



Boletim



**SOCIEDADE PORTUGUESA
DE ESTATÍSTICA**

Publicação semestral

primavera de 2015



Estatística no Desporto

Séries Temporais no Desporto

Ana Diniz e Paula Marta Bruno 25

Investigação em Ciências do Desporto: dos testes de hipótese nula à necessidade de interpretações com significância prática e/ou clínica

Rui Marcelino e Jaime Sampaio 28

Estatística de Extremos em Desporto

Lígia H. Rodrigues, M. Ivette Gomes e Dinis Pestana 36

Editorial	2
Mensagem da Presidente	4
Notícias	5
<i>Enigmística</i>	16
SPE e a Comunidade	17
Pós-Doc	45
Ciência Estatística	49
Edições SPE – Minicursos	51
Prémios “Estatístico Júnior”	52

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



SOCIEDADE PORTUGUESA
DE ESTATÍSTICA

Prémio SPE 2015

Está aberto, até **30 de junho de 2015**, o concurso para atribuição do **Prémio SPE 2015**, de acordo com o seguinte regulamento:

1. Pretendendo dar destaque ao XXII Congresso da SPE, a principal reunião científica organizada pela Sociedade Portuguesa de Estatística, é instituído o **Prémio SPE 2015**.
2. Este prémio destina-se a estimular a atividade de estudo e investigação científica em Probabilidades e Estatística entre os jovens que trabalham nestas áreas.
3. **Prémio SPE 2015** é constituído por uma quantia de 1000 euros.
4. Ao **Prémio SPE 2015** podem concorrer trabalhos originais sobre temas de Probabilidades e Estatística, desde que não tenham sido objeto de qualquer prémio atribuído por outra instituição.
5. Os autores dos trabalhos candidatos ao **Prémio SPE 2015** devem ser estudantes ou investigadores em alguma instituição portuguesa ou bolsheiros portugueses, devem ser sócios da SPE e não devem ter completado os 35 anos de idade até 31 de dezembro de 2015. Os autores não devem ter recebido o Prémio SPE nas quatro edições anteriores. O trabalho deve ser escrito em português e não poderá exceder 25 páginas A4.
6. As candidaturas deverão vir acompanhadas do trabalho concorrente e do *curriculum vitae* dos autores e ser dirigidas ao Presidente da SPE. Podem ser enviadas por correio electrónico para **spe@fc.ul.pt** ou, em carta registada, para a morada a seguir indicada. O carimbo do correio validará a data de entrega.

Sociedade Portuguesa de Estatística
Bloco C6, Piso 4 - Campo Grande
1749-016 LISBOA

7. A decisão de admissibilidade e a apreciação dos trabalhos submetidos a concurso é da competência de um júri, cuja constituição será da responsabilidade da Direção da SPE.
8. Os critérios de seleção pautar-se-ão pela exigência e precisão nos vários aspetos que o júri considerar pertinentes, nomeadamente: i) qualidade e clareza do texto; ii) inovação e rigor científico; iii) contribuição para o desenvolvimento da área de Probabilidades e Estatística nos planos teórico, metodológico e/ou aplicado.
9. O júri é soberano nas suas decisões, não havendo lugar a recurso.
10. O trabalho galardoado com o **Prémio SPE 2015** será apresentado em sessão plenária pelo seu autor ou autores no XXII Congresso da SPE e publicado nas respetivas Atas, com o devido formato editorial, de acordo com o processo de revisão dos artigos submetidos a estas.
11. A atribuição do **Prémio SPE 2015** será anunciada logo que conhecida a decisão do júri e a sua entrega formal será feita no XXII Congresso da SPE na sessão plenária da sua apresentação.
12. O júri reserva-se o direito de não atribuir o **Prémio SPE 2015**.



CONGRESSO
SPE2015

SOCIEDADE PORTUGUESA
DE ESTATÍSTICA

**Centro de Congressos
Ria Formosa**

7 a 10 de Outubro de 2015

<http://spe2015.mozello.com>

Editorial

... na Estatística, como instrumento de apoio à decisão, também no desporto...

1. O Boletim SPE, como se sabe, foi iniciado em fevereiro de 1979. Desde o outono de 2006 – primeiro, a convite das Direções SPE presididas por Carlos Braumann e, em seguida, também por convite da Direção presidida por Daniel Paulino – dei continuidade a esse projeto SPE que, assim o creio, é um instrumento importante de comunicação e “partilha de ciência” entre os sócios da Sociedade Portuguesa de Estatística e entre estes e a comunidade.

A atual Direção SPE presidida por Maria Eduarda Silva - empossada em 17 de janeiro e de que damos Notícia - teve a gentileza de me convidar para continuar... É uma honra que de imediato aceitei com o intuito de prosseguir o trabalho. Nele pode-se introduzir mais alguma reflexão sobre a “problemática editorial”. Para tal, a oportunidade de uma nova Direção é incentivadora.

O Boletim SPE está consolidado na sua maquete editorial. Ela assenta basicamente em duas secções: *Tema Central* e *SPE e a Comunidade*. Aquela foi iniciada no outono de 2006 e *SPE e a Comunidade* na primavera de 2008. A criação deste espaço, como escrevi em editorial, acrescentou matéria científica que podemos situar num objetivo vasto de divulgação da Estatística entre os sócios mas também destes para toda a comunidade. As novas páginas acrescentadas diversificaram e ampliaram as intenções de divulgação e abertura da SPE. Damos assim (também) notícia de “outras ações” da SPE para além da sua importante atividade tradicional. Nessa secção, aberta a todas as colaborações, contamos com textos dirigidos a uma comunidade específica, um texto de apoio a um determinado curso ou um “artigo de opinião” e/ou de “divulgação de matéria científica” para a comunidade estatística e não só para ela.

Foi assim há já 8 anos. O amadurecimento adquirido ao longo de muitos anos bem como a opinião interventiva que tenho recebido dos sócios e leitores do Boletim SPE, permitem concluir sobre o bom modelo editorial assim construído. Mas, tudo isto, sem prejuízo de um desiderato de melhor racionalização - por exemplo dos custos e da eficácia editoriais. Esta será também uma mais-valia do Boletim SPE em favor da SPE.

2. Ronaldo, o excelente desportista é, cada vez mais, um exemplo que, também no domínio da Estatística, nos conduz a diversos temas de análise e estudo de casos em *Estatística no Desporto*. É consensual que o sucesso que todos lhe reconhecem se apoia na sua capacidade única de fazer render todos os talentos. Na literatura da especialidade, o premiado jogador, também deve dar origem a (mais) um caso de reflexão sobre o chamado “efeito Mateus” - desde logo porque nasceu em fevereiro. O “fenómeno Ronaldo”, um acontecimento global é, de facto, um caso de estudo da sua capacidade física, da excelência dos resultados, da sua metodologia de treino, da sua disciplina de trabalho e até (e acima de tudo?) da dimensão do produto de marketing. É um conjugar de variáveis muito diferentes por trás das quais a Estatística e, muito especialmente, as estatísticas estão sempre à espreita e dispostas para as mais diversas análises. Também no desporto a Estatística é um importante instrumento de apoio à decisão.

Segundo notícias recentes, a singularidade de Cristiano Ronaldo levou a Universidade da Colúmbia Britânica, no Canadá, a ministrar um curso que pretende analisar como se “constrói a lenda futebolística” e o que “representa o jogador para a diáspora portuguesa”. Consideram que em domínios da Sociologia, a figura mediática do laureado jogador português é uma boa forma de explicar Portugal aos emigrantes de segunda geração. Para além disso, também é objetivo do curso, dar a conhecer aos

estudantes os fatores que representam a “sociedade em que vivemos hoje”. Decerto que todos estes objetivos se apoiam e consolidam com argumentos estatísticos. Sempre assim é!

E surge assim uma justificação para a importância da *Estatística no Desporto*.

3. Mourinho é um treinador de sucesso. No futebol atual, tudo é registado e contabilizado; corridas, passes, centros e, evidentemente, golos. Uma empresa de estatísticas desportivas, analisa um enorme número de variáveis em cada encontro e vende os seus estudos. No campeonato inglês, os clubes mais importantes, empregam analistas para decidirem sobre os principais ensinamentos dos dados estatísticos dos desafios de futebol. As suas prestações, inclusive, ajudam a definir os objetivos em matérias de transferências e a elaborar estratégias de formação para os iniciados. E aqui surgem, desde logo, aqueles que contestam os analistas quantitativos por reduzirem um jogo de futebol a números – desde as coordenadas GPS aos perfis cardíacos. Em sua defesa surgem os que alegam que as estatísticas são imparciais ao contrário dos seres humanos na avaliação das qualidades dos jogadores. Disse um treinador: “Queriam que eu acreditasse mais nos meus olhos do que nas estatísticas, mas não vou nisso. Já vi mágicos a tirarem coelhos do chapéu e sei muito bem que lá dentro não havia coelho nenhum”. O princípio essencial do desporto é: saber quem é o melhor. Recentemente foi elaborado um estudo onde cerca de 200 mil jogos de Inglaterra e País de Gales, foram analisados desde 1888, data da fundação da Federação Inglesa. Ponderando as mais diversas pontuações conseguiu-se determinar que o vencedor foi a equipa do Chelsea durante a época 2005/06, isto é, quando Mourinho era treinador. E, sabe-se como este estratega desportivo usa “a folha de cálculo” como apoio para as suas decisões e táticas no futebol. Para a discussão “do dia seguinte” fica: “Os adeptos tirarão sempre das estatísticas argumentos para puxarem pela equipa do seu coração”.

4. Com os dois exemplos apresentados e que enchem Portugal, o tema central deste Boletim fica, pois, introduzido e justificado, com uma certeza: Os textos científicos formam um contributo para ajudar a explicar muito no Desporto. Dificuldades de última hora impossibilitaram a colaboração do Prof. José Maia. O Boletim SPE, como sempre, fica à disposição para partilha científica dos interessados no tema. E, aqui também, a Estatística é muito importante!

5. Agradeço à Prof. Ivette Gomes o texto / notícia que publicamos neste Boletim. Traduzido do original publicado, por convite, no Boletim da ASA, é um relato pessoal que, assim, temos oportunidade de partilhar com os leitores do Boletim SPE. Para memória futura registamos as palavras “de um diário” que muito nos orgulha e que são reveladoras da grande atividade de uma sénior portuguesa internacionalmente reconhecida como grande embaixadora e líder da Ciência Estatística.

À Ivette - também como uma pedra basilar da SPE - que se afirma “jovem de espírito” que, bem o sabemos, tem “sido bem recebida em vários lugares do mundo” e que acaba de ser eleita para alto cargo no International Statistical Institute, devemos felicitar. “Embora já aposentada” como diz, para ela esta rotulagem social é um simples início de diário pois a sua atividade continua a ser um exemplo mesmo para os do ativo e, muito em especial, também para os jovens cientistas.

A Lusofonia vai liderar o ISI o que implica um grande orgulho para Brasil e Portugal, portanto, digamos também para ABE e SPE.

Muitos parabéns Pedro pela tua eleição como Presidente do ISI!

Muitos parabéns Ivette pela tua eleição como Vice-Presidente do ISI!

O tema central do próximo *Boletim* será *Estatística em Genética*.



Mensagem da Presidente

Caros sócios da SPE

Os novos órgãos administrativos da SPE, eleitos em Assembleia Geral que decorreu durante o I Encontro Luso-Galaico de Estatística em Ambiente e Ecologia na UTAD, tomaram posse no dia 17/01/2015 em sessão realizada na sede da SPE. Estamos, assim, no início do mandato e determinados a fazer todos os esforços para bem servir a Estatística. As direções que nos precederam deixaram uma SPE organizada e eclética nas ações em curso no âmbito dos 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 sector privado e na administração pública*”. De facto, em anos recentes o âmbito das atividades da SPE alargou-se, podendo mesmo ser uma surpresa para aqueles sócios que estão familiarizados apenas com o programa de encontros científicos. Há, no entanto, ainda muito a fazer em prol da imagem pública da Estatística como disciplina presente em todas as áreas da vida no século XXI. A nossa Sociedade tem de ter um papel mais ativo junto da sociedade em geral, dos meios de comunicação social, dos decisores a nível governamental. É de assinalar que as intervenções de vários sócios na Assembleia Geral do dia 11 de fevereiro e que desde já agradeço foram precisamente neste sentido: a SPE tem de reforçar a sua imagem e intervenção na sociedade de modo a sensibilizar o público em geral sobre a Estatística. Esta questão está intrinsecamente ligada ao apoio ao desenvolvimento da disciplina como área de conhecimento e, portanto, à intervenção da SPE em questões ligadas à educação a todos os níveis desde o ensino básico ao superior, à investigação e à interação com outras áreas científicas e de aplicação. Para além de consolidar as ações já em curso, temos de refletir sobre os desafios que enfrentamos e desenhar estratégias futuras para apoiar o desenvolvimento da Estatística. Esta é uma tarefa que requer o empenho de todos os sócios, pelo que desde já apelo ao envolvimento participado de todos.

O início do mandato tem sido pautado pela necessidade de responder a inúmeras solicitações, querendo desde já deixar um agradecimento especial aos sócios que aceitaram o nosso convite para representarem a SPE.

Neste boletim podem os sócios encontrar informação sobre o XXII Congresso da SPE a realizar no Algarve de 7 a 10 de Outubro, notícias frescas sobre AEVAE, um resumo dos primeiros 720 dias de atividade da Explorística e planos para o futuro desta iniciativa, e ainda informação sobre a constituição das comissões especializadas, e representação da SPE nas diversas entidades nacionais e instituições internacionais.

Porto, 3 de março de 2015

Cordiais saudações

Maria Eduarda Silva

Notícias

• Novos Órgãos Sociais da SPE

Em sessão realizada na sede da SPE, tomaram posse no dia 17 de janeiro de 2015, os elementos constituintes dos seus órgãos administrativos eleitos no passado dia 7 de novembro.

A constituição dos novos órgãos administrativos da SPE para o triénio 2015 – 2017 é a seguinte:

Direcção

Presidente: Maria Eduarda Silva, sócia 193 (Universidade do Porto)

Vice-Presidente: Isabel Simões Pereira, sócia 45 (Universidade de Aveiro)

Tesoureira: Patrícia de Zea Bermudez, sócia 185 (Universidade de Lisboa)

Primeiro Vogal: Cláudia Nunes Philippart, sócia 48 (Universidade de Lisboa)

Segundo Vogal: Maria Esmeralda Gonçalves, sócia 92 (Universidade de Coimbra)

Mesa da Assembleia Geral

Presidente: Maria Antónia Amaral Turkman, sócia 6 (Universidade de Lisboa)

Primeiro Vogal: Carlos Marcelo, sócio 112 (Instituto Nacional de Estatística) e

Segundo Vogal: Russell Alpizar-Jara, sócio 569 (Universidade de Évora)

Conselho Fiscal

Presidente: Marília Antunes, sócia 146 (Universidade de Lisboa)

Primeiro Vogal: Carla Henriques, sócia 511 (Instituto Politécnico de Viseu)

Segundo Vogal: Tiago Marques, sócia 722 (University of St.Andrews)



A Direção da SPE para o triénio 2015 -2017

Linhas programáticas

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 pelas direções anteriores 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.g., conceções pedagógicas atuais e seus resultados práticos; criação de associações setoriais e interface com a SPE) 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 e-boletim para membros e não membros;
- re-avaliar a programação de encontros científicos: reuniões magnas e encontros de abrangência mais reduzida.

• Comissões Especializadas e Representações na SPE

1. *Secção de Biometria*

Pedro Oliveira- Presidente
Giovani Silva- Secretário

2. *Comissão Especializada de Educação – CEE*

Eugénia Martins – Coordenadora
Cristina Rocha - *Representante na Comissão Científica do IAVE*
Manuela Neves
Andreia Hall
Fernanda Otília- *Auditoria de exames*
Cláudia Nunes

3. *Comissão Especializada de Nomenclatura Estatística- CENE*

Daniel Paulino, Dinis Pestana e João Branco

4. *Representação na Comissão Nacional de Matemática*

Isabel Pereira

5. *Representação no Centro Internacional de Matemática*

Esmeralda Gonçalves

6. *Representação na Rede Portuguesa de Matemática para a Indústria*

Cláudia Nunes

7. *Representação na FENSTATS e ISI*

Maria Eduarda Silva

8. *Representação no IASE*

Manuel Scotto

9. *Representação no Espaço Matemático em Língua Portuguesa EMeLP*

Eugénia Martins

10. *Explorística*

Pedro Campos - Coordenador
Conceição Rocha

11. *AEVAE – A Estatística vai à Escola*

Tiago Marques – Coordenador
Carla Henriques
Fátima Brilhante
Sandra Mendonça

• XXII Congresso SPE

O XXII Congresso da Sociedade Portuguesa de Estatística decorrerá de 7 a 10 de outubro no Algarve, sendo evento organizado pelo Departamento de Matemática da Faculdade de Ciências e Tecnologia, e Departamento de Engenharia Civil do Instituto Superior de Engenharia da UAlg.



O Congresso da SPE será antecedido por um minicurso intitulado **“Introdução à Estatística Bayesiana Computacional”** assegurado pelos Professores Antónia Turkman (Departamento de Estatística e Investigação Operacional da Faculdade de Ciências) e Carlos Daniel Paulino (Departamento de Matemática do Instituto Superior Técnico) da Universidade de Lisboa. O programa científico do congresso será composto por sessões temáticas organizadas, comunicações livres (orais ou posters) e por quatro sessões plenárias proferidas pelos conferencistas convidados:

- James W. Taylor (Said Business School, University of Oxford)
- Luzia Gonçalves (Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa)
- Manuel Scotto (Departamento de Matemática, Universidade de Aveiro)
- Peter Müller (Department of Mathematics, University of Texas at Austin).

Para além de um clima ameno e muito sol, no Algarve encontra uma diversidade de maravilhas naturais. Desde as praias de excelente qualidade, com areais a perder de vista até às falésias esplendorosas, à serra algarvia onde as tradições se mantêm.

Para uns dias de intenso convívio estatístico é com enorme entusiasmo que convidamos todos a participarem no XXII Congresso da SPE a realizar no Algarve onde, para “além do mistério do mar e do milagre do sol” (Miguel Torga), se “vêm as estrelas enormes reluzindo através das amendoeiras” (Raúl Brandão).

Para mais informações consulte: <http://spe2015.mozello.com>

A Comissão Organizadora Local

• JOCLAD2015

A Associação Portuguesa de Classificação e Análise de Dados (CLAD) e a Escola Superior de Tecnologia do Barreiro do Instituto Politécnico de Setúbal (ESTBarreiro/IPS) têm o prazer de organizar as XXII Jornadas de Classificação e Análise de Dados (JOCLAD 2015) que decorrerão nos dias 10 e 11 de abril de 2015 no Barreiro, Portugal.

No Programa Científico da edição de 2015, estão previstos:

- Dois Minicursos em paralelo, 10 de Abril de 2014 (manhã), ministrados pelos Professores convidados Gilles Celleux, INRIA, França e Cristina Rocha, FCUL;
- Três Sessões Plenárias proferidas pelos Professores convidados Gilles Celeux, INRIA, França; Ruy Ribeiro, FMUL e ALAMOS,EUA; Fernanda Figueiredo, FEP;
- Quatro Sessões Temáticas (INE, Banco de Portugal; Educação, Controle de Qualidade);
- Comunicações Livres selecionadas, organizadas em Sessões Paralelas Oraís e Posters.

Os participantes e seus acompanhantes registados são convidados a participar no Programa Social das Jornadas: Jantar ou Movimento IGNITE.

Informamos ainda que os trabalhos apresentados nas JOCLAD2015 podem ser submetidos para publicação no próximo livro CLAD, sendo sujeitos a avaliação por referee.

Mais informações em: <http://joclad2015.estbarreiro.ips.pt>

• As minhas experiências na África do Sul e na SASA 2014

Embora já aposentada, sinto-me ainda jovem de espírito e adoro viajar. Tenho sido bem recebida em vários lugares do mundo, mas a hospitalidade dos Sul-Africanos foi na realidade espetacular e deixou-me excelentemente impressionada. Depois de um convite informal de Lizanne Raubbenheimer, *Chair* da *56th Annual Conference of the South African Statistical Association (SASA)*, seguida de um convite formal de Paul Mostert, Presidente da SASA, o meu calendário levou-me à Universidade de Rhodes, *Grahamstown*, para um *workshop* de um dia e uma lição plenária na SASA 2014. Estas tarefas foram seguidas de visitas à Universidade de Stellenbosch e à Universidade *the Free State, Bloemfontein*, para seminários, e à Universidade de Pretoria para uma versão abreviada do curso em *Grahamstown*. Segue-se um relatório breve dessa visita, tradução do que me pediram para preparar para a SASA Newsletter.

Sexta/Sábado, Outubro 24/25: Quando cheguei a *Johannesburg* às primeiras horas da manhã de sexta feira, dia 24, a minha bagagem tinha ficado em Londres ... Mas mesmo assim, decidi não declinar o convite da Lizanne, e fomos diretas para o *Addo Elephant National Park*, onde tive a oportunidade de conhecer alguns simpáticos colegas da Universidade de *Port Elizabeth* (P.E.), entre os quais menciono Garry Sharp e Chantelle Clohessy. O safari guiado foi espetacular, e conseguimos ver uma grande diversidade de animais selvagens, em liberdade total. Mas no Sábado de manhã o passeio no carro da Lizanne foi também muito especial: quase tomámos um segundo duche ao permanecermos com a janela do carro aberta a menos de 2 metros de um enorme elefante, que estava calmamente a beber água numa poça perto da berma do caminho. Algumas fotos, tiradas pela Lizanne, estão acessíveis em <https://www.dropbox.com/sh/aaftpqbjja61f91/AABgc7r683Vcw4B-ByKsttSRa?dl=0>. Tivemos um almoço/jantar tardio já em *Grahamstown*, e a Lizanne deixou-me na *High Corner Guest House*, onde a hospitalidade foi deveras gratificante.

Domingo, Outubro 26: De manhã, tive a oportunidade de fazer uma breve visita a *Grahamstown*. Fui ao *Monument*, à Catedral, à Câmara, onde assisti a uma cerimónia cheia de cor, e a alguns outros lugares. Depois de um belo almoço de lulas, onde tive o grande prazer de conhecer Sarah Radlof, o seu marido, Tim, e alguns outros colegas da África do Sul, fomos até à *Pumba Game Reserve*. E apesar de ‘Pumba’ ter em Português uma conotação com uma ‘grande queda’, não caímos, mas a experiência foi bastante radical, em caminhos cheios de bossas ... Mas a visita foi mais uma vez fantástica. E se quiserem ter uma experiência de alguns dos animais que vimos, visitem <http://youtu.be/-2kytTtAwhg>, onde podem encontrar um vídeo que Michael Greenacre colocou no seu *site* CARME (Correspondence Analysis and Related Methods). Tal como Michael diz, o vídeo cobre um ‘*related method*’ ... O jantar foi delicioso e pela primeira vez consegui provar o já tão falado e maravilhoso pudim de malva. Fiquei realmente encantada com o passeio e o chamado ‘espírito do leão branco’, sentindo-me diferente do usual quando regressei, particularmente depois do ‘boma braai’, de uma canção sugerida também pelo Michael no fim do jantar, intitulada ‘Pumbaya My Lord’... e de um céu repleto de estrelas superluminosas.

