



SOCIEDADE PORTUGUESA
DE ESTATÍSTICA

PROGRAMA E LIVRO DE RESUMOS

Edições SPE

Ficha Técnica:

Programa e Livro de Resumos

Isabel Pereira, Adelaide Freitas, Cláudia Neves, Eugénio Rocha,
Manuel Scotto, Maria Eduarda Silva, Nélia Silva

Editora: Sociedade Portuguesa de Estatística

Capa: Carina Sousa

Impressão: Instituto Nacional de Estatística

Tiragem: 200 exemplares

ISBN: 978-972-8890-31-5

Depósito Legal: 366448/13

COMISSÃO ORGANIZADORA:

Isabel Pereira, CIDMA & Universidade de Aveiro

Adelaide Freitas, CIDMA & Universidade de Aveiro

Cláudia Neves, CEAUL & Universidade de Aveiro

Eugénio Rocha, CIDMA & Universidade de Aveiro

Manuel Scotto, CIDMA & Universidade de Aveiro

Maria Eduarda Silva, CIDMA & Universidade do Porto

Nélia Silva, CIDMA & Universidade de Aveiro

COMISSÃO CIENTÍFICA:

Carlos Daniel Paulino, Instituto Superior Técnico, Universidade de Lisboa

Esmeralda Gonçalves, Faculdade de Ciências, Universidade de Coimbra

Helena Ferreira, Universidade da Beira Interior

Denisa Mendonça, Instituto de Ciências Biomédicas Abel Salazar

Isabel Pereira, Universidade de Aveiro

Sexta-feira, 29 de novembro de 2013

09:00	Registo e entrega de documentação		
09:30	Mini-Curso Análise de Valores Extremos: uma introdução M. Ivette Gomes (CEAUL) Presidente: Cláudia Neves Sala: Santiago		
11:00	Pausa para café		
11:30	Mini-Curso (<i>cont.</i>) Sala: Santiago		
13:00	Almoço		
14:30	Mini-Curso (<i>cont.</i>) Sala: Santiago		
16:30	Pausa para café		
17:00	Sessão de abertura Sala: S. Pedro 2		
17:15	Sessão Plenária Visualising information: data, models and graphics Adrien Bowman (The University of Glasgow) Presidente: Maria Eduarda Silva Sala: S. Pedro 2		
18:15	Comunicações Oraís		
	Sessão: Estatísticas na Educação Presidente: Adelaide Freitas Sala: S. Pedro 1	Sessão: Inferência Não Paramétrica Presidente: Júlia Teles Sala: Santiago	Sessão: Séries Temporais 1 Presidente: Magda Monteiro Sala: S. João
19:15	Receção de boas vindas Hotel Meliá		

Sábado, 30 de novembro de 2013

09:00	Sessões Temáticas		
	Estatística Bayesiana e Aplicações Presidente: Valeska Andreozzi Sala: S. Pedro 1	Extremos Presidente: Alexandra Ramos Sala: Santiago	Modelos Estocásticos Presidente: Fátima Ferreira Sala: S. João
10:40	Pausa para café e Sessão de Posters- Posters 1 Sala: S. Tomás		
11:30	Sessão Plenária Um regresso aos fundamentos da Estatística Maria Antónia Amaral Turkman (CEAUL, FCUL) Presidente: Carlos Daniel Paulino Sala: S. Pedro 2		
12:30	Almoço		
14:00	Explorística		
14:20	Entrega de Prémios Estatístico Júnior Sala: S. Pedro 2		
15:00	Programa Social: Passeio		
20:00	Programa Social: Jantar do Congresso		

Domingo, 1 de dezembro de 2013

09:00	Sessões Temáticas		
	Séries Temporais e Aplicações Presidente: K. F. Turkman Sala: S.Pedro 1	Dados Categorizados Presidente: Nuno Sepúlveda Sala: Santiago	Estatística Multivariada Presidente: Conceição Amado Sala: S.João
10:40	Pausa para café		
11:00	Sessão Plenária Modelos estocásticos de crescimento populacional Carlos Braumann (CIMA, Universidade de Évora) Presidente: Esmeralda Gonçalves Sala: S.Pedro 2		

(Continua)

12:00	Comunicações Oraís			
	Estatística Bayesiana Presidente: Marília Antunes Sala: S. Pedro 1	Controlo de Qualidade Presidente: Fernanda Figueiredo Sala: Santiago	Modelos Espaço -Temporais Presidente: Giovani Silva Sala: S. João	Inferência Estatística I Presidente: Fernando Rosado Sala: S. Pedro 2
13:00	Almoço			
14:15	Sessão Plenária Dimensionality: curse or blessing? An empirical assessment when estimating factors in DFM Esther Ruiz (Universidad de Carlos III de Madrid) Presidente: Manuel Scotto Sala: S. Pedro 2			
15:15	Comunicações Oraís			
	Estatística Multivariada Presidente: Jorge Cadima Sala: S. Pedro 1	Estatística e Aplicações 1 Presidente: Carla Henriques Sala: Santiago	Inferência Estatística 2 Presidente: Emilia Athayde Sala: S. João	Séries Temporais 2 Presidente: Nazaré Mendes-Lopes Sala: S. Pedro 2
16:35	Pausa para café e Sessão de Posters– Posters 2 Sala: S. Tomás			
17:30	Sessão em memória de Daniel Muller Sala: S. Pedro 2			
17:45	Sessão Comemorativa do Ano Internacional da Estatística Sala: S. Pedro 2			
19:15	Forum de discussão: Periodicidade dos Congressos da SPE Sala: S. Pedro 2			

Segunda-feira, 2 de dezembro de 2013

09:00	Sessões Temáticas			
	Risco e Atuariado Presidente: Alfredo Egídio dos Reis Sala: S. Pedro 1	Estatística e Genómica Presidente: Lisete Sousa Sala: Santiago	Estatística na Ecologia e Ambiente Presidente: Raquel Menezes Sala: S. João	
10:40	Pausa para café			
11:00	Comunicações Orais			
	Estatística e Aplicações 2 Presidente: Cecília Azevedo Sala: S. Pedro 1	Bioestatística Presidente: Paulo Infante Sala: Santiago	Estatística na Indústria Presidente: Alda Carvalho Sala: S. João	Processos Estocásticos Presidente: António Pacheco Sala: S. Pedro 2
12:20	Análise Fatorial Presidente: Russell Alpizar-Jara Sala: S. Pedro 1	Extremos Presidente: Ivette Gomes Sala: Santiago	Análise de Sobrevivência Presidente: Ana Amorim Sala: S. João	
13:20	Sessão de Encerramento Sala: S. Pedro 2			

Sessões Plenárias



Um Regresso aos Fundamentos da Estatística

Maria Antónia AMARAL TURKMAN – *Centro de Estatística e Aplicações da Universidade de Lisboa*

maturkman@fc.ul.pt

Resumo

Celebra-se durante este ano de 2013 o Ano Internacional de Estatística. A escolha de 2013 para esta celebração (Paulino 2013) não é com certeza alheia ao facto de, durante este ano, se comemorarem os 300 anos da publicação da obra *Ars Conjectandi* de Jacob Bernoulli e os 250 anos da publicação do trabalho de Thomas Bayes sobre probabilidade inversa, *An Essay towards Solving a Problem in the Doctrine of Chances*, peça fundamental no desenvolvimento da metodologia bayesiana. Como parte da celebração desta última efeméride, pela Royal Statistical Society, Tony O’ Hagan entrevistou Dennis Lindley, precursor e acérrimo defensor da metodologia bayesiana que, acidentalmente, também completou 90 anos de idade em Julho de 2013. Esta entrevista, recentemente colocada on line (<http://www.youtube.com/watch?v=cgclGi8yEu4>) na página da Royal Statistical Society, de imediato suscitou, via e-mail, pela sua natureza provocadora e controversa, uma interessante e apaixonante troca de impressões e opiniões, por parte de vários probabilistas e estatísticos de renome sobre aspectos fundamentais da estatística focados naquela entrevista.

Nesta apresentação passarei vários excertos da entrevista de Tony O’Hagan a Dennis Lindley, revisitarei aspectos fundamentais aí focados, nomeadamente princípios da verosimilhança, suficiência e condicionalidade e suas consequências na inferência quer clássica quer bayesiana. Apresentarei vários aspectos da controvérsia que foi gerada pelas opiniões apresentadas por Dennis Lindley, assim como farei uma revisão de alguns trabalhos que estão associados a estas polémicas.

Agradecimentos

Este trabalho foi parcialmente financiado pela Fundação para a Ciência e a Tecnologia FCT, Portugal, através dos projectos - PEst-OE/MAT/UI0006/2011 e PTDC/MAT/118335/2010.

Referências

- [1] Birnbaum, A. (1962) *On the Foundations of Statistical Inference (with discussion)*, Journal of the American Statistical Association, 57: 269–326.
- [2] Durbin, J. (1970) *On Birnbaum's Theorem and the Relation between Sufficiency, Conditionality and Likelihood*, Journal of the American Statistical Association, 65: 395–398.
- [3] Evans, M.J., Fraser, D. and Monette, G. (1986) *On Principles and Arguments to Likelihood*, The Canadian Journal of Statistics, 14: 181–194.
- [4] Evans, Michael (2013) *What does the proof of Birnbaum's theorem prove?*, arXiv:1302.5468v1
- [5] Lindley, D. V. (2000) *The Philosophy of Statistics*, The Statistician (JRSS D), 49: 293–337.
- [6] Mayo, B. (2010) *An Error in the Argument from Conditionality and Sufficiency to the Likelihood Principle*, in *Error and Inference: Recent Exchanges on Experimental Reasoning, Reliability and the Objectivity and Rationality of Science* (D Mayo and A. Spanos eds.), Cambridge: Cambridge University Press: 305–14.
- [7] Mayo, Deborah (2013) *On the Birnbaum Argument for the Strong Likelihood Principle*, arXiv:1302.7021v2
- [8] Paulino, D. (2013) *O Ano Internacional da Estatística*, Pagela dos CTT.
- [9] Robins, J., and Wasserman, L. (2000) *Conditioning, Likelihood, and Coherence: A Review of Some Foundational Concepts*, Journal of the American Statistical Association, 95: 1340–6
- [10] Savage, L. (1962), *Discussion on Birnbaum (1962)*, Journal of the American Statistical Association, 57: 307–8.

Visualising information: data, models and graphics

Adrian BOWMAN – *The University of Glasgow*
adrian.bowman@glasgow.ac.uk

Resumo

Statistical concepts and methods have made enormous contributions to understanding and practice in the scientific, industrial, social and many other arenas. One fundamental contribution is how to think about uncertainty and variation, while another is the provision of flexible and sophisticated models for observed data together with powerful and informative means of fitting these. This has led to many successes. While there will never be a shortage of challenging problems to tackle, we can justifiably celebrate the achievements of statistical thinking.

These achievements can be appreciated in depth by those who have good understanding of the concepts and methods involved in the analysis of sophisticated data structures. They are often also appreciated, in a different way but equally strongly, by those who have some difficulty in understanding the technical mechanisms by which this can be achieved but who realise the need for, and benefit of, powerful quantitative analysis. However, there is a continuing need to widen the appreciation, and the understanding, of statistical thinking at all levels, from those who are in charge of policy, to those whose scientific work involves data collection, to members of the general public.

This talk aims to reflect a little on the languages we use to explain, discuss and communicate statistical concepts, models and analysis, both within our own community and beyond it. Several types of analysis will be considered, mostly involving forms of flexible regression and focussed largely on different forms of spatiotemporal data. However, there will be a particular focus on the role of graphics, which can provide a powerful means of conceptual communication and give clear expressions of the insights provided by models, while remaining true to the issues associated with uncertainty.

The discussion will be focussed around three questions.

1. *How should we plot spatiotemporal data and models?*

This kind of data, where measurements are indexed by both space and time, is now very common. Environmental data, where geographically distributed monitors collect data over time, is one widespread example. Recent work on visualisation, such as `lattice` graphics (Sarkar, 2008) and `ggplot2` (Wickham, 2009) allow very attractive displays of two-dimensional spatial data be to constructed, using good principles of graphic design. However, with spatiotemporal data, animation is particularly useful in displaying the changing nature of two-dimensional data and models over time. Tools to do this are readily available in systems such as R, for example using the `rpanel` package (Bowman et al., 2007). Some tools to assess the suitability of simple spatial models, through examination of the variogram, will also be discussed.

2. *What graphics are good for communicating uncertainty?*

Simple devices such as shading and intensity, as used in ‘density strips’ (Jackson, 2008), can communicate uncertainty in a very intuitive, non-technical manner. The use of this kind of device will be explored in simple settings such as comparing groups and in flexible regression. A particularly tricky issue arises in displaying the uncertainty associated with surfaces and some suggestions about this will be made.

3. *How do we handle data over non-linear domains?*

We are very familiar with sophisticated graphical representations in map form, using colour, shading and a wide variety of forms of annotation. Full three-dimensional representations are particularly useful for large areas, where the curvature of the earth becomes an issue. Another example of non-linear domains arises in modelling surface shape in three dimensions, where combinations of animation and colour shading can be very effective.

These issues will be explored on a variety of datasets, including models for the dispersion of nitrates in the River Tweed and the modelling of facial shape.

Referências

- [1] Bowman, A., Crawford, E., Alexander, G. & Bowman, R.W. (2007). rpanel: Simple Interactive Controls for R Functions Using the tcltk Package. *Journal of Statistical Software*, 17(9), 1-18. URL <http://www.jstatsoft.org/v17/i09/>.
- [2] Jackson, C. H. (2008) Displaying uncertainty with shading. *The American Statistician*, 62(4), 340-347.
- [3] Sarkar, D. (2008) *Lattice: Multivariate Data Visualization with R*. Springer, New York.
- [4] Wickham, H. (2009) *ggplot2: elegant graphics for data analysis*. Springer, New York.

Modelos estocásticos de crescimento populacional

Carlos A. BRAUMANN – *Centro de Investigação em Matemática e Aplicações, Universidade de Évora*

braumann@uevora.pt

Resumo

As variações amostrais dos nascimentos e mortes (aleatoriedade demográfica) e as flutuações aleatórias das condições ambientais (aleatoriedade ambiental) afetam ambas o crescimento populacional e são habitualmente modeladas usando diferentes tipos de processos estocásticos. Os processos de ramificação ou os processos de nascimento e morte são usados para modelar a aleatoriedade demográfica, mas habitualmente assumem um ambiente constante. Pelo contrário, as equações diferenciais estocásticas são usadas para modelar a aleatoriedade ambiental mas usualmente ignoram a aleatoriedade demográfica.

Iremos examinar as semelhanças e diferenças entre estas duas abordagens no que respeita à extinção da população e ao comportamento local, usando como base de referência os modelos malthusianos (crescimento independente da densidade populacional), designadamente o processo de Galton-Watson e o processo de nascimento e morte simples, para a aleatoriedade demográfica, e o modelo de equações diferenciais estocásticas malthusiano, também conhecido por movimento browniano geométrico, para a aleatoriedade ambiental.

No entanto, o crescimento de populações naturais depende da densidade. Para os modelos de equações diferenciais estocásticas de crescimento populacional em ambiente aleatório já há estudos adequados para crescimento com dependência da densidade. Fazemos uma revisão dos principais resultados obtidos com esses modelos no que se refere à extinção e existência de densidade estacionária. Estes resultados são robustos com respeito à forma de dependência da densidade, visto usarmos modelos gerais e não modelos específicos (como o logístico) habitualmente usados na literatura. No que se refere a modelos de aleatoriedade demográfica dependentes da densidade, os estudos são muito escassos e para modelos muito específicos, pelo que nos limitamos a especular sobre resultados mais gerais.

E o que sucede se ambas as formas de aleatoriedade ocorrerem?

Dimensionality: curse or blessing? An empirical assessment when estimating factors in DFM

Esther RUIZ – *Department of Statistics, Universidad Carlos III de Madrid*
ortega@est-econ.uc3m.es

Pilar Poncela – *Departamento de Análisis Económico: Economía Cuantitativa, Universidad Autónoma de Madrid*

Resumo

Based on asymptotic arguments, several authors argue that the usual methods for estimation of factors in Dynamic Factor Models turn the curse of dimensionality into a blessing. On the contrary, there are an increasing number of authors that doubts that using an extremely large number of variables to estimate the factors always implies larger accuracy. In this paper, we check empirically how the point estimates of the underlying unobserved factors and their associated Mean Squared Errors change depending on the number of variables used in their estimation. We analyze the macroeconomic data set popularized by Stock and Watson (2012) and consider alternative procedures for factor extraction based on Principal Components and the Kalman filter. Our paper differs from previous papers in the related literature in several ways. First, we focus on the estimation of the factors rather than on prediction. Second, our comparisons are carried out by implementing the same procedures to the same data. Third, we are interested not only on point estimates but also on confidence intervals for the factors.

Sessões Temáticas



Dados Categorizados

Análise bayesiana semiparamétrica de respostas binárias com uma covariável contínua sujeita a omissão informativa

(Comunicação)

Carlos Daniel PAULINO – *Instituto Superior Técnico, Universidade de Lisboa*
dpaulino@math.ist.utl.pt

Frederico Poleto – *Instituto de Matemática e Estatística, Universidade de São Paulo*

Julio Singer – *Instituto de Matemática e Estatística, Universidade de São Paulo*

Geert Molenberghs – *I-BioStat, Universiteit Hasselt and Katholieke Universiteit Leuven, Belgium*

Resumo

Omissão em variáveis explicativas requer um modelo marginal para elas mesmo que o interesse se centre no modelo condicional das respostas dadas as covariáveis. Uma especificação incorreta de tais modelos marginais ou para o respetivo mecanismo de omissão pode conduzir a inferências enviesadas sobre os parâmetros de interesse. Em literatura já publicada usam-se para as covariáveis ou distribuições paramétricas com um mecanismo de omissão informativa (MNAR) ou modelos mais flexíveis semiparamétricos ou não paramétricos identificados com a suposição de omissão ao acaso (MAR).

Neste trabalho considera-se uma análise bayesiana de respostas binárias, combinando um modelo não paramétrico, baseado em mistura por um processo Dirichlet, para as covariáveis contínuas com um mecanismo de omissão MNAR. A ilustração é feita através de simulações e de um conjunto de dados reais sobre pacientes com suspeita de embolia pulmonar retirados de literatura médica.

Agradecimentos: Financiamento parcial de F.Poleto e J.Singer pela CAPES, FAPESP e CNPq, Brasil; C. D. Paulino pela FCT através do projeto PEST-OE/MAT/UI0006/2011; G. Molenberghs por IAP research Network P6/03 of the Belgian Government (Belgian Science Policy).

On the performance of a polytomous logistic regression model in tackling missing genotypes

(Comunicação)

Nuno SEPÚLVEDA – *London School of Hygiene and Tropical Medicine, London, United Kingdom, and CEAUL, Lisbon, Portugal*

nuno.sepulveda@lshtm.ac.uk

Alphaxard Manjurano – *London School of Hygiene and Tropical Medicine, London, United Kingdom, and National Institute for Medical Research, Mwanza Centre, Tanzania*

Chris Drakeley – *London School of Hygiene and Tropical Medicine, London, United Kingdom*

Taane Clark – *London School of Hygiene and Tropical Medicine, London, United Kingdom*

Abstract

The presence of missing genotypes is intrinsic to any genetic association analysis but not always tackled in practice. In fact, the common statistical strategy is to discard missing observations and perform the corresponding genetic association analysis using complete data only — the so-called complete case analysis. It is known that this strategy decreases statistical efficiency and power, and may introduce severe estimation bias when missing data is far from being completely at random. With the advent of the HapMap project together with the drastic drop in costs of genome-wide association studies, several statistical methods based on data imputation have been proposed [1]. In theory, these approaches take advantage of the underlying correlation structure between genetic markers (linkage disequilibrium) and the availability of a reference data, typically taken from the HapMap database, in order to replace missing genotypes by highly plausible guesses. When data under analysis comprises a set of poorly correlated genetic markers, such as in candidate gene approach studies, alternative imputation strategies may be required. Multiple imputation based on chained equations (MICE) seems a good candidate approach because data imputation can be carried out using different types of

information, not necessarily of genetic nature [2]. The main goal of the present study is to assess the performance of MICE in the study of the genetic association between the α -thalassemia locus and malaria parasite positivity in more than 7000 individuals from 24 villages in northern Tanzania. We demonstrate that the best imputation model is a polytomous logistic regression model that includes data of (i) 4 genetic markers highly correlated with the α -thalassemia locus, (ii) a set of phenotypes that might be under influence of that locus, and (iii) other putative confounding factors, such as altitude or ethnicity. By simulation we demonstrate that, when missing genotypes are generated completely at random, the genotype error rates are around 44% but the corresponding estimates for the average genotype and genetic effects are typically unbiased. More importantly, we observe little bias in the association signals. However, when missing data was assumed to come all from a specific village, the estimates of different parameters of interest may or not be biased depending on the village showing missing data. When MICE was applied to the whole data set, where one can find a mix between completely-at-random and non-random missing data patterns, the association signal for the effect of the α -thalassemia locus on malaria parasite positivity increased from 0.80 (complete case analysis, $n=2666$) to 2.70 (overall sample, $n=7048$). In turn, estimates for mean genotype and genetic effects were similar to those obtained from complete case analysis but showing smaller standard errors. Confronting these result with those from the simulation study, we argue that the association signal seems accurate and robust but the corresponding genetic effects are most likely to show some bias.

References

- [1] Marchini, J., Howie, B. (2010) *Genotype imputation for genome-wide association studies*, Nat. Rev. Genet. 11, 499–511.
- [2] Souverein, O. W., Zwinderman, A. H., Tanck, M. W. T. (2006) *Multiple Imputation of Missing Genotype Data for Unrelated Individuals*, Ann. Hum. Genet. 70, 372–381.

Análise de dados categorizados no contexto de testes diagnósticos

(Comunicação)

Ana SUBTIL – *CEMAT e Faculdade de Ciências da Universidade de Lisboa, Universidade de Lisboa*

asubtil@ihmt.unl.pt

M. Rosário de Oliveira – *CEMAT e Departamento de Matemática, Instituto Superior Técnico, Universidade de Lisboa*

P. Zea Bermudez – *CEAUL e Faculdade de Ciências da Universidade de Lisboa, Universidade de Lisboa*

Luzia Gonçalves – *CEAUL e Unidade de Saúde Pública Internacional e Bioestatística, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa*

Resumo

A análise de dados categorizados é frequente no contexto do estudo do desempenho de testes de diagnóstico. Mesmo no caso de testes de natureza quantitativa, é comum a categorização da resposta de acordo com um ou mais pontos de corte, de modo a obter categorias tais como negativo/positivo ou negativo/duvidoso/positivo. Em algumas situações, também pode ser interessante avaliar o desempenho dos testes segundo características específicas dos indivíduos (e.g., presença ou ausência de sinais clínicos), surgindo na análise outro tipo de dados categorizados.

A análise de classes latentes aplicada a dados categorizados resultantes do processo de diagnóstico laboratorial é utilizada com frequência na estimação de medidas de desempenho dos testes, tais como sensibilidades e especificidades, e na estimação de prevalências. Aproveitando a versatilidade destes modelos, têm sido propostas variantes que permitem a modelação de efeitos de factores externos, a adequação a desenhos amostrais diversos e a inclusão de eventuais dependências locais entre os testes.

As abordagens bayesianas de estimação do Modelo de Classes Latentes têm ganho importância nos últimos anos, devido às suas potencialidades em supe-

rar alguns problemas de não identificabilidade e à possibilidade de incorporar informação *a priori* sobre o problema, fornecida por peritos, não obstante a complexidade do diálogo entre os estatísticos e estes profissionais de saúde.

Neste trabalho focaremos a aplicação de Modelos de Classes Latentes a dados categorizados referentes ao diagnóstico de algumas doenças como a Malária [1], a *Dirofilariose* [2] e a *Leishmaniose* [3].

Agradecimentos

Este trabalho foi parcialmente financiado pela Fundação Nacional para a Ciência e Tecnologia (FCT), através dos projectos PEst-OE/MAT/UI0006/2011 e PTDC/MAT/118335/2010. Ana Subtil tem uma Bolsa de Doutoramento da FCT SFRH/BD/69793/2010.

Referências

- [1] Gonçalves, L., Subtil, A., Oliveira, M. R., Rosário, V., Lee, P., Shaio, M.-F. (2012). *Bayesian latent class models in malaria diagnosis*, PLoS ONE:e40630.
- [2] Gonçalves, L., Subtil, A., Brites, N., Oliveira, M. R., Alho, A. M., Meireles, J., Madeira de Carvalho, L., Belo, S. (2012). *Bayesian Latent Class Models in Veterinary and Human Epidemiology. 46th Scientific Meeting of the Italian Statistical Society, Rome, 22 June 2012.*
- [3] Encarnação, F., Gonçalves, L., Campino, L., Cristóvão, J.M. and Oliveira, M.R. (2007) Evaluation of the Accuracy of Diagnostic Tests using Latent Class Analysis. *International Journal Communications in Dependability and Quality Management*, 10, pp.33-40.

Uso de Regressão Ordinal em Análise de Custo-Efetividade e Farmacoeconomia: um Estudo de Caso

(Comunicação)

Björn VANDEWALLE – *Exigo Consultores*

bjorn.vandewalle@exigoconsultores.com

Maria João Silva – *Exigo Consultores*

João Almeida – *Exigo Consultores*

Jorge Félix – *Exigo Consultores*

Resumo

A farmacoeconomia é uma disciplina que permite confrontar o valor terapêutico e económico entre opções terapêuticas através da identificação, medição e comparação dos custos e dos resultados de saúde. Os custos são definidos como o valor monetário dos recursos consumidos. Os resultados de saúde são definidos como o efeito (positivo e negativo) sobre o estado de saúde decorrente da utilização das opções terapêuticas. Dependendo da perspectiva, estes resultados podem ter uma tradução económica ou clínica/humanística [1-2]. A análise de custo-efetividade é a técnica mais usual em análise farmacoeconómica. Nesta, as consequências são expressas em termos de resultados específicos de saúde, tal como ganhos em anos de vida ou anos de vida ajustados pela qualidade.

A análise de custo-efetividade é frequentemente baseada em modelos estatísticos estimados a partir dos resultados de eficácia e segurança de ensaios clínicos e/ou estudos observacionais. Estes modelos contemplam, muitas vezes, uma componente temporal, com o objectivo de simular a vida dos doentes em particular ou de uma população inteira desde o início da terapêutica até à morte, incluindo os estados de saúde pelos quais os doentes transitam ao longo do tempo como resultado do problema de saúde e das terapêuticas [3].

Além de criar a necessidade de extrapolação para horizontes temporais mais longos do que observados nos ensaios clínicos, este tipo de modelo também

implica a identificação e análise de todos os estados possíveis em que um doente pode encontrar-se em função do problema de saúde e das terapêuticas indicadas [3]. A análise de dados categorizados é uma das possibilidades metodológicas mais consistentes para um modelo em que a saúde de um doente pode assumir um e só um estado de um número limitado e fixo de estados possíveis.

Nesta comunicação pretende-se demonstrar o valor da análise de dados categorizados na análise de custo-efetividade e na farmacoeconomia através de um estudo de caso relativamente à utilização da substância activa ivabradina no tratamento da insuficiência cardíaca em Portugal, utilizando dados individuais dos doentes do estudo SHIFT (*Systolic Heart failure treatment with the I_f inhibitor ivabradine Trial*) [4].

Referências

- [1] Bootman, J.L., Townsend, R.J., McGhan, W.F. (2004) Principles of Pharmacoeconomics, Harvey Whitney Books.
- [2] DiPiro, J., Talbert, R.L., Yee, G., Matzke, G., Wells, B., Posey, L.M. (2004) Pharmacotherapy: A Pathophysiologic Approach, McGraw–Hill Medical.
- [3] Norman, R., Spencer, A.E., Feder, G. (2007) *Cost-Effectiveness Analysis: What you always wanted to know but were afraid to ask*, Family Violence Prevention and Health Practice.
- [4] Swedberg, K., Komajda, M., Böhm, M., Borer, J.S., Ford, I., Dubost-Brama, A., Lerebours, G., Tavazzi, L. (2010) *Ivabradine and outcomes in chronic heart failure (SHIFT): a randomised placebo-controlled study*, Lancet. 376, 875-85.

Estatística Bayesiana e Aplicações

Modelação da abundância de *Culex pipiens* em Portugal

(Comunicação)

Marília ANTUNES ⁽¹⁾ – *CEAUL e DEIO-FCUL, Universidade de Lisboa*
marilia.antunes@fc.ul.pt

Patrícia de Zea Bermudez ⁽²⁾ – *CEAUL e DEIO-FCUL, Universidade de Lisboa*

Maria da Conceição Proença ⁽¹⁾ – *LOLS e DF-FCUL, Universidade de Lisboa*
Maria Teresa Rebelo ⁽¹⁾ – *CESAM/FCUL e DBA-FCUL, Universidade de Lisboa*

Maria João Alves, Hugo Osório e equipa REVIVE ⁽¹⁾ – *Centro de Estudos de Vectores e Doenças Infecciosas / INSA e ARS's*

Resumo

Os arbovírus são vírus transmitidos por artrópodes, de que o *Culex pipiens* é a espécie mais abundante. Durante a refeição de sangue, a fêmea inocula no hospedeiro vírus que podem provocar um síndrome febril, sinais e sintomas neurológicos como encefalites e meningites e, mais raramente, febres hemorrágicas. Entre os arbovírus neurotrópicos está o vírus West Nile (WNV, flavivírus) e o vírus Toscana (TOSV, flebovírus) cuja presença em Portugal tem sido bem documentada nas últimas décadas. Só em 2010, foi diagnosticado laboratorialmente um caso humano de infecção por WNV, no concelho de Setúbal, e vários casos de infecção por TOSV em doentes com sintomas neurológicos na região do Algarve. Como vírus transmitidos por artrópodes, os arbovírus têm a sua distribuição relacionada com a distribuição geográfica das espécies vectoras e com as condições e alterações ambientais que influenciam essa distribuição. Com o objectivo de conhecer as espécies e a sua distribuição, o Instituto Nacional de Saúde (INSA) e as autoridades de saúde regionais e nacionais portuguesas têm a decorrer um programa nacional de vigilância de mosquitos onde, para além dos insectos colhidos vivos, são reunidos dados ecológicos, climáticos e de georeferência para cada um dos locais. O conhecimento das coordenadas geográficas dos locais de colheita permite a obtenção, através de um programa de Sistemas de Informação Geográfica (SIG), de um conjunto relevante de variáveis associadas com a presença e

abundância de mosquitos como a altitude, o tipo de utilização do solo e a proximidade a massas de água, entre outras [1].

Dispomos de dados referentes a capturas realizadas nos anos de 2006 a 2012, durante os meses de Maio a Outubro em diversas localizações em todo o país. Estes dados têm a particularidade de não se encontrarem organizados de forma regular espacial ou temporalmente, uma vez que as acções de captura não foram programadas de forma a que fossem igualmente espaçadas no tempo nem cobrindo a generalidade do território nacional.

Considerando que a dimensão da captura é reflexo da abundância, apresentamos uma primeira abordagem ao problema que consiste na construção de um modelo em que a dimensão esperada da captura é estimada através de um preditor linear, segundo uma abordagem bayesiana. A dependência espacial é introduzida através de uma componente aleatória estruturada em que se considera que os resultados obtidos numa determinada localização dependem dos resultados obtidos em locais vizinhos. Os modelos considerados pertencem ainda à classe dos modelos para dados de contagens com excesso de zeros com sobredispersão uma vez que a análise preliminar dos dados levou à constatação da existência de elevada dispersão e número de zeros [2].

Fazendo uso do modelo estimado, apresentamos mapas de abundância esperada para diversos cenários meteorológicos e temporais.

Referências

- [1] Rochlin, I., Harding, K., Ginsberg, H.S., Campbell, S.R. (2008) *Comparative analysis of distribution and abundance of West Nile and Eastern equine Encephalomyelitis virus vectors in Suffolk county, New York, using human population density and land use/cover data*, J. Med. Entomol. 45(3), 563–571.
- [2] Mendes, J.M., Turkman, K.F., Jardim, E. (2007) *A Bayesian hierarchical model for over-dispersed count data: a case study for abundance of hake recruits*, Environmetrics 18, 27–53.

