro

Estabelecimento de Ensino: Colégio Senhor Milagres, Leiria

Trabalho classificado em 1º lugar *ex-aequo* (3º ciclo do Ensino Básico)

Título: *Peso das Mochilas dos alunos da EB 1/2/3 PFMSB*

Autora: Lara Filipa Rodrigues Vinagre

Professor orientador: Jaime Vieira

Estabelecimento de Ensino: EB 1/2/3, Fajã Ovelha, Madeira

Trabalho classificado em 2º lugar *ex-aequo* (3º ciclo do Ensino Básico)

Título: *Sociedade Materialista*

Autoras: Inês Santos Sousa Ferreira e Patrícia Oliveira Lopes

Professor orientador: Bernardino Carneiro Andrade

Estabelecimento de Ensino: EB 2/3 Maia, Maia

Trabalho classificado em 2º lugar *ex-aequo* (3º ciclo do Ensino Básico)

Título: *Turismo Sustentável*

Autor: Rui Pedro Carvalho Dias

Professor orientador: José António Fernandes Freitas

Estabelecimento de Ensino: EB Caldas Vizela, Vizela

Nota: Não foi atribuído o 3º lugar (Ensino Básico)

Trabalho classificado em 1º lugar *ex-aequo* (Ensino Secundário)

Título: *Qualidade almoço na cantina da escola*

Autores: Beatriz Isabel Tanganha Bento Lopes, Jorge Manuel S. Dordio e Pedro Miguel M. Rebocho

Professora orientadora: Cristina Alexandra Lopes Pereira

Estabelecimento de Ensino: Escola Secundária Rainha Santa Isabel, Estremoz

Trabalho classificado em 1º lugar *ex-aequo* (Ensino Secundário)

Título: *Jogos Santa Casa: E se eu fosse milionário?*

Autora: Andreia Filipa Lopes Sousa

Professora orientadora: Maria Alice da Silva Martins

Estabelecimento de Ensino: Escola Artur Gonçalves, Torres Novas

Nota: Não foi atribuído o 2º lugar (Ensino Secundário)

Trabalho classificado em 3º lugar (Ensino Secundário—Cursos Profissionais)

Título: *Hábitos Alimentares*

Autores: Ana Margarida Jacinto Costa Neto, Samanta Pratusевичiute e Carlos Filipe M. Santos Nunes

Professora orientadora: Mary Cristina Ferreira Rocha

Estabelecimento de Ensino: EnsiGuarda - Escola Profissional Guarda

Prémio SPE 2017

Problemas de tempo ótimo de paragem com critério integral

Carlos Miguel dos Santos Oliveira, *oliveira.cmds@gmail.com*

Os problemas de tempo ótimo de paragem são um tópico clássico em Análise Estocástica e Processos Estocásticos. Uma explanação aprofundada deste tema pode ser encontrada, por exemplo, no livro de Peskir e Shiryaev [1]. Com o desenvolvimento de algumas áreas contíguas à Análise Estocástica e aos Processos Estocásticos, como a Matemática Financeira, este tópico passou a ter um interesse crescente junto da comunidade académica, quer pela sua aplicabilidade (por exemplo, na valorização de opções financeiras), quer pelo grau de complexidade que foi introduzido com o refinamento destas aplicações.

A motivação para este trabalho vem da Teoria das Opções Reais: diz-se que uma empresa, ou um projecto de investimento em geral, “tem uma opção real” quando existe o direito, mas não a obrigação, de tomar uma decisão no futuro, como abandonar o mercado, investir num novo produto, aumentar a capacidade produtiva da empresa, entre outras (vide Dixit e Pindyck [2] ou Trigeorgis [3]). Assim, tendo em vista a opção de saída e a opção de investimento, neste trabalho estuda-se um problema de tempo ótimo de paragem, com critério integral, onde o processo estocástico é a única solução clássica de uma equação diferencial estocástica homogénea, com coeficientes determinísticos.