Segunda, Outubro 27: Mas na vida também precisamos de trabalhar, algo que felizmente ainda me dá prazer, e tive de dar aulas durante todo o dia num ‘workshop’ intitulado “*Statistics of Extremes and Applications*”. Estavam registados sessenta participantes, e foi para mim deveras gratificante ver tanta gente interessada na área, depois de mais de 30 anos dedicados ao desenvolvimento deste tema. E mesmo depois de um belo almoço nas ‘*Drostdy Lawns*’, o número de participantes era ainda bastante elevado e capazes de estar bem atentos, colocando questões interessantes e passíveis de contribuir para o desenvolvimento da área. Tal como disse no epílogo do ‘workshop’, e parafraseando o epílogo do nosso livro editado pela Sociedade Portuguesa de Estatística e Instituto Nacional de Estatística, em colaboração com as colegas e amigas Isabel Fraga Alves e Cláudia Neves, espero ter conseguido aumentar o interesse dos participantes pelo tema, relativamente recente de um ponto de vista histórico, mas com inúmeras aplicações, tantas quantas as que cada um consiga conceber. Um convite da Comissão Executiva da SASA, permitiu-nos ter um jantar soberbo num restaurante nas vizinhanças da universidade.

SASA 2014 (de Terça 28 até Quinta 30): Depois da minha conferência plenária de Terça-feira, sobre “*Penultimate Approximations and Reliability of Large Coherent Systems: Past, Present ... and Future*” tive a oportunidade de ouvir várias palestras interessantes em áreas variadas da Estatística, não só de estatísticos experientes mas também de jovens estatísticos. E fiquei deveras admirada com um tão grande número de entusiasmados estatísticos jovens, e espantada com as apresentações profissionais de alguns estudantes de Mestrado. Fiquei com a certeza que a Estatística na África do Sul está no bom caminho, contrariamente com o

que infelizmente está hoje em dia a acontecer em Portugal, não só com a Estatística mas com a Ciência em geral. O programa científico da **SASA 2014** teve uma gama muito variada de contribuições estimulantes, quer orais quer na forma de ‘*posters*’, e as sessões a que assisti foram sempre colmatadas por discussões muito salutares. O ‘*catering*’ e o programa social foram também excelentes. O Congresso Anual da **SASA** dependeu de um esforço muito bem sucedido da Lizanne, da Sarah e de alguns outros colegas da Universidade de Rhodes, auxiliados por um número muito elevado de estudantes, que estavam sempre prontos a ajudar os participantes. Sei por experiência própria que a organização de conferências é cansativa, mas esta conferência foi sem dúvida um ‘*Big Success*’, e tenho a certeza que os organizadores se sentem agora totalmente recompensados pelo grande esforço que despenderam. O jantar da conferência e o convívio de jovens estatísticos (e menos jovens) no ‘*Rat and Parrot*’ foram também momentos altos do programa.

Sexta, Outubro 31: Logo pela manhã, Tim, o marido da Sarah, levou-nos de carro até P.E. e eu viajei para *Stellenbosch*, via *Cape Town*. Durante o voo tive a oportunidade de apreciar uma bela vista da costa entre P.E. e *Cape Town*, e quando estávamos quase a chegar, percebi na realidade o verdadeiro significado da denominação ‘*Table Mountain*’. Tertius de Wet estava à minha espera no aeroporto, e fomos no seu carro até *Stellenbosch*. Mais uma vez, a *Guest House* (Bonne Esperance) era muito acolhedora, e nessa mesma tarde tive a oportunidade de fazer uma breve visita à cidade e ao jardim botânico. À noite ainda tive o prazer de encontrar Lynnete, a mulher de Tertius. Fui na realidade extraordinariamente bem recebida e o jantar que ela cozinhou estava soberbo.

Sábado, Novembro 1: Este foi um dia de passeio: *Cape Town*, *Table Mountain*, *Kirstenbosch Botanical Gardens*, e a costa perto de *Cape Town*. A paisagem que se desfruta do topo da *Table Mountain* é realmente espetacular e inesquecível. E também nunca irei esquecer a enorme variedade de proteias em *Kirstenbosch* nem o ‘*Centenary Tree Canopy Walkway*’ que passa sobre o *Arboretum* (*The Boomslang*’ ou árvore cobra). Ao almoço, na *Kirstenbosch Tea Room*, provei uma saborosa *Bobotie*, e o jantar, no restaurante *Makaron*, já em *Stellenbosch*, não só com Tertius de Wet e Willie Conradie, mas também as esposas, foi mais uma vez excecional. Não tivemos a oportunidade de ir a *Cape Point*, e portanto, como Portuguesa obrigada a aprender a história das nossas descobertas e do que os nossos marinheiros sofreram até finalmente conseguirem ultrapassar o Cabo das Tormentas em 1488, sob a liderança de Bartolomeu Dias, tenho uma razão forte para voltar à África do Sul e ao *Cape Point*, para ter oportunidade de ver, entre outras coisas, a réplica em pedra da cruz de Vasco da Gama.

Domingo, Novembro 2: Para além de uma segunda visita ao jardim botânico de *Stellenbosch* e de um agradável passeio ao longo do rio, tive a oportunidade de assistir a um belo concerto de Natal, seguido de um jantar de pizza, a convite do Tertius.

Segunda, Novembro 3: Durante a manhã, discuti com Tertius tópicos de possível investigação conjunta futura, essencialmente na área de valores extremos. O seminário, intitulado “*Threshold Selection for Reduced-Bias Location-Invariant EVI-estimators: A Financial Application*”, teve uma audiência bastante interessada e interessante, com colegas não só da Universidade de *Stellenbosch*, mas também das Universidades de *Cape Town* e de *Western Cape*. Mais uma vez, para além de alguns seniores, gostaria de mencionar um grupo de jovens cientistas com muito dinamismo e entusiasmo.

Terça, Novembro 4: Conduzida por Tertius, voei ao fim da manhã de *Cape Town* para *Bloemfontein*. A minha visita à *University of the Free State* foi curta, mas foi outra experiência inesquecível, mais uma vez essencialmente devido à hospitalidade dos colegas e ao local onde permaneci uma única noite, *Aandmuzik Guest House*. A minha anfitriã em *Bloemfontein* foi Andrehette Verster, e a hospitalidade esteve bem acima das minhas expectativas. Para além disso a *Guest House* em *Bloemfontein* era excecional, e depois do meu seminário na Universidade, onde mais uma vez contei com a presença de colegas muito entusiastas da Estatística, e um outro belo jantar, tive a oportunidade de ter um merecido e relaxante descanso.

Quarta, Novembro 5: Andrehette levou-me até ao aeroporto, e voei para *Pretoria*, via *Johannesburg*, de manhã cedo. Cheguei ao aeroporto de *Joanesburgo* ao fim da manhã, e fui eficientemente conduzida até uma outra bonita *Guest House*, onde fiquei apenas durante duas noites. E o *Campus* de *Hatfield* em *Pretoria* foi outra excelente experiência. Tinha anteriormente encontrado, durante a 56th *Session of the International Statistical Society (ISI 2007)*, em Lisboa, Andriette Bekker e René Ehlers, e na tarde deste mesmo dia Andriette organizou uma visita guiada ao *Campus*, onde fiquei a conhecer as inúmeras facilidades existentes, as magníficas esculturas espalhadas pelo campus, uma belíssima exibição de porcelanas pintadas e os belos Jacarandás azúis, ainda em flor. E o número de estatísticos jovens que trabalham arduamente para o desenvolvimento da

Estatística no Departamento liderado por Andriette é realmente fora do comum. Entre outros, tive o prazer de encontrar a Inger, o Johan, a Marin e o Paul.

Quinta, Novembro 6: Durante a manhã, uma versão abreviada do curso breve ministrado na **SASA 2014** tinha mais uma vez uma assistência de jovens e de seniores, que claramente mostraram o seu interesse pela área. Depois de um belíssimo almoço no *Kream*, que é sem dúvida um dos bons restaurantes de *Pretoria*, ainda tive o prazer de falar com alguns investigadores, que trabalham essencialmente na área de Controlo Estatístico da Qualidade, outra das áreas da minha eleição.

Sexta, Novembro 7: O meu avião partiu de *Johannesburg* para Lisboa, via Londres, à noite mas nesse dia ainda fomos, a Inger, a Marin, a René e eu, a um '*Wild Rhino and Lions' Park*'. No caminho para o Parque tivemos a oportunidade de apreciar bonitas vistas de Pretoria, e passámos por um conjunto bastante longo de Jacarandas brancos, um tipo de árvore que eu não fazia a mínima ideia que existia. Antes da visita ao Parque, ainda tivemos um pacífico e excelente almoço em *Toadbury Hall*. E a nossa visita ao parque foi muito bem sucedida. Vimos uma percentagem muito elevada das espécies disponíveis, e na Crèche, estivemos em contacto direto com um leão de 4 meses e 2 tigres de 3 meses. Algumas fotos, tiradas pela Inger Fabris-Rotelli, estão disponíveis em: <https://www.dropbox.com/home/Rhino%20Lion%20Park%20SA>

Sábado, Novembro 8: Depois de dois voos bastante pacíficos, cheguei a Lisboa, para reencontrar a família e ter um belo almoço, agora em minha casa e cozinhado pelo meu marido.

Gostaria de terminar com um agradecimento a todos os colegas que encontrei na África do Sul, pela sua enorme hospitalidade. Foi bom lembrar amizades e fazer novos amigos, que espero encontrar em breve algures no nosso planeta, possivelmente em Lisboa ou no Rio de Janeiro, para o *ISI World Congress*, em Julho 2015, uma vez que sei que muitos desses investigadores irão lá estar.

M. Ivette Gomes

• Ivette Gomes, Vice-Presidente do ISI

Num processo eleitoral que decorreu entre 4 de Setembro e 27 de Outubro de 2014, a nossa colega Maria Ivette Gomes – grande dinamizadora da SPE, ex-Presidente da SPE e cientista de renome internacional – foi eleita vice-Presidente do ISI International Statistical Institute.

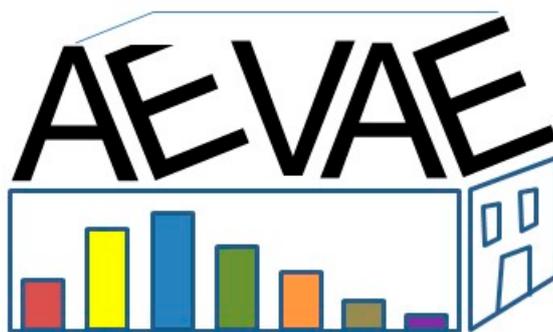
A aprovação formal dos resultados eleitoriais ocorrerá na Assembleia Geral do ISI por ocasião do WSC – World Statistics Congress que se vai realizar no Brasil em 31 de julho de 2015.

Os novos órgãos sociais do ISI para o período 2015-19, com o Presidente eleito Pedro Nascimento Silva iniciarão funções naquela atividade bienal do ISI.

Formulamos os votos do maior sucesso para esta nova equipa do ISI.

FR

• A Estatística vai à Escola

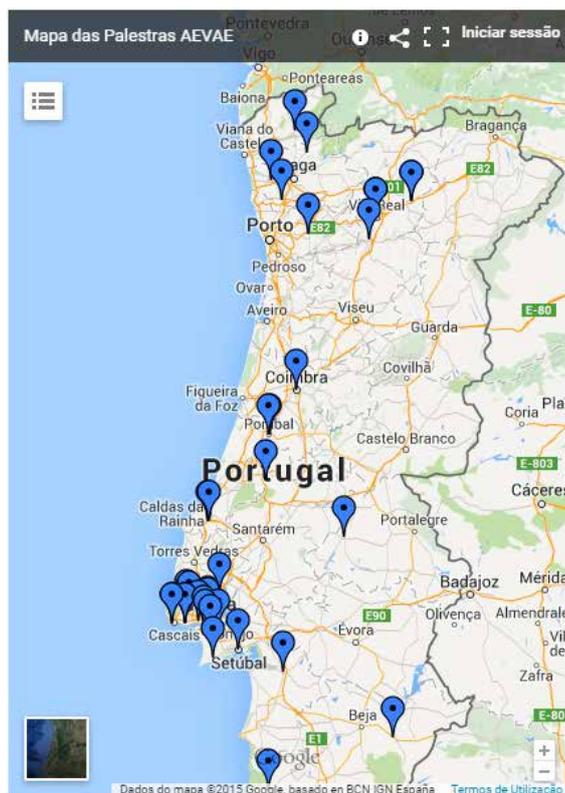


“A Estatística vai à Escola” (AEVAE) é um projeto da Sociedade Portuguesa de Estatística (SPE) em parceria com o Centro de Estatística e Aplicações da Universidade de Lisboa (CEAUL), que esteve em sintonia com o novo movimento internacional The World of Statistics.

Através da realização de palestras nas escolas de Ensino Básico e Secundário, o projeto visa a promoção da Estatística junto dos alunos, contribuindo para o “entusiasmo” pela Estatística, tanto dos mais novos como dos que se aproximam da decisão do curso/via profissional a seguir. A ideia é colocar os alunos em contacto com uma estatística aplicada e apelativa, pela voz de quem a pratica no terreno.

Continuando o trabalho desenvolvido em 2013, em 2014 a AEVAE chegou a cerca de 1500 alunos do secundário, através de 17 palestras dadas por 12 palestrantes. Os objectivos para 2015 passam por dar continuidade ao projecto, aumentando preferencialmente o número de palestras e palestrantes disponíveis, bem como a sua distribuição geográfica. O mapa ao lado ilustra os locais onde a AEVAE já chegou. Estamos convictos de que haverá um aumento *significativo* nos pontos assinalados no gráfico durante o decorrer de 2015.

A equipa do projeto vai ser renovada, saindo a Lígia Henriques (ex-Instituto Politécnico de Tomar) e o Giovanni Silva (Instituto Superior Técnico) e entrando a Carla Henriques (Instituto Politécnico de Viseu), a Fátima Brilhante (Universidade dos Açores) e a Sandra Mendonça (Universidade da Madeira), continuando a coordenação do projeto a cargo do Tiago Marques (University of St Andrews/Universidade de Lisboa).



As escolas ou professores interessados em receber as palestras da AEVAE podem contactar diretamente a comissão organizadora através do endereço eletrónico aevae.aie2013@gmail.com ou através do formulário de contacto através da página web <http://aevae-aie2013.weebly.com/>, onde também podem ter acesso à lista das palestras e palestrantes disponíveis.

TM



Os primeiros 720 dias

Em 2011, por iniciativa da Direção da SPE, começámos a desenvolver a Explorística, exposição itinerante concebida com o objetivo de levar os fundamentos da Estatística e das Probabilidades às comunidades educativas do ensino básico e secundário. Financiada pela Ciência Viva, a Explorística foi inaugurada na Escola Secundária de Tomaz Pelayo a 5 de Fevereiro de 2013 e, desde então, é verdadeiramente itinerante, tendo percorrido mais de 2000 quilómetros e tendo sido visitada por cerca de 3000 milhares de alunos em várias partes do país.

Cabe agora programar o futuro, definindo o que segue no imediato e quais os novos desafios.

O que se segue.

- Uma encomenda para uma cópia da Explorística em inglês destinada ao CSO, Central Statistical Office da Irlanda. Encontra-se em execução pelo mesmo autor/criativo da atual versão, o Jorge Ribeiro.
- Nova apresentação na conferência organizada pela IASE em 2015 e no Congresso Mundial de Estatística, organizado pelo ISI (ambos os eventos terão lugar no Rio de Janeiro, Brasil, em 2015).

Novos desafios.

- Não obstante o aparente sucesso, patente no número de pedidos para albergar a exposição, sentimos um natural desgaste nos módulos da Explorística (implicando substituição de materiais e assistência técnica) e também um desgaste humano (associado às viagens para montagem e apresentação). É por isso essencial reforçar equipa. Apelamos assim a todos os sócios e especialmente a sul do Mondego, para que a SPE possa encontrar alguém que ajude na coordenação e na dinamização da Explorística.
- Encontra-se a decorrer o concurso para o *Best Cooperative Project Award*, promovido pelo ISLP, para o qual estamos a pensar preparar uma candidatura da Explorística. O ALEA (primeiro lugar em 2007) e a Estatística Radical (segundo lugar em 2013) já conseguiram distinções neste concurso. Ver concurso em:
 - http://iase-web.org/islp/Competitions.php?p=Best_Cooperative_Project_2015
- Estamos também a tentar dar a conhecer a Explorística nos PALOP.
- Finalmente, faz sentido pensar numa nova versão da Explorística. Com novos módulos e novos materiais, a conceção da **Explorística 2.0** poderá ser motivo de financiamento pela Ciência Viva.

Itinerário:

Em Portugal, desde Fevereiro de 2013, altura da sua inauguração, a Explorística esteve na Escola Secundária de Tomaz Pelayo, Santo Tirso (Fevereiro 2013), Palácio de Cristal - Mostra da Universidade do Porto, (Março 2013), Universidade do Minho, (Abril 2013), Biblioteca de Torres Novas, (Abril/Maio 2013), Pavilhão do Conhecimento, Lisboa (Maio 2013), INE, Lisboa (Maio 2013), Faculdade de Ciências da Universidade de Lisboa, (Junho/Julho, 2013), UTAD, Vila Real (Julho, 2013), Centro Ciência Viva, Lousal (Agosto, 2013), FCT/UNL, Caparica (Setembro, 2013), Instituto Politécnico de Setúbal (Outubro 2013), Congresso SPE, Aveiro e Universidade de Aveiro (Novembro a Janeiro 2013), Escola Secundária Mem Ramires, Santarém (Fevereiro/ Março 2014), Instituto Politécnico de Beja, Beja (Abril-Julho), Universidade da Beira Interior – Covilhã (Outubro 2014), Instituto Politécnico da Guarda (Janeiro-Fevereiro de 2015), (**onde se encontra agora**), Instituto Politécnico de Portalegre, Portalegre (Março, de 2015). Há deslocações previstas para outros locais. Lagos - Centro Ciência Viva (Junho, 2015) a confirmar.

No estrangeiro, a Explorística foi apresentada em conferências internacionais: na Suíça (Neuchatel, IMAODBC, 2013), Hong Kong (ISI, 2013), Estados Unidos (ICOTS, Arizona, 2014).

Um abraço,

Pedro Campos

(coordenador da Explorística)



AGÊNCIA NACIONAL
PARA A CULTURA
CIENTÍFICA E TECNOLÓGICA

• Provas de Agregação - Maria Eduarda Silva

No passado mês de Janeiro, na véspera de tomar posse como Presidente da SPE, a nossa colega Maria Eduarda Silva, com o maior sucesso, terminou as provas de agregação na Universidade do Porto. Obteve o Título de Agregada em Matemática Aplicada na Faculdade de Economia da Universidade. O tema da lição foi: Modelling time series of counts – na INAR(1) approach.

FR

Enigmística de mefqa

r e l a ç ã o

S P E R A E S P E R
P E R A E S P E R A
E R A E S P E R A E

No Boletim SPE outono de 2014 (p.10):

r a z ã o i n

m a r t i n g a l a

o princípio da razão insuficiente

semi-martingala

Dando vida às relações estatísticas: Uma Metodologia Estocástico-Dinâmica aplicada à Ecologia

João Alexandre Cabral, *jcabral@utad.pt*

*Laboratório de Ecologia Aplicada
Centro de Investigação e Tecnologias Agro-Ambientais e Biológicas (CITAB)
Universidade de Trás-os-Montes e Alto Douro (UTAD)*

Introdução

Uma metodologia estocástico-dinâmica (Stochastic-Dynamic Methodology - StDM) tem sido desenvolvida pelo Laboratório de Ecologia Aplicada do CITAB-UTAD com base na premissa de que padrões estatísticos gerais, ao nível das relações causa-efeito dos fenómenos ecológicos, incluindo os de cariz aleatório (ou estocástico), são indícios da dinâmica de processos complexos à escala global (ou holística) dos ecossistemas. O presente artigo descreve, de forma simples e intuitiva, a base teórica e os procedimentos inerentes à StDM, incluindo os fundamentos de modelação e o enquadramento estatístico respetivo. Para efeitos de demonstração, foram selecionados casos de estudo que ilustram a aplicação académica e o uso efetivo de modelos de gestão no âmbito de programas de monitorização da integridade ecológica dos ecossistemas. A validação das simulações produzidas tem sido encorajadora pois tem demonstrado a capacidade da StDM em captar a dinâmica de processos holísticos dos ecossistemas estudados, prevendo com sucesso o padrão de comportamento de componentes chave em cenários ambientais muito complexos e variáveis.

No âmbito dos programas de gestão ambiental, os estudos de impacte ambiental (EIA) e a monitorização ecológica (ME) podem proporcionar informação relevante de suporte ao desenvolvimento sustentável e à conservação da diversidade biológica (Söderman, 2006). De facto, a aplicação dos EIA e ME está na base da implementação das estratégias nacionais para o desenvolvimento sustentável, considerando que os primeiros são usados para prever as consequências ambientais das acções humanas (também designadas antropogénicas) e que os segundos têm como objectivo promover correcções ou ajustes de modo a reduzir os impactes ecológicos que decorrem dessas acções. Por conseguinte, um dos grandes desafios em gestão ambiental é conseguir prever os efeitos das alterações ambientais induzidas pelas actividades antropogénicas sobre a abundância de espécies, grupos funcionais ou comunidades em ecossistemas perturbados (Bailey et al., 2007).

A utilização de modelos ecológicos na problemática ambiental tem assumido uma crescente importância nas últimas três décadas (Jørgensen, 2008). Em paralelo, o desenvolvimento contínuo de *hardware* e *software* permitiu que o uso de modelos se tornasse mais amigável e acessível. Quando desenvolvidos e testados adequadamente, os modelos ecológicos são capazes de simular condições que seriam difíceis ou impossíveis de compreender a partir de outras metodologias. Adicionalmente, a aplicação de modelos ecológicos consegue sintetizar a complexidade dos sistemas e do conhecimento sobre os fenómenos estudados, privilegiando as abordagens holísticas (Jørgensen, 2008). Uma abordagem holística incide sobre fenómenos de macro-escala, ao nível do próprio sistema, com propriedades que só se manifestam quando este é estudado globalmente, ou seja, que representam mais do que a simples soma das propriedades das suas componentes quando analisadas separadamente.