Agradecimentos:

Este trabalho é financiado pelos projectos da FCT

(1) PTDC/SAU-SAP/119199/2010

(2) PEst-OE/MAT/UI0006/2011 e PTDC/MAT/118335/2010

Comparison of Likelihood, PLS and Bayesian Estimators for Structural Equation Modeling. A Simulation with Customer Satisfaction Data

(Comunicação)

Jorge M. MENDES – *ISEGI-NOVA, NOVA University of Lisbon, Campus de Campolide, 1070-032 Lisbon, Portugal*

jmm@isegi.unl.pt

Pedro Simões Coelho – *ISEGI-NOVA, NOVA University of Lisbon, Campus de Campolide, 1070-032 Lisbon, Portugal*

Resumo

Covariance based methods are undoubtedly the most well-known methods to estimate Structural Equation Models (SEM), with the result that many social researchers use the terms (SEM and covariance-based methods) synonymously.

Partial Least Squares (PLS) methods constitute one alternative to estimating SEM. However in spite of the growing usage of PLS methods in several fields (for instance in customer satisfaction measurement), these methods are still often seen as ad hoc algorithms that have generally not been formally analyzed (McDonald, 1996). Several authors (e.g. Chin, 1998; Fornell and Bookstein, 1982; Hoogland and Boomsma, 1998) argue that PLS presents several advantages when compared to covariance-based methods. In fact it is argued that some conditions should be met in order for these later methods to produce consistent parameter estimates, namely the data should follow a specific multivariate distribution and have independence of observations. Moreover, indicators are typically required to be reflective and unique case values for latent variables cannot be obtained. On the top of that, the statistical properties of covariance-based methods are asymptotic and its robustness highly depends on the sample size.

A second alternative is the Bayesian approach to SEM which has received little attention of researchers working in SEM. In fact, contributions are only limi-

ted to factor analysis (e.g., Bartholomew, 1987). More recently the powerful tools in statistical computing for simulating from the posterior distributions have greatly enhanced the applicability of general Bayesian approach and a number of generalizations of the standard SEM have been developed. The basic attractiveness of a Bayesian approach is its flexibility to incorporate prior information for achieving better results. In many practical situations, the researcher might have notice of an high correlation between two latent variables or an high loading in a specific manifest variable. This prior information can be easily incorporated in the inference via prior elicitation. In the situations without accurate prior information, some type of non-informative priors can be used. The accuracy of Bayesian estimates is close to that of covariance-based methods.

There is no a formal proof in the framework of a realistic model that show the advantages of PLS and Bayesian techniques over covariance based methods. Also, few comprehensive simulation studies comparing the performance CBSEM, PLS and Bayesian approaches to SEM have been performed. Notable exceptions are the two well-known studies comparing covariance-based and PLS estimation in one hand (Fornell and Bookstein, 1982; Vilares, Almeida and Coelho, 2010) and the research comparing covariance-based and Bayesian estimation, on the other (Chumney, 2012).

In this talk, we will for the first time compare these 3 approaches. For that, we shall evaluate the effects of two assumptions typically present in CBSEM: the symmetry of the distribution and the reflective modeling of the indicators. Thus we compare how the three kinds of methods (covariance-based, PLS and Bayesian methods) perform both when these assumptions hold and when they are violated, i.e. when the distribution of the observations is skewed and some indicators follow a formative scheme. We shall perform this analysis in the framework of the ECSI (European Customer Satisfaction Index) model. In fact, the interest in the performance of the three methods in the context of skewness of response and the formative nature of some blocks in the model is particularly justified when we deal with customer satisfaction data. This formative nature of blocks is common in marketing applications and according to Hulland (1998) tends to result in slightly better overall model quality.

Referências

- [1] Bartholomew (1981), *Posterior analysis of the factor model*, British Journal of Mathematical and Statistical Psychology, 34, 93-99.
- [2] Chin, W. W. (1998), *The partial least squares approach to structural equation modeling* In G. A. Marcoulides (ed.), *Modern methods for business research*, Chap. 10, Lawrence Erlbaum Associates.
- [3] Fornell, C., and Bookstein, F. L. (1982), *The two structural equation models: LISREL and PLS applied to customer exit-voice theory*, Journal of Marketing Research, 19, 440-452.
- [4] Hoogland and Boomsma (1998), *Robustness studies in covariance structure modeling: an overview and meta analysis*, Sociological Methods and Research, 26, 329-368.
- [5] Hulland, J. (1998), *Use of partial least squares in strategic management research*, Comportamento Organizacional e Gestão, 4(1), 181-202.
- [6] McDonald, R. P. (1996), *Path analysis with composite variables*, Multivariate Behavioral Research, 31(2), 239-270.
- [7] Vilares, M.J., Almeida, M.H. and Coelho, P.S. (2010), *Comparison of Likelihood and PLS Estimators for Structural Equation Modeling: A Simulation with Customer Satisfaction Data* In V. Esposito Vinzi et al. (eds.), *Handbook of Partial Least Squares*, Chap. 13, Springer-Verlag Berlin Heidelberg.
- [8] Chumney, F. L. (2012), *Comparison of Maximum Likelihood, Bayesian, Partial Least Squares, and Generalized Structured Component Analysis Methods for Estimation of Structural Equation Models with Small Samples: An Exploratory Study*. Open Access Theses and Dissertations from the College of Education and Human Sciences, Paper 145.

A multilevel Bayesian model: an application using national discharge data in Portugal

(Comunicação)

Maria Fátima PINA – *Departamento de Epidemiologia Clínica, Medicina Preventiva e Saúde Pública, FMUP - Faculdade de Medicina da Universidade do Porto, INEB - Instituto de Engenharia Biomédica, ISPUP - Instituto de Saúde Pública da Universidade do Porto*

fpina@med.up.pt

Sandra Maria Alves – *INEB - Instituto de Engenharia Biomédica, ESTSP/IPP, Escola Superior de Tecnologia da Saúde do Porto, ISPUP - Instituto de Saúde Pública da Universidade do Porto*

Theodoros Economou – *College of Engineering, Mathematics and Physical Sciences, University of Exeter*

Marília Sá Carvalho – *Programa de Computação Científica - PROCC, Fundação Oswaldo Cruz, Rio de Janeiro, Brasil*

Trevor C. Bailey – *College of Engineering, Mathematics and Physical Sciences University of Exeter*

Resumo

Hip fractures represent a major public health problem due to high mortality and morbidity; the care provided during the acute treatment maybe determinant for a good outcome. The objectives of this study were:

1. model the probability of dying during a hospitalization for hip fracture, using a risk adjusted modelling framework;
2. quantify possible unobserved hospital effects on the probability of dying;
3. assess temporal trends in the probability of dying, associated with different hospitals;
4. quantify (spatial) area-level effects on the probability of dying.

A multilevel (patient/hospital/area) statistical model for the binary outcome death/survival was implemented. Risk adjustment was incorporated in the

model though the use of covariates and random effects (unobserved covariates), at each of the 3 levels or hierarchies of the model. Multilevel models, enable simultaneous analysis of the group-level (hospital, area) and individual level (patient) covariates (correcting for the biases in parameter estimates resulting from dependencies) and decomposition of the outcome at each level (Guo and Zhao 2000; Diez-Roux 2000). Nationwide secondary data in the period 2000-2008 were used. Variables related to patients were collected from the National Hospital Discharge Register; hospital variables from the National Hospital Inquiries and area variables from Statistics Portugal.

Our main findings were: variables with significant association with the probability of dying were (log-odds scale along with 95% Credible intervals): age (0.07 [0.06, 0.07]), sex (women compared with men -0.72 [$-0.80, -0.63$]), length of stay (-0.01 [$-0.02, -0.01$]), severity (highest versus lowest category (2.50 [2.30, 2.71])), conservative treatment had a highest increased in probability of dying versus other treatments, except internal fixation. The overall effect of time in the probability of dying was not significant (-0.02 [$-0.06, 0.00$]), however there was significant variation between hospitals.

In spite of the difficulty to assess quality of care, a multilevel approach seems appropriate. Our results point that improving the population health the in-hospital probability of dying can be reduced and some hospitals can still improve their performance.

Referências

- [1] Guo, Guang, and Hongxin Zhao. (2000). *Multilevel Modeling for Binary Data*. Annu. Rev. Sociol. 26, 441–462.
- [2] Diez-Roux, A. V. (2000). *Multilevel analysis in public health research*. Annu Rev Public Health 21, 171–92.

Acknowledgement

This work was financed by FEDER funds through the Programa Operacional Factores de Competitividade ? COMPETE and by Portuguese funds through FCT - Fundação para a Ciência e a Tecnologia within the framework of the project PEst-C/SAU/LA0002/2011 and by PTDC/SAU-EPI/113424/2009 grant and SFRH / BD / 40978 / 2007 fellowship. We would like to ack-

nowledge the Central Administration of Health Services (ACSS) for the data from the National Hospital Discharge Register, and to Statistics Portugal ? INE for the data from the National Hospital Inquiry (NHI). We would like also to acknowledge Carlos Alves (MD) for calculating the Charlson Index and help in the initial analysis of data, and José Lopes (MD), medical doctor responsible for the codification of discharges registers from Hospital São João for answering some questions related to coding of data from the National Hospital Discharge Register.

Meta-análise bayesiana: uma ferramenta indispensável na avaliação de tecnologias em saúde

(Comunicação)

Maria João SILVA – *Exigo Consultores*

mjoao.silva@exigoconsultores.com

Sara Rabiais – *Exigo Consultores*

Björn Vandewalle – *Exigo Consultores*

Jorge Félix – *Exigo Consultores*

Resumo

Uma das aplicações da estatística bayesiana mais comumente exploradas no âmbito da avaliação de tecnologias em saúde (ATS) é na meta-análise (MA), onde esta abordagem é a que mais se utiliza actualmente [1]. Neste sector, a MA é extremamente útil não só para agregar mas sobretudo para gerar evidência, através da comparação directa e indirecta de tratamentos.

Os medicamentos são a tecnologia da saúde mais regulada e avaliada. A sua aprovação depende da realização de ensaios clínicos, incluindo estudos comparativos de fase 3. No entanto, vários constrangimentos, incluindo os de natureza económica, resultam muitas vezes na inexistência de estudos experimentais que comparem directamente dois ou mais medicamentos de interesse. A MA, em particular na sua vertente bayesiana, tem vindo a preencher este tipo de lacunas. Para além de inúmeras ATS baseadas em resultados de MA, uma prova inequívoca da sua importância, por exemplo na farmacoeconomia, é o desenvolvimento de guidelines para a sua correcta implementação por parte de entidades com reconhecido mérito, nomeadamente o NICE [2] e a ISPOR [3-4].

Uma proporção significativa das avaliações económicas conduzidas pela Exigo Consultores é baseada no desenvolvimento de MA em áreas como o cancro ou HIV/SIDA [5]. Neste contexto, verifica-se que a utilização de software como o WinBUGS ou o OpenBugs proporciona uma versatilidade crucial à utilização desta metodologia, dado que requer a especificação da verosimilhança dos modelos pelo utilizador. A MA bayesiana, enquanto ferramenta, tem permitido

à Exigo abordar problemas complexos com aplicação e implicações directas no desempenho dos seus clientes.

Nesta comunicação, pretende-se evidenciar o contributo da estatística bayesiana, e em particular da MA baseada em inferência bayesiana, na resolução de problemas comuns na área da ATS.

Referências

- [1] Stegenga H, Rogers G, Kandaswamy P, Jones D. 2012. Network meta-analyses in national clinical guidelines: a description and quality assessment. Guidelines International Network. G-I-N Conference 2012. Berlin, 22.-25.08.2012. Düsseldorf: German Medical Science GMS Publishing House; DocO17
- [2] NICE DSU Technical Support Documents 1-7. 2011. Disponível em: <http://www.nicedsu.org.uk/>.
- [3] Jansen JP, Fleurence R, Devine B, Itzler R, Barrett A, Hawkins N, Lee K, Boersma C, Annemans L, Cappelleri JC .2011. Interpreting Indirect Treatment Comparisons and Network Meta-Analysis for Health-Care Decision Making: Report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: Part 1. Value Health; 14:417-28.
- [4] Hoaglin DC, Hawkins N, Jansen JP, Scott DA, Itzler R, Cappelleri JC, Boersma C, Thompson D, Larholt KM, Diaz M, Barrett A. 2011. Conducting Indirect-Treatment-Comparison and Network-Meta-Analysis Studies: Report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices – Part 2. Value Health; 14:429-37.
- [5] DeJesus E, Felix J, Vandewalle B et al. 2012. Indirect treatment comparison of efficacy, safety and resistance of EVG/COBI/FTC/TDF (Quad) vs. RAL+FTC/TDF in treatment-naïve HIV patients. Journal of the International AIDS Society; 15(Suppl 4):18239.

Estatística na Ecologia e Ambiente

Circular data analysis for environmental monitoring

(Comunicação)

Rosa M. CRUJEIRAS – *Universidade de Santiago de Compostela*

rosa.crujeiras@usc.es

María Oliveira – *Universidade de Santiago de Compostela*

Alberto Rodríguez-Casal – *Universidade de Santiago de Compostela*

Resumo

Circular data analysis has been approached both from parametric and nonparametric perspectives, generating a broad literature about parametric distribution models for circular variables. Some thorough and comprehensive reviews are the volumes by Mardia (1972), Fisher (1993), Mardia and Jupp (2000) and Jammalamadaka and SenGupta (2001). In these references, the classical parametric distribution models, such as the von Mises or the wrapped normal, jointly with uniformity and von-missity tests are revised. In practical problems, circular variables may be correlated with linear variables (for instance, wind speed and wind direction), and correlation coefficients for this setting have been also introduced, as well as parametric regression models, for linear response and circular covariate. Some examples of applications of these parametric circular models are the works by Jammalamadaka and Lund (2006), who analyzed the relation between wind direction and ozone levels, and SenGupta and Ugwuowo (2006), who propose an asymmetric circular-linear regression model to predict solar radiation and wind energy.

However, if there is not enough scientific motivation for specifying a parametric model (for density or regression), nonparametric tools may be more appealing. Density and regression estimation can be done by using local smoothers, based on kernel functions. There are some early works about kernel density estimation for the general case of directional data by Hall et al. (1987) and Bai et al. (1988). The problem of choosing a bandwidth parameter has been addressed by Hall et al. (1987), Taylor (2008) and Oliveira et al. (2012). For regression estimation with linear response and

circular covariate, Di Marzio et al. (2009) studied local linear smoothers. Additionally, and adaptation of the SiZer method, developed by Chaudhuri and Marron (1999), for the assessment of statistically significant features in the data structure from a circular smoother has been proposed by Oliveira et al. (2013).

The goal of this talk is to provide the audience with an introduction to circular data analysis, specifically from a nonparametric perspective, and its application to environmental problems. First, a brief review on the aforementioned circular parametric and nonparametric models, for density and regression, will be provided. The methods will be illustrated with some real data problems, from meteorology (wind direction and wind speed in the Galician coast) and environmental monitoring (temperature cycles in subantarctic areas).

Referências

- [1] Bai, Z.D., Rao, C.R. and Zhqo, L.C. (1988) *Kernel estimators of density function of directional data*. Journal of Multivariate Analysis, 27, 24–39.
- [2] Di Marzio, M., Panzera, A. and Taylor, C.C. (2009) *Local polynomial regression for circular predictors*. Statistics and Probability Letters, 79, 2066–2075.
- [3] Fisher, N.I. (1993) *Statistical Analysis of Circular Data*, Cambridge University Press.
- [4] Hall, P., Watson, G.P. and Cabrera, J. (1987) *Kernel density estimation for spherical data*. Biometrika, 74, 751–762.
- [5] Jammalamadaka, S.R. and Lund, U. (2006) *The effect of wind direction on ozone levels: a case study*. Environmental and Ecological Statistics, 13, 287–298.
- [6] Jammalamadaka, S.R. and SenGupta, A. (2001) *Topics in Circular Statistics*. World Scientific.
- [7] Klemela, J. (2000) *Estimation of densities and derivatives of densities with directional data*. Journal of Multivariate Analysis, 73, 18–40.
- [8] Mardia, K.V. (1972) *Statistics of Directional Data*. Academic Press.
- [9] Mardia, K.V. and Jupp, P.E. (2000) *Directional Statistics*. Wiley.

- [10] SenGupta, A. and Ugwuowo, F.I. (2006) *Asymmetric circular-linear multivariate regression models with applications to environmental data*. Environmental and Ecological Statistics, 13, 299–309.
- [11] Taylor, C.C. (2008) *Automatic bandwidth selection for circular density estimation*. Computational Statistics and Data Analysis, 52, 3493–3500.
- [12] Oliveira, M., Crujeiras, R.M. and Rodríguez-Casal, A. (2012) *A plug-in rule for bandwidths selection in circular density estimation*. Computational Statistics and Data Analysis, 56, 3898–3908.
- [13] Oliveira, M., Crujeiras, R.M. and Rodríguez-Casal, A. (2013) *CircSizer: an exploratory tool for circular data*. Environmental and Ecological Statistics. DOI 10.1007/s10651-013-0249-0.

Application of bootstrap approaches to environmental monitoring data

(Comunicação)

Pilar GARCÍA-SOIDÁN – *University of Vigo (Spain)*

pgarcia@uvigo.es

Raquel Menezes – *University of Minho (Portugal)*

Oscar Rubiños-López – *University of Vigo (Spain)*

Resumo

The risk of exceeding the maximum permitted levels of certain chemical elements can damage population health and pollute the environment. These are some of the reasons why in practice samples of these elements are periodically taken to measure their concentration at different locations. By using the prediction approaches in the spatial setting, the level of each element can be approximated at all the spatial sites, including the unobserved ones, or the probability that the pollutant exceeds (or does not exceed) a given threshold can be estimated. The resulting values allow us to build a prediction map of the observation region in the former case or a probability map showing the distribution function of the variable involved at a fixed maximum (or its complementary value) in the second one. In this respect, application of the parametric or the nonparametric prediction methodology demands previously address a variety of problems, which can require extra work to check the accuracy of the estimators employed, to derive their sampling distribution or to appropriately approximate the unknown terms involved. Some of the aforementioned problems have been solved in the curve estimation setting through a resampling approach referred to as Bootstrap (Efron, 1979).

The traditional Bootstrap techniques were designed for independent data, although in practice the parametric procedures can be easily adapted for spatial data, thus leading us to implement techniques for variogram assessment (Maglione and Diblasi, 2004) or construction of confidence intervals for the parameters of a distribution estimator (Goovaerts et al., 2005). When independence can be assumed from the residuals, the classic nonparametric methods can be applied for some issues, such as the approximation of the

variance of an estimator (Iranpanah et al., 2011). However, the design of specific strategies for resampling is necessary to guarantee consistency of the results, as the procedure based on resampling blocks of data (Hall 1985) or marks assigned to the spatial points (Loh, 2008).

In the current work, we have developed estimators of the joint distribution of more general use, which are distribution functions themselves, associated to discrete or continuous random variables, so they are used as the basis for implementation of a discrete or smooth Bootstrap approaches in the spatial setting. We should remark that the Bootstrap methodology must not be intended to replace other techniques designed for addressing specific issues, but for complementing them and adding extra information that can be required. From this perspective, these spatial Bootstrap approaches offer attractive alternatives for resampling, whose consistency can be proved by assuming stationarity from the random process or by relaxing the latter hypothesis to admit a deterministic trend. Numerical studies for simulated data and a real data set, obtained from environmental monitoring, are included to illustrate the application of the proposed Bootstrap methods, where the good behavior of the resampling methods can be observed, although the really important thing is that they help capture the main features of the underlying spatial process.

This work has been supported in part by the projects TEC2011-28683-C02-02 of the Spanish Ministry of Science and Innovation and PTDC/MAT/112338/2009 (FEDER support included) of the Portuguese Ministry of Science, Technology and Higher Education.

Referências

- [1] Efron, B. (1979) *Bootstrap methods: another look at the Jackknife*, Ann. Statist. 7, 1–26.
- [2] Govaerts, B., Beck, B., Lecoutre, E., Le Bailly, C., VandenEeckaut, P. (2005) *From monitoring data to regional distributions: a practical methodology applied to water risk assessment*, Environmetrics 16, 109–127.
- [3] Hall, P. (1985) *Resampling a coverage pattern*, Stoch. Proces. Applic. 20, 231–246.
- [4] Iranpanah, N., Mansourianb, A., Tashayob, B., Haghighic, F. (2011) *Spatial Semi-Parametric Bootstrap Method for Analysis of Kriging Predictor of*

Random Field, *Procedia. Environ. Sci.* 3, 81–86.

[5] Loh, J.M. (2008) *A valid and fast spatial Bootstrap for correlation functions*, *Astrophys. J.* 681, 726–734.

[6] Maglione, D.S., Diblasi, A.M. (2004) *Exploring a valid model for the variogram of an isotropic spatial process*, *Stoch. Envir. Res. Risk. Ass.* 18, 366–376.

Análise de extremos espaciais: Aplicação a dados de precipitação no sul de Portugal

(Comunicação)

Dora GOMES – *CMA e FCT-UNL*

dsrp@fct.unl.pt

M. Manuela Neves – *CEAUL e ISA-UL*

Elsa Moreira – *CMA e FCT-UNL*

Resumo

A estatística espacial integra um conjunto de procedimentos estatísticos nos quais as localizações espaciais desempenham um papel explícito na análise dos dados. Uma das sub-áreas da estatística espacial é a geoestatística, que se preocupa em modelar a variabilidade espacial de um processo contínuo, usando dados recolhidos em algumas localizações. Historicamente, os processos Gaussianos desempenham um papel central na modelação de processos espaciais. No entanto, acontecimentos de interesse tais como precipitação, ventos, tempestades . . . , nos quais o comportamento de valores extremos é preocupante e necessita de modelação adequada pelo impacte ou riscos que podem representar para as populações, não são adequadamente modelados por processos gaussianos. Quando os dados de que dispomos são geo-referenciados, variando a localização continuamente numa região do espaço, utilizam-se as ferramentas dos processos espaciais, $\{Y(\mathbf{s})\}$, vector aleatório numa localização $\mathbf{s} \in D \subset \mathcal{R}^d$. A análise dos dados extremos espaciais encontra-se no cruzamento de duas áreas estatísticas: a teoria de valores extremos e a geoestatística. Uma variedade de ferramentas estatísticas tem sido utilizada para a modelação espacial de extremos, incluindo os modelos bayesianos hierárquicos, cópulas e campos aleatórios max-estáveis. Uma excelente revisão recente pode ver-se em Cooley *et. al.* (2012) e Davison *et. al.* (2012).

Os processos max-estáveis são os modelos naturais para extremos espaciais. Na última década estes processos têm estado a ocupar um papel de relevo na modelação estatística de extremos espaciais. Mas a sua aplicação tem sido difícil pelo facto de não se ter a distribuição multivariada e portanto os métodos baseados na verosimilhança não permitirem fazer inferência. Vários

procedimentos alternativos têm estado a surgir: a verosimilhança composta, a verosimilhança a pares (do inglês *pairwise likelihood*), etc.

O estudo dos valores extremos da precipitação é de grande importância, pois se a ocorrência de valores muito elevados pode colocar as populações em risco, também as secas podem ter sérios impactes nos aspectos agrícolas e consequentemente económicos da região ou país. A modelação adequada da dependência entre extremos em várias localizações e a compreensão da forma como varia é crucial para a avaliação do risco.

Neste estudo é feita uma apresentação breve dos procedimentos mais usados na análise de extremos espaciais, aplicando a uma amostra de dados de precipitação mensal registados de 1941-2006 em várias localizações no Alentejo e Algarve. Outra característica deste trabalho é explorar algumas funções existentes no ambiente R, com ênfase especial para as incluídas no `SpatialExtremes`, Ribatet (2009).

Agradecimentos: Investigação parcialmente suportada por Fundos nacionais através da FCT—Fundação para a Ciência e a Tecnologia, projectos PEst-OE/MAT/UI0006/2011(CEAUL) e PEst-OE/MAT/UI0297/2011 (CMA).

Referências

- [1] Cooley, D., Cisewski, J., Erhardt, R.J., Jeon, S., Mannshardt, E., Omolo, B.O., and Sun, Y. (2012). *A survey of spatial extremes: Measuring spatial dependence and modeling spatial effects*, REVSTAT. 10, 135–165.
- [2] Davison, A., Padoan, S., and Ribatet, M. (2012) *Statistical modeling of spatial extremes*, Statistical Science. 27, 161–186.
- [3] Ribatet, M. (2009) *A User's Guide to the SpatialExtremes Package*, École Polytechnique Fédérale de Lausanne, Switzerland.

Modeling associations between outdoor air quality and birth weight: a geostatistical approach for assessing uncertainty

(Comunicação)

Maria João PEREIRA – *Centro de Recursos Naturais e Ambiente, Instituto Superior Técnico, Universidade Técnica de Lisboa*

maria.pereira@ist.utl.pt

Manuel Castro Ribeiro – *Centro de Recursos Naturais e Ambiente, Instituto Superior Técnico, Universidade Técnica de Lisboa*

Amílcar Soares – *Centro de Recursos Naturais e Ambiente, Instituto Superior Técnico, Universidade Técnica de Lisboa*

Resumo

Several environmental health studies suggest birth weight is associated with outdoor air pollution during gestation [1], [2]. In these studies, exposure assignments are usually based on measurements collected at air quality monitoring stations that do not coincide with health data locations. A possible solution to estimate exposure at health data locations involves the use of spatial models that incorporate spatial information on exposure. Then, exposure can be estimated at health data locations, provided that exposure estimates uncertainty are taken into account, otherwise results can be misleading [3]. In this work we conducted a semi-ecological study to analyze associations between air quality during gestation and birth weight. Air quality during gestation was measured using lichen diversity value (LDV) which is based on the number and frequency of lichen species occurring in each location. LDV is an integrative measure of all atmospheric pollutants and lichens can be found everywhere, allowing for a larger spatial resolution of exposure measurements. To our knowledge this is the first time that the association between air quality and birth weight is studied using lichens as biomonitors. To address exposure uncertainty at health locations, we applied geostatistical simulation on biomonitoring data that provided multiple equally probable realizations of biomonitoring data, with reproduction of observed histogram and spatial co-

variance while matching for conditioning data. Each simulation represented a measure of exposure at each location. The set of simulations provided a measure of exposure uncertainty at each location. To incorporate uncertainty in our analysis we used generalized linear models, fitted simulation outputs and health data on birth weights and assessed statistical significance of exposure parameter using non-parametric bootstrap techniques. We found a positive association between air quality and birth weight. However, this association was not statistically significant. We also found a modest but significant association between air quality and birth weight among babies exposed to gestational tobacco smoke.

Referências

- [1] Šrám, R. J., B. Binková, J. Dejmek, and M. Bobak (2005) *Ambient Air Pollution and Pregnancy Outcomes: A Review of the Literature.*, Environmental Health Perspectives 113:375-382. doi: 10.1289/ehp.6362.
- [2] Darrow, L., M. Klein, M. J. Strickland, J. a Mulholland, and P. E. Tolbert. (2011) *Ambient air pollution and birth weight in full-term infants in Atlanta, 1994-2004.* Environmental Health Perspectives 119:731-7. doi: 10.1289/ehp.1002785.
- [3] Gryparis, A., C. J. Paciorek, A. Zeka, J. Schwartz, and B. a Coull. (2009) *Measurement error caused by spatial misalignment in environmental epidemiology.* Biostatistics (Oxford, England) 10:258-74. doi: 10.1093/biostatistics/kxn033.

Estatística e Genómica

Será o ADN extraordinariamente simétrico?

(Comunicação)

Vera AFREIXO – *Universidade de Aveiro/CIDMA*

vera@ua.pt

Carlos A. C. Bastos – *Universidade de Aveiro/IEETA*

João M. O. S. Rodrigues – *Universidade de Aveiro/IEETA*

Resumo

Este trabalho surge no âmbito da análise de sequências de ADN. É apresentada uma técnica para análise do fenómeno das simetrias em cadeias simples de ADN no genoma humano (de notar que o fenómeno das simetrias está relacionado com a semelhança entre proporções de pares de palavras que são complemento invertido entre si).

Assumindo que os nucleótidos se ligaram de forma aleatória seria de esperar que as probabilidades de ocorrência dos oligonucleótidos (também designados por palavras) só dependessem da probabilidade de ocorrência dos nucleótidos. Será de esperar que alguns oligonucleótidos tenham igual probabilidade de ocorrência: aqueles com a mesma composição em termos de nucleótidos (e.g. CAA; ACA; AAC). Se adicionalmente, assumirmos como válida a segunda lei de Chargaff [1-3], que estabelece que as frequências dos nucleótidos complementares são semelhantes ($%A \sim %T$ e $%C \sim %G$), esperamos igual probabilidade de ocorrência em todas as palavras que contêm a mesma composição dos dois tipos de nucleótidos $T_1 = \{A, T\}$, $T_2 = \{C, G\}$ (e.g. CAA; ACA; AAC; GAA; AGA; AAG; CAT; ACT; ATC; CTA; TCA; TAC; GAT; AGT; ATG; GTA; TGA; TAG; CTT; TCT; TTC; GTT; TGT; TTG). A grupos de palavras com esta composição chamamos grupos de composição equivalente – GCE. Como facilmente se pode constatar, a semelhança entre proporções de pares de palavras que são complemento invertido entre si, é uma característica esperada até em sequências em que não existe estrutura de dependência entre os símbolos.

Quando existe semelhança entre todas as proporções de pares de palavras que são complemento invertido entre si, dizemos que se verifica o fenómeno de simetria básico. Quando dentro dos GCEs as palavras complementos invertidos

entre si têm frequência semelhante, mas distinta da dos restantes elementos do GCE, dizemos que se verifica o fenómeno de simetria não uniforme ou extraordinária.

Neste trabalho será introduzida uma medida para avaliar/caraterizar a simetria extraordinária. A medida é baseada numa razão entre a soma de desvios à uniforme nos GCEs e a soma de desvios ao fenómeno de simetria básico denotada por $R_{simetria}$. A expressão $1 - R_{simetria}$ poderá ser interpretada como uma medida da simetria não uniforme.

A amostra em estudo é o genoma humano de referência versão 37.1 disponível na página do National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>).

No genoma humano para os diferentes tamanhos de palavra em estudo ($k \in \{1, 2, \dots, 12\}$), o $R_{simetria}$ apresenta valores muito acima de 1 (valor esperado para os casos em que existe simetria básica e não existe simetria não uniforme). Também se observa que as palavras de tamanho 3 em relação às palavras em estudo parecem apresentar mais simetria não uniforme.

Dado que o ADN verifica o fenómeno da simetria, será a distribuição nos GCEs significativamente não uniforme? No sentido de responder a esta questão e avaliar a significância estatística do fenómeno de interesse são explorados testes de hipóteses. As hipóteses são testadas via método de Monte Carlo e naturalmente são usados dados simulados sob a validade da hipótese nula. De notar que a amostra em estudo é o genoma de referência com aproximadamente 3GB de tamanho e pretende-se estudar o fenómeno da simetria não uniforme para vários tamanhos de oligonucleótidos, o que requer um grande poder computacional.

Referências

- [1] Rudner, R., Karkas, J. D., Chargaff, E. (1964) *Separation of B. subtilis DNA into complementary strands, I. biological properties.*, Proceedings of the National Academy of Sciences of the United States of America 60(2), 630–635.
- [2] Karkas, J. D., Rudner, R., Chargaff, E. (1964) *Separation of B. subtilis DNA into complementary strands, II. template functions and composition as determined by transcription with RNA polymerase.*, Proceedings of the

National Academy of Sciences of the United States of America 60(3), 915–920.

[3] Rudner, R., Karkas, J. D., Chargaff, E. (1964) *Separation of B. subtilis DNA into complementary strands, III. direct analysis.*, Proceedings of the National Academy of Sciences of the United States of America 60(3), 921–922.

Identificação simultânea de biomarcadores em estudos genéticos de associação: Desafios estatísticos e computacionais

(Comunicação)

Pedro DUARTE SILVA – *Faculdade de Economia e Gestão e CEGE*

Universidade Católica Portuguesa/Porto

psilva@porto.ucp.pt

Resumo

Em estudos genéticos de associação é comum pesquisar a informação contida em grandes bases genéticas de dados com o objectivo de identificar um pequeno conjunto de marcadores relacionados com alguma doença ou uma característica genética de interesse. A disponibilidade de bases de dados contendo dezenas ou centenas de milhar de pequenas variações genéticas conhecidas como "Single Nucleotide Polymorphisms"(SNPs) tem permitido encontrar algumas dezenas de SNPs associados, ou casualmente ligados, a condições clínicas de risco tais como anomalias cardíacas, diabetes, ou vários tipos de cancro [4].