A solução deste tipo de problemas, usualmente denominada Função Valor, pode ser encontrada através da resolução de uma equação diferencial: a Equação de Hamilton-Jacobi-Bellman. O conjunto condições que a Função Valor deve satisfazer são geralmente conhecidas como Teoremas da Verificação.

No decorrer deste trabalho, demonstramos um Teorema da Verificação, onde se relaxa parte das condições de regularidade usualmente assumidas (como em Knudsen, Meister e Zervos [4]). Além disso, para o movimento Browniano geométrico, obtém-se, em alguns casos, uma solução analítica do problema em causa. Finalmente, dado que, em problemas económicos, é importante saber como varia a solução do problema de acordo com as variações de cada um dos parâmetros do modelo, realizamos uma análise de sensibilidade onde se estendem alguns resultados já conhecidos.

Referências

- [1] Peskir, Goran, and Albert Shiryaev. *Optimal stopping and free-boundary problems*. Birkhäuser Basel, 2006.
- [2] Dixit, Avinash K., and Robert S. Pindyck. *Investment under uncertainty*. Princeton university press
- [3] Trigeorgis Lenos. *Real options: Managerial flexibility and strategy in resource allocation*. MIT press, 1996.
- [4] Knudsen, Thomas S., Bernhard Meister, and Mihail Zervos. "Valuation of investments in real assets with implications for the stock prices." *SIAM journal on control and optimization* 36.6 (1998): 2082-2102.

Carlos Miguel dos Santos Oliveira, **galardoado com o Prémio SPE 2017**, é Licenciado em Matemática Aplicada à Economia e à Gestão pelo Instituto Superior de Economia e Gestão da Universidade de Lisboa, instituição onde também obteve o grau de Mestre em Matemática Financeira. Está a finalizar o Doutoramento em Matemática no Instituto Superior Técnico da Universidade de Lisboa sob a orientação da Professora Cláudia Nunes.

Prémio Carreira SPE 2017: um testemunho

O Prémio Carreira – SPE foi instituído em 2013 pela Sociedade Portuguesa de Estatística (SPE) e propõe-se reconhecer a atividade de estatísticos portugueses com papel de relevância no desenvolvimento científico, pedagógico e de divulgação da Estatística em Portugal.

Em 2017, o Prémio Carreira foi atribuído à **Doutora Maria de Nazaré Mendes Lopes**



numa sessão integrada no programa do XXIII Congresso da SPE (ISCTE, Lisboa) na tarde de 20 de Outubro.

Na breve e sucinta apresentação do seu percurso académico feita nessa sessão, pretendeu-se destacar a sua diversificada e intensa actividade docente, de investigação e gestão e reconhecer o seu papel preponderante na criação, desenvolvimento e afirmação do Grupo de Probabilidades e Estatística do Departamento de Matemática da Universidade de Coimbra, o que incluiu a sua própria formação académica, fortemente influenciada pela Escola Francesa, a sua capacidade de atrair novos elementos para esta área do saber e, também, todo o apoio dado para avançarem na sua formação científica.

Foi feita referência às teses de Doutoramento e Dissertações de Mestrado que orientou e à sua vasta produção científica, que inclui 3 livros publicados, edição de obras, capítulos de livros e múltiplos artigos em revistas e em *proceedings* de reuniões científicas nacionais e internacionais. Foi também referida a variedade dos seus interesses de investigação, que incidem sobre Estimacão funcional em processos estocásticos e sobre Modelacão estocástica de séries temporais, e realçada a sua contribuicão em aplicacões da Estatística às Ciências da Saude, com trabalhos colectivos sobre o Autismo em Portugal que mereceram a atribuicão de duas distincões, e à Astrofísica, com estudos sobre a Actividade Solar.

Foi ainda feita referência às suas permanentes actividades de gestão em órgãos universitários, destacando que tem sido, desde 2009, Directora do Departamento de Matemática da Universidade de Coimbra. A colaboracão sistemática com outras instituicões universitárias, em termos de docência, concursos, júris ou noutras matérias envolvendo a área de Probabilidades e Estatística, foi também salientada.