O desenvolvimento de modelos ecológicos requer um conhecimento consistente sobre o funcionamento dos ecossistemas e sobre os problemas ambientais que os afectam. No entanto, são reconhecidas as limitações e dificuldades em captar de forma satisfatória o cariz multifactorial e multi-escala dos processos alvo de estudo. A representação matemática dos processos contem parâmetros ou coeficientes, que podem ser considerados constantes para um ecossistema específico ou parte desse ecossistema (ex: taxas de crescimento, mortalidade, decaimento, etc.). A perspectiva reducionista ou analítica é um procedimento metodológico que, ao contrário da perspectiva holística, focaliza o problema no seu nível inferior da hierarquia de complexidade estudada (ex: a vida nas células é melhor explicada pelo estudo das suas características químicas e as plantas são melhor referenciadas pelo estudo dos seus aspectos florísticos). Em termos reducionistas, poucos são os parâmetros conhecidos exactamente, constituindo esta evidência uma das debilidades em modelação ecológica (Jørgensen, 2008).

Não obstante as limitações referidas, muitos fenómenos dos ecossistemas são de cariz holístico, que englobam aspectos ou propriedades à macro-escala. A principal vocação da metodologia estocástico-dinâmica (Stochastic-Dynamic Methodology - StDM) é reproduzir matematicamente estes fenómenos holísticos (Santos e Cabral, 2004). A StDM é um processo de modelação sequencial que tenta captar tendências holísticas reveladoras do estado ecológico de ecossistemas alterados por actividades antropogénicas. Esta metodologia, porque não depende do cálculo reducionista de parâmetros, tem-se revelado expedita e aplicável em vários tipos de ecossistemas e/ou contextos, tais como rios e ribeiras (e.g., Cabecinha et al., 2007), albufeiras (e.g., Cabecinha et al., 2009), agro-ecossistemas (e.g., Cabral et al., 2007), estuários (e.g., Silva-Santos et al., 2008), florestas (e.g., Silva-Santos et al., 2010), medidas conservacionistas (e.g., Bastos et al., 2012), invasões biológicas (e.g., Santos et al., 2011), impactes infra-estruturais (e.g., Bastos et al., 2013) ou na simulação do impacto de tendências socioeconómicas sobre espécies ou comunidades ameaçadas (e.g., Santos et al., 2007). O objectivo deste artigo será demonstrar, de forma ilustrativa, a aplicabilidade dos princípios da StDM em gestão ambiental.

Metodologia

Os estudos de comunidades e/ou da respectiva integridade ecológica dão usualmente origem a complexas bases de dados. Se o objectivo for encontrar padrões holísticos relevantes e tendências ecológicas a partir destas bases de dados, então é necessário condensar toda a informação de uma forma mais simplificada (Pardal et al., 2004). Um primeiro passo neste sentido poderá ser a determinação do factor ou factores causais que melhor explicam (ou influenciam) as tendências ecológicas estudadas. Se num universo de várias variáveis estudadas, uma das variáveis é claramente dependente das restantes (estas designadas variáveis independentes), então o fenómeno estudado será de cariz multivariado.

A StDM é um processo sequencial de modelação iniciado pela aplicação de modelos lineares generalizados (GLM). Porém, o facto de se considerar a existência de n potenciais variáveis independentes não implica, automaticamente, que todas elas tenham um efeito significativo no comportamento da variável dependente. Por esta razão, na StDM opta-se pelo chamado “best model” GLM (menor valor de AIC e maior valor de R^2 ajustado) para testar as relações causa-efeito entre as variáveis dependentes e independentes, com eliminação das variáveis independentes não significativas, ou pelo método “Multi-Model Inference” (MMI) em que, embora todas as variáveis independentes sejam consideradas, o seu peso relativo é ponderado pela relevância estatística que estas representam (Burnham e Anderson, 2002). Genericamente, as variáveis dependentes correspondem a indicadores ecológicos considerados relevantes, como, por exemplo, a riqueza específica, a diversidade e a abundância relativa de indivíduos, populações ou grupos funcionais. As variáveis independentes representarão os principais factores ambientais com potencial influência nas tendências evidenciadas pelos indicadores ecológicos face a perturbações antropogénicas. Deste procedimento resulta um conjunto de equações matemáticas, do tipo $Y = a + b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_nX_n$, que expressam as relações entre os indicadores seleccionados (Y) e as n variáveis ambientais independentes (X) que os influenciam. Os coeficientes parciais de regressão, $b_1, b_2, b_3, \dots, b_n$, expressam, de forma estatisticamente significativa, as relações holísticas entre as variáveis dependentes e independentes.

No entanto, em virtude do procedimento descrito, produzir resultados ou equações estáticas, um dos requisitos centrais da StDM prende-se com a necessidade das bases de dados cobrirem, de forma representativa, os gradientes temporais e/ou espaciais das alterações estudadas. Os gradientes podem ser, por exemplo, a sequência de alterações na paisagem agrícola de uma região, que abarque desde a ocupação tradicional do solo até às modernas práticas de intensificação da agricultura (Santos e Cabral, 2004; Cabral et al., 2007) ou a variação das características litológicas, topográficas e hidrológicas que ocorrem no seio de uma bacia hidrográfica, desde a nascente até à foz, incluindo condições pouco afectadas pelas actividades humanas e condições perturbadas em termos antropogénicos (Cabecinha et al., 2004, 2007). Deste modo, os factores tempo e espaço passam a estar presentes na base de dados o que confere potencial dinâmico às equações produzidas. Partindo do pressuposto de que o requisito anterior está presente na base de dados, então os coeficientes parciais de regressão assumem, na StDM, o papel de relevantes parâmetros ecológicos holísticos na construção dos modelos dinâmicos respectivos. Estes coeficientes passam a representar a influência global de um conjunto de variáveis ambientais com importância no funcionamento de processos ecológicos complexos, envolvendo, por exemplo, interações entre grupos biológicos e os recursos que os sustentam. Este é o cerne da filosofia StDM, cujos modelos têm sido desenvolvidos com recurso ao *software* STELLA[®] de modelação em Dinâmica de Sistemas (System Dynamics).

A unidade básica de modelação da StDM (Figura 1) consiste numa variável de estado, que representa uma entidade relevante do sistema (ex: um modelo sobre a qualidade da água de um lago terá naturalmente como variáveis de estado a concentração de fitoplâncton e a concentração de nutrientes no sistema), e as influências obtidas pela análise estatística prévia. As influências positivas são mediadas pelas constantes e coeficientes parciais de regressão positivos e as influências negativas pelas constantes e coeficientes parciais de regressão de valor negativo (Figura 1). Como o balanço entre influências positivas e negativas pode originar um efeito cumulativo, ao longo do tempo, na variável de estado (ex: uma contagem de 10 aves no dia 1 num determinado local, seguida de uma contagem de 15 aves no dia 2 no mesmo local, não deve somar 25 aves no dia 2!), a unidade base tem um fluxo de saída adicional, designado ajuste (Figura 1), que descarrega a variável de estado na transição entre os cálculos para períodos consecutivos.

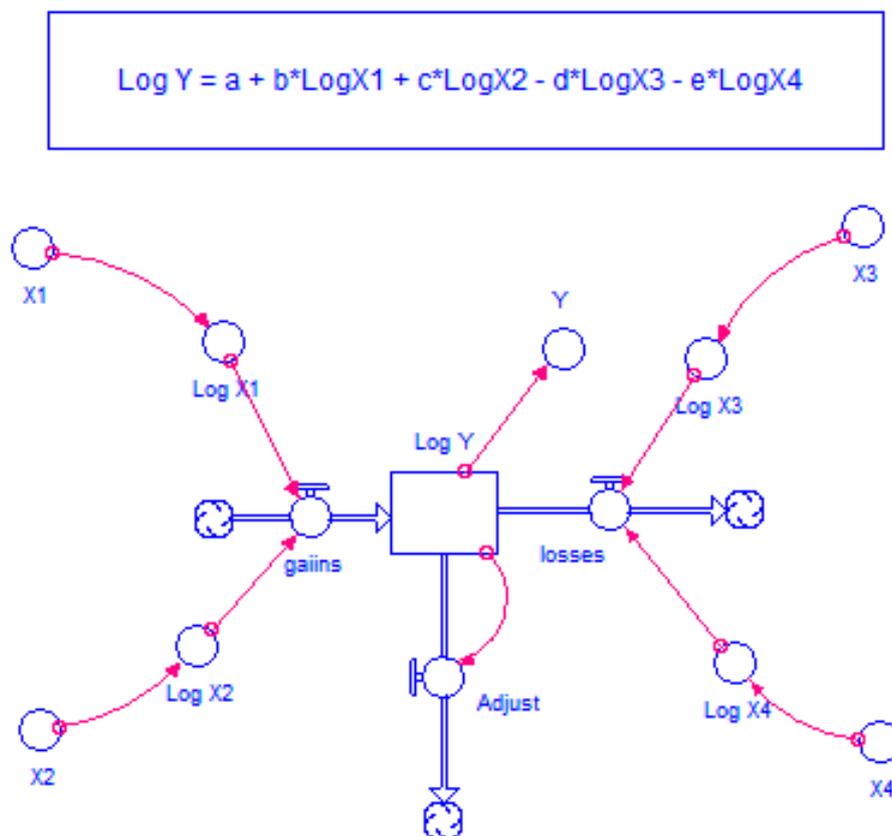


Figura 1

Na figura 1, apresenta-se a unidade base de modelação na StDM. Em linguagem conceptual de Dinâmica de Sistemas, o rectângulo representa a variável de estado, os fluxos que influenciam a variável de estado são representados por setas grossas, os círculos pequenos representam, variáveis ambientais, parâmetros ou constantes e as relações de auto-regulação da variável de estado, entre variáveis de estado ou com outras variáveis são representadas por setas finas.

Respeitando estes princípios básicos, a unidade base pode assumir diferentes arranjos (Figura 2) em função dos objectivos de cada estudo. Por exemplo, se o objectivo é contribuir para a avaliação da qualidade da água em rios, com ênfase nas interações entre macroinvertebrados aquáticos (como é o caso dos Trichoptera, grupo sensível à degradação da qualidade da água) e as condições físico-químicas prevaletentes na coluna de água (Cabecinha et al., 2004), então a unidade base pode assumir o formato primário (Figura 2a). No entanto, se o objectivo passar a ser a simulação da dinâmica dos produtores primários (ex: macro-algas) num ambiente estuarino sujeito a excesso de poluição orgânica, que gera grandes quantidades de nutrientes inorgânicos disponíveis (Silva-Santos et al., 2008), então o formato primário já não será suficiente. Neste caso, a produção primária depende das condições meteorológicas (ex: precipitação) para o seu desenvolvimento neste tipo de ambientes e, por conseguinte, serão necessários dois níveis complementares de influências (com as respectivas equações matemáticas) que resultem, por exemplo, da incidência alternada de meses secos e chuvosos ao longo do ano (Figura 2b). Por outro lado, se o objectivo consistir em simular o impacto dos fogos florestais na estrutura das comunidades avifaunísticas (Silva-Santos et al., 2010), dado que a abundância e riqueza específica de muitas aves (como é o caso das toutinegras ou Sylviidae) depende largamente dos seus ciclos anuais de reprodução e migração, então será conveniente considerar três níveis de influências, que capturem as condições dos períodos de hibernação, reprodução e migração (Figura 2c).

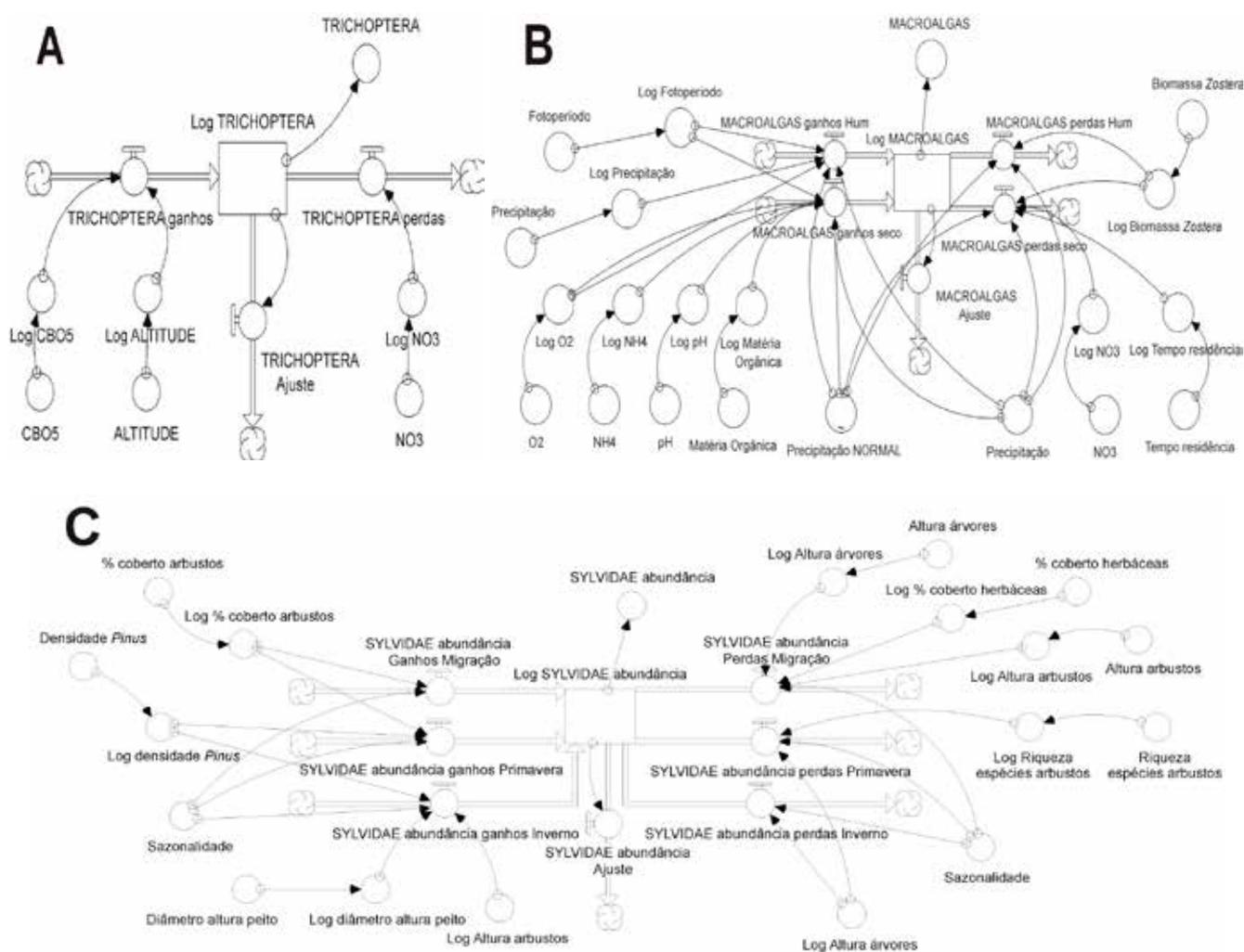


Figura 2

Na Figura 2 esquematiza-se a unidade base de modelação na StDM “desenhadas” de acordo com os objectivos de diferentes estudos: (A) avaliação da qualidade da água com base nas relações entre os macroinvertebrados aquáticos (ex: riqueza específica de Trichoptera) e as condições físico-químicas da coluna de água; (B) avaliação das tendências da biomassa de macroalgas em contextos de excesso de nutrientes em estuários; (C) avaliação do impacto dos fogos florestais sobre as comunidades avifaunísticas (ex: toutinegras ou Sylviidae).

A complexidade dos modelos StDM aumenta quando as variáveis de estado são seleccionadas como representativas de uma “cadeia alimentar” ou “cadeia trófica” (Figura 3). Neste caso, cada componente biológica interage com as restantes componentes biológicas (ex. interacções de competição ou predação) e com as características não biológicas do respectivo habitat. Estas interacções podem ser ilustradas com base num modelo StDM, construído para simular as tendências de três componentes representativas de uma cadeia alimentar estuarina afectada pela presença de excesso de nutrientes (fósforo e azoto), no âmbito de um estudo realizado no estuário do Mondego (Silva-Santos et al., 2006). As componentes consideradas são os produtores primários (dominados por macroalgas e macrófitas), os macroinvertebrados dos sedimentos (também designados bentónicos) e os predadores destes, as aves ribeirinhas (também designadas por limícolas) (Figura 3). Estas componentes estão todas enquadradas pelas condições físico-químicas com influência na dinâmica da cadeia alimentar em causa (Figura 3).

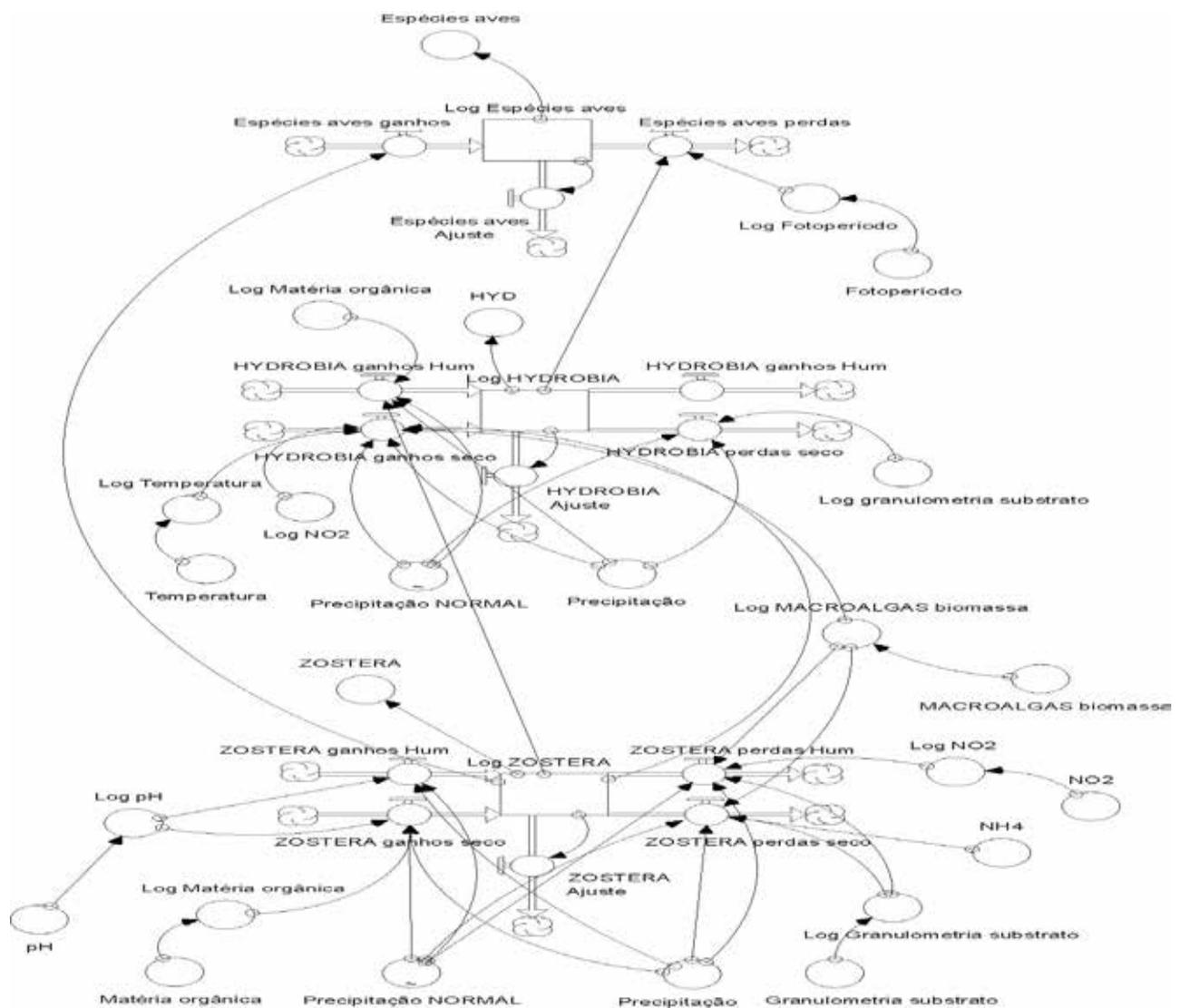


Figura 3 – Ilustração de interações encadeadas num submodelo StDM construído para prever alterações na cadeia alimentar desencadeadas por excesso de nutrientes no estuário do Mondego. Zosteras refere-se a um grupo de plantas aquáticas ou macrófitas. Hydrobia é um o nome genérico de um gastrópode estuarino, principal presa das aves limícolas.

Simulações e Conclusões

A performance da StDM tem sido analisada em dois tipos de situações distintas, em cenários determinísticos, em que as condições ambientais base são conhecidas e pré-determinadas no período de simulação, e em cenários estocásticos, em que as condições ambientais são geradas aleatoriamente entre limites realistas durante a simulação (Santos et al., 2013). A título de exemplo, um dos cenários determinísticos usados foi aquele que recriou a degradação da qualidade da água por poluição orgânica, monitorizada ao longo de uma década (1985-1995) no rio Corgo, inserido na bacia hidrográfica do Douro (Cortes, 1992; Sampaio, 1995). Este cenário previamente determinado foi utilizado para projectar as reacções de certos grupos de macroinvertebrados aquáticos, característicos dos rios de montanha, cuja sensibilidade à poluição orgânica está muito bem estudada (Cabecinha et al., 2004). Desta forma foi possível confrontar a reacção de dois grupos de macroinvertebrados distintos, os EPT (grupo que inclui os Ephemeroptera, Plecoptera e Trichoptera, todos muito sensíveis à poluição orgânica) e os Chironomidae, vulgarmente conhecidos por larvas de mosquitos (muito tolerantes à poluição orgânica), em dois contextos, o contexto de referência (que se registava em 1985, com boas condições ecológicas) e o contexto de degradação que ocorreu 10 anos mais tarde em 1995 (Figura 4). As simulações produzidas revelam que o modelo StDM conseguiu discriminar, de forma credível, a reacção negativa dos EPT quando a qualidade da água se degradou da reacção favorável dos Chironomidae, grupo tolerante, no mesmo cenário de poluição orgânica (Figura 4).