Em termos estatísticos, o problema central dos estudos de associação genética pode ser formalizado como um problema de selecção de variáveis (SNPs) em modelos de previsão para uma condição clínica de interesse. Este problema tem no entanto características específicas, nomeadamente devido ao elevado número de variáveis potencialmente preditivas, que é tipicamente muito superior ao número de observações disponíveis.

Devido às dificuldades computacionais envolvidas, muitas das metodologias inicialmente tentadas nestes problemas recorrem a técnicas univariadas, tais como o estabelecimento de rankings baseados em correlações marginais [2]. Essas abordagens revelam-se incapazes de identificar combinações de SNPs que só actuam quando agrupadas em conjunto, e mais recentemente várias técnicas multivariadas (ver, por exemplo, [1][3][5][6][7]) tem vindo a ser propostas e aplicadas com sucesso.

Nesta comunicação far-se-á uma revisão de propostas recentes para o problema da identificação simultânea de biomarcadores em estudos genéticos de associação, discutindo-se as vantagens e limitações das principais abordagens propostas, e identificando-se alguns dos desafios em aberto nesta área.

Referências

- [1] Ding, Y., Wilkins, D. (2006) *Improving the Performance of SVM-RFE to Select Genes in Microarray Data*, BMC Bioinformatics. 7.Suppl 2: S12.
- [2] Foulkes, A.S., (2009). Applied statistical genetics with R. New York: Springer.
- [3] Guyon, I., Weston, J., Barnhill, S., Vapnick, V. (2002) *Gene selection for cancer classification using support vector machines*. Machine Learning. 46, 389-422.
- [4] Hindorf, L., Sethupathy, P., Junkins, H.A., Ramos, E.M., Metha, J.P., Collins, F.S. Manolio, T.A. (2009) *Potential etiologic and functional implications of genome-wide association loci for human diseases and traits*. Proc. Natl. Acad. Sci. USA.
- [5] Hoggart, C.J., Whittaker, J.C., De Iorio M., Balding D.J. (2008) *Analysis of all SNPs, in genome-wide and re-sequencing association studies*. PLoS Genetics. 4:e1000130.
- [6] Zhang, C.H. (2010) *Nearly unbiased variable selection under minimax concave penalty*, *Annals of Statistics*. 38, 894-942
- [7] Zuber, B., Duarte Silva, A.P., Strimmer, K. (2012) *A novel algorithm for simultaneous SNP selection in high-dimensional genome-wide association studies*, BMC Bioinformatics. 13:284.

Detecção de iniciação intragénica em dados de NGS

(Comunicação)

Miguel PEREIRA – *Hospital de Santa Maria | Faculdade de Ciências da Universidade de Lisboa*

miguelpereira@gmail.com

Ana Rita Grosso – *IMM, Faculdade de Medicina da Universidade de Lisboa*

Lisete Sousa – *DEIO e CEAUL, Faculdade de Ciências da Universidade de Lisboa*

Resumo

As tecnologias de *Next Generation Sequencing* (NGS) permitem a sequenciação rápida do genoma de células e a obtenção de dados sobre a expressão de todos os genes dessas células (transcriptoma). Os dados da sequenciação correspondem às contagens de *reads* (pequenos fragmentos de DNA produto da sequenciação) que alinham com o genoma ou o transcriptoma (Shendure e Ji, 2008). O objectivo deste trabalho é explorar métodos de análise de dados obtidos através de NGS, para detecção de iniciação intragénica, isto é, o início de expressão génica num exão que não o primeiro. Considera-se uma base de dados de 4 linhas celulares de carcinoma renal de células claras com mutações no gene SETD2 - um gene supressor de tumor que se pensa estar envolvido no aumento da iniciação intragénica (Duns *et al.*, 2010) - e 2 controlos.

O pré-processamento dos dados permite normalizar as contagens das *reads* que alinham com cada exão. Os níveis de expressão de cada exão são representados pela proporção de *reads* que alinham com cada um desses exões, tendo em conta o total de alinhamentos possíveis com cada exão. Neste trabalho são aplicados dois procedimentos para detetar a iniciação intragénica, ou seja, para detetar se os níveis de expressão dos exões de determinado gene aumentam a partir de um exão que não o primeiro. O primeiro procedimento consiste no teste de comparação de duas proporções, em que para cada gene se testam as hipóteses de igualdade dos níveis de expressão de cada par de exões contíguos. Posteriormente, dado o problema da multiplicidade de testes, os valores-p são corrigidos recorrendo ao método de Benjamini-Hochberg. O segundo proce-

dimento é o método de Marascuilo (Marascuilo, 1966). Este método permite a comparação de múltiplas proporções e incorpora a correção para os testes múltiplos. As 6 linhas celulares - 4 com mutações no gene SETD2 e 2 controlos - foram analisadas separadamente utilizando ambos os métodos estatísticos, comparando-se posteriormente os resultados obtidos entre as linhas mutadas e os controlos.

Os resultados são depois comparados com os resultados obtidos por Grosso (2013, resultados não publicados) utilizando o teste exacto de Fisher. Este método tem sido utilizado na deteção de iniciação intragénica, no entanto, em vez de pares de exões, estudam-se os níveis de expressão relativos a um único exão em amostras diferentes (Carvalho *et al.*, 2008).

Agradecimentos: Este trabalho é financiado por Fundos Nacionais através da FCT no âmbito dos projectos PEst-OE/MAT/UI0006/2011 e PTDC/MAT/118335/2010.

Referências

- [1] Carvalho, S., Raposo, A. C., Martins, F. B., Grosso, A. R., Sridhara, S. C., Rino, J., Carmo-Fonseca, M., *et al.* (2013) *Histone methyltransferase SETD2 coordinates FACT recruitment with nucleosome dynamics during transcription*, *Nucleic Acids Research* 41(5), 2881–2893.
- [2] Duns, G., Van den Berg, E., Van Duivenbode, I., Osinga, J., Hollema, H., Hofstra, R. M. W., Kok, K. (2010) *Histone methyltransferase gene SETD2 is a novel tumor suppressor gene in clear cell renal cell carcinoma*, *Cancer Research* 70(11), 4287–4291.
- [3] Marascuilo, L. (1966) *Large-sample multiple comparisons*, *Psychological Bulletin* 65(5), 280–290.
- [4] Shendure, J., Ji, H. (2008) *Next-generation DNA sequencing*, *Nature Biotechnology* 26(10), 1135–1145.

Comparação de métodos não-paramétricos na seleção de genes que apresentam subtipos moleculares em dados de *microarrays*

(Comunicação)

Carina SILVA-FORTES – *Escola Superior de Tecnologia da Saúde de Lisboa-IPL e CEAUL*

carina.silva@estesl.ipl.pt

Maria Antónia Amaral Turkman – *FCUL e CEAUL*

Lisete Sousa – *FCUL e CEAUL*

Resumo

A tecnologia de *microarrays* permite monitorizar em simultâneo milhares de genes num único *chip*, tornando possível aos investigadores perceberem as suas interações. Embora a base química dos *microarrays* não seja recente, o estudo simultâneo de centenas de genes transformou os *microarrays* numa ferramenta de análise global muito importante com aplicações na biologia e na medicina.

Um dos objetivos da tecnologia de *microarrays* é medir a expressão de milhares de genes e identificar mudanças nas expressões entre diferentes estados biológicos. Assim, de um elevado número de genes em análise é possível obter um número substancialmente mais reduzido de genes com a(s) característica(s) de interesse para posterior análise.

A análise de genes diferencialmente expressos (DE) baseia-se fundamentalmente em dois aspetos: ordenação e seleção. A ordenação requer a especificação de uma estatística (ou medida) que evidencie o nível de expressão de cada gene; a seleção requer a especificação de um procedimento que arbitre o que constitui ser DE.

É do conhecimento geral que diferentes métodos de seleção conduzem a listas de genes DE (parcialmente) diferentes. Nesse sentido, é importante comparar os diferentes métodos e identificar as diferenças e analisar quais as consequências em utilizar diferentes métodos.

Em Silva-Fortes *et al.* (2012) [1] foi proposto um gráfico para a seleção de genes DE a partir da análise conjunta de duas estatísticas – *Arrow plot*. Este

gráfico permite selecionar, para além de genes com regulação positiva e negativa, genes com um comportamento biológico de interesse, designados por genes mistos. Estes genes têm a particularidade de revelarem a presença de subtipos moleculares, como por exemplo subtipos de cancro. A construção do *Arrow plot* baseia-se na representação das estimativas do coeficiente de sobreposição entre duas densidades (OVL) e da área abaixo da curva ROC (AUC) de todos os genes pertencentes à experiência. A estimação da AUC e do OVL foi realizada numa abordagem não-paramétrica, onde o OVL foi estimado com base em funções de densidade de probabilidade estimadas pelo método do núcleo e a AUC pelo método empírico.

Neste trabalho comparam-se resultados na seleção de genes considerando considerando a metodologia *Arrow plot* quando a AUC é estimada pelo método empírico e pelo método do núcleo. Embora se tenha verificado que, quando as distribuições das variáveis de diagnóstico são normais, o estimador da AUC pelos dois métodos tenha um viés semelhante, quando se aplica ao contexto da estimação de genes os resultados são bastante diferentes. Por exemplo, o número de genes selecionado com regulação positiva e regulação negativa é diferente quando se consideram os mesmos pontos de corte no *Arrow plot*, nomeadamente quando se considera a AUC empírica o número de genes selecionado é substancialmente maior. O *Arrow plot* com a AUC estimada pelo método empírico revela-se mais otimista, no sentido em que permite selecionar um número mais elevado de genes, num entanto presumivelmente menos realista, uma vez que biologicamente espera-se selecionar um número muito reduzido de genes DE. Relativamente ao número de genes mistos o número de genes selecionado considerando os dois métodos de estimação é semelhante.

Os resultados foram obtidos usando dados simulados e dados disponíveis publicamente e com recurso ao R.

Referências

[1] Silva-Fortes, C., Amaral Turkman, M. A. e Sousa, L. (2012). Arrow plot: a new graphical tool for selecting up and down-regulated genes and genes differentially expressed on sample subgroups. *BMC Bioinformatics*, 13:147.

Agradecimentos

Trabalho financiado por fundos nacionais através da FCT no âmbito dos projetos PEst-OE/MAT/UI0006/2011 e PTDC/MAT/118335/2010

Estadística Multivariada

Reconhecimento de Objectos em Imagens com Dados de Treino Incompletos

(Comunicação)

Alexandre BERNARDINO – *ISR/IST Lisboa*

alex@isr.ist.utl.pt

Ricardo Cabral – *ISR/IST Lisboa*

João Paulo Costeira – *ISR/IST Lisboa*

Resumo

O reconhecimento automático de objectos em imagens digitais consiste em localizar regiões nas imagens que contêm objectos de interesse e identificar as suas classes. Tem importantes aplicações em automação, robótica, sistemas de segurança, entretenimento e informática. Em particular existe grande interesse em métodos de procura automática de imagens em bases de dados digitais com informação semântica. Este é, no entanto, um problema extremamente complexo uma vez que não é fácil obter modelos analíticos que descrevam a aparência de objectos nas imagens digitais. Certos aspectos muito difíceis de modelar, como efeitos da iluminação, reflectância dos materiais e artefactos do processo de aquisição de imagem (ruído, saturação), impedem a criação de modelos físicos da formação de imagens. Por este motivo, a maioria das abordagens de reconhecimento de objectos actuais opta por “aprender” o modelo de forma estatística a partir de conjuntos de dados representativos da variabilidade da aparência dos objectos nas condições de funcionamento pretendidas. A obtenção destes conjuntos de dados é hoje em dia potenciada pela existência de ferramentas de procura (google) que contêm bases de dados extensas de imagens de objectos. O desenvolvimento de algoritmos de reconhecimento de objectos tipicamente requer dados de treino que (i) contenham a identificação da classe dos objectos na imagem, e (ii) a sua localização precisa. Embora o primeiro tipo de informação seja relativamente simples de fornecer, o segundo ponto é bastante mais moroso uma vez que requer que um humano delimite manualmente a região envolvente a cada um dos objecto de interesse na imagem. Nesta comunicação apresenta-se um método que permite utilizar apenas a informação das classes de objectos na imagem, sem necessitar da sua localização, para “aprender” representações de objectos e treinar classificad-

res para detectar, localizar e identificar objectos da mesma classe em imagens de teste. Cada imagem de treino pode ter múltiplas anotações correspondendo a múltiplas instancias de objectos. Através da co-ocorrência estatística entre características discriminantes na imagem e as anotações efectuadas, é possível obter representações para cada um dos objectos de interesse. Este problema tem sido abordado por diversos autores utilizando principalmente extensões a múltiplas classes de classificadores discriminativos [1],[2],[3]. No entanto as abordagens discriminativas são muito sensíveis a erros nas anotações e dados incompletos porque não permitem incorporar modelos arbitrários de ruído. Para abordar este problema, o nosso método explora a linearidade da representação de imagens baseada em histogramas (o histograma de uma imagem é a combinação linear dos histogramas dos objectos constituintes, adicionado a ruído de fundo) para propor um algoritmo baseado na minimização da característica de uma matriz de dados com entradas incompletas e ruidosas. Esta matriz concatena em colunas os histogramas representativos das imagens, quer de treino quer de teste, com as anotações fornecidas para as imagens de treino. O método proposto, inspirado em [4], permite estimar as classes de objectos presentes numa imagem de teste completando as entradas da matriz de dados correspondente às anotações desconhecidas. De forma análoga, o método permite estimar o histograma “canónico” de uma classe de objectos através da reconstrução das entradas desconhecidas do histograma pretendido. Apesar da minimização da característica de uma matriz com dados incompletos ser um problema não-convexo, resultados recentes [5] demonstram que, sob condições fracas, a solução pode ser obtida de forma convexa através da minimização da norma nuclear (soma dos valores singulares). Desta forma, o método apresentado propõe a minimização de um funcional convexo que inclui a norma nuclear da matriz de dados e termos que penalizam o desvio das entradas conhecidas da matriz de dados original face às entradas estimadas. O algoritmo de minimização desenvolvido aplica um método de ponto fixo [6] ao funcional proposto considerando as restrições típicas das representações em histograma (positividade e normalização). Em trabalho anterior foi comprovada a convergência do método [7] e várias experiências efectuadas em bases de dados de imagens de validação mostram um desempenho equivalente ou superior a métodos de estado-da-arte, sem necessitar de anotações relativas à localização dos objectos nas imagens de treino.

Referências

- [1] Lavrenko, V., Manmatha, R., Jeon, J. (2003) *A model for learning the semantics of pictures*, NIPS.
- [2] Wang, H., Ding, C., Huang, H. (2010) *Multi-label linear discriminant analysis*, ECCV.
- [3] Zhou, Z., Zhang, M. (2006) *Multi-instance multi-label learning with application to scene classification*, NIPS.
- [4] Goldberg, A., Zhu, X., Recht, B., Xu, J., Nowak, R. (2010) *Transduction with matrix completion: Three birds with one stone*, NIPS.
- [5] Candès, E., Recht, B. (2009) *Exact matrix completion via convex optimization*, Found Comput Math 9:717–772.
- [6] Ma, S., Goldfarb, D., Chen, L. (2011) *Fixed point and Bregman iterative methods for matrix rank minimization*, Mathematical Programming 128:321–353.
- [7] Cabral, S., De la Torre, F., Bernardino, A., Costeira, J. (2011) *Matrix completion for multi-label image classification*, NIPS.

109 anos de Análise Factorial

(Comunicação)

João A. BRANCO – *CEMAT-DM-IST*

jbranco@math.ist.utl.pt

Resumo

A apresentação da Análise Factorial à comunidade científica em geral aconteceu em 1904, com a publicação do artigo seminal de Charles Spearman intitulado “General Intelligence, Objectively Determined and Measured”. Decorridos 109 anos a Análise Factorial continua a provar a sua utilidade como ferramenta preferida em muitos estudos nos mais variados domínios científicos. Apesar do saber acumulado sobre Análise Factorial ser muito vasto, tenta-se, nesta breve comunicação, interpretar o segredo do sucesso deste método de Análise Multivariada e resumir os passos mais marcantes da sua evolução até aos nossos dias.

Testes robustos para modelos lineares generalizados com respostas omissas

(Comunicação)

Isabel M. RODRIGUES – *Departamento de Matemática e CEMAT, Instituto Superior Técnico, Technical University of Lisbon (TULisbon), Lisboa, Portugal*

irodrig@math.ist.utl.pt

Ana M. Bianco – *Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and CONICET, Argentina*

Graciela Boente – *Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and CONICET, Argentina*

Resumo

Em muitas situações em que os dados seguem um modelo linear generalizado a média das respostas é bem modelada considerando uma função (de ligação) linear das covariáveis. Com o objectivo de construir testes de hipóteses robustos para os parâmetros de regressão, neste trabalho são estudados novos estimadores robustos desses parâmetros, para conjuntos de dados com respostas completas ou incompletas. O comportamento assintótico dos estimadores robustos foi estudado baixo a hipótese nula e baixo alternativas contíguas, o que permitiu obter uma versão robusta da estatística de Wald. O grau de robustez desta estatística foi estudado através da função de influência do funcional correspondente. O comportamento da nova estatística, para amostra finitas, foi analisado num estudo de simulação de Monte Carlo.

Introdução

Os modelos lineares generalizados (MLG) são uma técnica popular para a modelação de uma vasta variedade de dados. Assumem que as observações (y_i, \mathbf{x}_i) , para $i = 1, \dots, n$, com $\mathbf{x}_i \in R^p$, são independentes com a mesma distribuição de $(y, \mathbf{x}) \in R^{p+1}$, e a distribuição condicional de $y|\mathbf{x}$ pertence à

família exponencial canónica

$$\exp \{ [y\theta(\mathbf{x}) - B(\theta(\mathbf{x}))] / A(\tau) + C(y, \tau) \} ,$$

onde A , B e C são funções conhecidas. Nesta situação, se denotarmos por B' a derivada de B , a média $\mu(\mathbf{x}) = E(y|\mathbf{x}) = B'(\theta(\mathbf{x}))$ é modelada linearmente à com uma função de ligação conhecida, g , i.e., $g(\mu(\mathbf{x})) = \theta(\mathbf{x}) = \mathbf{x}^T \boldsymbol{\beta}$. Procedimentos robustos para modelos lineares generalizados foram considerados, entre outros, por Stephanski, Carroll e Ruppert (1986), Künsch, Stefanski e Carroll (1989), Bianco e Yohai (1995), Cantoni e Ronchetti (2001), Croux e Haesbroeck (2003) e Bianco, García Ben e Yohai (2005). Recentemente, testes robustos para os parâmetros de regressão do modelo logístico foram estudados por Bianco e Martínez (2009).

Na prática, pode ocorrer que algumas variáveis resposta sejam incompletas e os métodos anteriormente referidos, delineados para conjuntos de dados completos, são inadequados. Neste trabalho desenvolvemos procedimentos inferenciais robustos para conjuntos de dados com observações incompletas da variável resposta mas com a covariável \mathbf{x} completamente observada. São introduzidos procedimentos robustos para estimar o parâmetro $\boldsymbol{\beta}$, sob o MLG, o qual inclui para o caso dos dados completos a família de estimadores anteriormente referida. Mostrou-se que o estimador robusto de $\boldsymbol{\beta}$ é consistente e assimpoticamente normal (com taxa de convergência \sqrt{n}). Um teste de hipóteses robusto, do tipo Wald, para testar $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ foi também estudado. Foi deduzida a função de influência para o funcional relacionado com a estatística de teste, o que permitiu avaliar o grau de robustez. Um estudo de simulação de Monte Carlo, com diferentes esquemas de contaminação, permitiu comparar o comportamento da nova estatística de teste, para amostras finitas, com o de outras estatísticas robustas e o da estatística clássica.

Referências

- [1] Bianco, A., García Ben, M. e Yohai, V. (2005). Robust estimation for linear regression with asymmetric errors. *Canad. J. Statist.* **33**, 511-528.
- [2] Bianco, A. e Martínez, E. (2009). Robust testing in the logistic regression model. *Comp. Statist. Data Anal.* **53**, 4095-4105.
- [3] Bianco, A. e Yohai, V. (1995). Robust estimation in the logistic regression model. *Lecture Notes in Statistics*, **109**, 17-34. Springer-Verlag, New York.

- [4] Cantoni, E. e Ronchetti, E. (2001). Robust Inference for Generalized Linear Models. *Journal of the American Statistical Association* , **96**, 1022-1030.
- [5] Croux, C. e Haesbroeck, G. (2003). Implementing the Bianco and Yohai estimator for logistic regression. *Comp. Statist. Data Anal.*, **44**, 273-295.
- [6] Künsch, H., Stefanski, L. e Carroll, R. (1989). Conditionally unbiased bounded influence estimation in general regression models with applications to generalized linear models. *J. Amer. Assoc.* **84**, 460-466.
- [6] Stefanski, L., Carroll, R. e Ruppert, D. (1986). Bounded score functions for generalized linear models. *Biometrika* **73**, 413-424.

Validação em classificação hierárquica: da teoria às aplicações

(Comunicação)

Fernanda SOUSA – *Faculdade de Engenharia e CITTA, Universidade do Porto*
fcsousa@fe.up.pt

Resumo

Associado à metodologia de classificação hierárquica (ascendente) está um procedimento que constrói uma sucessão de partições encaixadas, sobre o conjunto de elementos a classificar, constituída por partições com um número sucessivamente menor de classes. De uma forma geral assume-se que a partição inicial é composta exclusivamente por singletões, em cada etapa da hierarquia há lugar à reunião de um (ou mais) par de classes. Não havendo critério de paragem o processo termina quando todos os elementos estão reunidos numa só classe. Trata-se assim de um procedimento que, sendo automático a produzir a hierarquia de partições, deixa ao utilizador as tarefas de escolha do nível de corte da hierarquia, correspondente à “melhor partição” e de avaliação da qualidade/validade das classes que a constituem, individualmente ou no contexto da partição [1].

Acresce ainda o facto do resultado de um método de classificação hierárquica depender não só do conjunto de dados a que é aplicado, mas também da medida usada para comparar pares de elementos e, de forma mais acentuada, do critério que conduz à agregação das classes consideradas mais semelhantes. A dependência, que o resultado de uma classificação apresenta relativamente às medidas de comparação usadas é bem conhecida e a sua avaliação é um passo importante na análise dos resultados, [3] e [4]. Esta problemática insere-se na área da validação em classificação hierárquica, que tem como objectivo avaliar a qualidade das estruturas produzidas, sejam classes, partições ou hierarquias, suportando a tomada de decisões que conduzam à obtenção de estruturas em classes que reflectam verdadeiras relações de afinidade ou afastamento entre os elementos do conjunto a classificar, [2] e [3].

Neste trabalho serão apresentados i) resultados teóricos, no domínio da com-

paração de estruturas de classificação, dos índices de avaliação de classes, de partições e de ajuda à escolha da “melhor partição” ii) abordagens para a combinação de diferentes classificações, sobre um mesmo conjunto de dados, com vista à obtenção de uma classificação mais estável e mais consistente iii) estudos empíricos que visam ser uma contribuição para esta problemática iv) alguns exemplos de aplicação.

Referências

- [1] Gordon, A.D. (1999) *Classification*, 2nd ed., Chapman & Hall, London.
- [2] Halkidi M., Batistakis Y., Vazirgiannis M. (2001) *On Clustering Validation Techniques*, *Journal of Intelligent Information Systems*, 17, 2-3, 107–145.
- [3] Sousa, F. (2000) *Novas metodologias e validação em classificação hierárquica ascendente*, Dissertação de Doutoramento, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa.
- [4] Sousa, F., Tendeiro, J. (2005) *A Validation Methodology in Hierarchical Clustering*, *International Symposium of Applied Stochastic Models and Data Analysis (ASMDA 2005)*, Editors: J. Janssen, P. Lenca, 396–403, Brest (France).

Extremos

Dependência extremal: risco de contágio de valores extremos

(Comunicação)

Marta FERREIRA – *Centro de Matemática da Universidade do Minho/DMA*
msferreira@math.uminho.pt

Helena Ferreira – *Departamento de Matemática da Universidade da Beira Interior*

Resumo

O fenómeno da globalização, juntamente com um relaxamento na supervisão dos mercados financeiros, tornou-os mais vulneráveis e mais dependentes entre si. A ocorrência de grandes perdas em mercados fortes acaba por se reflectir ao nível das principais bolsas mundiais, e vice-versa. A necessidade de medir esta interdependência extremal conduziu ao aparecimento de diversos coeficientes no seio da teoria multivariada de valores extremos. Neste trabalho apresenta-se um novo coeficiente para a dependência extremal entre dois vetores aleatórios, estendendo medidas existentes na literatura (Ferreira e Ferreira, 2012). A estimação será também abordada e uma ilustração do conceito será feita com dados reais.

Referências

[1] Ferreira, H., Ferreira, M. (2012) *On extremal dependence of block vectors*, *Kybernetika* 48(5), 988–1006.

Dicotomia extremal para sistemas caóticos

(Comunicação)

Jorge Milhazes FREITAS – *Centro de Matemática e Faculdade de Ciências da Universidade do Porto*

jmfreita@fc.up.pt

Resumo

Considere-se um sistema dinâmico discreto dado por uma aplicação mensurável $f : M \rightarrow M$ onde $(M, \mathcal{B}, \mathbb{P})$ é um espaço de probabilidade tal que \mathbb{P} é uma probabilidade definida nos boreleanos, \mathcal{B} , de uma variedade compacta, M . Por hipótese \mathbb{P} é invariante, i.e., $\mathbb{P}(A) = \mathbb{P}(f^{-1}(A))$ para todo $A \in \mathcal{B}$. Seja X_0, X_1, \dots um processo estocástico estacionário gerado por este sistema do seguinte modo:

$$X_i = \varphi \circ f^i,$$

para todo $i \geq 0$, onde $\varphi : M \rightarrow \mathbb{R}$ é uma variável aleatória contínua que atinge um máximo global em $\zeta \in M$.

Considere-se o processo pontual

$$N_n([0, t]) = \sum_{j=0}^{v_n t} \mathbf{1}_{\{X_j \geq u_n\}}$$

onde $(u_n)_n$ é tal que $n\mathbb{P}(X_0 > u_n) \rightarrow \tau$ para algum $\tau \geq 0$ e $(v_n)_n$ é dado por $v_n = (\mathbb{P}(X_0 > u_n))^{-1}$.

Certos sistemas caóticos têm boas propriedades misturadoras, i.e., perdem memória rapidamente. Uma forma de medir a rapidez da perda de memória é através do cálculo de taxas de decaimento de correlações. Veremos que para sistemas com boas taxas de decaimento de correlações é possível provar uma dicotomia no que respeita ao comportamento de N_n relativamente à periodicidade (ou não) do ponto ζ . Nomeadamente, veremos que se ζ é um ponto não-periódico (ou seja, $f^i(\zeta) \neq \zeta$, para todo $i = 1, 2, \dots$) então N_n converge em distribuição para um processo de Poisson simples de intensidade 1. No caso de ζ ser um ponto periódico repulsor de período p ($f^p(\zeta) = \zeta$ e

$f^i(\zeta) \neq \zeta$ para $i = 1, 2, \dots, p - 1$) então N_n converge para um processo de Poisson composto em que os eventos se distribuem na linha real de acordo com um processo de Poisson de intensidade $\theta < 1$ e têm uma distribuição de multiplicidade geométrica. O parâmetro θ é o Índice Extremal e pode ser calculado pela expansão de f em ζ .

Estes resultados foram demonstrados em [1, 2].

Referências

- [1] Freitas, A.C.M., Freitas, J.M., Todd, M. (2013) *The Compound Poisson Limit Ruling Periodic Extreme Behaviour of Non-Uniformly Hyperbolic Dynamics*, Comm. Math. Phys. 321, 483–527.
- [2] Aytaç, H., Freitas, J.M., Vaienti, S. (2012) *Laws of rare events for deterministic and random dynamical systems*, Preprint 2012, arXiv:1207.5188.

Índice extremal e periodicidade em sistemas dinâmicos caóticos

(Comunicação)

Ana Cristina MOREIRA FREITAS – *CMUP e FEP, Universidade do Porto*
amoreira@fep.up.pt

Resumo

Neste trabalho estudamos a distribuição assintótica do máximo parcial de variáveis aleatórias observáveis avaliadas ao longo das órbitas de certos sistemas dinâmicos. Consideremos um sistema dinâmico discreto (X, B, μ, f) , em que X é uma variedade Riemanniana de dimensão d , B é a σ -álgebra de Borel, $f : X \rightarrow X$ é uma aplicação mensurável e μ uma medida de probabilidade f -invariante. Consideremos um observável $\varphi : \mathcal{X} \rightarrow \mathbb{R} \cup \{\pm\infty\}$ que atinge um máximo global em $\zeta \in \mathcal{X}$ e o processo estocástico estacionário X_0, X_1, \dots dado por

$$X_n = \varphi \circ f^n, \quad \text{para cada } n \in \mathbb{N}. \quad (1)$$

Suponhamos que φ atinge um máximo global em $\zeta \in \mathcal{X}$ e que o acontecimento $\{x \in \mathcal{X} : \varphi(x) > u\} = \{X_0 > u\}$ corresponde a uma bola topológica centrada em ζ (veja-se, por exemplo, [2]).

Definamos o máximo parcial

$$M_n := \max\{X_0, \dots, X_{n-1}\}.$$

Dizemos que temos uma *Lei de Valores Extremos* (LVE) para M_n se existir uma f.d. não-degenerada $H : \mathbb{R} \rightarrow [0, 1]$ com $H(0) = 0$ e, para todo $\tau > 0$, existir uma sucessão de níveis $u_n = u_n(\tau)$, $n = 1, 2, \dots$, tais que

$$n\mathbb{P}(X_0 > u_n) \rightarrow \tau, \quad \text{as } n \rightarrow \infty, \quad (2)$$

e para os quais se tem:

$$\mathbb{P}(M_n \leq u_n) \rightarrow \bar{H}(\tau), \quad \text{as } n \rightarrow \infty.$$

Como em [4], denotamos as condições de dependência por $D(u_n)$ e $D'(u_n)$. Em [1] propusemos uma versão mais fraca de $D(u_n)$, que denotamos por $D_2(u_n)$, que é consequência imediata do decaimento de correlações suficientemente rápido para observáveis que são de variação limitada ou Hölder contínuos.

Diz-se que $D_2(u_n)$ se verifica para a sucessão X_0, X_1, \dots se para quaisquer inteiros ℓ, t e n

$$\left| \mu(\{X_0 > u_n\} \cap \{\max\{X_t, \dots, X_{t+\ell-1}\} \leq u_n\}) - \mu(\{X_0 > u_n\})\mu(\{M_\ell \leq u_n\}) \right| \leq \gamma(n, t),$$

onde $\gamma(n, t)$ é não-crescente em t para cada n e $n\gamma(n, t_n) \rightarrow 0$ quando $n \rightarrow \infty$ para alguma sucessão $t_n = o(n)$.

Diz-se que $D'(u_n)$ se verifica para a sucessão X_0, X_1, \dots se $\lim_{k \rightarrow \infty} \limsup_{n \rightarrow \infty} n \sum_{j=1}^{\lfloor n/k \rfloor} \mu(\{X_0 > u_n\} \cap \{X_j > u_n\}) = 0$.

O resultado principal de [1] estabelece que, se $D_2(u_n)$ e $D'(u_n)$ se verificam para o processo X_0, X_1, \dots e para uma sucessão de níveis que satisfazem (2), então $\bar{H}(\tau) = e^{-\tau}$.

Consideremos agora que $\zeta \in \mathcal{X}$ é um ponto periódico repulsor, de período $p \in \mathbb{N}$, onde μ é uma medida absolutamente contínua com respeito a Lebesgue. Neste caso, $\mu(\{X_0 > u\} \cap \{X_p > u\}) \sim (1 - \theta)\mu(X_0 > u)$, para algum $\theta \in (0, 1)$, o que implica que $D'(u_n)$ não é satisfeita.

De modo a provar a existência de lei limite para o máximo na presença de um fenómeno periódico que cria clustering, introduzimos duas novas condições, que denotamos por $D^p(u_n)$ e $D'_p(u_n)$, essencialmente substituindo nas condições $D_2(u_n)$ e $D'(u_n)$ o papel das “excedências” (que correspondem a entradas em bolas) por aquilo a que chamamos “escapes” (que correspondem a entradas em anéis) (veja-se [3]).