A concluir, quero referir que trabalhar com a Doutora Nazaré Mendes Lopes tem sido um privilégio para mim. Tenho tido oportunidade, ao longo de anos, de apreciar o seu saber, as suas qualidades científicas, o seu rigor matemático e também as suas qualidades de lideranca. Pelos momentos de trabalho, de convívio e de amizade, deixo o registo da minha enorme gratidão à Doutora Nazaré Mendes Lopes.

Esmeralda Gonçalves

Edições SPE - Minicursos

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

Retrospectiva do Boletim SPE

O *Boletim SPE* através dos seus “Tema Central”

- Primavera de 2017 - Destaque: Incerteza em Engenharia
- Outono de 2016 - Destaque: O Tema Central da Estatística
- Primavera de 2016 - Destaque: Séries Temporais e suas aplicações
- Outono de 2015 - Destaque: Estatística em Genética
- Primavera de 2015 - Destaque: Estatística no Desporto
- Outono de 2014 - Destaque: Estatística no Ensino Básico e Secundário
- Primavera de 2014 - Destaque: (Um) Ano Internacional da Estatística
- Outono de 2013 - Destaque: A "Escola Bayesiana" em Portugal
- Primavera de 2013 - Destaque: Estatística não-paramétrica
- Outono de 2012 - Destaque: Métodos Estatísticos em Medicina
- Primavera de 2012 - Destaque: Estatística no Ensino Superior Politécnico
- Outono de 2011 - Destaque: Análise de Sobrevivência
- Primavera de 2011 - Destaque: Sondagens e Censos
- Outono de 2010 - Destaque: Estatística Espacial
- Primavera de 2010 - Destaque: Data Mining - Prospecção (Estatística) de Dados
- Outono de 2009 - Destaque: Modelos Económétricos
- Primavera de 2009 - Destaque: Investigação (em) Estatística
- Outono de 2008 - Destaque: Processos Estocásticos
- Primavera de 2008 - Destaque: ALEA - Um sítio do nosso mundo
- Outono de 2007 - Destaque: Bioestatística
- Primavera de 2007 - Destaque: A "Escola de Extremos" em Portugal
- Outono de 2006 - Destaque: Ensino e Aprendizagem da Estatística

também disponíveis em <http://www.spestatistica.pt/index.php/publicacoes-57/boletins>



**SOCIEDADE PORTUGUESA
DE ESTATÍSTICA**

www.spestatistica.pt

Índice

Editorial	1
Mensagem da Presidente	5
Notícias	6
<i>Enigmística</i>	25
O Tema Central da Estatística	
“Estatística, para ver longe”	
<i>Emília Athayde</i>	26
A Estatística, a indústria e a inovação	
<i>Cláudia Nunes</i>	27
<i>Data Science, Big Data</i> e um novo olhar sobre a Estatística	
<i>Isabel Fraga Alves</i>	29
Estatística - “Espelho meu, espelho meu, que futuro terei eu?”	
<i>Conceição Amado</i>	32
A Revolução dos Dados	
<i>Pedro Campos</i>	35
A tirania dos jargões	
<i>Regina Bispo</i>	38
Novo olhar sobre a Estatística, imaginar o mundo	
<i>Lisete Sousa</i>	42
Desafios da Estatística para o século XXI	
<i>Tiago A. Marques</i>	45
A minha utopia sobre o Tema Central da Estatística	
<i>Nuno Sepúlveda</i>	49
“... com um novo olhar”	
<i>Emília Athayde</i>	56
Ciência Estatística	
<i>Teses de Mestrado</i>	57
<i>Tese de Doutoramento</i>	57
<i>Livros e Capítulos de Livros</i>	57
Prémios “Estatístico Júnior 2017”	58
Prémio SPE 2017	61
Prémio Carreira SPE 2017	62
Edições SPE – Minicursos	63
Retrospectiva	64