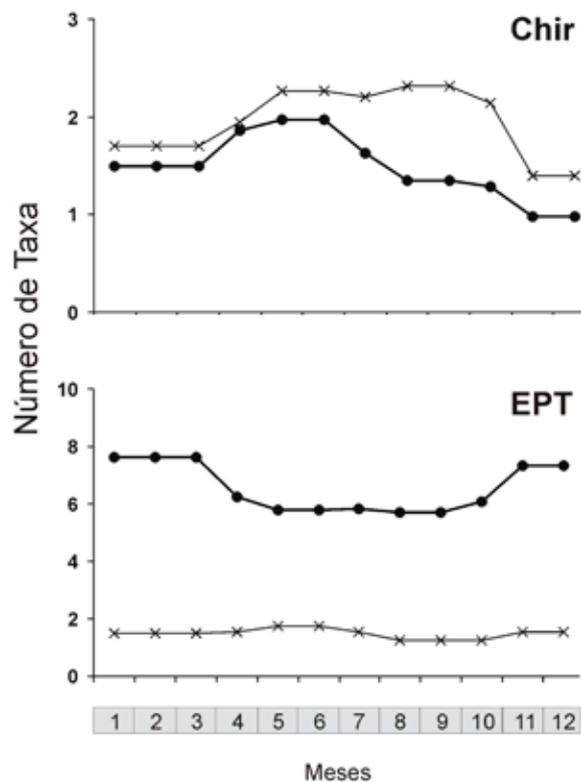


Figura 4 - Simulações para as tendências sazonais da riqueza específica do grupo Chironomidae (Chir) e EPT em dois contextos distintos: em condições de referência, registadas em 1985 (●), e em condições perturbadas, registadas 10 anos mais tarde em 1995 (x), ambas no rio Corgo (Norte de Portugal).

Em modelação ecológica em geral, e na StDM em particular, o objectivo passa frequentemente por antecipar situações que ainda não tiveram lugar, ou seja, simular cenários futuros não conhecidos previamente. Neste caso, as condições ambientais desconhecidas (por exemplo as condições meteorológicas) deverão ser projectadas de forma estocástica ou aleatória na sua gama realista de valores sazonais. Neste contexto, em termos de demonstração académica, são apresentados os resultados que tiveram por base a possível sucessão temporal das condições ambientais, antes e depois

das acções de gestão da qualidade da água, implementadas a partir de 1998 no estuário do Mondego (Silva-Santos et al., 2008). Estas acções visaram corrigir o enriquecimento anómalo de nutrientes (de origem agrícola) deste estuário (fenómeno conhecido por eutrofização) e promover a recuperação de uma importante planta/macrófita aquática, a *Zostera noltii*, em declínio por ser muito afectada pela eutrofização. As simulações, de base estocástica, em cenários em que as acções de gestão tiveram lugar e em cenários em que a situação de eutrofização se agravou continuamente, revelaram tendências completamente distintas para a biomassa de *Zostera noltii* (Figura 5). Estas tendências, que resultaram da média de 10 simulações estocásticas, reproduziram de forma muito consistente as tendências monitorizadas na realidade por Lillebø et al. (2005) no estuário do Mondego.

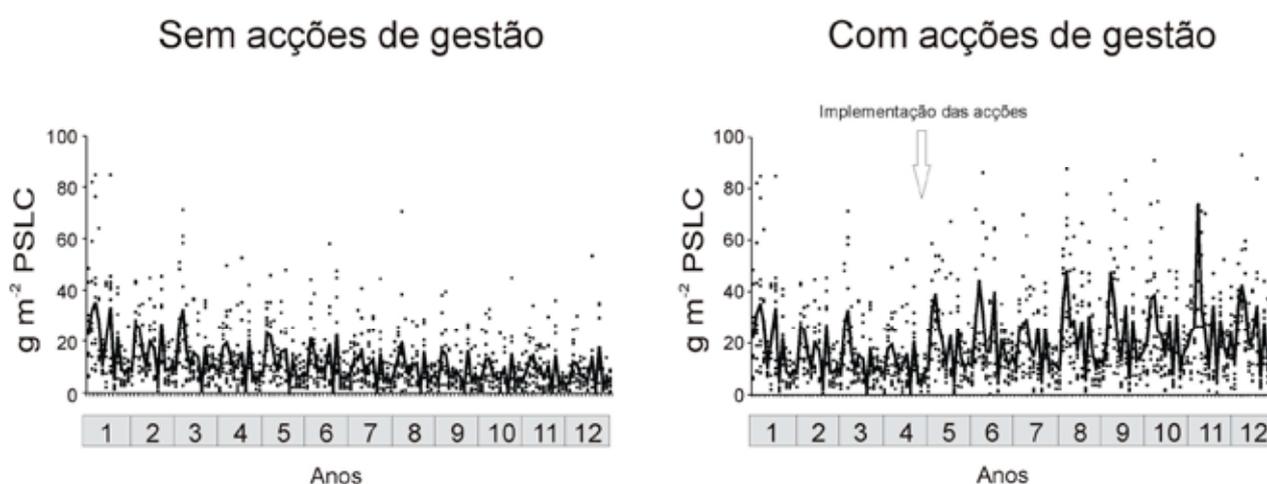


Figura 5 – Simulações das tendências da biomassa da macrófita *Zostera noltii* (expressa Peso Seco Livre de Cinzas, PSLC) com e sem a implementação de acções de gestão ambiental no estuário do Mondego (ao longo de um período de 12 anos). A linha representa a média de 10 simulações estocásticas mensais.

Os resultados obtidos são encorajadores pois parecem indiciar a capacidade da StDM em captar a dinâmica holística de fenómenos estudados em condições e cenários complexos. Esta metodologia tem os requisitos para poder ser integrada em estratégias de gestão ambiental, desde que sustentadas por bases de dados rigorosas, em virtude de se basear em relações causa-efeito de macro-escala, determinadas estatisticamente a partir de análises padronizadas, e por isso de concepção expedita e viável economicamente. As simulações produzidas em diferentes ecossistemas e problemas ambientais, mostraram a relevância de determinadas variáveis de estado como indicadores ecológicos, nomeadamente em contextos em que situações pouco perturbadas se degradaram posteriormente. Muitas das tendências simuladas foram corroboradas/validadas por observações reais em estudos realizados sobre as consequências das alterações ambientais induzidas por actividades humanas. Globalmente, os principais resultados demonstraram, tal como em qualquer processo científico complexo, que é válido, interessante e instrutivo apostar em metodologias holísticas de previsão, focadas nas interacções entre componentes chave de ecossistemas em alteração, como é o caso da StDM.

Os últimos desenvolvimentos da StDM têm privilegiado a criação de interfaces com os Sistemas de Informação Geográfica (SIG) que permitem a projecção espacial das tendências temporais produzidas pelas simulações dinâmicas (Bastos et al., 2012). Por conseguinte, o aperfeiçoamento desta proposta metodológica tem impulsionado o desenvolvimento de técnicas mais globais no âmbito da modelação dinâmico-espacial (Santos et al., 2013), tornando a StDM mais instrutiva, intuitiva e credível para os decisores e gestores ambientais.

Referências Bibliográficas

- Bailey, R. C., Reynoldson, T. B., Yates, A.G., Bailey, J. and Linke, S. (2007). Integrating stream bioassessment and landscape ecology as a tool for land use planning. *Freshwater Biology*, 52, 908–917.
- Bastos, R., Santos, M., Ramos, J. A., Vicente, J., Guerra, C., Alonso, J., Honrado, J., Ceia, R. S., Timóteo, S., Cabral, J. A. (2012). Testing a novel spatially-explicit dynamic modelling approach in the scope of the laurel forest management for the endangered Azores bullfinch (*Pyrrhula murina*) conservation. *Biological Conservation*, 147, 243-254.
- Bastos, R., Santos, M., Cabral, J. A. (2013). A new stochastic dynamic tool to improve the accuracy of mortality estimates for bats killed at wind farms. *Ecological Indicators*, 34, 428-440.
- Burnham, K. P., Anderson, D. R. (2002) Model Selection and Multi Model Inference: A Practical Information-Theoretic Approach, second ed. Springer.
- Cabecinha, E., Cortes, R. and Cabral, J. A. (2004). Performance of a stochastic-dynamic modelling methodology for running waters ecological assessment. *Ecological Modelling*, 175 (3), 303-317.
- Cabecinha E., Silva-Santos P., Cortes R. and Cabral J. A. (2007). Applying a stochastic-dynamic methodology (StDM) to facilitate ecological monitoring of running waters, using selected trophic and taxonomic metrics as state variables. *Ecological Modelling*, 207, 109-127.
- Cabecinha E, Lourenço M, Moura JP, Pardal MA, Cabral JA. (2009). A multiscale approach to modelling spatial and dynamic ecological patterns for reservoir's water quality management. *Ecological Modelling*, 220, 2559-2569
- Cabral J. A., Rocha A., Santos M. and Crespi A. L. (2007) A stochastic dynamic methodology (StDM) to facilitate handling simple passerine indicators in the scope of the agri-environmental measures problematics. *Ecological Indicators*, 7, 34-47.
- Cortes, R. M. V. (1992). Seasonal pattern of benthic communities along the longitudinal axis of river systems and the influence of abiotic factors on the spatial structure of those communities. *Archiv fur Hydrobiologie*, 126, 85-103.
- Jørgensen, S. E. (2008). Overview of the model types available for development of ecological models. *Ecological Modelling*, 215, 3-9.
- Lillebø, A. I., Neto, J. M., Martins, I., Verdelhos, T., Leston, S., Cardoso, P. G., Ferreira, S. M., Marques, J. C. and Pardal, M. A. (2005). Management of shallow temperate estuary to control eutrophication: The effect of hydrodynamics on the system's nutrient loading. *Estuarine, Coastal and Shelf Science*, 65, 697-707.
- Pardal, M. A., Cardoso, P. G., Sousa, J. P., Marques, J. C. and Raffaelli, D. (2004). Assessing environmental quality: a novel approach. *Marine Ecology Progress Series*, 267, 1-8.
- Sampaio, A. C. R. (1995). *Avaliação do impacto dum mini-hídrica no funcionamento dum ecossistema lótico: caso particular do Terragido do rio Corgo*. MSc. Thesis. Universidade de Trás-os-Montes e Alto Douro, Vila Real.
- Santos, M. and Cabral J.A. (2004). Development of a stochastic dynamic model for ecological indicators' prediction in changed Mediterranean agroecosystems of north-eastern Portugal. *Ecological Indicators*, 3, 285-303.
- Santos, M., Freitas, R., Crespi, A. L., Hughes, S. J., Cabral, J. A. (2011) Predicting trends of invasive plants richness using local socio-economic data: An application in North Portugal. *Environmental Research*, 111, 960-966.
- Santos, M., Vaz, C., Travassos, P. and Cabral, J. A. (2007). Simulating the impact of socio-economic trends on threatened Iberian wolf populations (*Canis lupus signatus*) in North-eastern Portugal. *Ecological Indicators*, 7, 649-664.
- Santos, M., Bastos, R., Cabral, J. A. (2013). Converting conventional ecological datasets in dynamic and dynamic spatially-explicit simulations: current advances and future applications of the Stochastic Dynamic Methodology (StDM). *Ecological Modelling*, 258, 91-100.
- Silva-Santos, P., Pardal, M. A., Lopes, R. J., Múrias, T. and Cabral J. A. (2006). A Stochastic Dynamic Methodology (StDM) to the modelling of trophic interactions, with a focus on estuarine eutrophication scenarios. *Ecological Indicators*, 6, 394-408.
- Silva-Santos, P., Pardal, M. A., Lopes, R. J., Múrias, T. and Cabral J. A. (2008). Testing the Stochastic Dynamic Methodology (StDM) as a management tool in a shallow temperate estuary of south Europe (Mondego, Portugal). *Ecological Modelling*, 210, 377-402.
- Silva-Santos, P., Valentim, H., Luís, A., Queirós, L., Travassos, P. and Cabral J. A. (2010). A Stochastic Dynamic Methodology (StDM) to simulate the effects of fire on vegetation and bird communities in *Pinus pinaster* stands. *Ecological Indicators*, 10, 206-211.
- Söderman, T. (2006). Treatment of biodiversity issues in impact assessment of electricity power transmission lines: A Finnish case review. *Environmental Impact Assessment Review*, 26, 319-338.

Séries Temporais no Desporto

Ana Diniz, *adiniz@fmh.ulisboa.pt*
Paula Marta Bruno, *pbruno@fmh.ulisboa.pt*

*Faculdade de Motricidade Humana
Universidade de Lisboa*

Introdução

A área das Séries Temporais aliada ao domínio do Desporto tem vindo a desenvolver-se gradualmente, dado o conjunto considerável de dados disponíveis e a necessidade crescente de encontrar respostas precisas para problemas reais. Alguns exemplos interessantes dizem respeito a séries de distâncias ou de velocidades em provas desportivas, séries de frequências cardíacas ou de sinais eletromiográficos em exercícios físicos realizados por atletas, etc. Os métodos clássicos de análise de dados temporais relativos ao desporto são baseados em medidas descritivas, como a média e o desvio padrão, e ignoram a estrutura de correlação ao longo do tempo. Em contraste, os métodos das séries temporais, no domínio tempo e no domínio frequência, consideram o comportamento dinâmico ao longo do tempo e permitem fazer modelação e inferência. A literatura mais recente sobre este tema já vai contendo alguns estudos científicos com uma componente de modelação. No entanto, esses trabalhos revelam ainda algumas limitações metodológicas.

Aplicações

A regata é uma prova náutica de velocidade efetuada por vários barcos num dado percurso. De entre as diversas variantes de regata encontra-se a regata “match racing” na qual só se defrontam dois barcos. Numa regata “match racing”, o período de pré-início é crucial, já que nessa fase cada barco tenta continuamente ganhar posição de vantagem sobre o outro para obter um “ponto ótimo” na linha de partida. Alguns estudos revelam que o facto de um barco passar a linha de partida em primeiro lugar conduz frequentemente a um resultado vitorioso.

Considerando, no período de pré-início, as séries temporais referentes às distâncias de cada barco à posição ótima de partida e utilizando métodos estatísticos, no domínio tempo e no domínio frequência (e.g., correlações e periodogramas), é possível recolher informação relevante sobre a dinâmica da pré-regata.

Em Araújo *et al.* (2015), quinze velejadores com idade média de 12.1 anos (e desvio padrão de 1.6 anos), organizados em pares, realizaram regatas “match racing” com períodos de pré-início de 2 minutos. A distância (em metros) de cada barco ao “ponto ótimo” de partida foi medida a cada 0.2 segundos durante um período máximo de 120 segundos. A figura em baixo exhibe as séries de distâncias de um par de barcos numa regata, assim como alguns resultados estimados (Figura 1).

Observa-se que as duas séries estão correlacionadas, com valores predominantemente positivos e alguns negativos; no total tem-se valores $|r| \geq 0.7$, em 88.4% do tempo. Por outro lado, verifica-se que a função de autocorrelação reflete um padrão cíclico, com valores significativos; o periodograma tem um pico significativo na frequência 0.00518, equivalente ao período $1/0.00518 = 193$. Isto significa que os dois barcos têm trajetórias coordenadas na fase de pré-início, resultantes de interação contínua co-adaptativa. Em paralelo, um dos barcos tem uma trajetória cíclica de aproximação-afastamento ao

ponto ótimo de partida, repetindo-se a cada 193 unidades, i.e., 38.6 segundos. Nas restantes regatas observadas, cada regata revelou um padrão coordenado e cíclico particular.

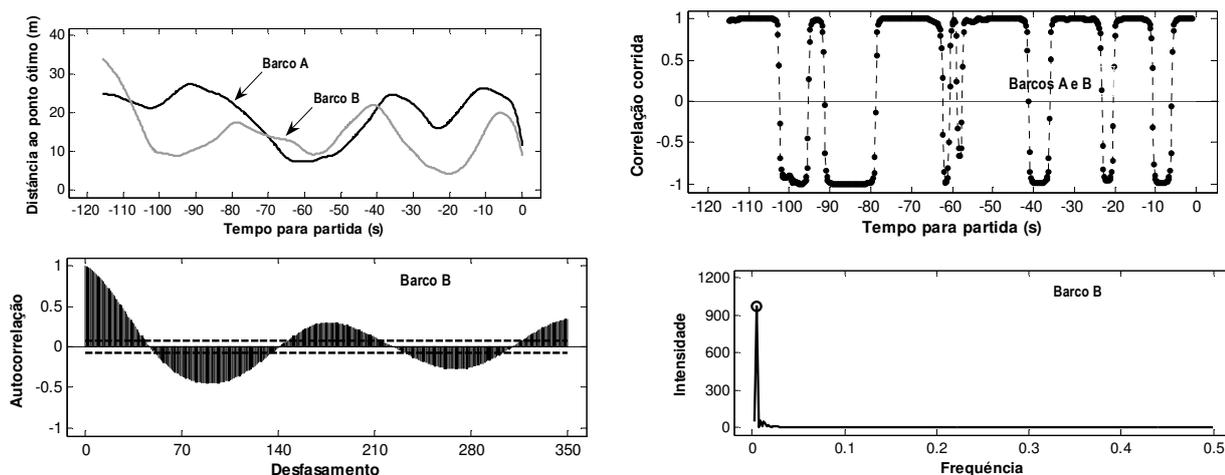


Figura 1. (a) Séries de distâncias ao ponto ótimo de partida de um par de barcos numa regata; (b) Correlações corridas; (c) Função de autocorrelação amostral; (d) Periodograma normalizado.

Bruno *et al.* (2011) utilizaram análise fatorial dinâmica no estudo de séries temporais na área da fisiologia pediátrica. Rapazes de 11 e 12 anos, com pelo menos dois anos de prática em futebol ou natação, foram avaliados no teste anaeróbio de Wingate. Este teste tem uma duração de 30 s e tem como objetivo pedalar o maior número possível de vezes contra uma resistência fixa (Figura 2a). Cada série temporal é constituída pela potência (W/kg) de um participante, registada segundo a segundo (Figura 3a).

A aplicação da análise fatorial dinâmica permitiu discriminar dois padrões (Figura 3b): os nadadores tendem a atingir o pico de potência mais cedo (Figura 3c) que os futebolistas (Figura 3d).

O teste de Wingate pode também ser adaptado para membros superiores (Figura 2b). Neste contexto, uma análise semelhante foi realizada para braços, a qual revelou que o padrão que a criança realiza é semelhante independentemente do membro.

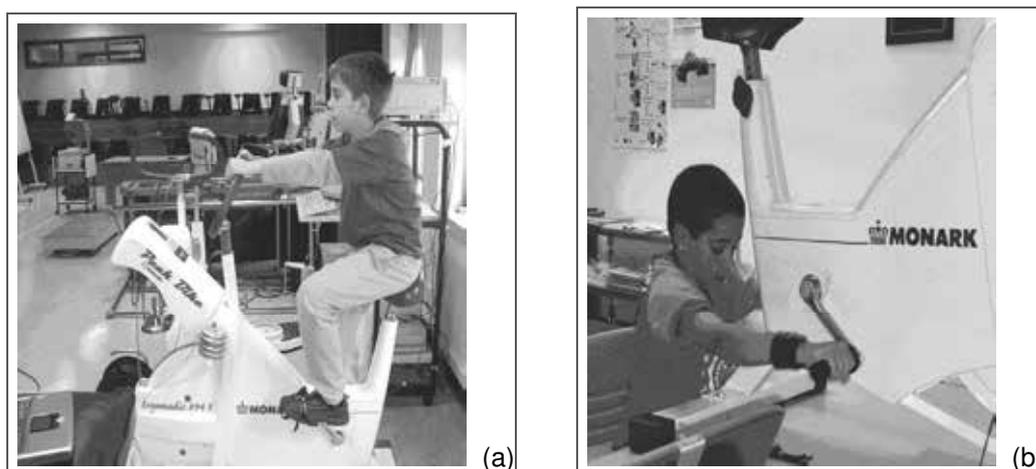


Figura 2. Teste de Wingate para (a) membros inferiores; (b) membros superiores.

Em Lames (2006), analisou-se algumas modalidades desportivas como o ténis e o andebol no contexto da interação dinâmica interpessoal e intergrupar, respetivamente. No caso do ténis, trinta sequências de pancadas de bola foram examinadas e as séries temporais de deslocamentos laterais de cada par de jogadores foram registadas. A partir das séries e recorrendo à transformada de Hilbert, a fase relativa entre as trajetórias de cada par de jogadores foi determinada. O comportamento da fase relativa ao longo do tempo revelou detalhes importantes da tática no ténis.

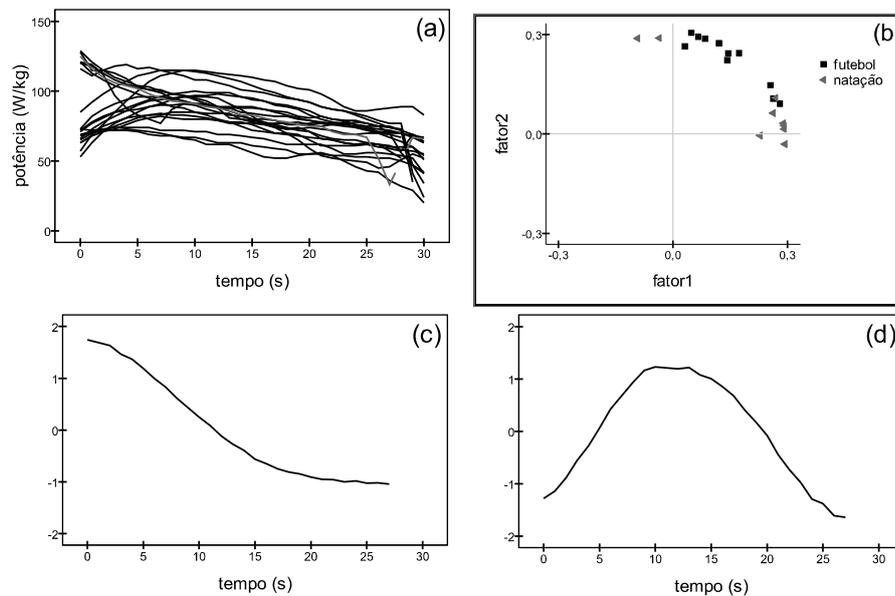


Figura 3. (a) Séries de potência de cada um dos participantes; (b) *Factor loadings* associados aos padrões comuns; (c) Padrão estimado com pico de potência no início; (d) Padrão estimado com pico de potência atrasado.

Outro exemplo, da utilização de séries temporais no desporto, é o estudo relativo à pancada de esquerda no ténis proposto por Riek *et al.* (1999), no qual foram consideradas séries de eletromiografia. A utilização de uma equação diferencial de primeira ordem permitiu-lhes relacionar o padrão de ativação muscular apresentado por jovens tenistas com as respetivas lesões sofridas no cotovelo.