Teorema. Seja $(u_n)_{n \in \mathbb{N}}$ tal que $n\mathbb{P}(X > u_n) = n(1 - F(u_n)) \rightarrow \tau$, quando $n \rightarrow \infty$, para algum $\tau \geq 0$. Consideremos um processo estocástico estacionário X_0, X_1, X_2, \dots definido por (1), em que φ atinge um máximo global num ponto periódico repulsor $\zeta \in \mathcal{X}$, de período $p \in \mathbb{N}$, e $\theta \in (0, 1)$. Assumamos ainda que $D^p(u_n)$ e $D'_p(u_n)$ se verificam. Então,

$$\lim_{n \rightarrow \infty} \mathbb{P}(M_n \leq u_n) = e^{-\theta\tau}.$$

Referências

- [1] Freitas, A.C.M., Freitas, J.M. (2008) *Hitting time statistics and extreme value theory*, Stat. & Prob. Letters 78, 1088–1093.
- [2] Freitas, A.C.M., Freitas, J.M., Todd, M. (2010) *On the link between dependence and independence in extreme value theory for dynamical systems*, Probab. Theory Relat. Fields 147, 675–710.
- [3] Freitas, A.C.M., Freitas, J.M., Todd, M. (2012) *Hitting Time Statistics and periodicity*, Adv. Math. 231, 2626–2665.
- [4] Lindgren, G., Leadbetter, M.R., Rootzén, H. (1983) *Extremes and related properties of random sequences and processes*, Springer Series in Statistics, Springer-Verlag, New York-Berlin.

Modelação de grandes incêndios em Portugal

(Comunicação)

Alexandra RAMOS – *FEP e CMUP, Universidade do Porto*

aramos@fep.up.pt

Resumo

Em Ramos e Ledford (2009) foi construído um modelo paramétrico para a cauda conjunta de uma distribuição bivariada. Usando este flexível modelo paramétrico, analisamos a dependência temporal de curta duração numa série temporal estacionária, caracterizando o comportamento extremal da série. Este tipo de caracterização tem sido feita, na sua grande maioria, com base no pressuposto da série temporal ter a propriedade de Markov, visto a classe das cadeias de Markov de ordem d ser suficientemente geral e tratável. A modelação da cauda conjunta da distribuição de um par de valores consecutivos (X_i, X_{i+1}) de uma cadeia de Markov estacionária de primeira ordem será então aqui realizada através do modelo descrito em Ramos e Ledford (2009).

Os grandes incêndios podem ter um efeito devastador no meio ambiente, economia e na vida humana. Recentemente, os incêndios têm causado danos significativos no clima e no eco-sistema. Com o objectivo de estudar a distribuição da dimensão dos incêndios em Portugal, aplicamos essa abordagem a dados consistindo em áreas ardidas. Esta análise requer a avaliação conjunta dos extremos da estrutura de dependência temporal, bem como da cauda da distribuição marginal. Em particular, estudamos o grau de dependência extremal na cauda conjunta superior de duas observações consecutivas e o peso da cauda da série.

Referências

- [1] Ramos, A. and Ledford, A. (2009) *A new class of models for bivariate joint tails*, J. R. Statist. Soc., B, 71(1), 219–241.

Modelos Estocásticos

Modelação do crescimento individual em ambiente aleatório

(Comunicação)

Patrícia A. FILIPE – *CIMA, Universidade de Évora*

pasf@uevora.pt

Carlos A. Braumann – *CIMA, Universidade de Évora*

Nuno Brites – *Instituto Português do Mar e da Atmosfera e CIMA, Universidade de Évora*

Clara Carlos – *ESTBarreiro, Instituto Politécnico de Setúbal e CIMA, Universidade de Évora*

Resumo

A modelação do crescimento individual de organismos deve ter em conta o efeito das flutuações aleatórias do ambiente que afectam a taxa de crescimento. Os modelos de equações diferenciais estocásticas (EDE) são construídos precisamente de forma a incorporar a dinâmica do processo de crescimento e o efeito das flutuações aleatórias ambientais nessa dinâmica.

A classe de modelos de EDE apresentada conduz a versões estocásticas de um modelo determinístico geral simples e flexível que inclui como casos particulares alguns dos modelos de crescimento determinísticos mais utilizados, como é o caso, por exemplo, do modelo monomolecular, do modelo logístico, do modelo de Gompertz e do modelo de Bertalanffy-Richards. Esta classe de modelos de EDE possui um termo adicional, que incorpora o efeito das flutuações aleatórias do ambiente. Considera-se o caso em que existe uma única forma funcional para descrever a dinâmica média da curva de crescimento completa (modelo monofásico), que nos conduz ao modelo de Ornstein-Uhlenbeck generalizado ou modelo de Vasicek conhecido no contexto da dinâmica de taxas de juro e para o qual podemos encontrar algumas aplicações ao crescimento individual de organismos, como por exemplo, ao crescimento de árvores[1] ou peixes[2]. Apresenta-se a generalização ao caso em que o coeficiente de crescimento assume valores diferentes para diferentes fases da vida do indivíduo (modelo multifásico)[3]. Também é abordado o caso em que se considera que o tamanho médio assintótico varia aleatoriamente de indivíduo para indi-

víduo, já que é de esperar que existam diferenças genéticas e outras entre eles (modelo de efeitos mistos). São apresentadas as principais propriedades que caracterizam cada um destes três casos, em particular, as soluções dos modelos e correspondentes função densidade, função densidade assintótica e função densidade de transição. Em termos de estimação, foram aplicados três métodos: o método de máxima verosimilhança, métodos bootstrap (paramétrico e não-paramétrico) e um método de estimação não-paramétrica[4]. Para cada um destes métodos considera-se o caso em que trabalhamos com os dados de uma trajectória (um indivíduo) e a extensão ao caso em que temos dados de várias trajectórias (vários indivíduos). Um estudo de previsão foi realizado com base nos modelos de EDE monofásico e bifásico. Mostrou-se que os modelos de EDE são os mais adequados em termos de previsão quando comparados com os correspondentes determinísticos para os quais são usualmente aplicados métodos de regressão clássicos[5].

Para ilustração dos resultados foram utilizados dados do peso de bovinos de raça Mertolenga. A aplicação dos modelos desenvolvidos pode ser de grande utilidade em diversas áreas, e em particular na área de criação de gado. Um maior conhecimento acerca do crescimento dos animais, permite aos criadores, por exemplo, do ponto de vista económico, otimizar o lucro médio de venda dos animais. Com este objetivo, foi abordada a questão de optimização do lucro médio de venda de um animal. Por um lado, com base nos nossos modelos, pode ser calculado o lucro esperado para diferentes idades do animal, e em particular, podemos determinar a idade óptima de venda do animal de forma a que esse lucro seja máximo. Pode também ser obtida a distribuição de probabilidade do lucro de venda e calcular probabilidades envolvendo esse lucro. Por outro lado, sabendo qual o peso do animal exigido pelo mercado, podem ser estudadas as propriedades do tempo que o animal demora a atingir esse peso pela primeira vez. Obteve-se a média, a variância destes tempos, assim como, expressões para a sua função densidade de probabilidade e função distribuição. Com base nestes resultados pode ser determinado qual o peso óptimo de venda do animal em termos de maximização do lucro médio de venda[6].

Referências

- [1] Garcia, O. (1983). *A stochastic differential equation model for the height of forest stands*, *Biometrics*, 39:1059–1072.
- [2] Qiming, Lv. and Pitchford, J. (2007). *Stochastic Von Bertalanffy models, with applications to fish recruitment*, *Journal of Theoretical Biology*, 244: 640–655.
- [3] Filipe, P.A., Braumann, C.A. e Roquete, C.J. (2012). *Multiphasic individual growth models in random environments*, *Methodology and Computing in Applied Probability*, 14(1):49-56.
- [4] Filipe, P.A., Braumann, C.A., Brites, N.M. e Roquete, C.J. (2010). *Modelling animal growth in random environments: An Application Using Non-parametric Estimation*, *Biometrical Journal*, 52(5):653-666.
- [5] Filipe, P.A., Braumann, C.A, Brites, N.M. e Roquete, C.J. (2013). *Prediction for individual growth in a random environment*, *Recent Developments in Modeling and Applications in Statistics, Studies in Theoretical and Applied Statistics*. Oliveira, P.E., Graça Temido, M., Henriques, C. e Vichi, M. (Eds.), Springer Berlin Heidelberg (ISBN 978-3-642-32419-2), 193-201.
- [6] Braumann, C.A, Filipe, P.A., Carlos, C. e Roquete, C.J. (2011). *Stochastic differential equations general models of individual growth in uncertain environments and application to profit optimization in livestock production*, *Bulletin of the International Statistical Institute, Proceedings of the 58th World Statistics Congress 2011, Dublin (ISBN: 978-90-73592-33-9)*, 4450-4455.

Estimação das características do tráfego numa fila M/G/1 utilizando sondas

(Comunicação)

Gonçalo JACINTO – *CIMA-UE e ECT da Universidade de Évora*
gjcj@uevora.pt

Nelson Antunes – *CEMAT e FCT da Universidade do Algarve*
nantunes@ualg.pt

António Pacheco – *CEMAT e Departamento de Matemática, Instituto Superior Técnico, Universidade de Lisboa*
apacheco@math.ist.utl.pt

Resumo

Motivado pela estimação de tráfego na Internet utilizando medições activas, consideramos neste trabalho a estimação da taxa de chegadas e dos momentos do tempo de serviço de uma fila M/G/1 utilizando sondas, isto é, quando clientes especiais entram no sistema. Os tempos entre chegadas das sondas são independentes e identicamente distribuídos e os tempos de serviço seguem uma distribuição geral positiva. As únicas observações usadas são os tempos de chegada, os tempos de serviço e os tempos de partida das sondas. São derivadas as equações principais a partir das quais as quantidades de interesse podem ser estimadas. No final, são comparados dois tipos particulares de tempos entre chegadas das sondas, determinístico e Poisson.

Referências

- [1] Baccelli, F., Kauffmann, B and Veitch, D. (2009). *Inverse problems in queueing theory and Internet probing*, Queueing Systems, 63(3), 59-107.
- [2] Nam, S., Kim, S. and Sung, D. (2009). *Estimation of available bandwidth for an M/G/1 queueing system*, Appl. Math. Model., 8(33), 3299-3308.
- [3] Novak, A., Taylor, P. and Veitch, D. (2006). *The distribution of the number of arrivals in a subinterval of a busy period in a single server queue*, Queueing Systems, 3(53), 105-114.
- [4] Sharma, V. and Mazumdar, R. (1998). *Estimating traffic parameters in queueing systems with local information*, Performance Evaluation, 32(3), 217-230.

Value of an option to exit the market

(Comunicação)

Carlos OLIVEIRA – *IST and CEMAT*

carlos_m_oliveira88@hotmail.com

Manuel Guerra – *ISEG and CEMAPRE*

Cláudia Nunes – *IST and CEMAT*

Resumo

Assume that a company is producing an established product, in a declining market. This company has the option to exit the market, paying a sunk cost. This decision is irreversible and once decided to exit the market, its profits are zero. The question is to derive the optimal decision of exiting the market, meaning that we want to derive when it is optimally to the company to take this decision. We assume that the source of uncertainty is the demand, that follows a geometric brownian motion with known drift and volatility.

This problem is in fact an optimal stopping problem. For such a problem we derive the Hamilton-Jacobi-Bellman equation, and we solve it, providing then the optimal decision of the company.

The decision can be stated as the existence of two regions (one for producing - *continuation region*) and one for stopping - *stopping region*), and the boundary represents the value that triggers the exit decision.

Multi-type linear-fractional branching processes

(Comunicação)

Maria Conceição SERRA – *Centro de Matemática da Universidade do Minho*
 mcserra@math.uminho.pt

Serik Sagitov – *Chalmers University of Technology e Gothenburg University*

Resumo

Branching processes is an active research area, with applications in various other sciences, such as biology, medicine, genetics, etc.

The simplest version of this processes are the Binaymé-Galton-Watson (BGW) processes, which describe populations of individuals living one unit of time (usually generations) and, at moment of death, give birth to a random number of individuals, always according to the same reproduction law $\{p(k), k \in \mathbb{N}_0\}$ and independently of each other.

In a linear-fractional BGW process the probability generating function of the reproduction law is of the form

$$\phi(s) = h_0 + \frac{h_1 s}{1 + m - ms}, \quad s \in [0, 1],$$

with $0 \leq h_0 \leq 1$, $h_1 = 1 - h_0$ and $m > 0$. This is a special class of BGW since it is possible to derive explicit calculations for various important entities regarding these processes, which are not possible for the most general processes. In this talk we present some examples of such calculations.

In the multi-type setting, individuals still reproduce independently of each other, but the reproduction law may depend on the mother's type. We present a multi-type version of the linear-fractional BGW process, which relies on the theory of BGW-processes with countably many types. Finally, we present results concerning the time it takes to observe an individual of a certain large type N , as $N \rightarrow \infty$.

Risco e Atuariado

Measuring the impact of a bonus malus system in finite and continuous time ruin probabilities, for large portfolios in motor insurance

(Comunicação)

Lourdes Belchior AFONSO – *CMA and FCT, Universidade Nova de Lisboa*
lbafonso@fct.unl.pt

Rui Cardoso – *CMA and FCT, Universidade Nova de Lisboa*

Alfredo Egídio dos Reis – *CEMAPRE and ISEG, Universidade Técnica de Lisboa*

Gracinda Rita Guerreiro – *CMA and FCT, Universidade Nova de Lisboa*

Resumo

We are going to consider the posterior ratemaking (experience rating) in automobile insurance. Premium calculation is based on past experience bringing a bigger variation in the premia during the lifetime of the portfolio. We are addressing the problem of the ruin probability of the automobile portfolio. Ruin probability is usually computed using the classical Cramér-Lundberg model that considers the premium to be paid continuously at a constant rate. Afonso *et al.* [1] consider a model applicable to large portfolios where a varying premium is used by means of a mix of calculation and simulation. That differs from the usual literature and allow us to obtain fast results for finite horizon in continuous time. The ideas in the model by Afonso *et al.* [1] can be brought and applied in motor insurance ratemaking, for two main reasons: First, because premium calculation is usually applied for large portfolios where the model applies, second, because premium calculation is based on the claim record. However, the model needs to be changed to fit in the common features in motor insurance. Usually, in motor insurance, experience rating changes the premium, not as a function of the aggregate claim record but as a function of the claim number record. This is usually approached by a Markov chain procedure. The number of claims is essential to determine the next year rating class and calculate the applicable premium. That, together with the aggregate claims, is necessary to compute ruin probabilities for the

portfolio. We will measure the impact of a bonus malus system (BMS) in the ruin probabilities, considering different known optimal scales as well as real commercial scales. In all scenarios we will use real data from automobile third-party liability portfolio of a Portuguese insurer.

Referências

- [1] Afonso, L.B., Egídio dos Reis, A.D., Waters, H.R. (2004) *Calculating continuous time ruin probabilities for a large portfolio with varying premiums*, ASTIN Bulletin 39 (1) , 117–136.

On a Sparre-Andersen risk model with $\text{PH}(n)$ interclaim times

(Comunicação)

Agnieszka BERGEL – *CEMAPRE and ISEG, Technical University of Lisbon, Portugal*

agnieszka@iseg.utl.pt

Alfredo D. Egídio dos Reis – *CEMAPRE and ISEG, Technical University of Lisbon, Portugal*

Resumo

For actuarial applications we consider the Sparre–Andersen risk model when the interclaim times follow a Phase–Type distribution, $\text{PH}(n)$.

Lundberg’s equation, is a major subject of study for the computation of ruin probabilities. This equation came first into light for the study for the well known Lundberg’s inequality and adjustment coefficient, being the former as an upper bound for the ultimate ruin probability. Nowadays, the roots of the Lundberg’s equation play an important role in the calculation of many quantities that are fundamental in risk and ruin theory. Namely, the ultimate and finite time ruin probabilities, the probability of arrival to a barrier prior to ruin, severity of ruin and its maximum, the expected discounted future dividends, among others.

All those calculations depend on the nature of the roots of the Lundberg’s equation, particularly its roots with positive real parts. There are several papers that have been devoted to the subject, namely Albrecher and Boxma (2005), Dickson and Waters (2004), Li and Garrido (2004a,b), Ren (2007) and some others. But in all those works it is always assumed that the roots are distinct.

Our interest is to address two problems: First, to determine the cases when multiple roots can arise, with the highest possible level of accuracy; Second, to overcome the issues that can arise when the assumption of having different roots is removed. We will then be able to compute the quantities discussed above for the case of multiple roots.

For this purpose we will use many ideas that come from complex analysis, the theory of integro-differential equations and linear algebra. We first address the problem of solving an integro-differential equation that is satisfied by the probability that the surplus reaches a given level b from u without first falling below zero, and then show a method to solve such equation.

This is done by considering the roots with positive real parts of the generalized Lundberg's equation, and establishing a one-one relation between them and the solutions of the integro-differential equation mentioned above. We study the two cases of existence or non-existence of multiple roots.

Afterwards, we apply our findings above in the computation of the distribution of the maximum severity of ruin.

We illustrate by finding explicit formulae for some examples and values for the parameter n of the $\text{PH}(n)$ family, and some particular claim amount distributions.

Finally, we introduce interest force to the model, consider the calculation of the expected discounted dividends prior to ruin, by finding and solving an integro-differential equation. Numerical examples are also provided for illustration.

Referências

- [1] Albrecher, H., Boxma, O. J. (2005) *On the discounted penalty function in a Markov-dependent risk model*, Insurance: Mathematics and Economics 37, 650–672.
- [2] Bergel, A., Egídio dos Reis, A. D. (2013) *Further developments in the Erlang(n) risk model*, Scandinavian Actuarial Journal.
- [3] Dickson, D. C.M. (2005) *Insurance Risk and Ruin*, Cambridge University Press.
- [4] Dickson, D. C. M., Waters, H. R. (2004) *Some optimal dividends problems*, ASTIN Bulletin 34, 49–74.
- [5] Ji, L., Zhang, Ch. (2011) *Analysis of the multiple roots of the Lundberg fundamental equation in the $\text{PH}(n)$ risk model*, Applied Stochastic Models in Business and Industry.
- [6] Li, S. (2008) *A note on the maximum severity of ruin in an Erlang(n) risk*

process, Bulletin of the Swiss Association of Actuaries 167–180.

[7] Li, S., Garrido, J. (2004a) *On ruin for the Erlang(n) risk process*, Insurance: Mathematics and Economics 34(3), 391–408.

[8] Li, S., Garrido, J. (2004b) *On a class of renewal risk models with constant dividend barrier*, Insurance: Mathematics and Economics 35, 691–701.

[9] Ren, J. (2007) *The joint distribution of the surplus prior to ruin and the deficit at ruin in a Sparre Andersen model*, North American Actuarial Journal 11(3), 128–136.

Statistical Approach to Open Bonus Malus Systems

(Comunicação)

Gracinda Rita GUERREIRO – *Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa*

grg@fct.unl.pt

João Tiago Mexia – *Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa*

Maria de Fátima Miguens – *Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa*

Resumo

Following an open portfolio approach, we show how to estimate a Bonus Malus System evolution.

Considering a model for the number of new annual policies, we obtain ML estimators, asymptotic distributions and confidence regions for the expected number of new policies entering the portfolio in each year, as well as for the expected number and proportion of insureds in each bonus class, by year of enrollment.

Confidence regions for the distribution of policyholders result in confidence regions for optimal bonus scales.

Our treatment is illustrated by an example with numerical results.

Referências

- [1] Feller, W. (1966) *An Introduction to Probability Theory and its applications*, 2nd Edition, John Wiley and Sons, Inc.
- [2] Guerreiro, G.R., Mexia, J.T. (2004) *An Alternative Approach to Bonus Malus*, *Discussiones Mathematicae, Probability and Statistics*, 24(2), 197–213.
- [3] Guerreiro, G.R., Mexia, J.T. (2008) *Stochastic Vortices in Periodically Re-classified Populations*, *Discussiones Mathematicae, Probability and Statistics*,

28(2), 209–227.

[4] Guerreiro, G.R. and Mexia, J.T. and Miguens, M.F. (2010) *A Model for Open Populations Subject to Periodical Re-classifications*, Journal of Statistical Theory and Practice, 4(2), 303–321.

[5] Guerreiro, G.R. and Mexia, J.T. and Miguens, M.F. (2012) *Stable Distributions for Open Populations Subject to Periodical Re-classifications*, Journal of Statistical Theory and Practice, 6(4), 621–635.

[6] Guerreiro, G.R. and Mexia, J.T. and Miguens, M.F. (2013) Preliminary Results on Confidence Intervals for Open Bonus Systems, Advances in Regression, Survival Analysis, Extreme Values, Markov Processes and Other Statistical Applications, Springer.

[7] Guerreiro, G.R. and Mexia, J.T. and Miguens, M.F., *Statistical Approach to Open Bonus Malus Systems*, ASTIN Bulletin, accepted for publication.

[8] Lemaire, J. (1995) *Bonus-Malus Systems in Automobile Insurance*, Kluwer Academic Publishers.

[9] Schott, J. (1997) *Matrix Analysis for Statistics*, Wiley Series in Probability and Statistics.

[10] Tiago de Oliveira, J. (1982) *The Delta-method for Obtention of Asymptotic Distributions; Applications*, Publications de l'Institut de Statistique de l'Université de Paris, 27, 49–70.

The Erlang(n) dual risk model, discounted dividends, moments and optimal expectations

(Comunicação)

Eugenio V. RODRÍGUEZ-MARTÍNEZ – *CEMAPRE and ISEG, Technical University of Lisbon, Portugal*

evrodriguez@gmail.com

Alfredo D. Egídio dos Reis – *CEMAPRE and ISEG, Technical University of Lisbon, Portugal*

Rui M.R. Cardoso – *CMA and FCT, New University of Lisbon*

Resumo

The dual risk model assumes that the surplus of a company decreases at a constant rate over time, and grows by means of upward jumps which occur at random times and sizes. It has applications to companies with economical activities involved in research and development. This model is *dual* to the well known Cramér-Lundberg risk model with applications to insurance

Existing results on the study of the dual model assume that the random waiting times between consecutive gains follow an exponential distribution, as in the classical Cramér-Lundberg risk model.

We make generalizations to other compound renewal risk models where such waiting times are Erlang(n) distributed. In a previous work, Rodríguez-Martínez *et al.* (2012), using the roots of the fundamental and the generalized Lundberg's equation, performed calculations to obtain expressions for the ruin probability and the Laplace transform of the time of ruin for an arbitrary single gain distribution. For the computation of the expected discounted dividends they assumed that the gains follow a Phase-Type, PH(n), distribution.

In the present work we keep the assumption that the gains follow a Phase-Type distribution, and extend the computation to higher moments for the discounted dividends. We also develop a numerical method to determine the dividend barrier that maximizes the expected discounted dividends.

Finally, we perform illustrations working some examples for some particular gain distributions and obtain numerical results.

Referências

- [1] Afonso, L. B., Cardoso R. M. R., Egídio dos Reis, A. D. (2011) *Dividend problems in the dual risk model*, preprint.
- [2] Albrecher, H., Badescu, A., Landriault, D. (2008) *On the dual risk model with tax payments*, *Insurance: Mathematics and Economics*, 42(3), 1086–1094.
- [3] Avanzi, B., Gerber, H. U. (2008) *Optimal dividends in the dual model with diffusion*, *ASTIN Bulletin*, 38(2), 653–667.
- [4] Avanzi, B., Gerber, H. U., Shiu, E. S. W. (2007) *Optimal dividends in the dual model*, *Insurance: Mathematics and Economics* 41(1), 111–123.
- [5] Cramér H. (1955) *Collective Risk Theory: A Survey of the Theory from the Point of View of the Theory of Stochastic Process*, Ab Nordiska Bokhandeln, Stockholm.
- [6] Cheung, E. C. K., Drekić, S. (2008) *Dividend problems in the dual risk model: exact and approximate approaches*, *ASTIN Bulletin*, 38(2), 399–422.
- [7] Dickson, D. C. M. (2005) *Insurance risk and ruin*, Cambridge University Press.
- [8] Gerber, H. U. (1979) *An introduction to Mathematical Risk Theory*, S. S. Huebner Foundation for Insurance Education, University of Pennsylvania, Philadelphia, Pa. 19104, USA.
- [9] Li, S., Garrido, J. (2004) *On ruin for the Erlang(n) risk process*, *Insurance: Mathematics and Economics* 34(3), 391–408.
- [10] Ng, A. C. Y. (2009) *On a dual model with a dividend threshold*, *Insurance: Mathematics and Economics* 44(2), 315–324.
- [11] Rodríguez-Martínez E. V., Cardoso R. M. R., Egídio dos Reis, A. D. (2012) *Some advances on the Erlang(n) dual risk model*, preprint.
- [12] Seal, H. L. (1969) *Stochastic Theory of a Risk Business*, Wiley, New York.
- [13] Takács, L. (1967) *Combinatorial Methods in the Theory of Stochastic Processes*, Wiley, New York.

Séries Temporais e Aplicações

Testing for Broken Trends in Multivariate Time Series

(Comunicação)

Luis CATELA NUNES – *Universidade Nova de Lisboa*

lcnunes@novasbe.pt

Nuno Sobreira – *INSPER*

Resumo

In this paper we develop the first tests for the presence of common broken trends in multivariate time series which do not require knowledge of the form of serial correlation in the data and are robust as to whether the shocks are stationary or have a unit-root, cointegrated or not-cointegrated. Our approach extends the test procedure proposed in [1] to a multivariate setting and is based on weighted averages of a sequence of Wald statistics for testing the existence of breaks in the trend. We start by considering the case of known break dates and show that the proposed test has a chi-square limiting distribution, regardless of the number of stochastic trends. Next, we propose tests when the number of trend breaks is known but the break dates are unknown. Here, the estimated break dates are global maximizers of the sequence of Wald statistics evaluated at all admissible partitions. Finally, we propose a sequential procedure that may be used to estimate the number of breaks. We present a Monte Carlo simulation experiment to study the finite sample properties of the proposed tests. We also provide an application of our test procedures to analyze the evolution of worldwide income inequality over the 20th century.

Referências

- [1] Harvey, David I. and Leybourne, Stephen J. and Taylor, A.M. Robert (2009) *Simple, Robust, And Powerful Tests Of The Breaking Trend Hypothesis*, *Econometric Theory*, 25(4), 29, 995–1029.

Efficiency, roughness and long memory in the PSI20 index returns: wavelet and entropy analysis

(Comunicação)

Rui PASCOAL – *School of Economics, University of Coimbra*

ruiapsp@fe.uc.pt

Ana Margarida Monteiro – *School of Economics, University of Coimbra*

ammm@fe.uc.pt

Resumo

In this study, features of financial returns of PSI20 index are captured using wavelet and entropy based techniques. This characterization includes the following points. First, the detection of long memory, associated to low frequencies, and a global measure of the time series: the Hurst exponent estimated by several methods including wavelets. Second, the degree of roughness, or regularity variation, associated to the Hölder exponent, fractal dimension and estimation based on multifractal spectrum. Finally, the behavior on the tails of the return distribution, this one assumed to be q -Gaussian.

These aspects may also be studied through the concepts of non-extensive entropy and distribution using, for instance, the Tsallis q -triplet. They allow to study the existence of efficiency in the financial market.

On the other hand, the study of local roughness is performed by relating the pointwise Hölder exponent, which allows to detect time series singularities related to the wavelet leaders based entropy, quantified at the same point. In fact, the wavelet coefficients are computed from a multiresolution analysis, and the wavelet leaders are defined by the local suprema of these coefficients, near the point we are considering. The resulting entropy is more accurate in that detection than the Hölder exponent. These procedures enhance the capacity to forecast the occurrence of crashes.

Referências

- [1] Jaffard, S. (2004) *Wavelet techniques in multifractal analysis*, Fractal Ge-

ometry and Applications: A Jubilee of Benoit Mandelbrot, M. Lapidus and M. van Frankenhuysen Eds., Proceedings of Symposia in Pure Mathematics (AMS) 72 Part 2, 91–151.

[2] Kristoufek, L. , Vosvrda, M. (2013) *Measuring capital market efficiency: Global and local correlations structure*, Physica A. 392(1), 184–193.

[3] Queirós, S. , Moyano, L., de Souza, J., Tsallis, C.(2013) *A nonextensive approach to the dynamics of financial observables*, The European Physical Journal B 55, 161–167.

[4] Rosenblatt, M., Serrano, E., Figliola, A. (2012) *An entropy based in wavelet leaders to quantify the local regularity of a signal and its application to analyze the Dow Jones Index*, International Journal Wavelets Multiresolution and Information Processing 10, DOI:10.1142/S0219691312500488

[5] Yalamova, R. (2006) *Wavelet test of multifractality of Asia-Pacific index price series*, Asian Academy of Management Journal of Accounting and Finance 2, 63–83.

A new class of bivariate binomial autoregressive models

(Comunicação)

Manuel SCOTTO – *Departamento de Matemática e CIDMA, Universidade de Aveiro, Portugal*

mscotto@ua.pt

Christian Weiß – *Fächerguppe Mathematik/Statistik, Helmut-Schmidt-Universität, Alemanha*

Maria Eduarda Silva – *Departamento de Economia e CIDMA, Universidade do Porto, Portugal*

Isabel Pereira – *Departamento de Matemática e CIDMA, Universidade de Aveiro, Portugal*

Resumo

In this talk a new class of bivariate time series models useful to model count data time series with a finite range of counts is presented. A new bivariate thinning operation (which marginally behaves as the usual binomial thinning operation) is introduced and explained in detail. The new thinning operation has a number of advantages including that it allows for both positive and negative cross-correlation. Based upon this new thinning operation, a bivariate extension of the binomial autoregressive model of order one is introduced. Basic probabilistic and statistical properties of the model are discussed. Parameter estimation and forecasting are also covered. The performance of these models is illustrated through an empirical application to a set of rainy days time series collected from 2000 up to 2010 in the German cities of Bremen and Cuxhaven.

Using an Approximate Bayesian Computation (ABC) algorithm for estimating the parameters of Bilinear Processes *

(Comunicação)

Patricia DE ZEA BERMUDEZ – *CEAUL and FCUL, Portugal*

pcbermudez@fc.ul.pt

M. A. Amaral Turkman – *CEAUL and FCUL, Portugal*

K. F. Turkman – *CEAUL and FCUL, Portugal*

Resumo

The class of bilinear models [5] plays an important role in modeling nonlinearity for various reasons, such as the fact that it is an obvious generalization of ARMA models. Under fairly general conditions, bilinear processes approximate finite order Volterra series expansions to any desired order of accuracy over finite time intervals. Volterra series expansion are a dense class within the class of nonlinear time series. Therefore, bilinear processes are also a dense class within non linear processes, approximating any nonlinear process to a desired level of accuracy. Due to their capacity of producing clusters of large values, bilinear models are often suggested. However, they are not frequently used in practice due to inference problems.

In general, the conditions of stationarity and invertibility can not be written in terms of the model parameters in a easily verifiable form. The likelihood cannot be written explicitly, and particularly for heavy-tailed data, conditional least squares and quasi-likelihood methods do not seem to give good results. However, it is relatively easy to simulate such series.

Approximate Bayesian Computation (ABC) algorithms arise as ways to deal with inference problems associated with likelihood functions which are analytically difficult to handle or even intractable (see for instance [4]). As such, establishing an ABC protocol for such nonlinear systems seems to be a good idea. The major challenge is to find a set of statistics capable of representing the nonlinear dynamics of the system. In this work, we suggest seven statistics, namely a portmanteau statistic that captures the linear time dyna-

mics through the empirical autocorrelations, the Moments' estimator (see [3]), which quantifies the tail heaviness and the extremal index (see *e.g.* [2]) that measures the degree of clustering of large values. The implemented method is based on the recently proposed method by [1], which used k -nearest neighbor techniques. We apply these strategies to several bilinear models and different independent and identically distributed innovation processes. The results are compared through a simulation study. The approach will be applied to some real data set.

Referências

- [1] Biau, G., Cérou, F. and Guyader, A. (2012). New insights into Approximate Bayesian Computation (submitted).
- [2] Coles, S.G., 2001. An Introduction to Statistical Modeling of Extreme Values. Springer-Verlag, London.
- [3] Dekkers, A. L. M., Einmahl, J. H. J. and de Haan, L. (1989). A moment estimator for the index of an extreme value distribution. *Ann. Stat.*, **17**, 1833-1855.
- [4] Marjoram, P., Molitor, J., Plagnol, V., and Tavaré, S. (2003). Markov chain Monte Carlo without likelihoods. *Proc Natl Acad Sci U S A*, **100 (26)**, 15324 - 15328.
- [5] Subba Rao, T. and Gabr, M. M. (1984). An Introduction to Bispectral Analysis and Bilinear Time Series Models. Springer-Verlag, New York.

* Este trabalho foi financiado pelos projectos da FCT: PEst-OE/MAT/UI0006/2011 e PTDC/MAT/118335/2010

Comunicações Orais



Coeficientes de correlação ordinal ponderados dando ênfase à concordância em ambos os extremos

(Comunicação)

Sandra M. ALEIXO – *CEAUL e Área Departamental de Matemática, Instituto Superior de Engenharia de Lisboa – ISEL*

sandra.aleixo@adm.isel.pt

Júlia Teles – *CIPER e Seção de Métodos Matemáticos, Faculdade de Motricidade Humana, Universidade Técnica de Lisboa*

Resumo

Em muitas situações práticas, o objetivo é avaliar a concordância entre os *ranks* superiores (respetivamente, inferiores), não devendo ser tão valorizada a discordância nos *ranks* inferiores (respetivamente, superiores). Nestes casos, devem ser usados coeficientes de correlação ponderados, como os que são apresentados em Iman e Conover [3], Maturi e Abdelfattah [5], Blest [1] e Shieh [6].

Noutros casos, pode ser necessário ponderar simultaneamente os *ranks* inferiores e superiores, dando-lhes a mesma relevância. Iman e Conover [3] sugeriram a utilização do coeficiente de correlação de van der Waerden (Hájek and Šidák [2]), para uma maior ponderação nos *ranks* de ambos os extremos em detrimento dos *ranks* centrais. Esta estatística consiste no cálculo do coeficiente de correlação de Pearson com os *scores* de van der Waerden. Neste trabalho são propostos dois coeficientes de correlação ordinal ponderados, os quais são determinados usando o coeficiente de correlação de Pearson com os *scores* de Klotz e de Mood (Klotz [4]) modificados. Estes novos coeficientes possibilitam dar maior ênfase aos *ranks* extremos do que o coeficiente de van der Waerden. Para cada um destes coeficientes, foi deduzida a respetiva distribuição assintótica, e foram calculados os quantis exatos e os quantis aproximados. Com vista a ilustrar a utilidade e avaliar o desempenho destes coeficientes, são apresentados alguns exemplos e os resultados de um estudo de simulação.

Referências

- [1] Blest, D.C. (2000) *Rank correlation – An alternative measure*, Australian & New Zealand Journal of Statistics 42(1), 101–111.
- [2] Hájek, J., Šidák, Z. (1967) *Theory of Rank Tests*, Academic Press, New York.
- [3] Iman, R.L., Conover, W.J. (1987) *A measure of top-down correlation*, Technometrics 29(3), 351–357.
- [4] Klotz, J. (1962) *Nonparametric tests for scale*, Annals of Mathematical Statistics 33, 498–512.
- [5] Maturi, T.A., Abdelfattah, E.Z. (2008) *A new weighted rank correlation*, Journal of Mathematics and Statistics 4(4), 226–230.
- [6] Shieh, G.S. (1998) *A weighted Kendall's tau statistic*, Statistics & Probability Letters 39, 17–24.

Amostragem por distâncias em ungulados de montanha: simulação baseada num caso de estudo

(Comunicação)

Russell ALPIZAR-JARA – *Centro de Investigação em Matemática e Aplicações, Universidade de Évora, Portugal*

email: alpizar@uevora.pt

Afonso, Anabela – *Centro de Investigação em Matemática e Aplicações, Universidade de Évora, Portugal*

Pérez, Jesús M. – *Departamento de Biología Animal, Biología Vegetal y Ecología, Universidade de Jaén, Espanha*

Resumo

A amostragem por distâncias foi inicialmente desenvolvida para regiões de estudo planas ou pouco acidentadas. Nesta técnica a probabilidade de deteção é estimada a partir de uma função ajustada às distâncias observadas (perpendiculares ou radiais) entre os animais e o observador, sendo esta função designada por função de deteção (Buckland et al., 2001).

Devido à sua simplicidade de aplicação, no final da década de 1980 a amostragem por distâncias começou a ser utilizada também em terrenos montanhosos (Escós e Alados, 1988). Mas, as populações de ungulados de montanha têm tendência para formar agrupamentos de dimensões bastante variáveis, i.e., de um modo geral observam-se vários grupos com um pequeno número de animais e alguns grupos de elevada dimensão. Estes grupos tanto são detetados perto como afastados da linha percorrida pelo observador, e não se observa associação linear entre as dimensões dos grupos e as distâncias a que se encontram em relação ao observador. É muito vulgar a observação de grupos a distâncias atípicas o que dificulta o processo de estimação da função de deteção e, conseqüentemente, da densidade ou abundância de animais na região de estudo.

Na estimação da função de deteção podemos optar por várias estratégias: i) truncar ou não truncar os dados de distâncias, ii) agrupar ou não agrupar as distâncias em classes, ou iii) combinar as estratégias anteriores. Neste trabalho pretendemos avaliar se existem diferenças significativas nos resultados obtidos com as várias estratégias. Em particular, estamos interessados em avaliar o efeito de truncar grupos de indivíduos eventualmente localizados a distâncias extremas. Para tal simulamos populações agrupadas, das quais serão retiradas amostras com várias dimensões: a) inferior a 20 indivíduos por grupo, b) de 20 a menos de 60 indivíduos em cada grupo, e c) pelo menos 60 indivíduos por grupo. Esta simulação está baseada numa aplicação com dados reais de várias populações de ungulados de montanha utilizados por Pérez et al. (submetido).

Referências

- [1] Buckland ST., Anderson DR., Burnham KP., Laake JL., Borchers DL., Thomas L. (2001) Introduction to Distance Sampling. Estimating abundance of biological populations. Oxford University Press, Oxford.
- [2] Escós J., Alados CL. (1988) *Estimating mountain ungulate density in Sierras de Cazorla y Segura*, Mammalia. 52, 425–428.
- [3] Pérez, JM., Sarasa, M., Moço, G., Granados, JF., Crampe, JP., Serrano, E., Maurino, L., Meneguz, P.M., Afonso, A., Alpizar-Jara, R. (submitted to Animal Biodiversity and Conservation). The effect of data analysis strategies in density estimation of mountain ungulates using Distance Sampling.

On the estimation of Kendall's Tau for successive survival times

(Comunicação)

Ana AMORIM – *University of Minho*

apamorim@math.uminho.pt

J. de Uña-Álvarez – *University of Vigo*

Resumo

Successive survival times appear when, for each individual, two or more successive events are under study. One important goal in this field is the estimation of the correlation (or association degree) among the successive inter-event times. In this work we consider a new estimator of the Kendall's Tau for two successive survival times based on presmoothing ideas. The estimator is defined as the difference between the estimated probability of concordance and discordance between two representative pairs, which may be computed from the semiparametric estimator of the bivariate distribution introduced in de Uña-Álvarez and Amorim (2011). Asymptotic properties of the introduced estimator as consistency and distributional convergence will be discussed. The finite sample performance of the estimator will be explored by simulations. Comparison to other possible estimators will be done, to explore the relative benefits of presmoothing.

Referências

- [1] de Uña-Álvarez, J., Amorim, A.P. (2011) *A semiparametric estimator of the bivariate distribution function for censored gap times*, *Biometrical Journal* 53. 1, 113–127.

Ajustamento à distribuição de Birnbaum-Saunders generalizada

(Comunicação)

Emilia ATHAYDE – *Universidade do Minho*

mefqa@math.uminho.pt

Resumo

Neste trabalho considera-se a distribuição de Birnbaum-Saunders generalizada, $GBS(\alpha, \beta; g)$, introduzida por Díaz-García e Leiva (2005), correspondente a uma transformação de uma variável aleatória Z simétrica com função densidade de probabilidade (f.d.p.) g , dada por $T = \beta \left[\frac{\alpha Z}{2} + \sqrt{\left\{ \frac{\alpha Z}{2} \right\}^2 + 1} \right]^2$, em que α é um parâmetro de forma e β é um parâmetro de escala. Esta família de distribuições consiste numa generalização do modelo de Birnbaum-Saunders clássico (Birnbaum e Saunders, 1969) no qual g corresponde à f.d.p. normal reduzida. Analisa-se esta família de distribuições quanto a uma propriedade que a caracteriza, dada à custa da simetria de uma transformação logarítmica, seguindo a linha desenvolvida por Seshradi (1964). A partir daí sugerem-se diversas abordagens para avaliar se uma dada amostra de dados provém de uma distribuição de Birnbaum-Saunders generalizada, passando pelo problema da estimação do parâmetro β , que nesta família representa também a mediana. Investiga-se a performance de alguns dos processos sugeridos por meio de estudos de simulação e exemplifica-se com dados reais. Finalmente discute-se ainda a relação desta família de distribuições com outras famílias conhecidas.

Referências

- [1] Birnbaum, Z. W., Saunders, S. C. (1969) *A new family of life distributions*, J. Appl. Probab. 6, 319–327.
- [2] Díaz-García, J.A., Leiva, V. (2005) *A new family of life distributions based on elliptically contoured distributions*, J. Stat. Plan. Infer., 128, 445–457.
- [3] Saunders, S. C. (1974) *A family of random variables closed under reciprocation*, J. Amer. Stat. Assoc. 69, 533–539.
- [4] Seshadri, V. (1964) *On random variables that have the same distribution as their reciprocals*, Can. Math. Bull. 8, 819–824.

Estimação de tendências comuns em séries temporais multivariadas através da análise factorial dinâmica e aplicações em recursos pesqueiros

(Comunicação)

Nuno Miguel BRITES – *Instituto Português do Mar e da Atmosfera, I.P. é Centro de Investigação em Matemática e Aplicações, Universidade de Évora, Portugal*

nbrites@ipma.pt

Manuela Azevedo – *Instituto Português do Mar e da Atmosfera, I.P., Portugal*

Russell Alpizar-Jara – *Centro de Investigação em Matemática e Aplicações, Universidade de Évora, Portugal*

Resumo

Na análise de recursos marinhos, mais precisamente no estudo da dinâmica populacional de um recurso pesqueiro, é frequente a existência de várias séries temporais de índices de abundância. Existindo padrões comuns nas trajetórias da abundância, podem ser estimadas tendências comuns (TCs) entre as séries de dados. Ao contrário dos métodos mais comuns de análise de séries temporais, como a análise espectral e os modelos ARIMA, a Análise Factorial Dinâmica (AFD) [1,2] permite analisar séries temporais curtas, incompletas e não-estacionárias, características comuns em séries temporais de abundância de recursos pesqueiros [3], permitindo a estimação de TCs.

Neste estudo, comparam-se as estimativas dos parâmetros de modelos de avaliação de recursos pesqueiros quando são utilizadas todas as séries temporais de abundância ou apenas as TCs dessas séries, obtidas com a AFD. Esta abordagem é aplicada ao stock ibérico de pescada (*Merluccius merluccius*), usando as séries temporais de seis índices da pesca comercial e três de campanhas de investigação, obtidas em águas continentais portuguesas e espanholas, no período 1982-2012. A pescada é um recurso muito importante para a economia do sector e a modelação da sua dinâmica, validada pelo ICES (International Council for the Exploration of the Sea), é usada para aconselhar a União Europeia relativamente às oportunidades de pesca.

Referências

- [1] Zuur, A.F., et al. (2003) *Estimating common trends in multivariate time series using dynamic factor analysis*, *Environmetrics*. 17, 665–685.
- [2] Jungbacker, B., et al (2012) *Smooth Dynamic Factor Analysis with Application to the U.S. Term Structure of Interest Rates*, *Journal of Applied Econometrics* (Forthcoming).
- [3] Azevedo, M., et al. (2008) *Application of dynamic factor analysis in the assessment of Iberian anglerfish stocks*, *ICES Journal of Marine Science*. 65, 1362–1369.

Distâncias de Mahalanobis, variáveis originais e componentes principais

(Comunicação)

Jorge CADIMA – *Centro de Estatística e Aplicações da Universidade de Lisboa (CEAUL) e Matemática-DCEB, Instituto Superior de Agronomia, Universidade Técnica de Lisboa*

jscadima@isa.utl.pt

Resumo

As distâncias de Mahalanobis são, desde há cerca de oito décadas, um conceito consagrado na análise de dados multivariada. A sua interpretação é bem conhecida, quando uma matriz $X_{n \times p}$ de dados é representada da forma usual: n pontos num espaço p -dimensional.

No entanto, resultados fundamentais sobre as distâncias de Mahalanobis foram recentemente descobertos (Gath & Hayes, 2006; Branco & Pires, 2011), revelando que as distâncias de Mahalanobis possuem ainda numerosas facetas pouco conhecidas. Cadima (2012) mostrou que estes resultados recentes têm uma interpretação geométrica natural quando a representação tradicional dos dados é substituída pela representação no chamado “espaço das variáveis”, ou seja, quando uma matriz de dados $n \times p$ é representada num espaço n -dimensional (cada eixo correspondente a um indivíduo e cada variável representada por um vector). Mais concretamente, a distância (ao quadrado) de Mahalanobis de um indivíduo ao centro é dada por $d_i^2 = (n-1) \cos^2 \theta_i$, onde θ_i é o ângulo entre o eixo correspondente ao indivíduo i e o subespaço $\mathcal{C}(X_c)$, gerado pelas colunas da matriz *centrada* de dados, X_c . Analogamente, a distância de Mahalanobis entre um par (i, j) de indivíduos é dada por $d_{ij}^2 = 2(n-1) \cos^2 \theta_{ij}$, onde θ_{ij} é o ângulo entre a diferença $\mathbf{e}_i - \mathbf{e}_j$ dos vectores da base canónica de \mathbb{R}^n associados aos indivíduos i e j e o subespaço $\mathcal{C}(X_c)$. Assim, as distâncias de Mahalanobis medem a inclinação do subespaço $\mathcal{C}(X_c)$ face ao sistema de eixos e a bissectrizes de ortantes de \mathbb{R}^n .

Uma base do subespaço $\mathcal{C}(X_c) \subset \mathbb{R}^n$ é dada pelo conjunto das p variáveis

originais centradas. Duas bases (ortogonais) alternativas, são dadas pelas componentes principais do conjunto de dados e pelas componentes principais do conjunto de dados normalizados (isto é, as componentes principais “sobre a matriz de correlações”). Nesta comunicação exploram-se as propriedades das distâncias de Mahalanobis analisando a geometria associada a cada uma das bases do referido subespaço. A discussão é acompanhado de exemplos de ilustração.

Referências

- [1] Branco, J.A., Pires, A.M. (2011) *Travelling through multivariate data spaces with Mahalanobis distance*, JOCLAD 2011, Vila Real.
- [2] Cadima, J. (2012) *A geometria das distâncias de Mahalanobis no espaço das variáveis*, XX Congresso SPE 2012, Porto (submetido para publicação nas Actas).
- [3] Gath, E.G., Hayes, K. (2006) *Bounds for the largest Mahalanobis distance*, Linear Algebra and its Applications 419, 93–106.

Análise do jogo em desportos coletivos: o estatístico como treinador de bancada

(Comunicação)

Ana Isabel CARITA – *CIPER e Secção de Métodos Matemáticos, Faculdade de Motricidade Humana, Universidade Técnica de Lisboa*

acarita@fmh.utl.pt

Resumo

Uma das atribuições da estatística aplicada é a de dar apoio no tratamento e análise da dados contribuindo para a qualidade da investigação em diversas áreas uma das quais o desporto.

A temática da análise do jogo desportivo coletivo engloba inúmeros focos de interesse com vista a encontrar e desenvolver metodologias adequadas e cada vez mais centradas nas respostas às questões do jogo.

Apresentamos exemplos de aplicação de várias metodologias estatísticas em análise de jogos coletivos, como o andebol, o basquetebol e o futebol. Entre os trabalhos desenvolvidos abordamos a utilização de modelos de regressão logística (Hosmer e Lemeshow [1]) para a explicação de variáveis de performance do jogo de futebol e para a deteção de talentos no basquetebol e no futsal, modelos de análise de sobrevivência para evolução do marcador no jogo de andebol, modelos de análise de variância multivariada e análise discriminante para o estudo do impacto da modificação das alterações às regras oficiais de basquetebol e algumas metodologias mais simples para estudos sobre *home advantage* (Pollard [2]) no futebol.

Referências

[1] Hosmer, D.W., Lemeshow, S. (2000). *Applied Logistic Regression* (second edition), Wiley, New York.

[2] Pollard, R. (1986). *Home advantage in soccer: A retrospective analysis*, *Journal of Sport Sciences*, 4(3), 237–248.

Modelos de crescimento populacional com efeitos de Allee em ambiente aleatório

(Comunicação)

Clara CARLOS – *Escola Superior de Tecnologia do Barreiro, Instituto Politécnico de Setúbal*

clara.carlos@estbarreiro.ips.pt

Carlos A. Braumann – *Centro de Investigação em Matemática e Aplicações, Universidade de Évora*

Resumo

Os modelos clássicos que não incorporam os efeitos de Allee pressupõem que a população cresce sempre mesmo para densidades populacionais muito baixas. À primeira vista parece aceitável, uma vez que abundam os recursos para os poucos indivíduos existentes, mas uma reflexão mais cuidadosa levamos a questionar a legitimidade desse pressuposto. É natural que para certas populações haja uma densidade mínima abaixo da qual a população não consegue repor o seu tamanho e acima desse valor a população possa crescer. Por esse motivo é interessante considerar modelos de crescimento populacional em que a taxa de crescimento possa ser negativa para tamanhos da população próximos de zero, conhecidos por modelos com efeitos de Allee fortes, em homenagem a Warder Clyde Allee, um zoólogo e ecologista, que estudou o comportamento animal e descreveu pela primeira vez em 1949 os chamados efeitos de Allee (ver [1]). Pode ver-se em [3] um estudo mais recente sobre os efeitos de Allee em ecologia e conservação.

Existem dois tipos de efeitos de Allee: os efeitos de Allee fortes e os efeitos de Allee fracos. Temos um efeito de Allee forte quando a população precisa de ultrapassar um determinado número de indivíduos para ter taxa de crescimento positiva, caso contrário a taxa de crescimento é negativa e a população extingue-se. Quando a taxa de crescimento é positiva mas reduzida para densidades populacionais muito baixas, dizemos que temos um efeito de Allee fraco.

Pretendemos estudar o efeito das flutuações aleatórias do ambiente. De forma

a que os resultados a que chegamos sejam robustos com respeito à escolha do modelo, apresentamos um modelo geral estocástico de crescimento populacional que inclui os efeitos de Allee fortes e um termo estocástico que descreve o efeito daquelas flutuações na taxa de crescimento. Estudamos a existência e unicidade de solução, o comportamento das fronteiras do espaço de estados, garantimos a inexistência de densidade estacionária e determinamos expressões para a média e a variância do tempo de extinção (baseadas nas expressões obtidas em [2]).

Para estudar o comportamento do tempo de extinção, obtemos expressões explícitas para a média e desvio padrão do tempo de extinção e apresentamos gráficos ilustrativos do seu comportamento para várias combinações de parâmetros de um modelo particular. Trata-se de um caso particular do modelo geral inspirado no modelo logístico estocástico tradicional, a que chamamos modelo logístico estocástico com efeitos de Allee.

Referências

- [1] Allee, W. C., Emerson, A. E., Park, O., Park, T. e Schmidt, K. P. (1949) *Principles of Animal Ecology*, Saunders, Philadelphia.
- [2] Carlos, C., Braumann, C. A e Filipe, P. A. (2013) *Models of individual growth in a random environment: study and application of first passage times*, Advances in Regression, Survival Analysis, Extreme Values, Markov Processes and Other Statistical Applications. Lita da Silva, J., Caeiro, F., Natário, I., Braumann, C., Esquível, M.L. e Mexia, J.T. (Eds.). Physica-Verlag, em impressão (ISBN 978-3-642-34903-4), p.91–98
- [3] Courchamp, F., Berec, L. e Gascoigne, J. (2008) *Allee Effects in Ecology and Conservation*, Oxford University Press, New York.

Cartas de controlo com intervalos de amostragem dinâmicos inicialmente predefinidos para um risco cumulativo constante

(Comunicação)

Manuel do CARMO – *Universidade Europeia, Laureate International Universities e CIMA-UE*

manuel.carmo@europeia.pt

Paulo Infante – *CIMA-UE e DMAT, ECT da Universidade de Évora*

Jorge Mendes – *ISEGI-NOVA, Universidade Nova de Lisboa e CEAUL-FCUL*

Resumo

Actualmente, qualquer organização, de produção ou serviços, faz uma avaliação rigorosa da qualidade dos bens que disponibiliza ao consumidor, recorrendo a técnicas estatísticas adequadas.

Para monitorizar uma, ou mais, das suas características, a carta de controlo tem sido uma das ferramentas mais utilizadas, pois permite a distinção entre a variação intrínseca do processo e a variação com origem em causas externas. O desempenho das cartas de controlo está directamente relacionado com a dimensão amostral e com o modo como são definidos os instantes de recolha das amostras. Para melhorar o desempenho do esquema clássico (amostras de tamanho fixo recolhidas em intervalos igualmente espaçados no tempo) surgiram métodos que definem o tamanho das amostras e/ou os instantes de amostragem em função do valor da estatística amostral (designados dinâmicos) e outros em que são determinados no início do controlo do processo em função da distribuição do tempo de vida (designados predefinidos).

Se por um lado é desejável analisar o menor número de amostras (para reduzir os custos de inspecção), por outro lado é fundamental que a carta seja eficaz a detectar as alterações na qualidade (de modo a que sejam reduzidos os custos de mau funcionamento).

Considerando uma carta de controlo para médias, Rodrigues Dias (2002) apresentou um esquema de amostragem cujos instantes são definidos, no início do controlo do processo, de modo a que a taxa cumulativa de risco seja constante entre dois quaisquer instantes consecutivos, denominado

PSI (Predetermined Sampling Intervals). Rodrigues Dias & Infante (2008) mostram que este método é sempre mais eficaz que o esquema periódico clássico, sendo tanto mais eficaz quanto menos amostras são recolhidas no período de controlo, quanto menor for a magnitude da alteração e quanto mais acentuadamente crescente for a taxa de risco do sistema. Quando comparado com outros esquemas de amostragem, revelou-se igualmente eficaz a detectar reduzidas e grandes alterações, em particular para sistemas com taxas de risco crescente.

Em Infante & Rosmaninho (2007) é apresentado um esquema de amostragem que combina um esquema adaptativo, *NSI (Normal Sampling Intervals)*, com o esquema *PSI*. Neste esquema os instantes de amostragem são definidos pela média ponderada entre os instantes obtidos com *NSI* e *PSI*. Considerando o uso simultâneo de carta para médias e carta para amplitudes, os resultados obtidos, em termos de *AATS*, demonstraram um enorme potencial do método quando comparado com os métodos *FSI (Fixed Sampling Intervals)*, *VSI (Variable Sampling Intervals)* e *VSSI (Variable Sample Sizes and Sampling Intervals)*.

Em Carmo *et al.* (2012) foi apresentado e analisado um esquema de amostragem adaptativo. Neste esquema, denominado *LSI (Laplace Sampling Intervals)*, os instantes de amostragem são actualizados ao longo do processo, dependendo da informação recolhida na amostra anterior, segundo a função densidade da distribuição de *Laplace standard*. Este método apresenta um bom desempenho, em particular para alterações moderadas da média, quando comparado com os métodos *FSI* e *VSI*.

Neste trabalho, analisamos a eficiência estatística de um novo método que define os instantes de amostragem com base numa média ponderada dos esquemas *PSI* e *LSI*, dando maior peso ao método *LSI* para alterações moderadas (onde *PSI* é menos eficaz) e maior peso ao método *PSI* nos restantes casos (onde *LSI* 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. Os resultados, já obtidos, revelam um muito bom desempenho deste método de amostragem quando se usam cartas de controlo para a média do tipo *Shewhart*.

Referências

- [1] Carmo, M., Infante P., Mendes, J. (Submetido). *Statistical Properties and Robustness of a New Adaptive Sampling Method for Quality Control*.
- [2] Rodrigues Dias, J., Infante P. (2002). *Amostragem em Controlo de Qualidade com Intervalos Diferentes Predefinidos: Uma Nova Abordagem*, JOCLAD02, IX Jornadas de Classificação e Análise de Dados, Lisboa, 10 p.
- [3] Rodrigues Dias, J., Infante P. (2008). *Control Charts with Predetermined Sampling Intervals*, International Journal of Quality and Reliability Management, 25(4), 423–435.
- [4] Rosmaninho, E.; Infante P. (2007). *Métodos de Amostragem com Paramâmetros Predefinidos Adaptáveis: Uma análise estatística e económica*, Em: Estatística: Ciência Interdisciplinar - Actas do XIV Congresso Anual da Sociedade Portuguesa de Estatística, (Ferrão, M.E.; Nunes, C. e Braumann, C.A., eds), Edições da SPE, 659–708.

Aproximação da matriz de correlações à família das matrizes Toeplitz - aplicação à reconstrução de imagens coloridas

(Comunicação)

Eunice CARRASQUINHA – *CEMAT, Instituto Superior Técnico*

eunicecarrasquinha@gmail.com

Ana M. Pires – *CEMAT, Instituto Superior Técnico*

Conceição Amado – *CEMAT, Instituto Superior Técnico*

Resumo

As características de interesse do processamento de sinais abarcam sons, sinais de telecomunicações, imagens, séries temporais e muitos outros. Por isso, o processamento de sinais é uma importante área de investigação quer na engenharia, na biomedicina e também na matemática e na estatística.

Algumas das técnicas estatísticas utilizadas para o processamento de sinal são, a Análise de Componentes Principais (ACP), muitas vezes também denominada por transformada de Karhunen-Loève, e a transformada de Fourier, nomeadamente a transformada de Fourier discreta. Apesar de na literatura ser possível encontrar inúmeros estudos de processamento de imagens e sinais onde a metodologia utilizada é a ACP ou a transformada de Fourier são raros os trabalhos onde se estabelece uma relação entre estas duas técnicas. Este trabalho é um passo nesse sentido, apresentando-se uma metodologia que relaciona estes dois tipos de técnicas. O método proposto tem por base uma aproximação da matriz de correlações através de uma matriz circulante (caso particular das matrizes Toeplitz) usando o método dos mínimos quadrados. A metodologia proposta foi aplicada a um conjunto de imagens. A comparação dos resultados obtidos pelo método ACP com os resultantes da transformada de Fourier discreta permitiu verificar o bom desempenho da nova metodologia.

Referências

- [1] Kouassi, R., Gouton, P. (2001) *Approximation of the Karhunen-Loève transformation and its application to colour images*, Signal Processing: Image Communication. 16, 541–551.
- [2] Petrou, M., Bosdlogianni, P. (2000) *Image Processing: The Fundamentals*, John Wiley and Sons, Inc..
- [3] Tismenetsky, M. (1991) *A decomposition of Toeplitz matrices and optimal circulant preconditioning*, Linear Algebra and its Applications. 154-156, 105–121.
- [4] Unser, M. (1984) *On the approximation of the discrete Karhunen-Loève transform for stationary processes*, Signal Processing. 7, 231–249.

Variabilidade do comportamento mecânico em compósitos de fibras naturais

(Comunicação)

Alda CARVALHO – *GI-MOSM/ISEL & CEMAPRE/ISEG*

acarvalho@adm.isel.pt

Sérgio Silva – *GI-MOSM/ISEL*

Fábio Damásio – *GI-MOSM/ISEL*

Maria Amélia Loja – *GI-MOSM/ISEL & IDMEC/IST*

José Miranda Guedes – *IST/DEM & IDMEC/IST*

Resumo

A previsão das propriedades características de um material é de extrema relevância para habilitar os profissionais de engenharia a uma adequada concepção de uma dada estrutura ou de uma sua componente. Nas actividades de projecto que correntemente se desenvolvem neste contexto, é comumente aceite a utilização de valores médios para essas propriedades, valores estes que são baseados em resultados de natureza experimental mas igualmente em pressupostos de natureza teórica que poderão circunstancialmente não se verificarem. Esta incerteza coloca-se de forma ainda mais premente quando em alternativa aos materiais “tradicionalis” são utilizados materiais compósitos, e mais ainda quando falamos de compósitos de fibras naturais em que a heterogeneidade material ao nível das suas micro-estruturas é uma realidade ainda mais efectiva. Com o presente trabalho, pretende-se apresentar um estudo sobre a variabilidade das propriedades[1] destes compósitos de fibras naturais, e dos efeitos que essa variabilidade pode provocar no comportamento e no colapso de uma estrutura.

Numa primeira fase foram utilizadas técnicas de simulação de Monte Carlo para reproduzir as arquitecturas consideradas como simplificações representativas da microestrutura do material compósito. Tendo em consideração alguns parâmetros que caracterizam estas microestruturas, é possível calcular as propriedades mecânicas homogeneizadas do material, como se de um material único se tratasse. Estas propriedades homogeneizadas são os módulos de elasticidade, os coeficientes de Poisson e os módulos de distorção.

Um dos objectivos deste trabalho é identificar a relação que se estabelece entre os parâmetros (variáveis) que definem a configuração da microestrutura e as respectivas propriedades elásticas. Uma vez que se pretende prever eventuais situações de colapso do material, elege-se ainda como objectivo deste estudo, a identificação das distribuições de parâmetros caracterizadores do comportamento mecânico do material, designadamente do estado de tensão e do estado de deformação num ponto, analisado numa perspectiva multi-escala[2].

Referências

- [1] Goldsmith, M., Sankar, B., Haftka, R., Goldberg, R. (2013) *Effects of Microstructural Variability on Thermo-Mechanical Properties of a Woven Ceramic Matrix Composite*, Report NASA/TM-2013-217817.
- [2] Loja, M. A., R. Rodrigues, H. C., Guedes, J.M. (2013) *Towards an Optimization Strategy for Natural Fibre Laminates*, Congreso de Métodos Numéricos en Ingeniería - CMN 2013, 25-28 Junho 2013, Bilbao, Espanha.

Compreensão dos Conceitos de Probabilidade Condicionada e Acontecimentos Independentes no Ensino Secundário

(Comunicação)

Maria José CARVALHO – *Escola Básica de Santa Marinha, Agrupamento de Escolas António Sérgio*

Adelaide FREITAS – *Departamento de Matemática & CIDMA, Universidade de Aveiro*

adelaide@ua.pt

Resumo

Partindo de uma análise descritiva e interpretativa, às respostas dadas por 43 alunos do 12.º ano a duas situações-problema em prova escrita, esta investigação pretendeu identificar e analisar dificuldades reveladas pelos alunos na compreensão dos conceitos de probabilidade condicionada e de acontecimentos independentes. Para tal, cada resposta escrita foi avaliada segundo duas perspetivas: desempenho e rigor. Estas duas perspetivas determinaram a criação de duas escalas qualitativas ordinais, em escalas de Likert, que foram designadas por grau de desempenho (GD), a variar entre 0 (não responde) e 5 (resposta totalmente correta), e grau de rigor (GR), a variar entre 0 (nenhum rigor) e 3 (com total rigor). Ambas as escalas tiveram como objetivo quantificar a qualidade de uma resposta, uma em termos de desempenho e outra em termos de rigor. O desempenho de uma resposta escrita está relacionado com a indicação dos passos básicos para a obtenção do resultado final. O rigor está relacionado com a indicação explícita dos passos fundamentais para a obtenção do resultado final correto tomando em linha de conta a linguagem, a argumentação e os procedimentos usados.

Da análise global dos resultados, constatam-se: i) conflitos nos conceitos de acontecimentos independentes e acontecimentos incompatíveis, sendo calculada uma probabilidade para justificar a incompatibilidade; ii) imensa dificuldade em formalizar o problema, ou por má interpretação do enunciado ou por incapacidade de formulação matemática ou execução esquemática do enunciado.

ado, conduzindo a uma resposta errada ou ao seu abandono; iii) confusão entre probabilidade condicionada e probabilidade conjunta, tanto na compreensão intuitiva como na sua aplicação.

Da análise individual das respostas, constata-se erros a nível dos conceitos, dos procedimentos e da linguagem.

As recomendações emergentes do estudo apontam para uma maior atenção à leção do conceito de independência e uma maior prática na formulação matemática de enunciados envolvendo aqueles dois conceitos.