Referências

- Araújo, D., Davids, K., Diniz, A., Rocha, L., Santos, J.C., Dias, G., Fernandes, O. (2015). Ecological dynamics of continuous and categorical decision-making: The regatta start in sailing. *European Journal of Sport Science*, 15 (3), 195-202.
- Bruno, P.M., Pereira, F.D., Fernandes, R., Mendonça, G.V. (2011). Analysis of Power Output Time Series in Response to Supramaximal Exercise: An Approach Through Dynamic Factor Analysis. *Pediatric Exercise Science*, 23, 3-16.
- Lames, M. (2006). Modelling the interaction in game sports – relative phase and moving correlations. *Journal of Sports Science and Medicine*, 5, 556-560.
- Riek, S., Chapman, A.E. Chapman, Milner, T. (1999). A simulation of muscle force and internal kinematics of extensor carpi radialis brevis during backhand tennis stroke: implications for injury. *Clinical Biomechanics*, 14, 477-483.



Investigação em Ciências do Desporto: dos testes de hipótese nula à necessidade de interpretações com significância prática e/ou clínica

Rui Marcelino, rmarcelino@utad.pt
Jaime Sampaio, ajaime@utad.pt

*CreativeLab, Centro de Investigação em Desporto, Saúde e Desenvolvimento Humano
Universidade de Trás-os-Montes e Alto Douro*

It is easy to lie with statistics. It is hard to tell the truth without it.

Andrejs Dunkels (1939–1998)

Matemático Sueco, professor de Matemática e escritor.

Introdução

Em ciência, o conhecimento é adquirido com recurso ao método científico, a partir do qual se testam e confirmam repetidamente as evidências através da observação, experimentação (Popper, 1972) e inferência estatística. A confirmação das teorias à luz dos dados e o progresso no conhecimento dependem diretamente da forma como os processos inferenciais são entendidos e aplicados. O paradigma dominante, apesar de não exclusivo, da inferência estatística na investigação em Desporto é o de Neyman-Pearson¹ – através da aceitação ou não-aceitação da hipótese nula com base no valor de p .

Por convenção, os novos conhecimentos são publicados sob a forma de artigos em revistas científicas, nos quais se apresentam também todas as evidências que os suportam, tentando convencer os leitores de que a matéria tratada é verdadeira. Em 2005, o estatístico John Ioannidis publicou um artigo com grande impacto na comunidade científica (citado mais de 1400 vezes e visto mais de 1 milhão de vezes), no qual argumenta que a maioria dos resultados científicos publicados são falsos, apontando como principais problemas: i) a investigação publicada ser uma seleção enviesada da totalidade de investigação produzida; ii) as análises de dados e os resultados apresentados serem também seletivos e enviesados; iii) em muitas áreas de investigação, os estudos raramente serem replicados, pelo que persistem as falsas conclusões. De acordo com este autor, as conclusões apresentadas nas investigações científicas são fundamentadas quase exclusivamente nos resultados estatisticamente significativos (obtidos através de decisões baseadas no valor de p).

Perspetiva histórica das análises quantitativas em Ciências do Desporto

A evolução da ciência aplicada ao Desporto sempre esteve associada à evolução da tecnologia. Os efeitos mais visíveis deste impacto da tecnologia estão relacionados com o desenvolvimento de novos instrumentos de medição. Se nos primórdios das análises quantitativas em Desporto os processos de

¹ Apesar dos conceitos de ‘Hipótese Nula’ (representado por H_0), de ‘significância’, de ‘graus de liberdade’, assim como de distinção entre população e amostra serem da autoria do Sir Ronald Fisher (1890-1962), tal como a sugestão do valor arbitrário de $p < 0.05$ para tomar decisões sobre a H_0 (R. Fisher, 1925), foram os matemáticos Jerzy Neyman e Egon Pearson que sugeriram pela primeira vez os conceitos de inferência estatística, assim como demonstraram os procedimentos específicos e a lógica de interpretação dos testes de hipótese nula, que perduram até aos dias de hoje (Neyman, 1933). Foram igualmente estes últimos autores que introduziram a ‘Hipótese Alternativa’ (representada por H_1), ‘statistical power’, falsos positivos e falsos negativos.

recolha de dados eram muito morosos - consistindo maioritariamente em notações manuais com recurso a análise de vídeo (Hughes & Franks, 1997; Marcelino, Sampaio, & Mesquita, 2011) -, atualmente, o volume de dados recolhidos em eventos desportivos é cada vez maior, sendo o tempo necessário para os obter cada vez menor. Dados posicionais de jogadores, recolhidos a frequências até 15 Hz com recurso a dispositivos de posicionamento global (GPS), são frequentes em equipas desportivas e em grupos de investigação (Aughey, 2011; Cummins, Orr, O'Connor, & West, 2013). Simultaneamente, o desenvolvimento de sistemas comerciais de *tracking* automático de jogadores, com a possibilidade de quantificar diversos parâmetros técnicos e físicos, tem permitido o acesso em tempo real a uma quantidade assinalável de dados (Castellano, Alvarez-Pastor, & Bradley, 2014; Dellaserra, Gao, & Ransdell, 2014). Correntemente, é prática comum as organizações de competições desportivas, como sejam os Jogos Olímpicos ou os Campeonatos do Mundo, disponibilizarem nas suas páginas da internet informações quantitativas detalhadas de todos os eventos ocorridos no decurso das competições.

O fácil e rápido acesso a estes dados, em paralelo com o surgimento de softwares poderosos para o processamento e a análise dos mesmos, tem seguramente contribuído para um aumento importante da produção científica de análises quantitativas em Desporto (Marcelino, Sampaio, et al., 2011).

Concordamos com David McConless em relação ao papel dos dados na sociedade contemporânea. Na TED talk intitulada “On the beauty of data visualization”, afirma que os dados não são apenas o novo petróleo mas também o novo solo (*data is the new oil e data is the new soil*). Um solo onde é possível que germinem informações passíveis de trazer novos conhecimentos. Quer a adequação das técnicas estatísticas aplicadas, quer o estilo escolhido para a visualização dos dados, em muito contribuirão para a disseminação desses conhecimentos.

“As estatísticas são apenas ferramentas, e, como a maioria das ferramentas, são moralmente neutras; é a forma como são utilizadas que determinará se estas são boas ou más, benéficas ou prejudiciais.”

(Fisher & Lerman, 2014. p243)

Em função da crescente capacidade computacional dos equipamentos tecnológicos, os estudos científicos tornaram-se mais complexos e os tamanhos das amostras aumentaram marcadamente. Aproximadamente há 20 anos atrás, os métodos estatísticos comuns na literatura científica cingiam-se a análises descritivas, *t*-test (Student's) de medidas independentes e teste do qui-quadrado. Contudo, o panorama contemporâneo dos métodos estatísticos alterou-se radicalmente. Atualmente um consumidor de investigação desportiva necessita de estar familiarizado com um amplo leque de métodos estatísticos (ver exemplos na tabela 1). As análises quantitativas envolvem frequentemente técnicas de regressão (linear, múltipla, ordinal, binomial, logística binária, multifatorial,...), análises da função discriminante, análises de componentes principais e análises de clusters, entre outras. São também cada vez mais frequentes técnicas de processamento de sinais não lineares, como sejam as análises da entropia aproximada (*ApEn*) – para identificar, por exemplo, regularidade em padrões de movimento de jogadores – ou o cálculo de valores de fase relativa – para identificar, por exemplo, coordenação de movimentos inter-jogador (Sampaio & Macas, 2012). Esta tendência suporta-se na necessidade, veiculada pela generalidade das investigações mais recentes, de estudar os comportamentos desportivos enquanto fenómenos não lineares. Assim, numa assunção mais teórica, assume-se que os sistemas complexos, como os que se verificam no Desporto, obedecem ao *caos* e, por isso, embora apresentem comportamentos regulares e previsíveis, podem sofrer súbitas mudanças estocásticas em resposta ao que parecem ser pequenas modificações (Marcelino, Sampaio, et al., 2011).

Tabela 1: Exemplos de métodos estatísticos não centrados no paradigma dos testes de hipótese nula, e respetivos objetivos, no âmbito da investigação recente em Ciências do Desporto.

Estudo	Métodos Estatísticos	Objetivos
(Goncalves, Figueira, Macas, & Sampaio, 2014)	Análise da entropia aproximada e fase relativa	Identificar diferenças em indicadores fisiológicos e padrões de movimento entre jogadores de futebol em função da posição de jogo (defesas, médios, avançados).
(Sampaio, Lago, Goncalves, Macas, & Leite, 2014)	Análise da função discriminante	Comparar variáveis posicionais e frequência cardíaca em função do ritmo (lento, normal e rápido), resultado (a ganhar, a perder) e equilíbrio (em superioridade, em inferioridade) dos jogos de futebol.
(Sankaran, 2014)	Análise de <i>Clusters</i>	Investigar a relação entre as performance individuais de jogadores de críquete e o valor de mercado dos mesmos.
(Moura, Martins, & Cunha, 2014)	Análise de componentes principais	Identificar as relações entre as estatísticas de jogo em futebol e o resultado final no campeonato do mundo de 2006.
(Fewell, Armbruster, Ingraham, Petersen, & Waters, 2012)	Análise de redes (<i>Network Analysis</i>)	Caraterizar jogos de basquetebol da NBA - determinando diferentes estratégias ofensivas - numa perspetiva de Networks, considerando os jogadores como os <i>nodes</i> e os movimentos da bola como os <i>links/edges</i> .
(Neiman & Loewenstein, 2014)	Aprendizagem operacional, com permutações de Monte Carlo	Utilizar informação sobre a localização espacial dos locais de lançamento em basquetebol para estudar como é que os jogadores adaptam as estratégias de ataque em tempo real, de acordo com os eventos recentes do jogo.
(Pierson, Addona, & Yates, 2014)	Construção e calibração de Modelos Matemáticos do comportamento	Modelar o efeito da idade relativa no sucesso desportivo no hóquei-no-gelo canadiano.
(Medeiros, Mesquita, Marcelino, & Palao, 2014)	Diferenças de médias estandardizadas com Intervalos de confiança	Analisar os efeitos da técnica de execução, da idade e do papel funcional (defesa, blocador) na eficácia das ações de jogo em voleibol de praia.
(Marcelino, Sampaio, & Mesquita, 2012)	Entropia aproximada de Duplas médias móveis	Examinar a evolução da performance do ataque e do serviço em voleibol, nos inícios e finais dos <i>sets</i> , em função da qualidade de oposição.
(Marcelino, Mesquita, & Sampaio, 2011)	Regressão Logística Multinomial	Analisar os efeitos da qualidade da oposição e do <i>match status</i> nos indicadores técnico-táticos em Voleibol.

A análise de redes (*Network Analysis*), tanto a análise de redes sociais - SNA, ex: (Fewell et al., 2012) - como a análise de redes neuronais - *Dynamically Controlled Network*, ex: (Perl, 2004) -, têm-se apresentado como uma metodologia capaz de descodificar interessantes padrões comportamentais até então desconhecidos. Indicadores estatísticos, como medidas de centralidade (*betweenness, closeness, in-degree, out-degree*), distâncias do caminho, coeficientes de clusters e entropias (tanto dos ‘nodes’ individualmente, com recurso a entropias das cadeias de Markov, como das redes, com recurso às entropias de Shannons - para detalhes técnicos consultar (Fewell et al., 2012)), surgem com cada vez mais frequência nas investigações que procuram entender os fenómenos desportivos enquanto elementos constituintes de sistemas dinâmicos e complexos.

Para além das vantagens já enumeradas, as Análises de Networks conferem um grande impacto visual na forma como são apresentados os resultados. São bem conhecidas as potencialidade de técnicas

recentes, como as SNA, para a visualização de dados na ciência (Banks, 2014), particularmente na análise de uma grande quantidade de dados (Skiba, 2014).

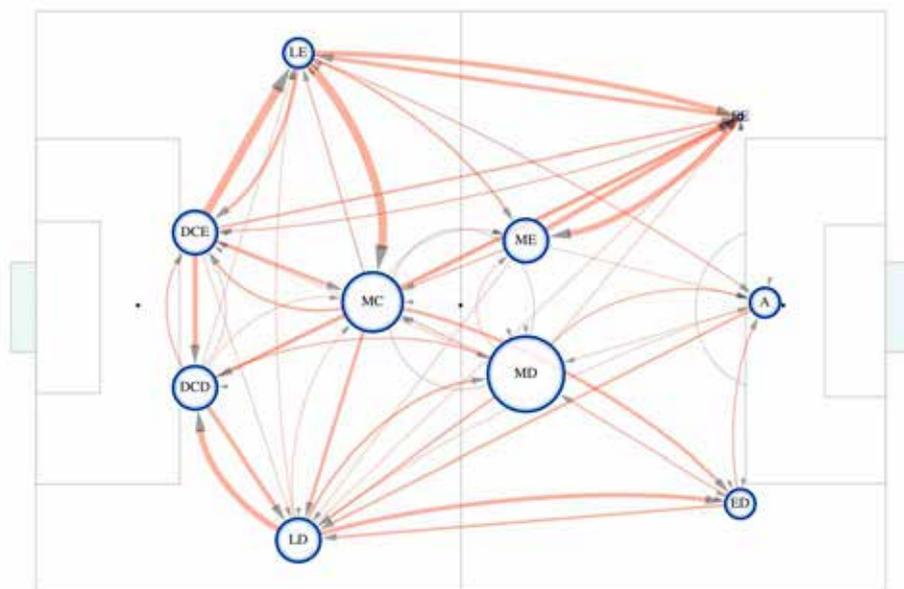


Figura 1: Exemplo da representação de uma SNA. As setas representam o número de passes efetuados pelos jogadores de futebol durante um jogo completo. Quanto mais espessa a seta, maior o número de passes realizado. A dimensão dos *Nodes* representa a medida de centralidade de *Closeness*. Quanto maior a dimensão do *Node*, maior é a suscetibilidade deste ser solicitado pelo sistema.

Atualmente observa-se uma mudança no paradigma de suporte às análises estatísticas realizadas em ciências do Desporto. A utilização das “Novas Estatísticas” (Cumming, 2014), também chamadas de “progressivas” (Hopkins, Marshall, Batterham, & Hanin, 2009), sustentadas genericamente pela utilização de tamanhos dos efeitos (*Effect Sizes*) e intervalos de confiança, em detrimento do valor de *p* e do “estatisticamente significativo”, começa a ser uma realidade (ver exemplo na Figura 2). A necessidade de os investigadores adotarem estes modelos de análise e interpretação dos dados nos seus artigos científicos tem sido referida recentemente por editores de revistas relevantes, como o *Journal of Sports Sciences* (Winter, Abt, & Nevill, 2014) ou o *International Journal of Sports Physiology and Performance* (Pyne, 2014).

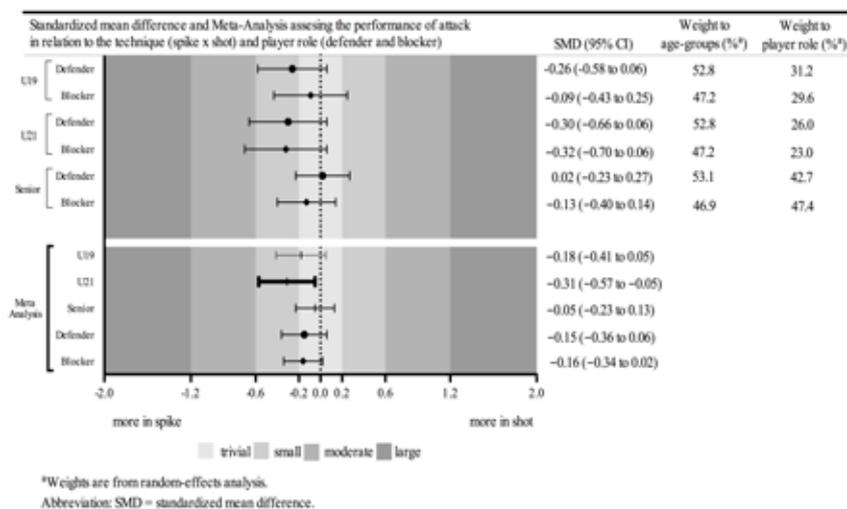


Figura 2

A figura 2, apresenta um exemplo de resultados obtidos com recurso às estatísticas progressivas. O gráfico apresenta os Tamanhos dos Efeitos, neste caso são as diferenças de médias estandardizadas, e os respetivos intervalos de confiança a 95%. As áreas sombreadas apresentam a magnitude das diferenças de forma qualitativa, ou seja, triviais, pequenas, moderadas ou grandes. Na parte inferior do gráfico, são igualmente apresentados os valores dos Tamanhos dos Efeitos produzidos através de meta-análise dos valores apresentados na parte superior, sendo a sua importância relativa (em percentagem) apresentada de forma numérica.

Interpretações abusivas dos testes de hipótese nula e valores de p

Apesar da controvérsia acerca da possibilidade dos testes de hipótese nula responderem às questões da investigação ter-se iniciado nos anos 60 (Meehl, 1967), a verdade é que este tem sido o paradigma dominante na investigação em Desporto. Cohen (1994), Kline (2004) e Cummings (2012) apresentam explicações detalhadas sobre os problemas/limitações dos testes de hipótese nula. Sinteticamente,

- valor de p não é a probabilidade da hipótese nula;
- rejeitar a hipótese nula não prova que a hipótese alternativa é verdadeira;
- não rejeitar a hipótese nula não prova que a hipótese alternativa é falsa;
- o valor de p não dá indicação do tamanho do efeito.

De acordo com Gardner e Altman, utilizar o valor arbitrário de 5% como nível de significância estatística para definir duas possibilidades – significativo ou não significativo – “não é útil e encoraja o pensamento preguiçoso” (Gardner & Altman, 1986, p.746).

Independentemente da interpretação do valor de p , os testes de hipóteses são ilógicos, porque a hipótese nula de não relação ou não diferença é sempre falsa, uma vez que não há verdadeiros efeitos zero na natureza (Batterham & Hopkins, 2006).

Na formulação de um problema ou de uma questão de pesquisa, os investigadores têm normalmente fundamentos para acreditar que os efeitos serão diferentes de zero. Assim, mais relevante do que saber se o efeito existe é perceber quão grande ele é.

Infelizmente, o valor de p por si só não fornece qualquer tipo de informação sobre a direção ou tamanho do efeito (Shakespeare, Gebski, Veness, & Simes, 2001). Dependendo, por exemplo, do tamanho e da variabilidade da amostra, um resultado estatístico com $p < 0.05$ pode representar um efeito que é clinicamente irrelevante. Ao contrário, um resultado não significativo ($p > 0.05$) não implica necessariamente que o efeito não deva ser tido em consideração, isto porque a combinação de amostras de tamanhos reduzidos com elevadas variabilidades pode mascarar efeitos importantes (Cohen, 1994).

Significância estatística vs. Significância prática/clínica com recurso à inferência estatística

Na investigação em Desporto, tal como em todas as outras áreas de conhecimento, deve haver uma distinção clara entre a significância estatística e a prática/clínica. A *significância estatística* – principal foco dos testes de hipótese nula – está relacionada com a possibilidade dos resultados serem atribuídos à sorte ou à variabilidade da amostra, enquanto que a *significância prática/clínica* está relacionada como a utilidade dos resultados (Kirk, 2007).

Considerando, a título de exemplo, o teste acerca da eficácia de um programa de desenvolvimento de força, os investigadores poderão questionar-se: “o programa funciona?” e “quão bem funciona?”, ou seja, “quais os ganhos esperados ao nível do aumento da força?”. Para além destas questões deverão igualmente questionar-se: “qual é a precisão da estimativa sobre os ganhos ao nível da força?”.

Precisão e Significância no Mundo Real

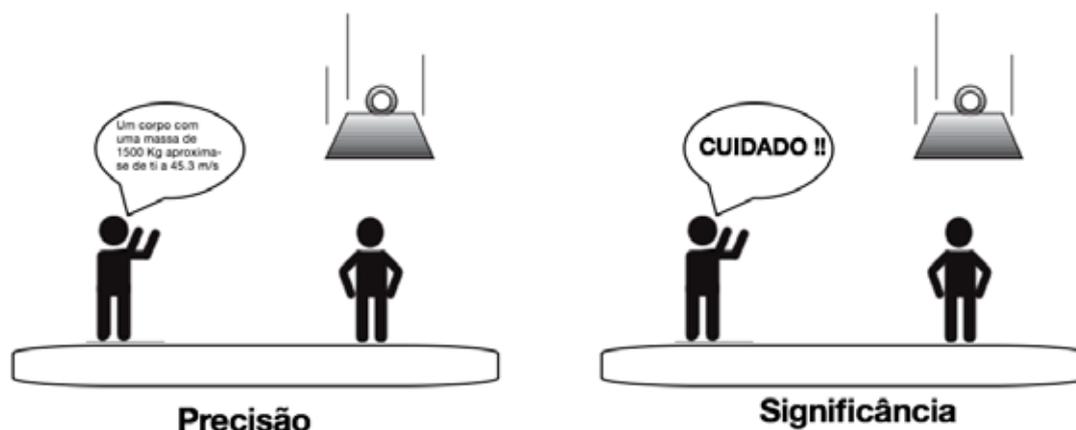


Figura 3: Retirado e adaptado de <http://www.mathworks.com/help/fuzzy/what-is-fuzzy-logic.html>

Um reduzido tamanho da amostra, ou uma grande variabilidade individual, poderão resultar numa incerteza acerca da eficácia do programa de treino, apesar de, hipoteticamente, encontrarmos uma grande diferença nas médias das performances entre os grupos intervencionados e não intervencionados. Os intervalos de confiança apresentam-se como uma medida capaz de captar a imprecisão da medida (ou seja, do tamanho do efeito), representando uma amplitude plausível na qual o valor verdadeiro (mas desconhecido) da população se situa (Cumming, 2012).

A questão mais difícil de responder será, eventualmente: “a intervenção – aplicação do programa de desenvolvimento de força – é rentável, i.e. custo-efetiva?”

Apesar da dificuldade em responder de forma inequívoca a esta questão, ela deverá ser considerada na investigação. Uma possibilidade de avaliar a eficácia de uma intervenção poderá ser através da definição prévia da diferença clínica/prática mínima para considerar a intervenção efetiva. Esta diferença é frequentemente considerada como sendo equivalente a um tamanho do efeito de 0.20 (Cohen, 1992).

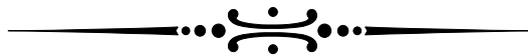
Em 2001, Shakespeare e colaboradores (2001), num artigo publicado na revista *The Lancet*, defenderam que o processo de interpretação dos resultados da investigação médica e a sua tradução nas práticas clínicas deveria ser efetuado com recurso a intervalos de confiança, curvas de significância clínica e análises de risco-benefício, e não apenas nos valores de p . Estas mesmas estratégias, particularmente a utilização de intervalos de confiança para os tamanhos dos efeitos (como apresentados na Figura 2), têm sido recorrentemente referidas como potenciadora do fornecimento de informação acerca da significância prática/clínica nas investigações em Desporto (Batterham & Hopkins, 2006; Hopkins et al., 2009; Pyne, 2014; Winter et al., 2014).