Agradecimentos

Trabalho subsidiado pelo FEDER, através do COMPETE, pelo CIDMA (Centro de Investigação e Desenvolvimento em Matemática e Aplicações) da Universidade de Aveiro e FCT, dentro do projecto PEst-C/MAT/UI4106/2011 com número COMPETE FCOMP-01-0124-FEDER-022690.

Determinantes de matrizes de elementos uniformes inteiros: é a singularidade rara ?

(Comunicação)

Miguel CASQUILHO – *Instituto Superior Técnico, Universidade Técnica de Lisboa; Ave. Rovisco Pais, IST, 1049-001 Lisboa, Portugal*

mcasquilho@ist.utl.pt

João Luís de Miranda – *Escola Superior de Tecnologia e Gestão, Instituto Politécnico de Portalegre; P7300-110, Portalegre, Portugal*

Jorge Buescu – *Faculdade de Ciências, Universidade de Lisboa; Campo Grande, 1749-016 Lisboa, Portugal*

Resumo

Ao tratar da inversão de matrizes e do cálculo do seu determinante, tanto no âmbito académico como em computação (informática), surge a questão da eventual singularidade da matriz. A imprevisibilidade sobre se a matriz é invertível ou singular provém quer do possível imprevisto académico na construção dum exemplo, quer, na computação, do curso da execução dos cálculos. *A priori*, em comparação com o caso de elementos reais (contínuos), a questão põe-se de forma mais aguda quando os elementos da matriz (discretos) são "números inteiros e pequenos".

Em estudo anterior, que versava variáveis contínuas ([1]), afluímos a questão dos determinantes de matrizes de ordem n de elementos aleatórios com distribuição uniforme 0–1. A escolha desta distribuição "simples" foi crucial na obtenção de resultados que certamente ficariam ocultos se se tivesse optado por uma distribuição de domínio ilimitado, como a gaussiana. Da dificuldade da questão, previsível com qualquer distribuição, resulta a sua complexidade computacional, sugerindo o recurso à simulação pelo método de Monte Carlo. Por esta via, verificámos: que, expectavelmente, a densidade do determinante é simétrica em torno de zero; e que junto deste valor se concentra a maior probabilidade. Mantemos essa computação e a do caso presente acessíveis na Internet ([2] e [3]), em conformidade com a nossa prática de desde há mais de uma década, como vimos advogando (*v. g.*, [4]), de disponibilizar na Internet ao utilizador a computação relativa às matérias que investigamos.

No caso contínuo mencionado, conseguimos deduzir, analiticamente, a função densidade de probabilidade para $n = 2$; porém, os casos de $n > 2$ —objectivo que, realmente, completaria o estudo— revelaram-se intratáveis. Viríamos, posteriormente, a determinar, mas apenas numericamente (até $n = 12$), os valores extremos do determinante, com o resultado intuitivamente convincente de $\pm n$.

No estudo presente, que trata de variáveis discretas (matrizes de elementos inteiros), há maior aproximação à prática académica. Usou-se a mesma distribuição: uniforme, mas de variável inteira, no intervalo $[-M, +M]$, com $M = 10$. (O intervalo $[-9, +9]$, por ter um só algarismo, seria mais típico, mas torna os resultados menos sugestivos.) Mostrar-se-ão, graficamente, para determinantes de ordem 2: a 'pdf' teórica, $f_{theo.}$, de determinantes de elementos uniformes contínuos 0–1 (uma curva com aspecto aparentado a $exp(-|x|)$); um histograma da simulação, $f_{sim.}$, que se lhe “sobrepeõe”, e a 'cdf' deste, $F_{sim.}$ (sua acumulada); e o histograma da simulação de 'pf', função de probabilidade, f , de determinantes de elementos uniformes discretos em $[-10, +10]$ (similar à contínua), e a sua acumulada, F . Notar-se-á a (não inesperada) suavidade de $F_{sim.}$ face à irregularidade da 'pdf'. Esta irregularidade deve-se a uma simulação propositadamente pouco extensa para que se possam visualizar duas curvas sobrepostas (e não uma). Uma simulação mais longa, como se faria apenas para efeitos práticos, produziria duas curvas sobrepostas indistinguíveis.

Verificar-se-á que a nova curva, simulação da 'pf' (homóloga da 'pdf'), com extremos (obviamente) $\pm 2M^2 (= \pm 200)$, tem andamento compatível (i.é, aparentado à referida $exp(-|x|)$) com a do caso contínuo. A probabilidade do determinante nulo continua, embora máxima, pequena. Para maiores ordens dos determinantes, $n > 2$, verificou-se a mesma tendência, i.é, curvas sempre do tipo da referida. No caso apresentado, a singularidade é rara: revela-se baixa a probabilidade (cerca de 1 %) de determinantes nulos.

Referências

- [1] Casquilho, M., Buescu, J. (2012) *Determinantes de matrizes de elementos uniformes 0–1: ordem 2 e (algo) mais*, XX Congresso anual da SPE, Porto (Portugal), 28-Set., MP2 (“Modelação Paramétrica II”), comunicação oral.

- [2] Casquilho, M. (2013) *Random determinants*, <http://web.ist.utl.pt/mcasquilho/compute/or/Fx-randDeterm.php>, acedido em 2013-Jun-15.
- [3] Casquilho, M. (2013) *Random determinants*, <http://web.ist.utl.pt/mcasquilho/compute/or/Fx-randDetermInt.php>., acedido em 2013-Jun-15.
- [4] Casquilho, M. (2013) *Computação científica, Internet, Indústria*, 1.st Portuguese Meeting on Mathematics for Industry, Faculdade de Ciências, Universidade do Porto, Porto (Portugal), 06-08-Jun., comunicação oral.

Modelação das taxas de captura de tintureira pela frota Portuguesa de palangre de superfície a operar no Atlântico Sul

(Comunicação)

Rui COELHO – *Instituto Português do Mar e da Atmosfera (IPMA, I.P.)*

rpcoelho@ipma.pt

Paulo Infante – *Centro de Investigação em Matemática e Aplicações (CIMA-UE) e Dep. Matemática, ECT - Univ. Évora*

Miguel N. Santos – *Instituto Português do Mar e da Atmosfera (IPMA, I.P.)*

Resumo

A pescaria Portuguesa de palangre de superfície opera numa vasta área do Oceano Atlântico e tem como espécie alvo o espadarte, *Xiphias gladius*. Os tubarões oceânicos são uma componente importante desta pescaria, sendo a tintureira, *Prionace glauca*, a espécie mais capturada [1]. Os tubarões possuem ciclos de vida típicos de predadores de topo, com taxas de crescimento lentas, maturidade tardia e baixa fecundidade, pelo que as suas populações são de um modo geral mais vulneráveis à sobre-pesca [2]. Assim, a modelação das taxas de captura destas populações é particularmente importante, uma vez que permite uma melhor compreensão das variáveis que influenciam as suas capturas. Com este trabalho pretendeu-se modelar as taxas de captura de tintureira capturada pela frota Portuguesa de palangre de superfície a operar no Atlântico Sul. Os dados foram recolhidos por observadores de bordo do *Instituto Português do Mar e da Atmosfera (IPMA, I.P.)* entre Outubro de 2008 e Dezembro de 2011, tendo sido recolhida informação de 533 lances de pesca (728.254 anzóis). Tomando como variável resposta a taxa de captura em número (CPUE, Capturas Por Unidade de Esforço, $n/1000$ anzóis), considerou-se uma série de possíveis variáveis explicativas, incluindo localização espacial (latitude e longitude), ano, estação, material do estralho, navio, fase lunar, iluminação lunar, temperatura da superfície do mar, clorofila, salinidade e velocidade da corrente. A maioria dos dados ambientais foi obtida *a posteriori* a partir de dados de satélite [3], utilizando as coordenadas e data das pescas. Devido à dependência dos dados relativamente à própria

pesca, foram utilizadas duas abordagens de modelação comparativas, usando modelos lineares generalizados (GLM) e modelos lineares generalizados com efeitos mistos (GLMM), neste último caso com a utilização do navio de pesca como efeito aleatório. Consideraram-se os modelos Gama, Poisson, Binomial Negativo (BN) e Tweedie, sendo que nos Poisson e BN as capturas foram modeladas em contagens (n) com utilização do número de anzóis como variável *offset*. Os modelos candidatos foram comparados quanto à qualidade do ajustamento pelo pseudo- R^2 , AIC e por validação cruzada. Dos 533 lances de pesca analisados ocorreram capturas positivas de tintureira em 525, pelo que 1,52% dos lances possuíam CPUE=0. Nos lances positivos as capturas variavam entre 1 e 184, com uma média igual a 28,8 e um desvio padrão igual a 25,7. As variáveis selecionadas para o modelo final incluíram o material do estralho, ano, estação do ano, latitude, longitude, clorofila e temperatura, tendo-se obtido fatores de expansão generalizados (GVIF) <10 em todos os casos, o que significa uma provável ausência de colinearidade entre as variáveis. Em todos os modelos testados a análise de resíduos produziu resultados razoáveis, exceto nos modelos Poisson em que se observou um aumento de variância ao longo dos valores previstos, resultado de uma sobre-dispersão nos dados (parâmetro dispersão = 7.18). Observou-se que em todos os modelos os parâmetros estimados foram relativamente semelhantes, sendo o GLMM Gama o que apresentou o melhor ajustamento. Com os modelos finais concluiu-se que os valores esperados das capturas variam sazonalmente (menos capturas no primeiro trimestre e mais no terceiro), em termos ambientais (maiores capturas associadas a valores elevados de temperatura e menores capturas em zonas de maior concentração de clorofila de águas mais frias), em termos espaciais (capturas superiores a Sul e a Este da zona de estudo) e em termos operacionais (capturas superiores quando utilizados de estralhos de aço no aparelho de pesca). Foram criados alguns cenários teóricos de possíveis operações de pesca em diversas condições sazonais, espaciais, ambientais e operacionais, e os valores previstos das capturas foram estimados e comparados. Os resultados deste trabalho apresentam informação nova e importante relativamente à modelação das taxas de captura desta espécie no Atlântico Sul, e podem agora ser utilizados para prever as taxas de capturas da pescaria em diferentes cenários da atividade da pesca, contribuindo assim para uma gestão mais sustentável desta espécie no Oceano Atlântico.

Referências

- [1] Coelho, R., Fernandez-Carvalho, J., Lino, P.G., Santos, M.N. (2012) *An overview of the hooking mortality of elasmobranchs caught in a swordfish pelagic longline fishery in the Atlantic Ocean*, Aquatic Living Resources 25, 311–319.
- [2] Stevens, J.D., Bonfil, R., Dulvy, N.K., Walker, P.A. (2000) *The effects of fishing on sharks, rays, and chimaeras (chondrichthyans), and the implications for marine ecosystems*, ICES Journal of Marine Science 57, 467–494.
- [3] Roberts, J.J., Best, B.D., Dunn, D.C., Treml, E.A., Halpin, P.N. (2010) *Marine Geospatial Ecology Tools: An integrated framework for ecological geoprocessing with ArcGIS, Python, R, MATLAB, and C++*, Environmental Modelling & Software 25, 1197–1207.

Sistemas de alarme ótimos com aplicação ao modelo INAPARCH

(Comunicação)

Maria da Conceição COSTA – *Departamento de Matemática e CIDMA, Universidade de Aveiro*

lopescosta@ua.pt

Manuel Scotto – *Departamento de Matemática e CIDMA, Universidade de Aveiro*

Isabel Pereira – *Departamento de Matemática e CIDMA, Universidade de Aveiro*

Resumo

Neste trabalho, preve-se se uma série temporal de valores inteiros ultrapassa um determinado nível fixo e é emitido um alarme sempre que a probabilidade de tal acontecimento for crítica. A aplicação é feita a séries de contagem relacionadas com dados financeiros. Dadas as propriedades interessantes do modelo APARCH(p, q), de Ding *et al.* (1993), na modelação de séries de dados financeiros, é proposta a introdução do seu homólogo para a modelação de séries temporais de contagens: o modelo ARCH de valores INteiros e Potência Assimétrica, INAPARCH(p, q). São apresentadas as propriedades probabilísticas do modelo INAPARCH(1, 1) e desenvolve-se o método de máxima verosimilhança condicional para a estimação dos parâmetros do modelo. A implementação do sistema de alarme ótimo é realizada para o modelo INAPARCH(1, 1) e é feita a aplicação a uma série de dados relativa ao número de transacções em *stocks*.

Referências

[1] Ding, Z., Granger, C.W.J. and Engle, R.F. (1993) *A long memory property of stock market returns and a new model*, J. Emp. Finance 1, 83–106.

Multivariate Markov Chains: Estimation and Forecast

(Comunicação)

Bruno DAMÁSIO – *ISEG - Departamento de Matemática*

bdamasio@iseg.utl.pt

João Nicolau – *ISEG - Departamento de Matemática*

Resumo

Desde o início do século XXI que variados estudos têm cientificado em torno da temática das Cadeias de Markov Multivariadas (CMM). Podemos afirmar que foi com Ching et al (2002), inspirado em Raftery (1985), que se deu um grande avanço nos métodos de estimação das CMM pois, até então, a possibilidade de se estimarem CMM não passava de uma ideia remota. Neste trabalho, expõem-se alguns métodos que viabilizam a estimação de CMM, salientando brevemente alguns dos problemas que tornam impraticável a referida estimação de forma tradicional, isto é, como se de Cadeias de Markov univariadas se tratassem. O denominador comum da panóplia de material que existe sobre as CMM é o facto de que esse material versar as CMM enquanto variáveis dependentes, ou seja, enquanto um fim em si mesmo. Neste estudo, abordam-se conceptualmente as CMM enquanto regressores, ou seja, enquanto processos estocásticos que causam, a la Granger, uma certa variável dependente. Demonstra-se que, em termos potencial de previsão da referida variável dependente, a utilização de CMM conduz a melhores resultados do que outros métodos econométricos tradicionais concorrentes, resultados, esses, atestados através de estudos de simulação de Monte Carlo. Por fim, ilustra-se ainda a pertinência e a proficiência do método que propomos com uma aplicação económica empírica.

Referências

- [1] Ching, W., Fung, E. and Ng, M. (2002) *A Multivariate Markov Chain Model for Categorical Data Sequences and its Applications in Demand Predictions*, Journal of Management Mathematics 13, 187-199.
- [2] Raftery, A. (1985) *A model for high-order Markov chains.*, Journal of the Royal Statistical Society - Series B 47, 528.539.

Análise da fiabilidade de centros de maquinaria – um caso de estudo

(Comunicação)

Maria João DIAS – *Departamento de Matemática, Universidade de Aveiro*
moreiradias@ua.pt

Adelaide Freitas – *Departamento de Matemática & CIDMA, Universidade de Aveiro*

Constantino Pinto – *Renault C.A.C.I.A., Aveiro*

Resumo

A necessidade premente de melhorar a produtividade leva as empresas a uma constante procura de técnicas que permitam conhecer o correcto funcionamento dos equipamentos e, conseqüentemente, definir as melhores práticas e processos de optimização. Nesse sentido, a caracterização da fiabilidade dos equipamentos, tendo em conta o plano de ações de manutenção preventiva executado, providencia um instrumento de avaliação valioso. Neste trabalho são exploradas metodologias estatísticas e processos estocásticos adequados à análise do comportamento da fiabilidade, em termos da manutenção, de um conjunto de máquinas existentes numa unidade elementar de trabalho (UET) da unidade fabril da Renault C.A.C.I.A., com vista à determinação de melhores práticas preventivas a associar a cada máquina. Em particular, a regressão localmente ponderada (*lowess*, [1]) é proposta para estimar a curva de ajustamento do tempo médio sem intervenção entre duas preventivas consecutivas, em função do tempo entre preventivas. Esta ferramenta não paramétrica providencia uma representação gráfica da relação entre a fiabilidade e o plano de manutenção de cada máquina. Analisa-se ainda trajetórias dos processos $\{f(T(t)), t > 0\}$, para diversas funções f , onde $T(t)$ representa o número de intervenções (corretivas e preventivas) ocorridas num intervalo de amplitude t . Quando $f(T(t)) = 1 - \frac{x(t)}{T(t)}$, onde $x(t)$ representa o número (determinístico) de intervenções preventivas realizadas até ao instante t , aquelas trajetórias fornecem uma visualização da variação da percentagem de paragens da máquina devida a falhas de funcionamento (corretivas), por t unidades de tempo.

Agradecimentos

Trabalho subsidiado pelo FEDER, através do COMPETE, pelo CIDMA (Centro de Investigação e Desenvolvimento em Matemática e Aplicações) da Universidade de Aveiro e FCT, dentro do projecto PEst-C/MAT/UI4106/2011 com número COMPETE FCOMP-01-0124-FEDER-022690.

Referências

- [1] Cleveland, W., Devlin, S. (1988) *Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting*, Journal of the American Statistical Association, 403, 596–610.

Perfil dos doutorados em Portugal

(Comunicação)

Joana DUARTE – *Direção-Geral de Estatísticas em Educação e Ciência*

joana.duarte@dgeec.mec.pt

Rui Banha – *Direção-Geral de Estatísticas em Educação e Ciência*

Ricardo Cotrim – *Direção-Geral de Estatísticas em Educação e Ciência*

Luísa Canto e Castro – *Direção-Geral de Estatísticas em Educação e Ciência*

Resumo

Este estudo tem como objetivo apresentar as principais componentes descritoras do perfil do doutorados em Portugal através da exploração de dados provenientes do Inquérito aos Doutorados e a metodologias estatísticas de análises de doutorados multivariadas. No pressuposto de que as variáveis caracterizadoras dos respondentes ao inquérito CDH podem ser de algum modo determinantes para as respetivas carreiras e percursos profissionais, começa-se por utilizar técnicas de análise de dados multivariados para, de forma exploratória, reduzir a dimensionalidade e identificar alguns perfis de doutorados. Procedede-se, posteriormente, a uma análise contextualizada das respostas às questões relacionadas com a carreira profissional dos doutorados, participação em projetos de investigação e mobilidade entre instituições e empresas.

Referências

- [1] Schwabe, M. (2011) The Career Paths of Doctoral Graduates in Austria, *European Journal of Education*. 46, No.1, 153–168.

Alunos Portugueses no PISA: Modelos de Regressão Multinível

(Comunicação)

Susana FARIA – *Departamento de Matemática e Aplicações. CMAT, Centro de Matemática. Universidade do Minho, Portugal*

sfaria@math.uminho.pt

Liliana Faria – *Departamento de Matemática e Aplicações. Universidade do Minho*

Conceição Portela – *Universidade Católica Portuguesa, Porto*

Resumo

Os Modelos Multinível são modelos de regressão que se aplicam em situações em que os dados se estruturam hierarquicamente (por níveis).

No contexto educacional, este tipo de estrutura é muito usual: os alunos estão agrupados em turmas, as turmas em diferentes escolas e as escolas em áreas geográficas distintas. Dada a estrutura hierárquica dos dados escolares, a utilização de modelos clássicos de regressão podem resultar em inferências erradas, uma vez que assumem a independência entre os indivíduos. Os modelos multinível são modelos de regressão que, na sua construção, incorporam os diferentes níveis observacionais dos dados, produzindo assim inferências mais fiáveis.

Muitos são os fatores que podem influenciar o desempenho escolar dos alunos: fatores associados à família (características socioeconómicas e culturais), fatores escolares (infraestruturas, práticas didáticas, professores) e ainda aqueles relacionados ao próprio aluno (habilidade, motivação).

Neste trabalho pretende-se aplicar modelos de regressão multinível a dados de alunos de escolas nacionais, com o objetivo de identificar os principais fatores que podem determinar os resultados escolares dos alunos Portugueses relativamente à literacia em matemática e ciências, bem como estudar a contribuição das escolas portuguesas no desempenho escolar dos seus alunos, usando a base de dados do Programa Internacional de Avaliação de Estudantes (PISA) 2009. O Programa para Avaliação Internacional de Estudantes (PISA), coordenado

pela Organização para a Cooperação e Desenvolvimento Económico (OCDE), permite monitorizar os sistemas educativos em termos do desempenho dos alunos, no contexto de um enquadramento concetual aceite internacionalmente. Em Portugal, o programa PISA 2009 foi aplicado a 6298 alunos afectados a 212 escolas. Foram incluídos na população alvo todos os alunos de 15 anos a frequentarem a escola desde o 7 ano ao 11 ano de escolaridade.

Os fatores que, ao nível do aluno, foram considerados como podendo influenciar o seu desempenho escolar são: as características demográficas dos alunos (idade, sexo e situação de imigrante), as características familiares dos alunos (nível socioeconómico e cultural dos pais, estrutura familiar, ajuda pelos pais nos trabalhos de casa e frequência de explicações) e as características de aprendizagem/cognitivas dos alunos (uso de estratégias de controle e de elaboração dos alunos e situação de repetente). Ao nível da escola foram consideradas variáveis relacionadas com o contexto da escola (tais como localização, tipo de escola, dimensão da escola, percentagem de raparigas e média nível socioeconómico e cultural dos alunos), com os recursos da escola (tais como proporção de computadores ligados à internet, proporção de computadores por aluno e proporção de estudantes por professor) e com as políticas/estratégias adotadas na escola (características dos seus professores e alunos, características relacionadas com a admissão dos alunos na escola, com criação de turmas de melhor rendimento escolar e com as atividades extracurriculares oferecidas).

A principal conclusão deste trabalho é que existe um efeito significativo da escola a que os alunos pertencem na determinação dos seus resultados escolares, e o nível socioeconómico e cultural dos alunos é um dos principais determinantes do desempenho escolar dos alunos.

Agradecimentos

Este trabalho é financiado por Fundos FEDER através do Programa Operacional Factores de Competitividade - COMPETE e por Fundos Nacionais através da FCT - Fundação para a Ciência e a Tecnologia no âmbito do projecto PEst-C/MAT/UI0013/2011

Referências

- [1] Dias, V.M., Ferrão, M.E. (2006) *Modelo Multinível do Desempenho Escolar de Alunos Socialmente Desfavorecidos em Escolas Públicas/Privadas - Aplicação aos Dados Portugueses do PISA 2000* Psicologia e Educação 2, 63–77.
- [2] Gilleece, L., Cosgrove, J., Sofroniou, N.(2010)*Equity in Mathematics and Science Outcomes: Characteristics Associated with High and Low Achievement on PISA 2006 in Ireland.*, International Journal of Science and Mathematics Education 8, 475–496.
- [3] Goldstein, H. (2010) *Multilevel Statistical Models*, John Wiley and Sons.
- [4] Raudenbush, W., Bryk, A. S. (2002) *Hierarchical Linear Models: Applications and Data Analysis Methods* Newbury Park, CA: Sage.
- [5] Romão, P. (2012) *Alguns fatores determinantes dos resultados obtidos pelos alunos do 9 e 12 anos nos exames nacionais de português e matemática e o efeito escola. Tese de Doutoramento em Ciências da Educação da Universidade Católica Portuguesa, Porto.*

A Estatística ao serviço da Engenharia da Fiabilidade

(Comunicação)

Filipe Salvador FERNANDES – *ISEL*

filipedesalvadorfernandes@gmail.com

José Sobral – *ISEL*

Alda Carvalho – *GI-MOSM/ISEL & CEMAPRE/ISEG*

Resumo

Os estudos relativos à fiabilidade de bens têm vindo a tornar-se ao longo dos últimos anos uma área bastante importante e aliciante, pois permite que as organizações comecem a possuir um conhecimento mais profundo acerca da probabilidade de falha dos activos. Muitas vezes se observa que o sucesso ou insucesso de uma organização passa pela percepção atempada dos potenciais problemas, permitindo a tomada de decisão mais correcta e no momento mais adequado. Este tipo de decisões, quando bem suportadas, permitirá às organizações industriais obter significativas reduções de custos e um aumento bastante acentuado da disponibilidade dos seus bens.

De facto, é hoje muito comum encontrar na maior parte das indústrias um Departamento de Manutenção, onde a área da Fiabilidade é uma peça chave da gestão dos activos praticada diariamente. No entanto, o maior constrangimento encontrado quando se pretende realizar uma análise de fiabilidade é, por um lado, a inexistência de dados fidedignos relativos às falhas ocorridas e, por outro lado, a realização de um mau ajuste dos dados a distribuições estatísticas conhecidas, assim como uma deficiente determinação dos seus parâmetros.

O trabalho proposto visa mostrar uma aplicação que faz a análise, do ponto de vista da fiabilidade, a um conjunto de dados de falhas. Esta análise passa pela aplicação dos testes de ajustamento como o teste do qui-quadrado, Kolmogorv-Smirnov e Anderson-Darling às distribuições mais usuais como a Normal, Lognormal, Exponencial, Gama e Weibull. A aplicação foi desenvolvida usando o geogebra, a folha de cálculo do excel e e algumas macros do VBA, e tem como principal objectivo a ordenação das diferentes distribuições com base nos testes efectuados.

Estimativas robustas de localização e escala obtidas via bootstrapping

(Comunicação)

Fernanda Otilia FIGUEIREDO – *Faculdade de Economia, Universidade do Porto, e CEAUL*

otilia@fep.up.pt

Maria Ivette Gomes – *FCUL, DEIO e CEAUL, Universidade de Lisboa*

Resumo

As cartas de controlo constituem uma das principais ferramentas em Controlo Estatístico de Processos. A maior parte das cartas existentes são desenvolvidas sob a hipótese de que a distribuição subjacente aos dados do processo é conhecida, desconhecendo-se eventualmente alguns dos seus parâmetros. Recentemente grande ênfase tem sido dada à construção de cartas de controlo não paramétricas (ver, por exemplo, Chakraborti et al, 2011), que possivelmente serão mais adequadas quando a distribuição subjacente aos dados não é completamente conhecida, e também ao estudo do desempenho real de uma carta de controlo implementada com base na estimação de parâmetros adequados do processo (ver, entre outros, Jensen et al., 2006; Castagliola e Figueiredo, 2013). Do ponto de vista não paramétrico a monitorização de um processo incide essencialmente sobre o controlo de parâmetros de localização e de escala da distribuição subjacente aos dados. Neste trabalho iremos re-visitá-la a mediana total e a amplitude total, estatísticas amostrais definidas em Cox e Iguzquiza (2001) e Figueiredo e Gomes (2003, 2004), com o intuito de encontrar estimadores adequados para os parâmetros de interesse na fase I de implementação de uma carta de controlo. A partir de um conjunto de dados de referência, recolhidos de um processo considerado estável e sob controlo, iremos recorrer ao método de reamostragem bootstrap para definir estimadores consistentes para os parâmetros de interesse, e recorrer a simulações de Monte Carlo para avaliar a sua eficiência e compará-los com outros estimadores usualmente utilizados para o efeito.

Referências

- [1] Chakraborti, S., Human, S.W., Graham, M. A. (2011) *Nonparametric (Distribution-Free) quality control charts*, In Handbook of Methods and Applications of Statistics: Engineering, Quality Control, and Physical Sciences. N. Balakrishnan, Ed., John Wiley & Sons, New York; 298-329.
- [2] Castagliola, P., Figueiredo, F. (2013) *The Median Chart with Estimated Parameters*, European Journal of Industrial Engineering (Accepted for publication).
- [3] Cox, M.G., Iguzquiza, E.P. (2001) *The total median and its uncertainty*, In Advanced Mathematical and Computational Tools in Metrology 5, Ciarlini P et al. (eds). World Scientific Publishing Co. Singapore; 106-117.
- [4] Figueiredo, F. (2003) *Robust estimators for the standard deviation based on a bootstrap sample*, In Proc. 13th European Young Statisticians Meeting, 53-62.
- [5] Figueiredo, F., Gomes, M. I. (2004) *The total median in Statistical Quality Control*, Applied Stochastic Models in Business and Industry 20, 339-353.
- [6] Jensen, W. A., Jones-Farmer, L. A., Champ, C. H., Woodall, W. H. (2006) *Effects of parameter estimation on control chart properties: a literature review*, J. Quality Technology 38, 349-364.

Modelos Lineares Mistos Não Normais

(Comunicação)

Miguel FONSECA – *Centro de Matemática e Aplicações, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa*

fmig@fct.unl.pt

Elsa Moreira – *Centro de Matemática e Aplicações, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa*

João Tiago Mexia – *Centro de Matemática e Aplicações, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa*

Resumo

O objectivo deste trabalho é estudar a eficiência de estimadores quadráticos algébricos para componentes de variância. Os BQUE (Best Quadratic Unbiased Estimators) não requerem qualquer tipo de suposição em relação a distribuições, permitindo a utilização de outras distribuições para componentes aleatórias para além da ubíqua distribuição normal. É estudada a inferência para modelos lineares mistos não normais.

Referências

- [1] Fonseca, M., Mexia, J.T., Zmyślony, R. (2006) *Binary operations on Jordan algebras and orthogonal normal models*, Linear Algebra App. 417, 75–86.
- [2] Mueller, J. (1987) *Sufficiency and completeness in the linear model*, J. Multivariate Anal. 21, 312–323.
- [2] Seely, J. (1972) *Completeness for a family of multivariate normal distributions*, Ann. Math. Stat. 43, 1644–1647.

VIH/SIDA: monitorização da adesão ao tratamento

(Comunicação)

Fernando GOMES – *Escola de Ciências; Universidade do Minho*
parentegomes@hotmail.com

Cecília Azevedo – *Escola de Ciências; Universidade do Minho*

Resumo

Desde a década de 1990 que, com a introdução da terapêutica anti-retrovírica de alta eficácia no tratamento de doentes infetados com VIH, a SIDA passou a ser considerada uma doença crónica.

O tratamento desta doença em ambulatório exige que o paciente se dirija ao serviço de saúde para consultas, exames e levantamentos periódicos de medicação.

Através de variáveis tais como “Quantidade de medicamentos levantada pelo paciente” (podem ser vários tipos de medicação), “Quantidade de medicamentos, de determinado tipo, que o doente toma por dia”, do “Período de toma do medicamento”, “Número de dias de internamento” a que temos acesso, e do “Número de dias de duração do medicamento” que construímos a partir destas, propomos um sistema de vigilância que ajude o profissional de saúde a monitorizar a adesão dos doentes à terapêutica em ambulatório.

Definido esse indicador, tentar perceber, que doentes são mais propensos à não adesão, em função das suas características.

Os dados de que dispomos relativos aos doentes, tais como, idade, sexo, estado civil, distância de casa ao local de tratamento, subsistema de saúde e relativos a duas unidades de saúde distintas quer geograficamente, quer em número de doentes, permitem-nos uma caracterização dos doentes com esta patologia.

Embora as unidades de saúde não representem a população destes doentes na área geográfica em que se inserem, verificamos que não há diferenças estatisticamente significativas entre as principais características dos doentes em função da unidade de saúde onde estão a ser tratados.

Referências

- [1] Dias, A. M., Cunha, M., Santos, A., Neves, A., Pinto, A., Silva, A, Castro, S. (2011) Adesão ao regime Terapêutico na Doença Crônica: Revisão da Literatura, *Millenium*, 40, 201–219.
- [2] Pestana, D., Velosa, S. (2010) *Introdução à Probabilidade e à Estatística* (4a edição), Fundação Calouste Gulbenkian.
- [3] Thode Jr., H.C. (2002) *Testing for Normality*, Marcel Dekker, New York.
- [4] Wand, M.P; Jones, M.C. (1995) *Kernel Smoothing*. Chapman & Hall/CRC, London.

Estimação MOP de viés reduzido do parâmetro VaR

(Comunicação)

M. Ivette GOMES – CEAUL e DEIO, FCUL, Universidade de Lisboa

ivette.gomes@fc.ul.pt

Resumo

Um dos principais objectivos da *Estatística de Extremos Univariados* é a estimação de quantis extremos. Em muitas áreas de aplicação, como por exemplo na área financeira, pretendemos frequentemente estimar um quantil elevado, o chamado *Valor-em-Risco* num nível q (VaR_q), onde VaR é o acrónimo para o Inglês, “Value-at-Risk”. O nível q é usualmente próximo de zero, de modo a ser pequena a probabilidade de excedência do valor (VaR_q). Estamos pois interessados no parâmetro $\chi_{1-q} := F^{\leftarrow}(1 - q)$ associado a função de distribuição (f.d.) F subjacente aos dados, com $F^{\leftarrow}(y) = \inf \{x : F(x) \geq y\}$, a função inversa generalizada de F . Denotemos por $U(t)$ ta função inversa generalizada de $1/(1 - F)$. Então, para q pequeno, pretendemos estimar o parâmetro $\text{VaR}_q = U(1/q)$, $q = q_n \rightarrow 0$, $nq_n \leq 1$. Como estamos a trabalhar com uma probabilidade pequena, podemos-nos confinar à estimação da cauda. Para além disso, como em aplicações financeiras nos deparamos frequentemente com caudas pesadas, admitimos que a f.d. subjacente aos dados é tal que $1 - F(x) \sim c x^{-1/\gamma}$, quando $x \rightarrow \infty$, para reais positivos c e γ . Equivalentemente, e para algum $C > 0$, $U(t) \sim C t^\gamma$, quando $t \rightarrow \infty$, onde a notação $a(t) \sim b(t)$ significa que $a(t)/b(t) \rightarrow 1$, quando $t \rightarrow \infty$. O parâmetro γ é o *índice de valores extremos* (EVI, do Inglês “*extreme value index*”), o parâmetro fundamental de acontecimentos extremos (e elevados). Weissman (1978) propôs o estimador

$$Q_{\hat{\gamma}}^{(q)}(k) := X_{n-k:n} (k/(nq))^{\hat{\gamma}}, \quad (3)$$

onde $X_{n-k:n}$ é a $(k + 1)$ -ésima estatística ordinal de topo, $\hat{\gamma}$ é um qualquer estimador consistente de γ e Q é um acrónimo para quantil.

Tal como é usual na estimação semi-paramétrica de parâmetros de acontecimentos extremos, precisamos aqui também de trabalhar com sucessões de

inteiros *intermédias*, $k = k_n \rightarrow \infty$, $k \in [1, n)$, $k = o(n)$ quando $n \rightarrow \infty$. Para caudas pesadas, o estimador clássico do EVI-estimator, que é usualmente considerado em (3) para a estimação semi-paramétrica de um quantil, é o estimador de Hill $\hat{\gamma} = \hat{\gamma}(k) =: H(k)$ (Hill, 1975), com a expressão funcional, $H(k) := \sum_{i=1}^k V_{ik}/k$, $V_{ik} = \ln X_{n-i+1:n} - \ln X_{n-k:n}$, $1 \leq i \leq k$, o logaritmo da *média geométrica* (ou *média-de-ordem-0*) de $U_{ik} := X_{n-i+1:n}/X_{n-k:n}$, $1 \leq i \leq k < n$. Mais geralmente, Brillhante *et al.* (2013) consideraram como estatísticas básicas a *média-de-ordem-p* (MOP) de U_{ik} , $1 \leq i \leq k$, $p \geq 0$, que denotamos por $M_p(k)$, e a classe de estimadores MOP do EVI, os funcionais

$$H_p(k) \equiv \text{MOP}(k) \equiv \hat{\gamma}^{\text{H}_p}(k) := \begin{cases} \frac{1 - M_p^{-p}(k)}{p}, & \text{se } 0 < p < 1/\gamma, \\ \ln M_0(k) = H(k), & \text{se } p = 0, \end{cases} \quad (4)$$

com $H_0(k) \equiv H(k)$. O comportamento assintótico dos estimadores do EVI em (4) levou Gomes *et al.* (2013) a introduzir uma classe de estimadores MOP de viés reduzido do EVI, os chamados estimadores RBMOP do Inglês “reduced-bias MOP”, dados por

$$\text{RB}^*(k) \equiv \text{RB}_{\hat{\beta}, \hat{\rho}}(k) := \text{H}^*(k) \left(1 - \frac{\hat{\beta}(1 - \varphi(\hat{\rho}))}{1 - \hat{\rho} - \varphi(\hat{\rho})} \left(\frac{n}{k} \right)^{\hat{\rho}} \right), \quad (5)$$

onde $\text{H}^*(k)$ é o estimador MOP ótimo do EVI, sendo $(\beta, \rho) \in \mathbb{R} \times \mathbb{R}^-$ um vector de parâmetros de segunda-ordem na classe de modelos $U(t) = Ct^\gamma(1 + A(t)/\rho + o(t^\rho))$, $A(t) = \gamma \beta t^\rho$. Neste artigo estudamos o comportamento de novos estimadores de VaR_q , válidos para caudas pesadas, a nova classe de RBMOP estimadores do VaR baseados nos estimadores RBMOP do EVI, em (5), que são incorporados em (3). Desta forma, tencionamos melhorar o comportamento dos estimadores RB do VaR já existentes, estudados em Gomes and Pestana (2007). Derivamos as propriedades assintóticas da classe de estimadores proposta e comparamos os estimadores com estimadores alternativos, não só assintoticamente, mas também para amostras finitas, através de métodos de simulação de Monte-carlo. Fornecemos ainda aplicações na área das Finanças.

Referências

- [1] Brillhante, M.F., Gomes, M.I. and Pestana, D. (2013) *A simple generalisation of the Hill estimator*, J. Comput. Statist. and Data Anal. 57:1, 518–535.

- [2] Gomes, M.I. and Pestana, D. (2007) *A sturdy reduced bias extreme quantile (VaR) estimator*, J. Amer. Statist. Assoc. 102: 477, 280–292.
- [3] Gomes, Brillhante, M.F. and Pestana, D. (2013) *New Reduced-bias Estimators of a Positive Extreme Value Index*, Notas e Comunicações CEAUL 10/2013, submitted.
- [4] Hill, B.M. (1975) *A simple general approach to inference about the tail of a distribution*, Ann. Statist. 3, 1163–1174.
- [5] Weissman, I. (1978) *Estimation of parameters and large quantiles based on the k largest observations*, J. Amer. Statist. Assoc. 73, 812–815.

Metodologias estatísticas para estudo da interacção genótipo x ambiente em clones de videira

(Comunicação)

Elsa GONÇALVES – *Instituto Superior de Agronomia, Universidade de Lisboa*
elsagoncalves@isa.utl.pt

Antero Martins – *Instituto Superior de Agronomia, Universidade de Lisboa*

Resumo

Quando os materiais de propagação utilizados na viticultura são geneticamente homogéneos (clones), eles podem exibir sensibilidade à interacção genótipo×ambiente ($G \times E$). Isto significa que resultados da avaliação de um genótipo em determinados ambientes não são passíveis de generalização a outros ambientes, nos quais o genótipo irá ser posteriormente cultivado. Na videira o estudo interacção $G \times E$ não tem sido alvo de uma abordagem consentânea com a sua importância, encontrando-se as metodologias ainda insuficientemente desenvolvidas e sedimentadas.

No melhoramento de plantas em geral, as principais técnicas aplicáveis ao estudo deste fenómeno recorrem geralmente a modelos mistos (Smith et al. 2005), à análise de regressão dos rendimentos do genótipo sobre os sobre os índices ambientais (Finlay e Wilkinson, 1963), a métodos não paramétricos baseados na ordenação dos genótipos em diferentes ambientes (Nassar e Hühn, 1987; Hühn e Nassar, 1989) e a técnicas baseadas na análise em componentes principais (Kempton, 1984), este último tipo de abordagem com algumas variantes, dependendo da transformação aplicada aos dados antes de se fazer a decomposição em valores singulares.

O presente trabalho pretende reforçar metodologicamente o estudo da interacção genótipo \times ambiente em clones de videira. Utilizando dados reais de ensaios de selecção da videira, analisam-se várias das técnicas de interpretação da interacção acima referidas, discutindo a eficiência da sua aplicabilidade no contexto da selecção da videira.

Referências

- [1] Finlay, K., Wilkinson, G.(1963) *The analysis of adaptation in a plant breeding program*, Australian Journal of Agricultural Research 14, 742–754.
- [2] Hühn, M. , Nassar, R. (1989) *On tests of significance for nonparametric measures of phenotypic stability*, Biometrics 45, 997–1000.
- [3] Kempton, R.(1984) *The use of biplots in interpreting cultivar by environment interactions*, Journal of Agricultural Science 103, 123–135.
- [4] Nassar, R. , Hühn, M. (1987) *Studies on estimation of phenotypic stability: tests of significance for non parametric measures of phenotypic stability*, Biometrics 43, 45–53.
- [5] Smith, A., Cullis, B., Thompson, R. (2005) *The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches*, Journal of Agricultural Science 143, 449–462.

Modelação de erros em previsões de temperatura

(Comunicação)

Rui GONÇALVES – LIAAD INESC TEC e Universidade do Porto

rjasg@fe.up.pt

Resumo

Consideramos uma variável quase gaussiana X , que, após uma transformação de potência, X^c , onde $c = \{(2k+1)/(2j+1)\}$, $k, j = \{0, 1, \dots\}$, se torna Gaussiana. Os expoentes apropriados a cada caso podem ser encontrados usando a curtose da potência de uma variável Gaussiana de média zero. Aplicamos o método descrito a dados que consistem nos erros nas previsões diárias de temperatura máxima e mínima para o ano de 2011. No caso das temperaturas máximas encontramos os expoentes $9/11$ e $7/9$ como sendo apropriados para transformar os dados originais numa amostra Gaussiana o que significa que os dados originais podem ser bem ajustados por uma Gaussiana elevada à potência $11/9$ ou $9/7$.

Referências

- [1] Lefebvre, M. (2010) *Nearly Gaussian Distributions and Application/*, Communications in Statistics - Theory and Methods, 39, 823–836.
- [2] Stuart, A. and Ord, K. (1964) *Kendall's Advanced Theory of Statistics*, Oxford University Press.

Modelos de retorno financeiro

(Comunicação)

Maria do Carmo GUEDES – *CMUP e Faculdade de Ciências, Universidade do Porto*

mmguedes@fc.up.pt

Maria Eduarda Silva – *Faculdade de Economia, Universidade do Porto e CIDMA*

Jorge Brasil – *Faculdade de Ciências, Universidade do Porto*

Resumo

Observações empíricas ao longo do tempo levaram a concluir que nem a volatilidade tem um valor constante, nem os logo-retornos são normalmente distribuídos. Por isto o modelo de Black-Scholes tornou-se bastante limitado e novos modelos têm vindo a ser desenvolvidos ao longo dos anos. Neste trabalho faz-se um estudo comparativo sobre a performance de quatro modelos para o preço de opções financeiras: o modelo de Black-Scholes, o modelo de Heston em que à volatilidade é associado um comportamento estocástico e dois modelos denominados de Processo de Meixner e de Processo de Brown-Laplace cujos desenvolvimentos são baseados em processos estocásticos de Levy. A comparação entre os quatro modelos é feita com base em dezoito opções de compra sobre seis títulos cotados no mercado Norte Americano.

Referências

- [1] Reed, William J. (2007), *Brownian-Laplace motion and its use in financial modelling*, Communications in Statistics: theory and methods 36 (23), 473-484.
- [2] Coutinho, F., Barbachan, J.(2011) *Processo de Meixner: Teoria e aplicações no mercado financeiro brasileiro*, Estudos Econ., São Paulo, 41.

Validação Interna de Modelos Preditivos de Portadores de Mutação Genética em Síndrome Brugada

(Comunicação)

Carla HENRIQUES – *CMUC e Escola Superior de Tecnologia e Gestão, Instituto Politécnico de Viseu*

carlahenriq@estv.ipv.pt

Ana Cristina Matos – *Escola Superior de Tecnologia e Gestão, Instituto Politécnico de Viseu*

Luís Ferreira dos Santos – *Centro Hospitalar Tondela - Viseu*

Resumo

Num trabalho de Henriques et al. [2] foram estudadas formas de combinar linearmente alguns marcadores electrocardiográficos, com vista a obter um modelo de prognóstico para a presença de mutação genética em famílias afetadas por Síndrome Brugada (SB). A Síndrome Brugada é uma doença com carácter hereditário que predispõe à morte súbita cardíaca, sendo, habitualmente, o seu diagnóstico feito através de um electrocardiograma (ECG) onde se deteta o padrão típico, dito padrão de Brugada. A dificuldade deste meio de diagnóstico reside na intermitência deste padrão numa série de ECGs, podendo ser feitos vários ECGs a um paciente Brugada até que o padrão típico se manifeste. Em Henriques et al. [2] verificou-se que a combinação de algumas medidas electrocardiográficas, independentes do padrão típico, apresentava um bom desempenho quanto à capacidade de discriminação entre portadores e não portadores da mutação genética responsável pela doença. De facto, os modelos estudados apresentaram uma estimativa para a área sob a curva ROC (AUC) superior a 0.9, evidenciando uma excelente capacidade de discriminação. Contudo, é bem conhecido o otimismo do desempenho aparente dos modelos, isto é, do seu desempenho quando aplicados ao mesmo conjunto de dados que foi usado para os construir, face ao desempenho quando aplicados a outros conjuntos de dados (Steyerberg et al. [4]). A validação de modelos de prognóstico é, pois, fundamental para avaliar a sua habilidade para produzir

previsões fiáveis ou para discriminar de forma eficaz futuros casos. Podemos encontrar na literatura muitos procedimentos para avaliar o desempenho de um modelo na previsão. Steyerberg et al. [4] apresentam um estudo comparativo de muitos desses procedimentos e recomendam o Bootstrap para a validação interna de um modelo de regressão logística.

No caso em estudo em Henriques et al. [2], a reduzida dimensão amostral e o correlacionamento entre alguns dos marcadores, aumentam as preocupações com respeito ao desempenho preditivo dos modelos. Em situações como esta, as técnicas de contração dos coeficientes (shrinkage) são aconselhadas para melhorar as estimativas dos coeficientes do modelo e, assim, melhorar as previsões (Steyerberg et al. [3]). Um exemplo de aplicação destas técnicas a amostras de pequena dimensão no âmbito da vida selvagem é dado em Gude et al. [1], onde se registou uma notória melhoria das previsões em áreas geográficas onde as amostras eram pequenas.

Neste trabalho, adotando a metodologia Bootstrap, é avaliado o desempenho dos modelos de regressão logística considerados em Henriques et al. [2], comprovando-se a boa capacidade destes modelos na identificação de portadores de mutação genética, com vantagens evidentes face à deteção do padrão Brugada num ECG. Tal como em Gude et al. [1], são também aplicados métodos de contração (shrinkage) dos coeficientes do modelo, de modo a melhorar a sua capacidade preditiva.

Com este trabalho pretende-se salientar o potencial destes modelos no diagnóstico de Síndrome Brugada.

Referências

- [1] Gude, J.A., Mitchell, M.S., Ausband, D.E., Sime, C.A., Bangs, E.E. (2009) *Internal validation of predictive logistics regression models for decision-making in wildlife management*, *Wildlife Biology* 15: 352-369.
- [2] Henriques, C., Matos, A., Santos, L. F. *Brugada Syndrome Diagnosis – Three Approaches to Combining Diagnostic Markers*, (aceite para publicação em *Selected papers of the XIX Congress of the Portuguese Society of Statistics, International Book Series, Studies in Theoretical and Applied Statistics, Selected Papers of the Statistical Societies*).
- [3] Steyerberg, E.W., Eijkemans, M.J.C., Harrell, F.E. Jr., Habbema, J.D.F.

(2000) *Prognostic modeling with logistic regression analysis: a comparison of selection and estimation methods in small data sets*, *Statistics in Medicine* 19: 1059-1079.

[4] Steyerberg, E.W., Harrell, F.E. Jr., Borsboom, G.J.J.M., Eijkemans, M.J.C., Vergouwe, Y., Habbema, J.D.F. (2001) *Internal validation of predictive models: efficiency of some procedures for logistic regression analysis*, *Journal of Clinical Epidemiology* 54: 774-781.

Propriedade de Taylor no modelo TGARCH(1,1)

(Comunicação)

Joana LEITE – *Instituto Superior de Contabilidade e Administração de Coimbra, IPC, e CMUC*

jleite@iscac.pt

Esmeralda Gonçalves – *Departamento de Matemática, Universidade de Coimbra, e CMUC*

Nazaré Mendes-Lopes – *Departamento de Matemática, Universidade de Coimbra, e CMUC*

Resumo

Os factos estilizados de séries temporais financeiras têm assumido uma relevância crescente na valorização dos modelos adequados à sua descrição, como é o caso dos condicionalmente heteroscedásticos.

Entre os factos estilizados frequentemente associados a tais séries reais está o efeito de Taylor, presente quando a autocorrelação da série em valor absoluto é maior do que a autocorrelação, da mesma ordem, do seu quadrado (Taylor, 1986; Granger e Ding, 1995; Granger, 2005). Apesar deste facto estar presente tão frequentemente em séries reais, a análise da propriedade de Taylor, isto é, a contrapartida teórica do efeito de Taylor (He e Teräsvirta, 1999) nos modelos estocásticos de séries temporais, está ainda pouco desenvolvida na literatura conhecida. Tal dever-se-á seguramente ao facto das expressões teóricas das autocorrelações envolvidas serem de difícil obtenção ou, quando conhecidas, apresentarem elevado grau de complexidade.

Neste trabalho mostramos que a propriedade de Taylor está presente no modelo Threshold GARCH (TGARCH), alargando o âmbito do estudo feito por Gonçalves, Leite e Mendes-Lopes (2009) e por Haas (2009) no que respeita ao modelo e ainda às ordens das autocorrelações consideradas.

Exploramos ainda a extensão da região de parametrizações para as quais a propriedade é verificada, ilustrando-a para algumas distribuições marginais particulares do processo gerador do modelo.

Referências

- [1] Gonçalves, E., Leite, J., Mendes-Lopes, N. (2009) *A mathematical approach to detect the Taylor property in TAR_{CH} processes*, Statistics and Probability Letters. 79, 602–610.
- [2] Granger, C.W.J., Ding, Z. (1995) *Some properties of absolute return: an alternative measure of risk*, Annales D'Économie et de Statistique. 40, 67–91.
- [3] Granger, C.W.J. (2005) *The past and future of empirical finance: some personal comments*, Journal of Econometrics. 129, 35–40.
- [4] Haas, M. (2009) *Persistence in volatility, conditional kurtosis, and the Taylor property in absolute value GARCH processes*, Statistics and Probability Letters. 79, 1674–1683.
- [5] He, C., Teräsvirta, T. (1999) *Properties of moments of a family of GARCH processes*, Journal of Econometrics. 92, 173–192.
- [6] Taylor, S.J. (1986) *Modelling Financial Time Series*, John Wiley & Sons.

CreditRisk+: Uma Interpretação Bayesiana na Modelagem de Risco de Crédito

(Comunicação)

Raphael LOURENÇO – *Universidade Federal do Rio de Janeiro*

raphaelolourenco@gmail.com

Marco Aurélio Sanfins – *Universidade Federal Fluminense*

Gabriel Rodrigues Sanfins – *Universidade Federal do Rio de Janeiro*

Resumo

A demanda por metodologias robustas e mais poderosas de modelagem risco de crédito vem crescendo muito devido a instabilidade econômica financeira que se propaga pelo mundo. Portanto, as avaliações de risco de crédito se tornaram cada vez mais fundamentais para instituições financeiras e não-financeiras.

O modelo do CreditRisk+ é um modelo mais simples quando comparado aos outros modelos presentes no mercado de crédito. Este modelo foi desenvolvido pela Crédit Suisse Financial Products (CSFP) e é muito difundido no mercado de seguros pela sua facilidade já que não considera premissas sobre o motivo do default.

O objetivo principal é chegar a mensuração de perdas esperadas e não-esperadas em uma carteira de crédito. O CreditRisk+ considera que os pagamentos dos empréstimos são levados ao vencimento, ou seja, o pagamento ou o default é observado apenas na data do vencimento. O modelo considera apenas dois eventos para o devedor: inadimplente ou não.

Para mensurar os eventos de default o modelo sugere agrupamento dos devedores em faixas de exposição de tal forma que a distribuição de perda pode ser aproximada de uma *Poisson*(μ). Então é proposto por (CREDIT SUISSE FINANCIAL PRODUCTS, 1997) uma nova modelagem onde as incertezas, volatilidades, das taxas de default são incorporadas, trata-se de um modelo que assume uma distribuição Gama associada a essas incertezas.

Não é necessário um conhecimento mais profundo para perceber que a distribuição que se adequaria mais ao caso seria uma *Bernoulli*(p) para um devedor

ou uma $Binomial(n, p)$ para uma carteira de crédito.

Assumindo a volatilidade das taxas a distribuição de perda teórica é obtida através de uma fórmula fechada e o primeiro passo para essa distribuição é obter a distribuição de Poisson e Gama de cada faixa e realizar uma convolução. De acordo com as teorias de Estatística Bayesiana a multiplicação de uma Poisson por uma Gama é proporcional a uma Gama, cuja a distribuição preditiva é uma distribuição Binomial Negativa. A fórmula fechada define a partir de um cálculo recursivo sua distribuição de perda.

A partir da distribuição obtida é possível calcular o VaR (Value-at-Risk) de crédito assim como a distribuição de perda e algumas estimativas pontuais como a perda esperada em um período de tempo e a perda de capital.

O objetivo principal deste trabalho é mostrar que é possível reconstruir todo o modelo do CreditRisk+, usando as ferramentas da Inferência Bayesiana, obtendo os mesmos resultados para a distribuição das Perdas do Portfólio. Por fim, adaptaremos o modelo, usando a nova abordagem Bayesiana, substituindo as distribuições Poisson e Gama usadas inicialmente, por Binomial e Beta, respectivamente, calculando para esse novo caso, a distribuição de perdas, com a vantagem de que o novo modelo proposto esteja sendo mais fiel a realidade.

Referências

[1] Credit Suisse Financial Products (1997) CreditRisk+: A credit risk management framework.

Disponível em: <<http://www.ma.hw.ac.uk/~mcneil/F79CR/creditrisk.pdf>>.

[2] Sanfins, M. A. S., Clark, T. M. (2010) CreditRisk+.

Disponível em: http://www.ime.unicamp.br/sinape/sites/default/files/artigo2_0.pdf>.

A low-rank tensor method for large-scale Markov Chains

(Comunicação)

Francisco MACEDO – *CEMAT (Centro de Matemática e Aplicações) / EPFL (École Polytechnique Fédérale de Lausanne)*

francisco.quartin@gmail.com

António Pacheco – *CEMAT*

Daniel Kressner – *EPFL*

Resumo

A number of practical applications lead to Markov Chains with extremely large state spaces. Such an instance arises from models for calcium channels, which are structures in the body that allow cells to transmit electrical charges to each other. These charges are carried on a calcium ion which can travel freely back and forth through the *calcium channel*. The state space of a Markov process describing these interactions typically grows exponentially with the number of cells. More generally, Stochastic Automata Networks (*SANs*) are networks of interacting stochastic automata. The dimension of the resulting state space grows exponentially with the number of involved automata. Several techniques have been established to arrive at a formulation such that the transition matrix has Kronecker product structure. This allows, for example, for efficient matrix-vector multiplications. However, the number of possible automata is still severely limited by the need of representing a single vector (e.g., the stationary vector) explicitly. We propose the use of low-rank tensor techniques to avoid this barrier. More specifically, an algorithm will be presented that allows to approximate the solution of certain *SANs* very efficiently in a low-rank tensor format.

Uma aplicação da Análise de Dados Simbólicos

(Comunicação)

Carlos MARCELO – *Escola de Ciências; Universidade do Minho*

carlos.marcelo.pt@gmail.com

Cecília Azevedo – *Escola de Ciências; Universidade do Minho*

Resumo

No ano letivo 2010/2011 foi lançado pelo Conselho Pedagógico da Escola de Ciências da Universidade do Minho (ECUM) aos alunos do 1º Ciclo de estudos, um questionário com o objetivo principal de conhecer as razões que os levam ao abandono dos Cursos desta Escola. O conhecimento sobre as motivações, expectativas, desmotivações, associadas ao interesse de continuidade para o 2º Ciclo de estudos e nível posterior, são fatores determinantes para a obtenção de indicadores que podem auxiliar no planeamento e gestão da atividade da ECUM.

Numa análise multivariada clássica, cada indivíduo i é descrito por um conjunto de variáveis j , sendo que o elemento genérico ij da matriz é um valor único. No caso da Análise Dados Simbólicos (ADS), o elemento genérico pode ser uma distribuição de frequências ou mesmo um intervalo, dependendo do tipo da variável de interesse.

Em termos práticos, dispomos inicialmente de um conjunto de indivíduos (microdados ou objetos simbólicos de primeiro nível), e pretendemos efetuar um estudo com base na agregação desses registos de acordo com uma dada característica, obtida por uma ou várias variáveis (macrodados ou objetos simbólicos de 2º nível). Na análise clássica, o valor da célula da tabela é um valor único (uma frequência, uma média, . . .); nas células da tabela simbólica, tem-se uma distribuição de frequências ou um intervalo, caso as variáveis de interesse sejam, respetivamente, de tipo categórica ou quantitativa. Assim sendo, a análise simbólica é mais completa que a clássica, uma vez que, para além de outras características, integra a variação interna associada às variáveis em estudo.

Será efetuada uma aplicação da análise simbólica à base de dados de respostas do questionário.

Referências

- [1] Diday, E., Noirhomme-Fraiture, M. (2008) Symbolic Data Analysis and the SODAS Software, John Wiley & Sons, Ltd
- [2] H.-H. Bock, E. Diday (2000) Analysis of Symbolic Data - Exploratory Methods for Extracting Statistical Information from Complex Data, Springer

Modelação espaço-temporal. Aplicação a dados de monitorização ambiental

(Comunicação)

Luís MARGALHO – *Instituto Superior de Engenharia de Coimbra & CMAT-UM*

lmelo@isec.pt

Raquel Menezes – *CMAT - Universidade do Minho*

Inês Sousa – *CMAT - Universidade do Minho*

Resumo

Dados espaço-temporais resultam de observações efectuadas em diversas localizações, repetidamente ao longo do tempo. Para este tipo de dados, propomos um modelo espaço-temporal para dados gaussianos, definido, na localização $\mathbf{s} \in IR^2$ e no instante $t \in IN$, por

$$Y(\mathbf{s}, t) = \mu(\mathbf{s}, t) + \omega(\mathbf{s}, t) + \varepsilon(\mathbf{s}, t)$$

A decomposição deste modelo incorpora uma componente média $\mu(\mathbf{s}, t)$ utilizando possíveis covariáveis observadas $f_i(\mathbf{s}, t)$, na forma

$$\mu(\mathbf{s}, t) = \sum_{i=1}^n \beta_i f_i(\mathbf{s}, t)$$

e tal que $E[Y(\mathbf{s}, t)] = \mu(\mathbf{s}, t)$. A componente espaço-temporal $\omega(\mathbf{s}, t)$ e a componente de resíduos $\varepsilon(\mathbf{s}, t)$ são tais que

$$\omega(\mathbf{s}, t) \sim MVN(0, \Sigma(\theta)) \tag{6}$$

e

$$\varepsilon(\mathbf{s}, t) \sim N(0, \tau^2)$$

Em (6), sendo N o número de localizações observadas e T o número de campanhas, $\Sigma(\theta)$ é uma matriz simétrica $T \times T$ com elementos $\Sigma_{t_k, t_l}(\theta)$ matrizes $N \times N$, tais que a linha i e a coluna j representa $\text{Cov}[Y(\mathbf{s}_i, t_k), Y(\mathbf{s}_j, t_l)]$.

Este modelo assume separabilidade na estrutura de covariância, traduzindo-se na existência de funções de covariância puramente espacial e puramente temporal, Cov_S e Cov_T , por forma a que

$$\begin{aligned} \text{Cov} [Y(\mathbf{s}_i, t_k), Y(\mathbf{s}_j, t_l)] &= \text{Cov}_S (\|\mathbf{s}_i - \mathbf{s}_j\|) \times \text{Cov}_T (|t_k - t_l|) \\ &= \sigma^2 \rho(\|\mathbf{s}_i - \mathbf{s}_j\|; \phi) \xi_{kl} \end{aligned}$$

representando σ^2 e ϕ a variância total e o raio de influência, respetivamente. Nos casos em que o número de tempos observados é reduzido, a covariância temporal será modelada como efeitos fixos dados por ξ_{kl} , $k, l = 1, \dots, T$.

Como aplicação do modelo proposto, serão utilizados dados de biomonitorização recolhidos em Portugal continental em $N=146$ localizações comuns a $T=3$ campanhas. Serão obtidas estimativas da máxima verosimilhança para os parâmetros do modelo, que permitirão a construção de mapas de concentração predita para a campanha de amostragem mais recente. Como medida do erro cometido, serão apresentados mapas referentes à variância de interpolação.

Referências

- [1] de Cesare L., Meyers D.E., Posa D. (2001) *Estimating and modelling space-time correlation structures*, Statistics and probability letters, 51, 9–14.
- [2] Figueira R., Sérgio C., Ramalho C., Sousa A. J. (2002) *Distribution of trace metals in moss biomonitors and assessment of contamination sources in Portugal*, Environmental Pollution, 118, 153–163.
- [3] Kyriakidis P.C., Journel A.G. (1999) *Geostatistical Space-Time Models: A Review*, Mathematical Geology, 31 (6), 651–684.
- [4] Martins A., Figueira R., Sousa A.J., Sérgio C. (2012) *Spatio-temporal patterns of Cu contamination in mosses using geostatistical estimation*, Environmental Pollution, 170, 276–284.
- [5] Sherman, M. (2011) *Spatial statistics and spatio-temporal data. Covariance functions and directional properties*, John Wiley & Sons, Ltd

Probabilidade de não-contacto numa sondagem por telemóvel: uma abordagem multinível

(Comunicação)

Catarina MARQUES – *Instituto Universitário de Lisboa (ISCTE-IUL),
Business Research Unit*

cmam@iscte.pt

Paula Vicente – *Instituto Universitário de Lisboa (ISCTE-IUL), Business
Research Unit*

Resumo

Nos últimos anos as sondagens têm-se confrontado com um declínio nas taxas de resposta [1], fruto da tendência por parte das pessoas em recusar a colaboração, mas sobretudo da crescente dificuldade em contactar as pessoas. Groves et al. [2, p. 170] propõem um modelo explicativo da probabilidade de (não-)contacto nas sondagens que toma como variáveis explicativas factores relativos a: (a) contexto social dos indivíduos, (b) impedimentos de acesso, (c) atributos socio-demográficos dos indivíduos e (d) processo de contacto.

Os telemóveis são cada vez mais usados para realizar sondagens e importa por isso investigar as suas potencialidades e vantagens efectivas como modo de sondagem. Os telemóveis têm a particularidade de tornar o seu utilizador contactável em qualquer lugar, dia ou horário, o que, em teoria, favorece a sua utilização para a realização de sondagens. Porém, como em qualquer modo de sondagem, a probabilidade de contacto por telemóvel não é uniforme para todas as unidades a inquirir e conhecer os factores que potenciam o não-contacto torna-se relevante quer para um planeamento mais eficaz do desenho da sondagem (por exemplo, definição de esquemas de estratificação ou pós-estratificação) quer para uma gestão mais eficiente do trabalho de entrevista (por exemplo, agendamento de contactos e gestão da equipa de entrevistadores).

O objectivo deste estudo é investigar os factores determinantes dos não-contactos no contexto de uma sondagem por telemóvel. Para tal, é necessário

dispor de informação sobre as unidades não-contactadas. Uma das abordagens para estudar os "não-contactos" é considerar que as unidades que não se conseguiram contactar na primeira tentativa de contacto são representativas dos "não-contactos" [3]. Esta investigação adopta esta abordagem considerando os entrevistados no 1º contacto como "contactos" e os entrevistados após 2 ou mais contactos como "não-contactos". São utilizados dados de uma sondagem por telemóvel realizada em Portugal a indivíduos com 15 ou mais anos. Os números de telemóvel marcados foram gerados aleatoriamente tendo a amostra sido estratificada por operador de serviço móvel. A amostra final foi de 1501 entrevistas. Utiliza-se uma modelação logit multinomial multinível considerando como variável dependente a ordem de contacto dos indivíduos - 1º contacto, 2º contacto e 3º ou mais contactos. O período de contacto (dia horário) é a variável de 2º nível e como variáveis explicativas de 1º nível o modelo inclui factores relativos a (a) utilização do telemóvel e (b) características socio-demográficas dos indivíduos (pessoais e de envolvente social).

Os resultados apontam para diferenças significativas entre contactos e não-contactos sobretudo ao nível do perfil socio-demográfico.

Agradecimentos: Fundação para a Ciência e Tecnologia, PTDC/EGE-GES/116934/2010.

Referências

- [1] De Heer, W. (1999) *International response trends: results of an international survey*, Journal of Official Statistics 15, 129–142.
- [2] Groves, R., Fowler, F., Jr., Couper, M., Lepkowski, J., Singer, E., Tourangeau, R. (2004) *Survey methodology*. New York, Wiley.
- [3] Johnson, T., Cho, Y. (2004) *Understanding nonresponse mechanisms in telephone surveys*, Proceedings of the American Association for Public Opinion Research Conference. AAPOR, 4952–4959.