Referências

- Aughey, R. (2011). Applications of GPS Technologies to Field Sports. *International Journal of Sports Physiology and Performance*, 6(3), 295-310.
- Banks, M. (2014). Searching for beauty in data visualization. *Physics World*, 27(4), 10-10.
- Batterham, A., & Hopkins, W. (2006). Making Meaningful Inferences About Magnitudes. *International Journal of Sports Physiology and Performance*, 1(1), 50-57.
- Castellano, J., Alvarez-Pastor, D., & Bradley, P. S. (2014). Evaluation of research using computerised tracking systems (Amisco and Prozone) to analyse physical performance in elite soccer: a systematic review. *Sports Med*, 44(5), 701-712. doi: 10.1007/s40279-014-0144-3
- Cohen, J. (1992). A power primer. *Psychol Bull*, 112(1), 155-159.

- Cohen, J. (1994). The Earth Is Round (P-Less-Than.05). *American Psychologist*, 49(12), 997-1003. doi: Doi 10.1037/0003-066x.50.12.1103
- Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York, NY US: Routledge/Taylor & Francis Group.
- Cumming, G. (2014). The new statistics: why and how. *Psychol Sci*, 25(1), 7-29. doi: 10.1177/0956797613504966
- Cummins, C., Orr, R., O'Connor, H., & West, C. (2013). Global positioning systems (GPS) and microtechnology sensors in team sports: a systematic review. *Sports Med*, 43(10), 1025-1042. doi: 10.1007/s40279-013-0069-2
- Dellaserra, C., Gao, Y., & Ransdell, L. (2014). Use of integrated technology in team sports: a review of opportunities, challenges, and future directions for athletes. *J Strength Cond Res*, 28(2), 556-573. doi: 10.1519/JSC.0b013e3182a952fb
- Fewell, J., Armbruster, D., Ingraham, J., Petersen, A., & Waters, J. (2012). Basketball teams as strategic networks. *PLoS ONE*, 7(11), e47445. doi: 10.1371/journal.pone.0047445
- Fisher, R. (1925). *Statistical methods for research workers* (First Edition ed.). Edinburgh: Tweeddale Court: Oliver & Boyd.
- Fisher, W., & Lerman, D. (2014). It has been said that, "There are three degrees of falsehoods: lies, damn lies, and statistics". *J Sch Psychol*, 52(2), 243-248. doi: 10.1016/j.jsp.2014.01.001
- Gardner, M., & Altman, D. (1986). Confidence-Intervals Rather Than P-Values - Estimation Rather Than Hypothesis-Testing. *British Medical Journal*, 292(6522), 746-750.
- Goncalves, B., Figueira, B., Macas, V., & Sampaio, J. (2014). Effect of player position on movement behaviour, physical and physiological performances during an 11-a-side football game. *J Sports Sci*, 32(2), 191-199. doi: 10.1080/02640414.2013.816761
- Hopkins, W., Marshall, S., Batterham, A., & Hanin, J. (2009). Progressive statistics for studies in sports medicine and exercise science. *Med Sci Sports Exerc*, 41(1), 3-13. doi: 10.1249/MSS.0b013e31818cb278
- Hughes, M., & Franks, I. (1997). *Notational Analysis of Sport*. London: E & FN Spon.
- Ioannidis, J. (2005). Why most published research findings are false. *PLoS Medicine*, 2(8), e124. doi: 10.1371/journal.pmed.0020124
- Kirk, R. (2007). Effect magnitude: A different focus. *Journal of Statistical Planning and Inference*, 137(5), 1634-1646. doi: Doi 10.1016/J.Jspi.2006.09.011
- Kline, R. B. (2004). *Beyond Significance Testing: Reforming Data Analysis Methods in Behavioral Research*. Washington, DC: American Psychological Association.
- Marcelino, R., Mesquita, I., & Sampaio, J. (2011). Effects of quality of opposition and match status on technical and tactical performances in elite volleyball. *J Sports Sci*, 29(7), 733-741. doi: 10.1080/02640414.2011.552516
- Marcelino, R., Sampaio, J., & Mesquita, I. (2011). Investigação centrada na análise do jogo: da modelação estática à modelação dinâmica. *Rev Port Cien Desp*, 11(1), 481-499.
- Marcelino, R., Sampaio, J., & Mesquita, I. (2012). Attack and serve performances according to the match period and quality of opposition in elite volleyball matches. *J Strength Cond Res*, 26(12), 3385-3391. doi: 10.1519/JSC.0b013e3182474269
- Medeiros, A., Mesquita, I., Marcelino, R., & Palao, J. (2014). Effects of technique, age and player's role on serve and attack efficacy in high level beach volleyball players. *International Journal of Performance Analysis in Sport*, 14(3), 680-691.
- Meehl, P. (1967). Theory-Testing in Psychology and Physics: A Methodological Paradox. *Philosophy of Science*, 34(2), 103. doi: 10.1086/288135
- Moura, F., Martins, L., & Cunha, S. (2014). Analysis of football game-related statistics using multivariate techniques. *Journal of Sports Sciences*, 1-7. doi: 10.1080/02640414.2013.853130
- Neiman, T., & Loewenstein, Y. (2014). Spatial generalization in operant learning: lessons from professional basketball. *PLoS Comput Biol*, 10(5). doi: 10.1371/journal.pcbi.1003623
- Neyman, J. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London Series a-Containing Papers of a Mathematical or Physical Character*, 231, 289-337. doi: Doi 10.1098/Rsta.1933.0009
- Perl, J. (2004). A neural network approach to movement pattern analysis. *Hum Mov Sci*, 23(5), 605-620. doi: 10.1016/j.humov.2004.10.010

- Pierson, K., Addona, V., & Yates, P. (2014). A behavioural dynamic model of the relative age effect. *J Sports Sci*, 32(8), 776-784. doi: 10.1080/02640414.2013.855804
- Popper, K. (1972). *The logic of scientific discovery*. 6th ed. London: Hutchinson & Co Ltd.
- Pyne, D. (2014). Improving the Practice of Sports Science Research. *Int J Sports Physiol Perform*, 9(6), 899. doi: 10.1123/IJSPP.2014-0451
- Sampaio, J., Lago, C., Goncalves, B., Macas, V. M., & Leite, N. (2014). Effects of pacing, status and unbalance in time motion variables, heart rate and tactical behaviour when playing 5-a-side football small-sided games. *J Sci Med Sport*, 17(2), 229-233. doi: 10.1016/j.jsams.2013.04.005
- Sampaio, J., & Macas, V. (2012). Measuring tactical behaviour in football. *Int J Sports Med*, 33(5), 395-401. doi: 10.1055/s-0031-1301320
- Sankaran, S. (2014). Comparing Pay versus Performance of IPL Bowlers: An application of Cluster Analysis. *International Journal of Performance Analysis in Sport*, 14(1), 174-187.
- Shakespeare, T., Gebiski, V., Veness, M., & Simes, J. (2001). Improving interpretation of clinical studies by use of confidence levels, clinical significance curves, and risk-benefit contours. *Lancet*, 357(9265), 1349-1353. doi: Doi 10.1016/S0140-6736(00)04522-0
- Skiba, D. (2014). The connected age: big data & data visualization. *Nurs Educ Perspect*, 35(4), 267-268.
- Winter, E., Abt, G., & Nevill, A. (2014). Metrics of meaningfulness as opposed to sleights of significance. *J Sports Sci*, 32(10), 901-902. doi: 10.1080/02640414.2014.895118



Estatística de Extremos em Desporto

Lígia Henriques Rodrigues

Centro de Estatística e Aplicações da Universidade de Lisboa (CEAUL)

lcphjr@gmail.com

M. Ivette Gomes

Universidade de Lisboa, CEAUL, e DEIO-FCUL

Instituto de Investigação Científica Bento da Rocha Cabral, Portugal
migomes@fc.ul.pt

Dinis Pestana

Universidade de Lisboa, CEAUL, e DEIO-FCUL

Instituto de Investigação Científica Bento da Rocha Cabral, Portugal
dinis.pestana@fc.ul.pt

1 Introdução

É muito usual ouvirmos os comentadores desportivos apoiarem as suas opiniões e argumentos em factos estatísticos. Excelentes programas sobre as mais diversas modalidades, tais como corridas com ou sem barreiras, natação, salto em altura e salto à vara, entre outras, mostram que a Estatística é um instrumento indispensável para quem ambiciona alcançar uma medalha, sendo um instrumento que os instrutores de campeões têm obrigatoriamente de usar. As *estatísticas ordinais* (EO's) extremas, e em particular os máximos (ou mínimos) e os recordes, são pois as estatísticas de maior relevância.

Muitas questões da vida real, incluindo o Desporto, requerem a estimação sobre acontecimentos acerca dos quais os dados são inexistentes ou se existem são escassos, os chamados acontecimentos extremos ou raros. A *Teoria de Valores Extremos* (EVT, do Inglês “*extreme value theory*”) é um ramo probabilístico de suporte à Estatística que lida exatamente com tais situações, ajudando a descrever e a quantificar os ditos acontecimentos raros. Em particular, permite a estimação de probabilidades de acontecimentos para os quais não existem dados, ou como usualmente dizemos, permite *extrapolar para além da amostra*.

O *Teorema de Fisher-Tippett-Gnedenko* (o teorema fulcral dos tipos em EVT) é um resultado acerca da distribuição assintótica das EO's extremas. O teorema dos tipos extremos desempenha um papel análogo ao tão famoso *Teorema Limite Central* (TLC) para as médias (somas). Basicamente, estabelece que o máximo amostral linear e convenientemente normalizado converge nas mais variadas situações para uma *variável aleatória* (VA) com uma de 3 distribuições possíveis, a Gumbel, a Fréchet ou a Max-Weibull, que podem ser unificadas (Von Mises, 1936; Jenkinson, 1955) na forma funcional,

$$G_{\xi}(x) = \begin{cases} \exp\left(-(1 + \xi x)^{-1/\xi}\right), & 1 + \xi x > 0, \text{ se } \xi \neq 0 \\ \exp(-\exp(-x)), & x \in \mathfrak{R}, \text{ se } \xi = 0, \end{cases} \quad (1)$$

a chamada *função de distribuição* (FD) de *valores extremos* (EV, do Inglês “*extreme value*”). O parâmetro de forma ξ em (1) é o chamado *índice de valores extremos* (EVI, do Inglês “*extreme value index*”). Independentemente da forma do centro da distribuição, a *cauda assume formas sempre muito características* quando estamos suficientemente longe nessa cauda. O crédito deste resultado é devido essencialmente a Gnedenko (1943), embora versões anteriores tivessem sido estabelecidas por Fréchet (1927) e Fisher and Tippett (1928). Em abordagens paramétricas à Estatística de Extremos trabalhamos usualmente com o modelo G_{ξ} , em (1), ou com o modelo generalizado de Pareto (GP) associado aos excessos acima de um nível elevado, $GP_{\xi}(x) = 1 + \ln G_{\xi}(x)$, $x \geq 0$ (veja-se Gomes *et al.*, 2013).

Apesar da existência de abordagens paramétricas em Estatística de Extremos para aplicações a dados de Desporto, como se pode ver em Robinson and Tawn (1995) e Barão and Tawn (1999), que consideram os melhores tempos anuais em corridas de 3000 metros para mulheres, e em Smith (1988), que trabalha com dados da maratona, entre outros, enquadrar-nos-emos aqui essencialmente numa abordagem de índole semi-paramétrica à Estatística de Extremos. Trabalhamos então com estimadores de parâmetros de acontecimentos raros baseados nas k EO's de topo. O parâmetro fundamental continua a ser o EVI, ξ , em (1), que deve ser estimado de forma “precisa”, uma vez que é a base para a estimação de outros parâmetros de acontecimentos extremos, tais como o *limite superior do suporte* do modelo subjacente aos dados,

$$x^F = \sup\{x : F(x) < 1\}, \quad (2)$$

de importância fundamental na área do Desporto. Tendo em vista a aplicação a dados de melhores marcas em modalidades de atletismo, e em contexto semelhante ao usado em Einmahl and Magnus (2008), Gomes and Pestana (2009), Einmahl and Smeets (2011) e Henriques-Rodrigues *et al.* (2011), daremos atenção à estimação do índice de cauda, ξ , bem como à estimação do limite superior do suporte x^F , em (2), se finito, o “recorde mundial” possível face às condições atuais.

2 Alguns Resultados em EVT

Consideremos um qualquer acontecimento desportivo, denotemos as melhores marcas pessoais de n atletas por X_1, \dots, X_n e por $X_{1:n} \leq \dots \leq X_{n:n}$ as EO's ascendentes associadas. Admitamos ainda que, caso necessário, os dados são transformados de modo a podermos falar de máximos (e não de mínimos). Iremos pois trabalhar com EO's superiores. Por simplicidade, admitamos também que (X_1, \dots, X_n) podem ser consideradas como observações independentes e identicamente distribuídas (IID) de um modelo F , desconhecido.

Um dos resultados fundamentais em EVT tem a ver com a identificação das possíveis leis limites de $X_{n:n} := \max(X_1, \dots, X_n)$. Tem-se obviamente $X_{n:n} \xrightarrow[n \rightarrow \infty]{p} x^F$, com x^F dado em (2). Para se obter um possível comportamento limite não-degenerado, é necessário normalizar $X_{n:n}$. De forma análoga ao TLC, sabemos que se o máximo $X_{n:n}$, linearmente normalizado, convergir para uma VA não degenerada, existem sucessões de reais $\{a_n\}_{n \geq 1}$ ($a_n > 0$) e $\{b_n\}_{n \geq 1}$ tais que

$$\lim_{n \rightarrow \infty} P(a_n^{-1}(X_{n:n} - b_n) \leq x) = \lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G_\xi(x) \quad (3)$$

para algum $\xi \in \mathfrak{R}$ (Gnedenko, 1943), com $G_\xi(x)$ definida em (1). Dizemos então que F está no *max-domínio de atração* de G_ξ e usamos a notação óbvia $F \in D_M(G_\xi)$. O índice de valores extremos ξ mede pois essencialmente o peso da cauda direita $\bar{F} = 1 - F$:

- se $\xi < 0$, temos uma cauda direita leve, i.e. F tem limite superior de suporte finito;
- se $\xi > 0$, temos uma cauda direita pesada de tipo polinomial negativo, também chamada de tipo Pareto, i.e. F tem limite superior de suporte infinito;
- se $\xi = 0$, a cauda direita é de tipo exponencial e o limite superior do suporte de F pode ser finito ou infinito.

Observação 2.1. *Note-se que qualquer resultado obtido para máximos pode ser reformulado para mínimos, uma vez que $\min(X_1, \dots, X_n) = -\max(-X_1, \dots, -X_n)$.*

Observação 2.2. *Note-se ainda que dizer que $F \in D_M(G_\xi)$ é equivalente a dizer que para qualquer $x \in \mathfrak{R}$ tal que $0 < G_\xi(x) < 1$, tem-se $\lim_{n \rightarrow \infty} n \ln F(a_n x + b_n) = \ln G_\xi(x) = -(1 + \xi x)^{-1/\xi}$, tendo-se, equivalentemente,*

$$\lim_{n \rightarrow \infty} n(1 - F(a_n x + b_n)) = -\ln G_\xi(x) = (1 + \xi x)^{-1/\xi}. \quad (4)$$

Definamos

$$U(t) := F^{\leftarrow}(1-1/t) \quad (t > 1), \quad F^{\leftarrow}(x) := \inf \{y : F(y) \geq x\}. \quad (5)$$

É fácil de demonstrar (veja-se o Teorema 1.1.2 em de Haan and Ferreira, 2006) que, com $a_t \equiv a(t) := a_{\lfloor t \rfloor}$ e $b_t \equiv b(t) := b_{\lfloor t \rfloor}$, com $\lfloor t \rfloor =$ parte inteira de t , (a_n, b_n) definidos em (3), e denotando G_ξ^{-1} a função inversa da FD G_ξ em (1),

$$\lim_{n \rightarrow \infty} (U(tx) - b_t) / a_t = D(x) = G_\xi^{-1}(\exp(-1/x)) = (x^\xi - 1) / \xi \quad (6)$$

para todo o $x > 0$, podendo-se escolher $b_t = U(t)$, com $U(\cdot)$ definido em (5).

Observação 2.3. Quando $\xi = 0$, as funções $-\ln G_\xi(x) = (1 + \xi x)^{-1/\xi}$ e $G_\xi^{-1}(\exp(-1/x)) = (x^\xi - 1) / \xi$, em (4) e em (6), devem ser interpretadas como $\exp(-x)$ e $\ln x$, respetivamente.

3 Estimação semi-paramétrica de alguns parâmetros de interesse

Como estimar o índice de valores extremos ξ , a escala a , a localização b e o limite superior do suporte x^{F^*} ?

3.1 Estimadores do índice de valores extremos

Para $j \geq 1$, denotemos

$$L_{k,n}^{(j)} := \frac{1}{k} \sum_{i=1}^k \left\{ 1 - \frac{X_{n-k:n}}{X_{n-i+1:n}} \right\}^j, \quad M_{k,n}^{(j)} := \frac{1}{k} \sum_{i=1}^k \left\{ \ln \frac{X_{n-i+1:n}}{X_{n-k:n}} \right\}^j. \quad (7)$$

Neste estudo, consideraremos os estimadores seguintes, válidos para qualquer $\xi \in \mathfrak{R}$.

1. O estimador de *momentos* (M), introduzido em Dekkers *et al.* (1989), com a forma funcional

$$\hat{\xi}_{k,n}^M \equiv M_{k,n} := M_{k,n}^{(1)} + \frac{1}{2} \left\{ 1 - \left(M_{k,n}^{(2)} / [M_{k,n}^{(1)}]^2 - 1 \right)^{-1} \right\}. \quad (8)$$

2. O estimador *generalizado de Hill* (GH), introduzido em Beirlant *et al.* (1996) e estudado posteriormente em Beirlant *et al.* (2005). Tem-se

$$\hat{\xi}_{k,n}^{GH} \equiv GH_{k,n} := \hat{\xi}_{k,n}^H + \frac{1}{k} \sum_{i=1}^k \left\{ \ln \frac{\hat{\xi}_{i,n}^H}{\hat{\xi}_{k,n}^H} \right\}, \quad \hat{\xi}_{k,n}^H := \frac{1}{k} \sum_{i=1}^k \left\{ \ln \frac{X_{n-i+1:n}}{X_{n-k:n}} \right\}. \quad (9)$$

3. O estimador de *momentos mistos* (MM) (Fraga Alves *et al.*, 2009), com a forma funcional

$$\hat{\xi}_{k,n}^{MM} \equiv MM_{k,n} := \frac{\hat{\varphi}_{k,n} - 1}{1 + 2 \min(\hat{\varphi}_{k,n} - 1, 0)}, \quad \hat{\varphi}_{k,n} := \frac{M_{k,n}^{(1)} - L_{k,n}^{(1)}}{(L_{k,n}^{(1)})^2}. \quad (10)$$

Devido à especificidade dos dados, consideraremos também um estimador simples, válido só para $\xi < -1/2$,

4. o estimador introduzido em Falk (1995), denotado F , e dado por

$$\hat{\xi}_{k,n}^F \equiv F_{k,n} := \frac{1}{k} \sum_{i=1}^{k-1} \ln \frac{X_{n:n} - X_{n-1:n}}{X_{n:n} - X_{n-k:n}}. \quad (11)$$

Trata-se de estimadores que são consistentes desde que $k = k_n$ seja uma sucessão intermédia, i.e. uma sucessão de inteiros tais que

$$k = k_n \rightarrow \infty \text{ e } k_n = o(n), \text{ quando } n \rightarrow \infty. \quad (12)$$

Numa sub-classe vasta do domínio de atração, e sob condições adequadas em k , conseguimos ainda garantir a normalidade assintótica de qualquer destes estimadores, bem como dos estimadores de parâmetros de acontecimentos extremos a introduzir na Secção 3.2.

3.2 Estimação de outros parâmetros de interesse

3.2.1 Estimadores dos parâmetros de localização e escala

Já mencionámos que podemos escolher $b_t = U(t)$, com $U(\cdot)$ definida em (5). Por outro lado, a transformação uniformizante permite-nos garantir que $\forall F$ desconhecido e subjacente à VA X , $X = U(Y)$ com Y VA Pareto unitária, i.e., VA com FD $F_Y(y) = 1 - 1/y$, $y \geq 1$. Consequentemente,

$$X_{n-k:n} \stackrel{d}{=} U(Y_{n-k:n}), \text{ e como } Y_{n-k:n} \stackrel{p}{\sim} n/k, \text{ quando } n \rightarrow \infty,$$

é sensato considerar

$$\hat{b} = \hat{b}_{k,n} = \hat{U}(n/k) = X_{n-k:n}.$$

E qualquer que seja o estimador de ξ , denotado $\hat{\xi}_{k,n}^{\bullet}$, pode-se considerar

$$\hat{a}^{\bullet} = \hat{a}_{k,n}^{\bullet} = X_{n-k:n} M_{k,n}^{(1)} \left(1 - \min(0, \hat{\xi}_{k,n}^{\bullet}) \right)$$

com $M_{k,n}^{(1)}$, definido em (7).

3.2.2 Estimadores do limite superior do suporte, caso $\xi < 0$

É possível provar (de Haan, 1984) que $F \in D_M(G_{\xi})$ se e só se, para todo o $x > 0$,

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \begin{cases} \frac{x^{\xi} - 1}{\xi} & \text{se } \xi \neq 0 \\ \ln x & \text{se } \xi = 0, \end{cases}$$

com $U(\cdot)$ definida em (5). Para valores elevados de t , podemos pois escrever

$$U(tx) \approx U(t) + a(t)(x^{\xi} - 1) / \xi.$$

Mas $x^F = U(\infty)$ e para $\xi < 0$, $(x^{\xi} - 1) / \xi \rightarrow -1 / \xi$, quando $x \rightarrow \infty$. Fazendo $t = n/k$, podemos pois garantir que, sempre que $\hat{\xi}_{k,n}^{\bullet} < 0$,

$$x^F \approx U(n/k) - a(n/k) / \xi \Rightarrow \hat{x}_k^F := \hat{b} - \hat{a}^{\bullet} / \hat{\xi}_{k,n}^{\bullet}.$$

Para mais detalhes sobre o assunto, vejam-se os Capítulos 1 e 4 em de Haan and Ferreira (2006).

4 Análise de um conjunto de dados em atletismo

Como ilustração, procedemos a uma análise dos dados associados a corridas de 200 metros, para homens. Os dados foram recolhidos em <http://www.iaaf.org/statistics/toplists/index.htm> e http://hem.bredband.net/athletics/athletics_all-time_best.htm. As observações foram colecionadas até ao fim de 2007, e para cada atleta foi unicamente considerada a sua melhor marca. A dimensão da amostra é $n=352$. Tal como anteriormente referido, estamos interessados na cauda direita do modelo F subjacente aos dados. Consequentemente, convertemos *tempos de corrida* em *velocidades*, i.e., 20 segundos nos 200 metros (ou seja, nos 0.2 quilómetros) são transformados na velocidade de $3600 \times 0.2/20 = 36$ km/h. Com esta transformação é óbvio que quanto maior for a velocidade tanto melhor, e estamos pois interessados em valores máximos, usando unicamente o melhor resultado de cada um dos n atletas. Análises semelhantes de outros conjuntos de dados em Desporto são efectuadas em Gomes e Pestana (2009) e em Henriques-Rodrigues *et al.* (2011).

4.1 Análise paramétrica dos dados

Como os dados observados (após a mudança de escala atrás referida) já são valores máximos, possivelmente de um número pequeno de marcas dependentes associadas com cada um dos n atletas, mas a lei limite em (1) é robusta relativamente ao relaxamento quer da hipótese de independência quer de identidade distribucional, tentámos primeiro ajustar, através da máxima verosimilhança (ML, do Inglês “*maximum likelihood*”), um modelo de valores extremos $F(x;\lambda,\delta,\xi)=G_\xi((x-\lambda)/\delta)$, com $G_\xi(x)$ dada em (1). Usámos para isso o “package” EVIR, do “software” **R** e os resultados são apresentados na Tabela 1. Tal como esperado, a estimativa de ξ é negativa. Um pouco mais decepcionante é a estimação do limite superior do suporte, dada por $\hat{x}^F = \max(x_{n:n}, \hat{\lambda} - \hat{\delta} / \hat{\xi})$, que é igual ao valor máximo dos dados.

Modalidade	n	$(x_{1:n}, x_{n:n})$ em Km/h	$\hat{\xi}$ (IC a 95%)	$\hat{\lambda}$	$\hat{\delta}$	\hat{x}^F
200M	352	(33.72, 36.14)	-0.22 (-0.340,-0.101)	34.08	0.28	36.14

Tabela 1: Estimativas ML de ξ , (λ, δ) e x^F para o modelo $G_\xi((x-\lambda)/\delta)$, com $G_\xi(x)$ dada em (1)

A um nível de significância $\alpha=0.05$, o modelo (unificado) de *valores extremos* foi rejeitado pelo teste de Kolmogorov-Smirnov, tal como pode ser inferido graficamente da Figura 1, em que representamos à *esquerda* a FD empírica e a FD de valores extremos estimada, para o conjunto de dados em análise.

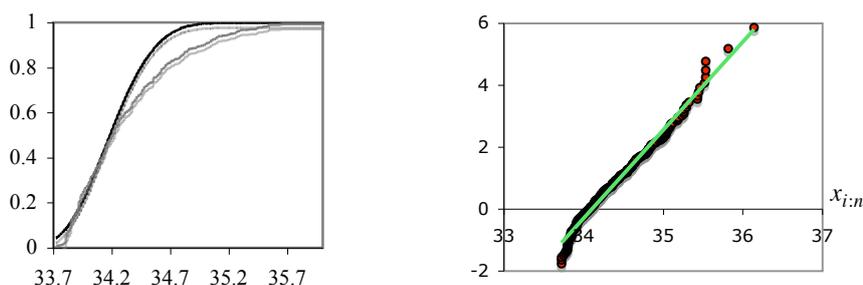


Figura 1: FD empírica (a cinzento) e FD de valores extremos (a preto) (*esquerda*), e Gumbel QQ-plot (*direita*) do conjunto de dados relativos a corridas de 200 metros, para homens

Antes de procedermos a uma análise em contexto semi-paramétrico, ilustramos, também na Figura 1, à *direita*, o gráfico em papel de probabilidade Gumbel associado aos dados em análise, com a marcação dos pontos $(x_{i:n}, p_i^A = -\ln(-\ln(i/(n+1))))$, $1 \leq i \leq n$, e o ajustamento de uma recta de mínimos quadrados. E este QQ-plot evidencia também um comportamento de cauda direita leve, i.e. um índice de cauda $\xi < 0$.

Face à rejeição do modelo EV, e de outros modelos paramétricos alternativos, sentimos pois a necessidade de uma análise de dados em contexto semi-paramétrico, a desenvolver em seguida.

4.2 Análise semi-paramétrica dos dados

4.2.1 Teste ao sinal do índice de valores extremos

Tal como mencionámos anteriormente, e sempre que nos colocamos em contexto semi-paramétrico, admitimos unicamente que $F \in D_M(G_\xi)$, sendo ξ o parâmetro fundamental de valores extremos. E em muitas áreas em que os extremos são relevantes, o caso mais frequente e simples é considerar $\xi = 0$. Além disso, se claramente pensamos que $\xi < 0$ ou que $\xi > 0$, temos procedimentos específicos para a

estimação ξ , frequentemente mais fiáveis do que os procedimentos válidos para $\xi \in \mathfrak{R}$. Antes de procedermos a uma análise mais aprofundada da cauda direita de F é pois sensato testar

$$H_0 : F \in D_M(G_\xi)_{\xi=0} \left(\text{ou } F \in D_M(G_\xi)_{\xi \geq 0} \right) \text{ versus } H_1 : F \in D_M(G_\xi)_{\xi < 0}. \quad (13)$$

Iremos aqui considerar duas estatísticas de teste baseadas nos excessos acima de um limiar elevado, $X_{n-k:n}$, com k intermédio, i.e. tal que se tem a validade de (12). A primeira estatística de teste foi introduzida em Greenwood (1946) e a segunda em Hasofer and Wang (1992). Estas estatísticas foram estudadas em contexto semi-paramétrico em Neves and Fraga Alves (2007) e são dadas por

$$G_{k,n} := \frac{\frac{1}{k} \sum_{i=1}^k (X_{n-i+1:n} - X_{n-k:n})^2}{\left(\frac{1}{k} \sum_{i=1}^k (X_{n-i+1:n} - X_{n-k:n}) \right)^2}, \quad W_{k,n} := \frac{1}{k(G_{k,n} - 1)}.$$

Sob a validade da hipótese nula H_0 , em (13), e condições extras de regularidade impostas na cauda direita de F e no crescimento de $k=k_n$, ambas as estatísticas têm um comportamento assintótico normal. Mais especificamente,

$$G_{k,n}^* := \sqrt{k/4} (G_{k,n} - 2) \Big|_{F \in D_M(G_0)} \xrightarrow[n \rightarrow \infty]{d} N(0,1) \quad (14)$$

e

$$W_{k,n}^* := \sqrt{k/4} (kW_{k,n} - 1) \Big|_{F \in D_M(G_0)} \xrightarrow[n \rightarrow \infty]{d} N(0,1). \quad (15)$$

Face à importante contribuição do máximo para a soma dos k excessos, $X_{n-i+1:n} - X_{n-k:n}$, $1 \leq i \leq k$, Neves *et al.* (2006) introduziram a estatística,

$$R_{k,n} := \frac{X_{n:n} - X_{n-k:n}}{\frac{1}{k} \sum_{i=1}^k (X_{n-i+1:n} - X_{n-k:n})},$$

também incluída nesta análise de dados. O comportamento assintótico de $R_{k,n}$ é Gumbel, i.e. $\Lambda = G_0$ é a FD da VA limite, com G_ξ definida em (1). Mais especificamente,

$$R_{k,n}^* := R_{k,n} - \ln k \Big|_{F \in D_M(G_0)} \xrightarrow[n \rightarrow \infty]{d} Z \sim G_0. \quad (16)$$

Como função de k , $G_{k,n}^*$ e $R_{k,n}^*$ tendem a ter uma inclinação com o sinal de ξ . A estatística $W_{k,n}^*$ funciona em sentido reverso.

Na Figura 2, e para os dados dos 200 metros, apresentamos as trajetórias amostrais das três estatísticas de teste, $G_{k,n}^*$, $W_{k,n}^*$ e $R_{k,n}^*$ em (14), (15) e (16), respectivamente. Nessa mesma figura também marcamos os quantis $(\chi_{0.025}^\Phi, \chi_{0.975}^\Phi)$ da normal padrão Φ , os valores $(-1.96, +1.96)$, e da Gumbel padrão, $\Lambda = G_0$, os valores $(-1.31, +3.68)$.

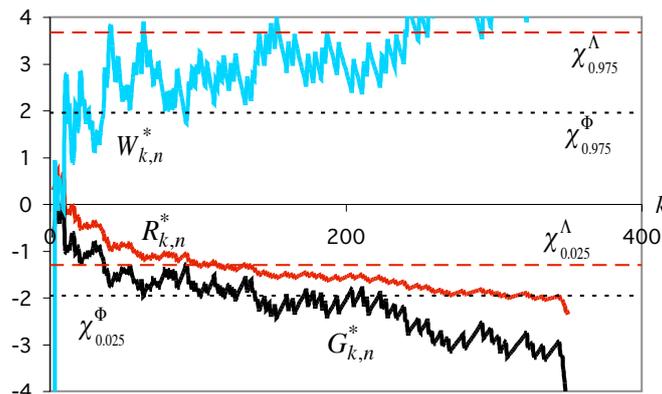


Figura 2: Estatísticas de teste para as corridas de 200 metros, para homens

Este gráfico mostra claramente uma tendência decrescente de $R_{k,n}^*$ e de $G_{k,n}^*$ (respectivamente abaixo de $\chi_{0,025}^A$ e de $\chi_{0,025}^\Phi$ para um grande número de valores de k), bem como uma tendência crescente de $W_{k,n}^*$ (acima de $\chi_{0,975}^\Phi$ desde valores moderados até valores elevados de k). Essa tendência está essencialmente relacionada com o viés, mas o viés está por sua vez fortemente relacionado com o sinal do EVI. Apesar disto note-se que para o ficheiro 200M (dos 200 metros, para homens), a trajetória amostral de $G_{k,n}^*$ e mais acentuadamente a de $R_{k,n}^*$ estão dentro do intervalo de confiança a 95% para uma zona vasta de valores de k . Tal situação era também esperada, uma vez que se sabe (veja-se, por exemplo Neves and Fraga Alves, 2008) que $R_{k,n}^*$ é um teste conservativo e o verdadeiro valor de ξ não está certamente muito longe de zero. Temos de qualquer modo uma indicação clara de um índice de valores extremos negativo, tal como esperado.

4.2.2 Estimativas semi-paramétricas do EVI e do limite superior do suporte

Dificuldades na estimação do valor ótimo de k , no sentido de erro quadrático médio mínimo, para $\xi < 0$, levaram-nos a considerar um método heurístico, apresentado em seguida, estudado em Henriques-Rodrigues *et al.* (2011). Seja $\hat{\xi}_{k,n}^{(i)}$, $i \in T = \{1, 2, 3, 4\}$, o conjunto de estimadores alternativos do EVI aqui considerados, i.e. os estimadores em (8), (9), (10) e (11). Sugerimos então a escolha

$$k_{\min}^* := \arg \min_k \sum_{(i,j) \in T, i \neq j} \left(\hat{\xi}_{k,n}^{(i)} - \hat{\xi}_{k,n}^{(j)} \right)^2 \quad (17)$$

e a consideração dos estimadores adaptativos

$$T_{\min}^* = T_{k_{\min}^*, n}, \text{ para } T=M, GH, MM \text{ e } F, \quad (18)$$

com $M_{k,n}$, $GH_{k,n}$, $MM_{k,n}$, $F_{k,n}$ e k_{\min}^* dados em (8), (9), (10), (11) e (17), respectivamente.

Na Tabela 2 apresentamos uma estimativa de ξ e associado intervalo de confiança a 95%. Esta estimativa de ξ foi obtida através do estimador MM em (10), calculado no valor k_{\min}^* em (17) também apresentado na última coluna da tabela. A escolha de MM_{\min}^* prende-se com um estudo de simulação desenvolvido para o modelo de *valores extremos* G_ξ em (1), que nos forneceu, para o viés absoluto e o erro quadrático médio (EQM) dos estimadores T_{\min}^* em (18), padrões já referidos em Gomes e Pestana (2009), do tipo do fornecido na Figura 3.

Modalidade	n	$(x_{1,n}, x_{n,n})$	MM* (IC a 95%)	k_{\min}^*
200M	352	(33.72, 36.14)*	-0.25 (-0.379,-0.129)	269

Tabela 2: Estimativas do índice de valores extremos: * km/h

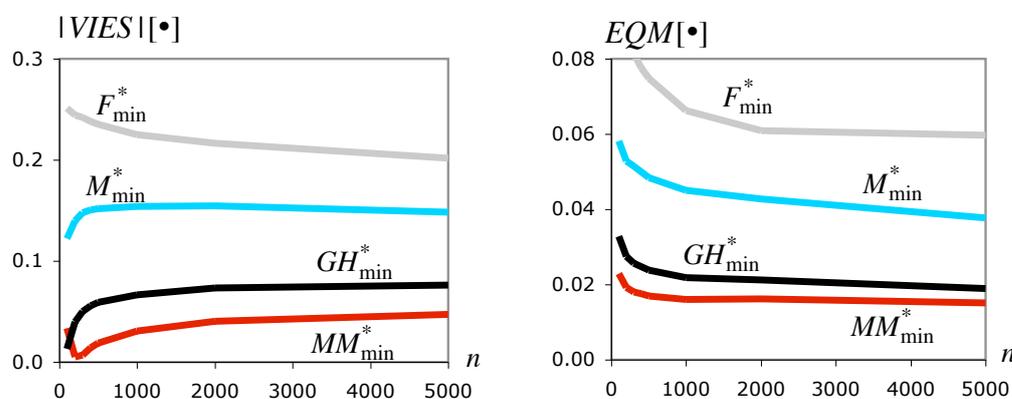


Figura 3: Valor absoluto do viés (*esquerda*) e erros médios quadráticos (*direita*) dos estimadores T_{\min}^* em (18), para um modelo de valores extremos com $\xi = -0.1$ e para dimensões de amostra $n=100, 200, 500, 1000$,

Finalmente, na Tabela 3, apresentamos a estimativa do limite superior do suporte obtida com base na estimativa de ξ associada ao estimador MM em (10), para as corridas de 200 metros, para homens.

Modalidade	$x_{n:n}$	k^{*F}	\hat{x}_{MM}^F
200M	36.14* (00:19.92)*	269	36.47* (00:19.74)*

Tabela 3: Estimativa do limite superior do suporte, *km/h, e conversão para o limite inferior, *minutos

O resultado apresentado na Tabela 3 não é surpreendente e mostra-nos que, nas condições atuais, existe um limite superior para a velocidade, que implica que haja um limite inferior para o tempo de corrida da modalidade analisada, tal como já tinha sido detectado em artigos anteriores sobre o assunto.

Agradecimentos

Investigação parcialmente financiada por Fundos Nacionais através da **FCT** --- Fundação para a Ciência e a Tecnologia, bolsa SFRH/BPD/77319/2011 e projeto PEst-OE/MAT/UI0006/2014.

Referências

- [1] Barão, M.I., and Tawn, J. (1999). Extremal analysis of short series with outliers: sea-levels and athletic records. *Applied Statistics* **48**, 469–487.
- [2] Beirlant, J., Dierckx, G., and Guillou, A. (2005). Estimation of the extreme-value index and generalized quantile plots. *Bernoulli* **11**:6, 949–970.
- [3] Beirlant, J., Vynckier, P., and Teugels, J. (1996). Excess functions and estimation of the extreme-value index. *Bernoulli* **2**, 293–318.
- [4] Dekkers, A.L.M., Einmahl, J.H.J. and Haan, L. de (1989). A moment estimator for the index of an extreme-value distribution. *Ann. Statist.* **17**, 1833–1855.
- [5] Einmahl, J., and Magnus, J.R. (2008). Records in athletics through extreme-value theory. *Journal of the American Statistical Association*, **103**, 1382–1391.
- [6] Einmahl, J., and Smeets, S.G.W.R. (2011). Ultimate 100-m world records through extreme-value theory. *Statistica Neerlandica* **65**:1, 32–42.
- [7] Falk, M. (1995). Some best parameter estimates for distributions with finite endpoint. *Statistics* **27**(1-2), 115–125.
- [8] Fisher, R.A., and Tippett, L.H.C. (1928). Limiting forms of the frequency of the largest or smallest member of a sample. *Proc. Cambridge Phil. Soc.* **24**, 180–190.
- [9] Fraga Alves, M.I., Gomes M.I., de Haan, L., and Neves, C. (2009). Mixed moment estimator and location invariant alternatives. *Extremes* **12**, 149–185.
- [10] Fréchet, M. (1927). Sur le loi de probabilité de l'écart maximum. *Ann. Société Polonaise de Mathématique* **6**, 93–116.
- [11] Gnedenko, B.V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.* **44**, 423–453.
- [12] Gomes, M.I., and Pestana, D. (2009). Caudas leves em desporto: estimação de parâmetros úteis. In Oliveira, I., M.M. et al. (eds.), *Estatística: Arte de Explicar o Acaso*, 307–318, Edições S.P.E.
- [13] Gomes, M.I., Fraga Alves, M.I., e Neves, C. (2013). *Análise de Valores Extremos: uma Introdução*. Edições S.P.E. and I.N.E.
- [14] Greenwood, M. (1946). The statistical study of infectious diseases. *J. Roy. Statist. Soc.* **A109**, 85–109.
- [15] Haan, L. de (1984). Slow variation and characterization of domains of attraction. In Tiago de

Oliveira, ed., *Statistical Extremes and Applications*, D. Reidel, Dordrecht, 31–48.

- [16] Haan, L. de, and Ferreira, A. (2006). *Extreme Value Theory: an Introduction*. Springer Science+Business Media, LLC, New York.
- [17] Hasofer, A., and Wang, J.Z. (1992). A test for extreme value domain of attraction. *J. Amer. Statist. Assoc.* **87**, 171–177.
- [18] Henriques-Rodrigues, L., Gomes, M.I., and Pestana, D. (2011). Statistics of extremes in athletics. *Revstat* **9**:2, 127–153.
- [19] Jenkinson, A.F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quart. J. Royal Meteorol. Society* **81**, 158–171.
- [20] Mises, R. von (1936). La distribution de la plus grande de n valeurs, *Revue Math. Union Interbalcanique* **1**, 141-160. Reprinted in *Selected Papers of Richard von Mises*, Amer. Math. Soc. **2** (1964), 271–294.
- [21] Neves, C., and Fraga Alves, M.I. (2007). Semi-parametric approach to Hasofer-Wang and Greenwood statistics in extremes. *Test* **16**, 297–313.
- [22] Neves, C., and Fraga Alves, M.I. (2008). Testing extreme value conditions — an overview and recent approaches. *Revstat* **6**:1, 83–100.
- [23] Neves, C., Picek, J., and Fraga Alves, M.I. (2006). The contribution of the maximum to the sum of excesses for testing max-domains of attraction. *J. Statist. Planning and Inference* **136** (4), 1281–1301.
- [24] Robinson, M.E., and Tawn, J. (1995). Statistics for exceptional athletic records. *Applied Statistics* **44**, 499–511.
- [25] Smith, R.L. (1988). Forecasting records by maximum likelihood. *J. American Statistical Association* **83**, 331–338.



Tese de doutoramento: Métodos Estatísticos de Screening e Classificação Supervisionada

(Boletim SPE primavera de 2011, p. 98)

Modelação flexível do problema de triagem

Sandra Ramos, *sfr@isep.ipp.pt*

Departamento de Matemática do Instituto Superior de Engenharia do Porto

e

Centro de Estatística e Aplicações da Universidade de Lisboa

Caros colegas e amigos,

Em 22 de Dezembro de 2010 concluí o meu doutoramento em Estatística e Investigação Operacional (especialidade em Probabilidade e Estatística) com uma tese sobre métodos de triagem bayesianos em classificação supervisionada. Uma das contribuições do trabalho desenvolvido foi um método de triagem semi-paramétrico baseado nos mais atuais métodos não paramétricos bayesianos. A perspetiva totalmente não paramétrica, num contexto bayesiano, do problema de triagem foi já desenvolvida no período pós-doc, e é sobre esta perspetiva que aqui faço algumas considerações.

Introdução

Os procedimentos de triagem são atualmente usados em vários contextos como a medicina, psicologia, educação, ambiente e controlo de qualidade. Estes procedimentos têm como objetivo a retenção de indivíduos da população de modo a que, para os indivíduos retidos, a probabilidade de uma determinada característica estar presente (a resposta Y pertencer a uma região conhecida C_Y) exceda um determinado valor pré-especificado δ . Os indivíduos que apresentam a característica são rotulados como *sucesso*. Em várias situações Y é de difícil obtenção direta, devendo ser observada apenas quando o indivíduo tem grande probabilidade de ser classificado como um *sucesso*. A avaliação indireta dessa probabilidade pode ser feita recorrendo à observação de um vetor d -dimensional ($d \geq 1$) X correlacionado com Y e de mais fácil observação, de modo a reter indivíduos uma alta probabilidade de serem *sucesso*. Assim, o procedimento de triagem envolve a construção de uma região de especificação C_X no espaço d -dimensional, de modo a que um indivíduo futuro com um vetor de características em C_X tenha maior probabilidade de ser identificado como um *sucesso*.

No contexto preditivo bayesiano a obtenção da região C_X é baseada num critério ótimo assente na maximização de $P(Y \in C_Y | X \in C_X; \mathcal{D})$, restringida à classe das regiões C_X , com probabilidade preditiva de triagem α . Em Turkman e Amaral Turkman (1989) é apresentado um critério de triagem ótimo de onde resulta a região:

$$C_X = \{\mathbf{x} \in \mathbb{R}^d : P(Y \in C_Y | \mathbf{x}; \mathcal{D}) > k\},$$

onde k é tal que $\alpha = P(X \in C_X | \mathcal{D})$ e $\mathcal{D} = \{(y_1, \mathbf{x}_1), \dots, (y_n, \mathbf{x}_n)\}$ é o conjunto de dados.

É prática corrente, tanto em contexto clássico como em contexto bayesiano, utilizar modelos paramétricos na construção de C_X . Porém, em muitas situações práticas os modelos paramétricos não conseguem descrever o processo que gera as observações, justificando-se assim a necessidade de relaxar o pressuposto paramétrico. Apresenta-se aqui um método de triagem bayesiano que não especifica qualquer modelo paramétrico, sendo a solução obtida por aplicação dos recentes métodos bayesianos não paramétricos de estimação de densidades. Esta solução tornou-se viável devido às implicações que os avanços computacionais da última década tiveram no desenvolvimento dos métodos de Monte Carlo por Cadeias de Markov (MCMC) usados na implementação de métodos bayesianos não paramétricos.

Método de triagem não paramétrico

Considerando a decomposição $[X, Y] = [Y|X][X]$ na modelação do par aleatório (X, Y) , a obtenção da solução não paramétrica bayesiana do problema de triagem e das estimativas das características operacionais (CO) (probabilidades preditivas que interessam ter em consideração no problema de triagem; veja-se Turkman e Amaral Turkman (1989) e Ramos e outros (2010) para uma descrição destas medidas) resume-se à estimação de $P(Y \in C_Y | \mathbf{x}; \mathcal{D})$ e $P(X \in C_X | \mathcal{D})$. Por sua vez, a obtenção dessas estimativas exige, por aplicação de métodos não paramétricos de estimação de densidades, a estimação das funções densidade de probabilidade subjacentes. Neste problema, a distribuição dos dados é desconhecida (vista como um “parâmetro” aleatório), procurando-se uma classe geral de modelos $\{F: F \in \mathcal{F}\}$, o que requer a especificação de distribuições *a priori* sobre o espaço de todas as medidas de probabilidade \mathcal{F} .

Dos vários métodos bayesianos não paramétricos descritos na literatura (veja-se, por exemplo, Müller e Quintana (2014)), o processo de Dirichlet (DP) é o mais amplamente usado devido à intensa atividade de investigação de que tem sido alvo.

O processo de Dirichlet

Ferguson (1973) introduziu o processo de Dirichlet (DP) como uma distribuição *a priori* sobre \mathcal{F} . Um processo de Dirichlet é definido por uma distribuição paramétrica de localização (ou centro) F_0 e por um parâmetro de precisão $\alpha > 0$. Diz-se então que F é distribuída segundo um processo de Dirichlet se, para qualquer partição mensurável (B_1, B_2, \dots, B_J) do espaço amostral,

$$(F_0(B_1), F_0(B_2), \dots, F_0(B_J)) \sim \text{Dirichlet}(\alpha F_0(B_1), \alpha F_0(B_2), \dots, \alpha F_0(B_J)).$$

Note-se que $E(F(\cdot)) = F_0$ e $\text{Var}(F(\cdot)) = [1 - F_0(\cdot)]F_0(\cdot)/(\alpha + 1)$. Fica pois claro que F_0 pode ser interpretada como a distribuição média do processo e que α determina o quanto F se desvia de F_0 .

Para uma amostra aleatória é comum usar-se a representação:

$$\begin{aligned} x_1, x_2, \dots, x_n &\sim F \\ F | \alpha, F_0 &\sim DP(\alpha, F_0) \end{aligned}$$

A representação quebra-vara do processo de Dirichlet tem sido muito explorada no âmbito do desenvolvimento de algoritmos de Monte Carlo por Cadeias de Markov (MCMC) eficientes. De acordo com Sethuraman (Sethuraman, 1994), se $F \sim DP(\alpha, F_0)$, então, com probabilidade um

$$F(\cdot) = \sum_{l=1}^{\infty} w_l \delta_{\tau_l}(\cdot),$$

com $w_1 = z_1$ e $w_l = z_l \prod_{i < l} (1 - z_i)$, $l = 2, 3, \dots$ onde $z_l \sim \text{Beta}(1, \alpha)$, $\alpha > 0$. $\tau_l \sim F_0$ para $l = 1, 2, \dots$.

O modelo estatístico

Sejam $\{y_i, x_i\}_{i=1}^n$ os dados onde Y e $X \in \mathbb{R}$. Admite-se a decomposição $[Y|X][X]$ na modelação de (X, Y) e que $Y_i | x_i \sim f_{y|x}(\cdot | x_i)$ e $X_i \sim f_x(\cdot)$ com $f_{y|x}(\cdot | x_i)$ e $f_x(\cdot)$ desconhecidas.

Para estimar a distribuição de X ajustamos um modelo de mistura por processos de Dirichlet de distribuições normais cuja representação hierárquica é

$$x_i \sim \int \phi(x_i | \mu, \sigma^2) dG(\mu, \sigma^2), i = 1, \dots, n$$

$$G | \alpha, F_0 \sim DP(\alpha, F_0)$$

$$\alpha \sim p(\alpha)$$

com $F_0 \equiv \phi(\mu | m, V) GI(\sigma^2 | a, b)$, onde $\phi(\cdot | \mu, \sigma^2)$ é a função densidade de probabilidade de uma variável normal de média μ e variância σ^2 , GI a distribuição gama invertida e m, V, a e b os hiperparâmetros aos quais são atribuídas distribuições a priori (ver Jara e outros (2011) para mais detalhes).

O ajustamento de um modelo de mistura por processos de Dirichlet em vez de um simples processo de Dirichlet deve-se à natureza discreta deste processo, o que inviabiliza a sua utilização na modelação de dados de natureza contínua. Note-se que uma mistura por processos de Dirichlet não é mais do que uma mistura entre uma distribuição contínua e uma medida de probabilidade aleatória que segue um processo de Dirichlet.

Para estimar a distribuição condicional desconhecida necessária para a adequada construção da região de especificação ótima C_X , ajustamos uma mistura por processo de Dirichlet dependente (DDP) de modelos de regressão gaussianos:

$$f_{y|x}(\cdot | x) = \int \phi(\cdot | x\beta, \sigma^2) dG(\beta, \vartheta^2),$$

$$G | \alpha, F_0 \sim DP(\alpha, F_0),$$

$$\alpha \sim p(\alpha)$$

com $F_0 \equiv \phi(\beta | \mu, \nu^2) \Gamma(\vartheta^{-2} | \tau_1/2, \tau_2/2)$, onde $\phi(\cdot | \mu, \sigma^2)$ é a função densidade de probabilidade de uma variável normal de média μ e variância σ^2 , Γ a distribuição gama e μ, ν^2, τ_1 e τ_2 os hiperparâmetros aos quais são atribuídas distribuições a priori (ver Jara e outros (2011) para mais detalhes).

Implementação computacional

Na execução de inferências no modelo de mistura por processos de Dirichlet consideraram-se métodos computacionais baseados na representação quebra-vara do processo de Dirichlet. As funções *LDDPdensity* e *DPdensity* da biblioteca *DPpackage* do R (Jara, 2007; Jara et al., 2011) estimam, com as devidas adaptações, as densidades desconhecidas de interesse ao problema de triagem.

Dado que as inferências são baseadas em iterações de algoritmos MCMC, houve a necessidade de desenvolver um procedimento para aproximar C_X . Assim, considerou-se uma discretização, suficiente fina, do suporte de X , $S = (x'_1, x'_2, \dots, x'_m)$. Para cada x'_j de $S, j = 1, \dots, m$, obtêm-se T iterações do algoritmo MCMC, $\{(y^t | x'_j), t = 1, \dots, T\}$, sendo $\hat{P}(Y \in C_Y | x'_j; \mathcal{D}) = N/T$, onde N representa o número de $y^t | x'_j$ que pertencem a C_Y . Finalmente

$$\hat{C}_X = \{x' \in S: \hat{P}(Y \in C_Y | x'; \mathcal{D}) > k\}.$$

Estudo de simulação

Com o objetivo de ilustrar o método apresentado, recorreu-se a um pequeno estudo de simulação com dados gerados a partir de distribuições gaussianas e não gaussianas. O desempenho do método proposto foi comparado com o desempenho do procedimento de triagem paramétrico onde se considera que o par (X, Y) tem distribuição normal bivariada. No estudo de simulação os valores dos hiperparâmetros foram fixados de forma a se obterem distribuições *a priori* vagas. Em todos os casos,

foram consideradas 2000 iterações do algoritmo MCMC, já livres das 2000 iterações do período de aquecimento.

Observou-se que, quando o par aleatório (X, Y) é bem modelado por uma distribuição normal bivariada, os dois métodos apresentam resultados muito semelhantes, sendo a região de especificação ótima igual a $\hat{C}_X = (-\infty, 0.47]$ no método não paramétrico e igual a $\hat{C}_X = (-\infty, 0.48]$ no método de triagem gaussiano paramétrico. Nos casos em que os dados não são provenientes de distribuições simétricas, os métodos paramétrico e não paramétrico diferem substancialmente, sendo o método de triagem não paramétrico o que conduz a regiões de triagem mais pequenas. É também neste método que se obtêm os melhores valores das CO.

Referências

- Boys, R.J. (1992). On a Kernel Approach to a Screening Problem. *J. R. Statist. Soc. B*, Vol.54, No.1, pp. 157–169.
- Jara, A. (2007). Applied Bayesian non- and semi-parametric inference using DPpackage. *Rnews*, pp. 17-26.
- Jara, A., Hanson, T.E., Quintana, F.A., Müller, P. e Rosner, G.L. (2011). DPpackage: Bayesian semi- and nonparametric modeling in R. *Journal of Statistical Software*, 40, pp. 1-30.
- Ramos, S., Amaral Turkman, M.A. e Antunes, M. (2010). Bayesian classification for bivariate normal gene expression. *Computational Statist and Data Analysis*, Vol.54, No.8 (2010), pp. 2012–2020.
- Turkman, K.F. e Amaral Turkman, M.A. (1989). Optimal screening methods. *J. R. Statist. Soc. B*, Vol.51, No.2, pp. 287–295.
- Müller, P. e Quintana, F. A. (2004). Nonparametric Bayesian data analysis. *Statistical Science*, 19, pp. 95–110.
- Ferguson, T.S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1, pp. 209-230.
- Sethuraman, J. (1994). A constructive definition of the Dirichlet process prior. *Statistica Sinica*, 19, pp. 639-650.

Saudações acadêmicas,
Sandra Ramos



• Artigos Científicos Publicados

- Araújo, D., Davids, K., Diniz, A., Rocha, L., Santos, J.C., Dias, G., Fernandes, O. (2015). Ecological dynamics of continuous and categorical decision-making: The regatta start in sailing. *European Journal of Sport Science*, 15 (3), 195-202.
- Filipe, P.A.; Braumann, C.A.; Carlos, C. (2014). Profit optimization for cattle growing in a randomly fluctuating environment. *Optimization: A Journal of Mathematical Programming and Operations Research*, published online on October 31, <http://dx.doi.org/10.1080/02331934.2014.974598>
- Ferreira, A. and de Haan, L. (2015). On the block maxima method in extreme value theory: PWM estimators. *Ann. Statist.* Volume 43, Number 1, 276-298.
- Ferreira, A. (2015). Spatial aggregation and high quantile estimation applied to extreme precipitation. *Statistics and Its Interface (SII)* Vol. 8.1. 33-43.

• Tese de Mestrado

Título: Análise de Sobrevivência com o R

Autora: Alexandra Isabel Monteiro Borges, alexandra.m.borges@gmail.com

Orientadora: Ana Maria Abreu

• Livros

Título: Estatística: A Ciência da Incerteza.

Atas do XXI Congresso da Sociedade Portuguesa de Estatística

Editores: Isabel Pereira, Adelaide Freitas, Manuel Scotto, Maria Eduarda Silva, Carlos Daniel Paulino

Ano: 2014. Edições SPE. ISBN: 978-972-8890-35-3

Título: New Advances in Statistical Modeling and Applications

Editores: Pacheco, A., Santos, R., Oliveira, R. M. e Paulino, C. D.

Ano: 2014. Springer. ISBN 978-3-319-05322-6.

Título: Políticas de Amostragem em Controlo Estatístico da Qualidade

Autor: Manuel do Carmo, *manuel.carmo@europaia.pt*

Orientadores: Paulo Infante e Jorge M. Mendes

Na minha tese apresentam-se e estudam-se, de uma forma crítica, dois novos métodos de amostragem adaptativa e uma nova medida de desempenho de métodos de amostragem, no contexto do controlo estatístico da qualidade.

Considerando como base uma carta de controlo para a média do tipo *Shewhart*, obtenho as principais propriedades estatísticas e realizo estudos comparativos, em termos do desempenho estatístico, com alguns dos métodos mais referenciados na literatura.

Começo por apresentar um método adaptativo de amostragem no qual os intervalos entre as amostras são obtidos com base na função densidade da distribuição de *Laplace* reduzida. Este método revela-se, particularmente, eficiente na deteção de moderadas e grandes alterações da média, pouco sensível à limitação do menor intervalo de amostragem e robusto face a diferentes situações consideradas para a não normalidade da característica da qualidade. Em determinadas situações, este método é sempre mais eficiente do que o método com intervalos de amostragem adaptativos, dimensões amostrais fixas e coeficientes dos limites de controlo fixos.

Com base no método de amostragem definido no ponto anterior e num método no qual os intervalos de amostragem são definidos antes do início do controlo do processo com base na taxa cumulativa de risco do sistema, apresento novo método que combina o método de intervalos predefinidos com o método de intervalos adaptativos. Neste método, os instantes de amostragem são definidos pela média ponderada dos instantes dos dois métodos, atribuindo-se maior peso ao método adaptativo para alterações moderadas (onde o método predefinido é menos eficaz) e maior peso ao método predefinido nos restantes casos (onde o método adaptativo é menos eficaz). Desta forma, os instantes de amostragem, inicialmente calendarizados de acordo com as expectativas de ocorrência de uma alteração tomando como base a distribuição do tempo de vida do sistema, são adaptados em função do valor da estatística amostral calculada no instante anterior. Este método é sempre mais eficiente do que o método periódico clássico, o que não acontece com nenhum outro esquema adaptativo, e do que o método de amostragem VSI (*variable sampling intervals*) para alguns pares de amostragem, posicionando-se como uma forte alternativa aos procedimentos de amostragem encontrados na literatura.

Por fim, apresento uma nova medida de desempenho de métodos de amostragem. Considerando que dois métodos em comparação têm o mesmo tempo médio de mau funcionamento, o desempenho dos métodos é comparado através do número médio de amostras recolhidas sob controlo. Tendo em conta o tempo de vida do sistema, com diferentes taxas de risco, esta medida mostra-se robusta e permite, num contexto económico, um melhor controlo de custos por unidade de tempo.

Manuel do Carmo

Edições SPE - Minicursos

- Título:** Análise de Valores Extremos:
Uma Introdução
Autoras: M. Ivette Gomes, M. Isabel Fraga
Alves e Cláudia 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



PRÉMIOS “ESTATÍSTICO JÚNIOR 2015”

Está aberto, até 1 de Junho de 2015, o concurso para atribuição de prémios “Estatístico Júnior 2015”, de acordo com o seguinte regulamento:

1. A atribuição de prémios “Estatístico Júnior 2015” é promovida pela Sociedade Portuguesa de Estatística (SPE), com o apoio da Porto Editora, e tem como objectivo estimular e desenvolver o interesse dos alunos do ensino básico e secundário pelas áreas da Probabilidade e Estatística.

2. Os candidatos aos prémios “Estatístico Júnior 2015” devem ser alunos do 3.º Ciclo do Ensino Básico, do Ensino Secundário, dos Cursos de Educação e Formação (CEF), ou dos Cursos de Educação e Formação de Adultos (EFA), no ano letivo 2014-2015.

3. As candidaturas podem ser individuais ou em **grupo com um máximo de 3 alunos**. Do grupo pode ainda fazer parte um professor, da categoria onde o prémio se insere, ao qual caberá o papel de orientador.

4. Os candidatos devem apresentar um trabalho cuja temática deve estar relacionada com a teoria da Probabilidade e/ou Estatística.

5. O trabalho deverá ser constituído por um texto escrito em Português com um máximo de 10 páginas A4 dactilografadas e um poster formato A2 que resuma os principais aspectos do trabalho. O trabalho (poster e texto escrito) deverá ser **enviado impresso em papel para efeitos da avaliação**.

6. Poderão ser atribuídos prémios “Estatístico Júnior 2015” a sete trabalhos: aos três primeiros classificados de entre os trabalhos candidatos do 3.º Ciclo do Ensino Básico, aos três primeiros classificados de entre os trabalhos candidatos do Ensino Secundário e um primeiro classificado de entre os trabalhos candidatos dos Cursos CEF-EFA. Os prémios são constituídos por produtos pedagógicos editados pela Porto Editora (à exceção de manuais escolares) no valor de 600 euros, 300 euros e 200 euros, a atribuir, respetivamente, aos grupos cujos trabalhos sejam classificados em 1.º, 2.º e 3.º lugares, para as categorias Ensino Básico e Secundário, e 600 euros para a categoria Cursos CEF-EFA.

7. Ao professor orientador do trabalho classificado em 1º lugar, em cada categoria, é ainda atribuída uma anuidade grátis como sócio da SPE, ajudas de custo para participação na Sessão de Entrega do Prémio e produtos pedagógicos editados pela Porto Editora (à exceção de manuais escolares) no valor de 500 euros.

8. Aos grupos proponentes dos trabalhos classificados em 1º lugar será também oferecida uma ampliação do correspondente poster que será exposto na Sessão de Entrega do Prémio.

9. O boletim de candidatura, acompanhado do trabalho concorrente, deverá ser dirigido ao Presidente da SPE para a morada abaixo indicada. O carimbo do correio validará a data de entrega.

Sociedade Portuguesa de Estatística – Bloco C6, Piso 4 – Campo Grande – 1749-016 Lisboa

O boletim de candidatura e este regulamento podem ser obtidos em

<http://www.spestatistica.pt/BoletimCandidaturaPEJ14.pdf>
<http://www.spestatistica.pt/RegulamentoPEJ14.pdf>

10. A admissibilidade e apreciação dos trabalhos submetidos a concurso é da competência de um júri, cuja constituição e nomeação será da responsabilidade da Direção da SPE.

11. O júri é soberano nas decisões, não havendo lugar a impugnação ou recurso.

12. A atribuição dos prémios “Estatístico Júnior 2015” será anunciada logo que conhecida a decisão do júri e a sua entrega formal será realizada numa Sessão expressamente dedicada a essa entrega.

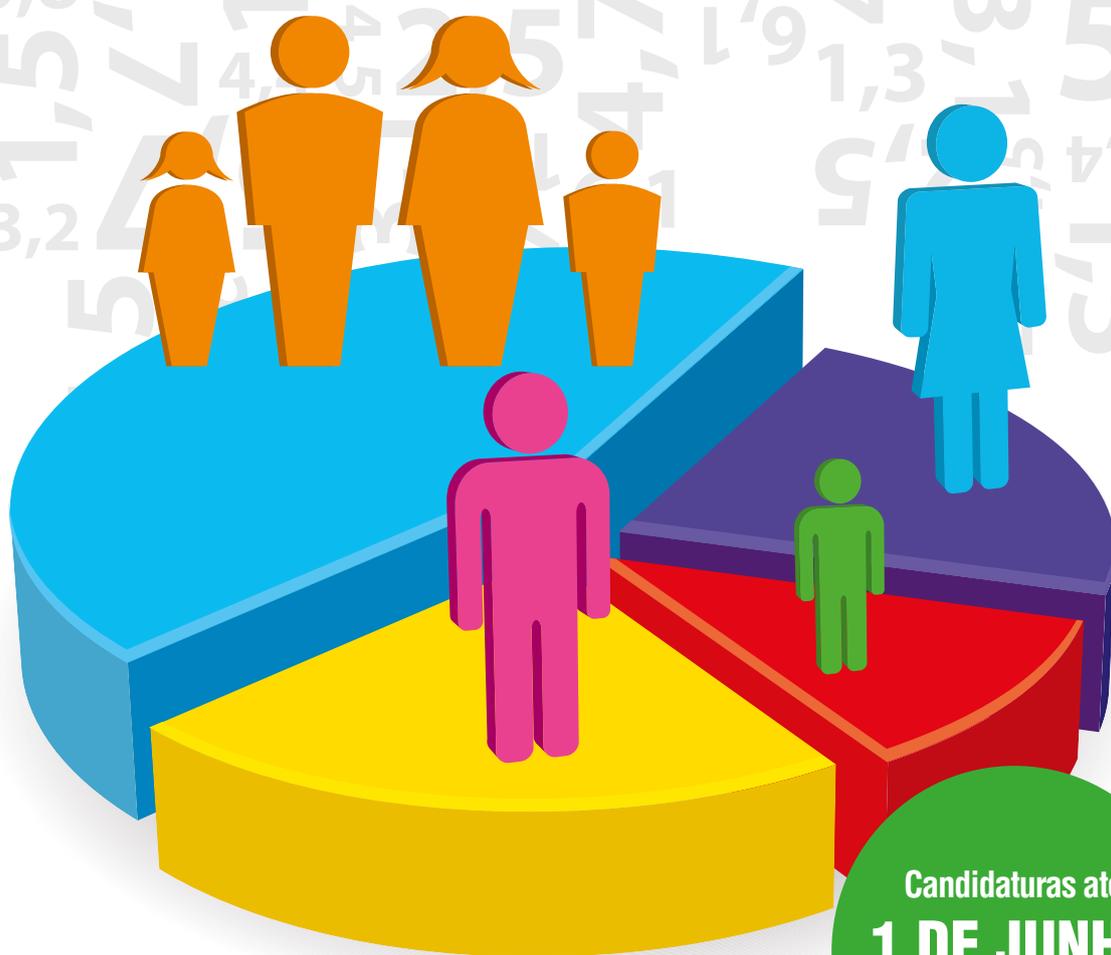
13. Os prémios “Estatístico Júnior 2015” poderão não ser atribuídos.

Apoio  Porto
Editora



SOCIEDADE PORTUGUESA
DE ESTATÍSTICA

PRÉMIO ESTATÍSTICO JÚNIOR 2015



Candidaturas até
**1 DE JUNHO
DE 2015**

CONTACTOS

Sociedade Portuguesa de Estatística
Bloco C6, Piso 4 – Campo Grande
1749-016 Lisboa
Telef. 21 750 01 20

www.spestatistica.pt
spe@fc.ul.pt

Com o apoio:

 Porto
Editora

Índice

Editorial	2
Mensagem da Presidente	4
Notícias	5
<i>Enigmística</i>	16
<i>SPE e a Comunidade</i>	
Dando vida às relações estatísticas: Uma Metodologia Estocástico-Dinâmica aplicada à Ecologia <i>João Alexandre Cabral</i>	17
<i>Estatística no Desporto</i>	
Séries Temporais no Desporto <i>Ana Diniz e Paula Marta Bruno</i>	25
Investigação em Ciências do Desporto: dos testes de hipótese nula à necessidade de interpretações com significância prática e/ou clínica <i>Rui Marcelino e Jaime Sampaio</i>	28
Estatística de Extremos em Desporto <i>Lígia H. Rodrigues, M. Ivette Gomes e Dinis Pestana</i>	36
<i>Pós-Doc</i>	
<i>Sandra Ramos</i>	45
<i>Ciência Estatística</i>	
<i>Artigos Científicos Publicados</i>	49
<i>Tese de Mestrado</i>	49
<i>Livros</i>	49
<i>Tese de Doutoramento</i>	50
Edições SPE – Minicursos	51
Prémios “Estatístico Júnior”	52