Representação assintótica de integrais de Kaplan-Meier com covariáveis

(Comunicação)

Jorge MENDONÇA – *Instituto Superior de Engenharia do Porto*

jpm@isep.ipp.pt

Jacobo de Uña Álvarez – *Departamento de Estatística e Investigação Operacional da Universidade de Vigo*

Resumo

O método Kaplan-Meier é usualmente utilizado em estudos de tempos de vida com censura aleatória. A consistência forte dos integrais de Kaplan-Meier foi provado em Stute e Wang (1993) e na presença de covariáveis em Stute (1993). Uma representação assintótica destes integrais foi introduzida em Stute (1995) e em Stute (1996) com a presença de covariáveis. No entanto, este método perde precisão quando existem muitos dados censurados. Neste caso, pode ser útil uma aproximação semiparamétrica desenvolvida em detalhe em Dikta (1998, 2000 e 2005). Neste cenário, a consistência forte dos integrais de Kaplan-Meier foi provada em Dikta (2000) e com a presença de covariáveis em Uña-Álvarez e Rodríguez-Campos (2004). No presente trabalho obtemos uma representação assintótica dos integrais de Kaplan-Meier com covariáveis, segundo um modelo de censura semiparamétrico. Como corolário apresentamos uma extensão do Teorema do Limite Central para integrais presuavizados.

Referências

- [1] de Uña-Álvarez J, Rodríguez-Campos C.,(2004) *Strong consistency of presmoothed Kaplan-Meier integrals when covariables are present* , *Statistics*. 38, 483–496.
- [2] Dikta G.,(1998): *On semiparametric random censorship models* , *Journal of Statistical Planning and Inference*. 66, 253–279.

- [3] Dikta G.(2000) *The strong law under semiparametric random censorship models* ,Journal of Statistical Planning and Inference. 83, 1–10.
- [4] Dikta G.,J. Ghoraï, and C. Schmidt.(2005) *The central limit theorem under semiparametric random censorship models* , Journal of Statistical Planning and Inference. 127, 23-51.
- [5] Stute, W.(1993) *Consistent estimation under random censorship when co-variables are present*, Journal of Multivariate Analysis. 45, 89–103.
- [6] Stute, W. and Wang J-L,(1993) *The strong law under random censorship*, The Annals of Statistics. 21, 1591 - 1607.
- [7] Stute, W.(1995) *The central limit theorem under random censorship*, The Annals of Statistics.23, 422–439.
- [8] Stute, W.(1996) *Distributional convergence under random censorship when covariates are present*, Scandinavian Journal of Statistics. 23, 461 - 471.

O processo bivariado periódico INAR(1) com uma aplicação ao número de fogos

(Comunicação)

Magda MONTEIRO – *ESTGA e CIDMA- Universidade de Aveiro*

msvm@ua.pt

Isabel Pereira – *Departamento de Matemática e CIDMA- Universidade de Aveiro*

Manuel Scotto – *Departamento de Matemática e CIDMA- Universidade de Aveiro*

Resumo

Nos últimos anos, o estudo de modelos para séries de contagem tem sido alvo de grande interesse com grande destaque para os modelos univariados. Mais recentemente Pedeli e Karlis (2011) apresentaram o modelo bivariado INAR(1) cuja generalização é feita neste trabalho através da introdução e estudo do modelo bivariado INAR(1) com estrutura periódica. Apresentam-se algumas propriedades probabilísticas e estatísticas do modelo realçando o caso em que as inovações têm distribuição Poisson bivariada periódica. Como a distribuição de Poisson não permite a sobredispersão dos dados é ainda analisado o caso em que as inovações têm outro tipo de distribuição bivariada periódica. O trabalho termina com a aplicação do modelo ao número mensal de fogos nos distritos de Aveiro e de Coimbra.

Referências

- [1] Pedeli, X., Karlis, D. (2011) *A bivariate INAR(1) process with application*, Statist. Model. 11, 325–349
- [2] Johnson, N. L., Kotz, S., Balakrishnan, N. (1997) *Discrete Multivariate Distributions*, John Wiley & Sons.

Testes F com amostras de dimensão aleatória. Análise de potência e aplicações

(Comunicação)

Elsa MOREIRA – *CMA, FCT, UNL*

efnm@fct.unl.pt

João Tiago Mexia – *CMA, FCT, UNL*

Christoph Minder – *University of Zurich*

Resumo

Suponha-se que existe um determinado prazo pré-fixado para recolha de observações num estudo de comparação, por exemplo, entre diversas patologias de pacientes que chegam a um hospital de forma sequencial e que os dados são recolhidos do paciente à medida que ele chega. Torna-se impossível conhecer à partida o tamanho das amostras e a repetição do mesmo estudo noutra altura para o mesmo prazo pré-fixado levaria provavelmente à recolha de um diferente número de pacientes por patologia. Nesta situação torna-se aconselhável considerar as dimensões das amostras dos níveis de uma ANOVA como realizações de variáveis aleatórias de Poisson. A estatística para um teste F, sob esta assumpção, terá então distribuição F condicional ao número de observações. Usando esta abordagem, testes mais potentes e levando a valores críticos ligeiramente inferiores são obtidos. Além disso, a abordagem pode também ser usada, numa fase de planeamento do estudo, para obter a duração para o período de recolha das observações de forma a assegurar uma pré-determinada potência para o teste.

Referências

Moreira, E. E., Mexia J.T., Minder C.E. (2013) *F tests with random sample size. Theory and applications*, Statistics & Probability Letters 83 (6), 1520–1526

Mexia, J.T., Moreira, E.E. (2010) *Randomized sample size F tests for the one-way layout* AIP Conference Proceedings, volume 1281- ICNAAM 2010- 8th International Conference of Numerical Analysis and Applied Mathematics,

1248–1251, doi:10.1063/1.3497917.

Mexia, J.T., Nunes, C., Ferreira, D., Ferreira, S., Moreira, E. (2011) *Orthogonal fixed effects ANOVA with random sample sizes*, WSEAS proceedings of the 5th International conference on Applied Mathematics, Simulation, Modelling, 84–90

On tail trend detection: modeling relative risk

(Comunicação)

Cláudia NEVES – *University of Aveiro & CEAUL*

claudia.neves@ua.pt

Laurens de Haan – *University of Lisbon & Erasmus University Rotterdam*

Albert Klein Tank – *KNMI, Royal Netherlands Meteorological Institute*

Abstract

The climate change dispute is about changes over time of environmental characteristics (such as rainfall). Some people say that a possible change is not so much in the mean but rather in the extreme phenomena (that is, the average rainfall may not change much but heavy storms may become more or less frequent). In this talk, we address changes over time in the probability that some high threshold is exceeded. The model is such that the threshold does not need to be specified, the results hold for any high threshold. For simplicity a certain linear trend is studied depending on one real parameter. Estimation and testing procedures (is there a trend?) are developed. Simulation results are presented. The method is applied to trends in heavy rainfall at 18 gauging stations across Germany and The Netherlands. A tentative conclusion is that the trend seems to depend on whether or not a station is close to the sea.

Research partially supported by ENES – Extremes in Space, project PTDC/MAT/112770/2009 and by national funds through the Fundação Nacional para a Ciência e Tecnologia, Portugal – FCT under the project PEst-OE/MAT/UI0006/2011.

Influência da dimensão da amostra na detecção de outliers em análise Forense

(Comunicação)

Maria Manuela NEVES – *Instituto Politécnico da Guarda CEAUL e UDI/IPG*
mfigueira@ipg.pt

Fernando Rosado – *Faculdade de Ciências da Universidade de Lisboa DEIO*
e CEAUL

Resumo

Na literatura de estatística forense são muitos os estudos que envolvem a análise de fragmentos de vidro e em particular a análise do índice de refração desses fragmentos. Para além do ADN os fragmentos de vidro são o tipo de prova mais estudado e referido na bibliografia sobre análise forense. Os fragmentos podem surgir de janelas ou portas a partir das quais foi forçada a entrada, de vidros de carros ou outros objetos usados durante um ato criminoso. A análise da prova relativa a fragmentos de vidro consiste, genericamente, na comparação de propriedades físicas e químicas dos fragmentos encontrados no suspeito com a sua possível origem. Este tipo de prova é de grande importância na associação do indivíduo que quebrou o vidro com a origem do vidro. Sendo os fragmentos de vidro uma das evidências mais encontradas na maior parte dos crimes, é de extrema importância a sua análise para a descoberta dos culpados.

O estudo de outliers, é um tópico de interesse e transversal em diversas áreas de investigação e tem um papel importante na análise de dados relativos a fragmentos de vidro. No contexto forense predomina a utilização de metodologias bayesianas enquanto que no estudo de outliers a abordagem é maioritariamente frequencista. Na dicotomia “frequencismo/bayesianismo” é muito importante enquadrar a problemática do estudo estatístico de outliers no contexto forense.

A característica presente nas observações outliers de apresentarem um certo afastamento em relação às restantes observações da amostra depende de diver-

sas componentes a que está ligado o valor aberrante e uma delas é a dimensão da amostra onde o valor foi observado, Rosado (2006). O aumento da dimensão da amostra apresenta uma maior possibilidade de aparecimento de outliers. No entanto, o efeito surpresa, que é uma característica desse tipo de observação, pode contrariar aquela possibilidade. Para investigar sobre o efeito da dimensão da amostra no aparecimento de observações outliers, vamos neste trabalho apresentar algumas aplicações do estudo de outliers a dados forenses. Mantem-se a controvérsia da utilização de métodos clássicos versus métodos bayesianos no estudo de observações outliers, que deve ser aprofundada

Referências

- [1] Aitken, C.G.G. e Lucy, D. (2004) *Evaluation of trace evidence in the form of multivariate data*, Applied Statistics, 53, 109–122.
- [2] Aitken, C. e Taroni, F. (2004) *Statistics and Evaluation of Evidence for Forensic Scientists*, 2ª ed. Wiley.
- [3] Barnettt, V. e Lewis, T. (1994) *Outliers in Statistical Data*, 3ª edição. Wiley.
- [4] Evett, I.W.(1977) *The interpretation of refractive index measurements*, Forensic Science, 209–217.
- [5] Rosado, F. (2006) *Outliers em Dados Estatísticos*. Edições SPE - Sociedade Portuguesa de Estatística.
- [6] Terry, K.W; Riessen, A.van e Lynch, B.F. (1983) *Identification of small glass fragments for forensic purposes*. Government Chemical Laboratories, Western Australia.

A New Model for Multivariate Markov Chains

(Comunicação)

João Nicolau – *ISEG/UL*

nicolau@iseg.utl.pt

Abstract

We propose a new model to estimate multivariate Markov chains of order one or higher. Through a numerical analysis, a Monte Carlo experiment and an empirical application, we show that the proposed specification is more precise than the mixture transition distribution (MTD) model.

References

- [1] Berchtold, A., Raftery, A. (2002). The Mixture Transition Distribution Model for High-Order Markov Chains and Non-Gaussian Time Series. *Statistical Science*, 17, 328-356.
- [2] Ching, W., Fung, E., (2002). A Multivariate Markov Chain Model for Categorical Data Sequences and Its Applications in Demand Predictions. *Journal of Management Mathematics* 13, 187-199.
- [3] McQueen, G., Thorley, S. (1991). Are Stock Returns Predictable? A Test Using Markov Chains. *Journal of Finance* 46, 239-263
- [4] Raftery, A. (1985). A model for high-order Markov chains. *Journal of the Royal Statistical Society - Series B* 47, 528–539
- [5] Zhu, D., Ching, W. (2010). A New Estimation Method for Multivariate Markov Chain Model with Application in Demand Predictions. In *BIFE '10 Proceedings of the 2010 Third International Conference on Business Intelligence and Financial Engineering* (eds L. Yu, K. Lai, S. Wang), Hong Kong.

Processos pontuais espaço-temporais aplicados ao estudo dos fogos florestais em Portugal

(Comunicação)

Paula PEREIRA – *ESTSetúbal, Instituto Politécnico de Setúbal e CEAUL-FCUL, Universidade de Lisboa*

paula.pereira@estsetubal.ips.pt

Kamil Feridun Turkman – *CEAUL-FCUL, Universidade de Lisboa*

Resumo

Os fogos florestais, particularmente em Portugal, são das catástrofes naturais mais graves, não só pela elevada frequência com que ocorrem e extensão que alcançam, como pelos efeitos destrutivos que causam, provocando grandes prejuízos económicos e ambientais. Os fogos florestais podem ser vistos como padrões espaço-temporais de pontos. Neste trabalho procurar-se-à modelar a localização dos fogos florestais em Portugal recorrendo a modelos hierárquicos Bayesianos para processos pontuais espaço-temporais. Para atingir este objetivo iremos recorrer aos processos de Cox log-Gaussianos [1] e à metodologia SPDE (equações diferenciais parciais estocásticas) introduzida por Lindgren et al. (2011) [2]. A base de dados é constituída por 13457 registos de fogos florestais que ocorreram em Portugal durante os anos de 1975 a 2005 e cuja a área ardida foi no mínimo de 35 hectares.

Referências

- [1] Møller, J., Syversveen, A., Waagepetersen, R. (1998) *Log Gaussian Cox Processes*, Scandinavian Journal of Statistics. 25, 451–482.
- [2] Lindgren, F., Rue, H., Lindstrøm, J. (2011) *An explicit link between Gaussian fields and Gaussian Markov random fields: the SPDE approach (with discussion)*., Journal of the Royal Statistical Society, Series B. 73, 423–498.

Modelação Espacial de Variáveis Latentes da Análise Fatorial - o caso da Biomonitorização Ambiental em Portugal Continental

(Comunicação)

Helena PIAIRO – *Universidade do Minho, Departamento de Matemática e Aplicações / Centro de Matemática*

helen.piairo@gmail.com

Raquel Menezes – *Universidade do Minho*

Inês Sousa – *Universidade do Minho*

Resumo

Determinados organismos vivos apresentam capacidades de acumulação de metais pesados e uma sensibilidade à poluição e, por este motivo, tem sido largamente utilizados para quantificar a concentração desses poluentes no ambiente [3] e [6]. Assim, o uso destes organismos como biomonitores, através da determinação da concentração acumulada de metais pesados, funciona como um indicador da concentração dos mesmos no ambiente, tornando-se uma metodologia bastante simples e com uma elevada resolução espacial e temporal, que irá proporcionar uma interpretação relevante em termos da qualidade ambiental [4].

Neste trabalho, pretende-se fazer um estudo estatístico, através da análise multivariada [2] e [7], dos dados amostrados através de uma rede de biomonitorização que abrangeu Portugal continental. Os dados foram obtidos ao longo de três campanhas, nomeadamente em 1992, 1997 e 2002, tendo sido recolhidas amostras de musgo em 173, 177 e 151 localizações, respetivamente, definidas numa escala nacional mas intensificadas em grandes zonas urbanas e industriais (tais como Lisboa/Vale do Tejo, Porto/Aveiro e Complexo Industrial de Sines).

É objetivo deste estudo o reconhecimento das relações lineares existentes entre os conjuntos múltiplos de dados a que temos acesso, com o intuito de encontrar diferentes fontes de poluição que estejam relacionadas com as concentrações dos metais pesados nos musgos, constituindo variáveis que não podem ser

medidas diretamente no meio ambiente - variáveis latentes. Desta forma, recorre-se ao método de Análise Fatorial, que tem vindo a ser utilizada em diversas áreas do conhecimento, como por exemplo, ciências ambientais [5] e [8]. Esta metodologia permite que estas variáveis possam ser tratadas como contínuas, resultando da separação de um número limitado de associações discretas, fornecendo um meio para testar se esta divisão é estatisticamente significativa.

Uma vez que esta análise não implica a ortogonalidade das variáveis latentes, potencia-se o valor indicador dos biomonitores, cujas concentrações de elementos químicos possam resultar, simultaneamente, da contaminação por diferentes fontes ambientais. Assim, ter-se-á o reconhecimento das diferentes fontes de poluição e respetivas áreas críticas, evidenciando-se o potencial valor deste tipo de análise.

Pretende-se, ainda, aplicar técnicas de modelação geoestatística [1] sobre os scores das variáveis latentes obtidas através da metodologia referida. Deste modo, consegue-se obter mapas de valores preditos das diferentes fontes de poluição, resumindo desta forma, a informação obtida através das concentrações dos vários elementos nos musgos, determinando-se quais os locais que padecem de determinada fonte de poluição.

Os dados analisados no presente estudo foram cedidos pelo Centro de Biologia Ambiental da Universidade de Lisboa, parceiro do projeto "Modelos conjuntos para processos espaço-temporais e respetivo desenho amostral, em Ciências do Ambiente e Saúde", PTDC/MAT/112338/2009, financiado por fundos nacionais através da FCT, no qual este trabalho de investigação se integra.

Referências

- [1] Diggle P., Ribeiro P., (2007) *Model-based Geostatistics*, Springer.
- [2] Everitt, B.S., Dunn G. (2001) *Applied Multivariate Data Analysis*, Second edition, Arnold.
- [3] Figueira R., Ferreira J., Sérgio C., Ramalho C., Sousa A.J. (2004) *Distribuição da deposição atmosférica de metais em Portugal analisada através de musgos biomonitores*, Atas da 8ª Conferência Nacional sobre a Qualidade do

Ambiente. 27 a 29 de Outubro de 2004, Lisboa.

- [4] Figueira R., Sérgio C., Ramalho C., Sousa A.J. (2002) *Distribution of trace metals in moss biomonitors and assessment of contamination sources in Portugal*, Environmental Pollution. 118, 153–163.
- [5] Kaplunovsky A. (2005) *Factor analysis in environmental studies*, Journal of Science and Engineering B2. 1–2, 54–94.
- [6] Martins A., Figueira R., Sousa A.J., Sérgio C. (2012) *Spatio-temporal patterns of Cu contamination in mosses using geostatistical estimation*, Environmental Pollution. 170, 276–284.
- [7] Morrison D. (2005) *Multivariate Statistical Methods*, Fourth edition, McGraw-Hill.
- [8] Xie Y., Hopke P., Paatero P., Barrie L., Li S. (1999) *Identification of Source Nature and Seasonal Variations of Arctic Aerosol by Positive Matrix Factorization*, Journal of the Atmospheric Sciences. 56, 249–260.

h-index: a stochastic approach

(Comunicação)

Rita PIMENTEL – *CEMAT and Instituto Superior Técnico*

rita.duarte.pimentel@gmail.com

António Pacheco – *CEMAT and Instituto Superior Técnico*

Resumo

We use bibliometric methods to explore the academic impact of a researcher or a set of researchers. In the line of Burrell (2007), we will introduce a stochastic model for an author's publication and citation processes with the aim of studying the *h*-index distribution and its moments. The *h*-index was created by Hirsch (2005) to quantify a scientist's publication productivity. This index is based on the set of most cited papers of an author and the number of citations that those papers received from other authors in other publications. The *h*-index is currently the most widely accepted index to rank scientists, appearing in the most popular databases. We assume a counting process for the publication of papers from an author and consider that each published paper leads to a new citation counting process, which counts the number of citations the paper receives from other authors. We will investigate the effect that changes in relevant parameters of the publication and citation processes, like mean publication rate and mean citation rate, have on the *h*-index. For a better understanding of our findings, we will illustrate numerically the results obtained on the *h*-index.

Referências

- [1] Hirsch, J.E. (2005) *An index to quantify an individual's scientific research output*, Proceedings of the National Academy of Sciences of the United States of America. 102(46), 16569–16572.
- [2] Burrell, Q. (2007) *Hirsch's h-index: A stochastic model*, Journal of Informetrics. 1(1), 16–25.

Testes estatísticos em tempo real em LabView

(Comunicação)

Iola M.S. PINTO – *Área Departamental de Matemática, Instituto Superior de Engenharia de Lisboa, Rua Conselheiro Emídio Navarro 1, 1950-062 Lisboa, Portugal*

ipinto@adm.isel.pt

Manfred Niehus – *Área Departamental de Física, Instituto Superior de Engenharia de Lisboa, Rua Conselheiro Emídio Navarro 1, 1950-062 Lisboa, Portugal*

João Dionísio – *Área Departamental de Engenharia de Electrónica e Telecomunicações e de Computadores, Instituto Superior de Engenharia de Lisboa, Rua Conselheiro Emídio Navarro 1, 1950-062 Lisboa, Portugal*

Tiago Mota – *Área Departamental de Engenharia de Electrónica e Telecomunicações e de Computadores, Instituto Superior de Engenharia de Lisboa, Rua Conselheiro Emídio Navarro 1, 1950-062 Lisboa, Portugal*

Resumo

Numa era informatizada as áreas de segurança informática e criptografia têm vindo a assumir papéis cada vez mais importantes na protecção de dados e preservação de privacidade [1]. A segurança dos cripto-sistemas mais avançados, sejam eles clássicos [2] ou quânticos [3], depende da disponibilidade de uma chave de encriptação que só pode ser utilizada uma única vez, *One-Time-Pad*, e cuja sequência terá que ser absolutamente imprevisível. A principal característica a ter em conta é a aleatoriedade que, por ser uma característica probabilística, pode ser avaliada pela realização de testes de aleatoriedade. A Estatística encontra nesta área uma grande variedade de aplicações, nomeadamente a avaliação e classificação do grau de aleatoriedade de sequências binárias, principalmente numa fase em que os tradicionais geradores de números aleatórios baseados em software estão a ser substituídos por geradores baseados em hardware, baseados nas leis da ótica quântica [4].

Neste trabalho apresenta-se uma aplicação informática para a realização de testes de aleatoriedade em sequências binárias, geradas e processadas em tempo real, por hardware. Escolheu-se a plataforma LabVIEW[®] (*Labora-*

tory *Virtual Instrument Engineering Workbench*) que é adequada ao processamento em tempo real e permite a criação de uma interface com experiências em ambiente laboratorial.

Foi implementado um conjunto de testes estatísticos de aleatoriedade, de acordo com a metodologia seguida pelo *NIST Test Suite* [5]. Estes testes permitem averiguar sobre a presença de uma variedade de padrões, que a existirem, evidenciam falhas de aleatoriedade. Permitem quantificar o nível de aleatoriedade das sequências geradas e assim avaliar a qualidade do respectivo gerador.

Para cada teste estatístico é produzido um conjunto de *p-values*, correspondendo ao conjunto de subseqüências testadas. A análise estatística dos resultados consiste por um lado na construção de intervalos de confiança para a proporção de *p-values* que indicam falha de aleatoriedade e por outro lado na realização de testes não paramétricos com o objetivo de averiguar se as amostras dos *p-values*, para cada teste, seguem a distribuição uniforme no intervalo $[0, 1]$, como é esperado para um gerador ideal, [6], [7].

No trabalho desenvolvido destaca-se o papel fundamental da estatística e o facto do processamento das sequências binárias e da respetiva quantificação do seu grau de aleatoriedade serem realizados em tempo real, abrindo caminho ao equilíbrio desejado entre o rigor na análise dos resultados e a rapidez de aquisição experimental.

Referências

- [1] Singh, S. (1999) *The Code Book: The Science of Secrecy from Ancient Egypt to Quantum Cryptography*, ed. Fourth Estate London.
- [2] Shannon, C. E. (1949) *Communication theory of secrecy systems*, Bell Syst. Tech. J. 28, p.656 - 715.
- [3] Gisin, N., Ribordy, G., Tittel, W. and Zbinden, H. (2002) *Quantum Cryptography*, Rev. Mod. Phys. 74, p. 145-193.
- [4] Branning, D., Bermudez, M. (2010) *Testing quantum randomness in single-photon polarization measurements with the NIST test suite*, J. Opt. Soc. of Am. 27.
- [5] Rukhin, A., [et al](2011) *A Statistical Test Suite for Random and Pseudo-random Number Generators for Cryptographic Applications*, NIST (National Institute of Standards and Technology).

- [6] Montgomery, C.D., Runger, C.G. (2003) Applied Statistics and Probability for Engineers, third edition, John Wiley and Sons.
- [7] Siegel, S., Castellan, J.N. (1988) Non parametric Statistics for the Behavioral Sciences, second edition, McGraw-Hill International Editions.

Estimação dos parâmetros de um modelo estocástico com mudança de regime

(Comunicação)

Ana Filipa PRIOR – *Instituto Superior de Engenharia de Lisboa, CMUP*
 anafpontos@gmail.com

Paula Milheiro de Oliveira – *Faculdade de Engenharia da Universidade do Porto, CMUP*

Resumo

Neste trabalho consideramos um modelo estocástico com mudança de regime, em tempo contínuo. Assumimos que, num dado instante conhecido, o modelo passa de um regime para outro. Esta mudança ocorre na matriz de deriva que caracteriza o modelo conforme se especifica a seguir. O modelo é representado pela equação diferencial estocástica

$$dX_t = AX_t dt + B^{\frac{1}{2}} dW_t, \quad (7)$$

onde $\{W_t, t \geq 0\}$ é um processo de Wiener bidimensional e A e $B^{\frac{1}{2}}$ são matrizes quadradas 2×2 . A matriz de deriva A e a matriz de difusão B são as matrizes

$$A = \begin{bmatrix} 0 & 1 \\ -\frac{k_t}{m} & -\frac{c}{m} \end{bmatrix} \text{ e } B = \begin{bmatrix} 0 & 0 \\ 0 & \sigma^2 \end{bmatrix}, \quad (8)$$

onde k_t é o coeficiente de rigidez e $c > 0$ é o coeficiente de amortecimento do modelo. A mudança de regime ocorre quando, num dado instante conhecido, o coeficiente de rigidez k_t passa de um determinado valor k_1 para k_2 , sendo $k_1, k_2 > 0$. A EDE (1)-(2) modela o movimento vibratório de estruturas sujeitas a acções aleatórias com mudança de regime. Em particular, descreve o cenário de uma estrutura que sofreu uma diminuição de rigidez após um determinado período de tempo. Considerando as observações em tempo discreto, descrevemos como obter estimativas de máxima verosimilhança dos parâmetros k_1 , k_2 e c do modelo estocástico antes e depois da mudança de regime. O parâmetro σ é suposto conhecido. Apresentamos um estudo de simulação para ilustrar o procedimento e os resultados obtidos. O software *Yuima* é usado para simular o modelo.

Referências

- [1] Lin, N., Lototsky, S. (2011) *Undamped harmonic oscillator driven by additive Gaussian white noise: A statistical analysis.*, Communications on Stochastic Analysis, Vol.5, No.1, pp. 233-250.
- [2] Rao, B., Basawa, I. (1980) *Statistical Inference for Stochastic Processes.* Academic Press, London.

Modelação flexível do problema de triagem

(Comunicação)

Sandra RAMOS – *ISEP/CEAUL*

sfr@isep.ipp.pt

Antónia Amaral Turkman – *FCUL/CEAUL*

Marília Antunes – *FCUL/CEAUL*

Resumo

O procedimento de triagem envolve a construção de uma região de especificação $C_{\mathbf{X}}$, no espaço d -dimensional, de modo a que um indivíduo futuro com um vector de características em $C_{\mathbf{X}}$ tenha maior probabilidade de ser identificado como um *sucesso* (a resposta Y pertence a uma região conhecida C_Y). Na abordagem preditiva bayesiana a obtenção da região $C_{\mathbf{X}}$ é baseada num critério óptimo assente na maximização de $P(Y \in C_Y | \mathbf{X} \in C_{\mathbf{X}}; D)$, restringida à classe das regiões $C_{\mathbf{X}}$ com probabilidade preditiva de triagem $\alpha = P(\mathbf{X} \in C_{\mathbf{X}} | D)$ fixa. Em literatura já publicada, a construção da região $C_{\mathbf{X}}$ baseia-se habitualmente em modelos paramétricos para (Y, \mathbf{X}) e ou modelos mais flexíveis semiparamétricos.

Neste trabalho, propõe-se uma abordagem não paramétrica que considera um modelo bayesiano, baseado em mistura por um processo de Dirichlet dependente para $Y | \mathbf{X}$ e em mistura por um processo de Dirichlet para \mathbf{X} . A metodologia é ilustrada através de simulações e de conjuntos de dados reais descritos na literatura.

Agradecimentos

Este trabalho foi parcialmente financiado pela Fundação para a Ciência e Tecnologia - FCT: Projecto *PTDC/MAT/118335/2010* e *Pest – OE/MAT/UI0006/2011*.

Modelação de dados espaço-temporais em segurança rodoviária

(Comunicação)

Conceição RIBEIRO – *ISE-UALG/CEAUL*

cribeiro@ualg.pt

Maria Antónia Amaral Turkman – *FCUL/CEAUL*

João Lourenço Cardoso – *LNEC*

Resumo

Os acidentes rodoviários são um grave problema de saúde pública mundial com consequências socio-económicas devastadoras. Apesar dos enormes esforços efetuados nos últimos anos, os valores de sinistralidade rodoviária ainda são inadmissíveis. As estatísticas registadas suscitaram o interesse de investigadores e de políticos, como peça fundamental para no sentido de compreender melhor a complexidade de fatores relacionados com os acidentes rodoviários e determinar que aspetos do problema de segurança rodoviária são locais (nacional ou regional) e quais são os partilhados por vários países. Assim, tem havido uma considerável investigação direcionada para o desenvolvimento de modelação estatística de dados de acidentes rodoviários ocorridos em áreas geográficas correspondentes a órgãos de administração pública ou gestão de infraestruturas de transporte, com o objectivo de encontrar modelos que possam servir como referência na implementação de ações corretivas destinadas a diminuir o número de acidentes rodoviários.

A estimação do número esperado de acidentes continua a ser um problema em aberto, não existindo um método que possa ser considerado ótimo. No entanto, é generalizada a referência à superioridade dos métodos bayesianos. Este trabalho foca-se na aplicação de modelos hierárquicos bayesianos a dados espaço-temporais em segurança rodoviária.

No que diz respeito à metodologia de inferência foi utilizada uma abordagem recente, proposta em [4], que consiste na inferência bayesiana aproximada via INLA, (*Integrated Nested Laplace Approximations*), e que é uma alternativa determinística aos métodos de simulação MCMC habitualmente usados. As-

sim, explorou-se no R , através do R-INLA, a implementação desta metodologia que torna o uso dos modelos bayesianos hierárquicos mais acessível, sendo uma alternativa muito competitiva relativamente ao WinBUGS para situações em que a distribuição da variável dependente pertence à família exponencial. Com a implementação desses modelos foi possível reduzir a variabilidade das estimativas dos riscos de acidentes em áreas pequenas e, simultaneamente, revelar padrões de dependência espacial, tendências temporais e interações espaço-temporais. Conseguiu-se assim uma suavização espacial dos dados, ainda que as regiões sob estudo possuam áreas pequenas. Para além disso, foi possível fazer a identificação dos concelhos de Portugal Continental de maior risco relativamente ao país e a análise da associação concelhia com potenciais fatores de risco ou de exposição. Espera-se que esta metodologia possa vir a permitir efetuar uma comparação no que respeita à segurança rodoviária, não só ao nível nacional ou regional, mas também ao nível internacional.

Referências

- [1] Agüero-Valverde, J., Jovanis, P. (2006) *Spatial analysis of fatal and injury crashes in Pennsylvania.*, Accident Analysis & Prevention 38(3), 618 – 625.
- [2] Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C. Ghislandi, M. and Songini, M. (2006) *Bayesian analysis of space-time variation in disease risk.*, Statist. Med. 14(21-22), 2433–2443.
- [3] Besag, J., York, J. and Mollié, A. (1991) *Bayesian image restoration, with two applications in spatial statistics.*, Annals of the Institute of Statistical Mathematics 43, 1–20.
- [4] Miaou, S.-P., Song, J. and Mallick, B. (2003) *Roadway traffic crash mapping: A space-time modeling approach.*, Journal of Transport Stat. 6, 33–57.
- [4] Rue, H. and Martino, S. (2009) *Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations.*, Journal of the Royal Statistical Society B 71 (2), 319–392.

Uma abordagem não paramétrica à previsão da dose individualizada de *atracurium*.

(Comunicação)

Conceição ROCHA – *Faculdade de Ciências da Universidade do Porto & CIDMA*

mnrocha@fc.up.pt

Maria Eduarda Silva – *Faculdade de Economia da Universidade do Porto & CIDMA*

Teresa Mendonça – *Faculdade de Ciências da Universidade do Porto & CIDMA*

Resumo

A definição da dose de fármaco, de acordo com a intensidade do estímulo cirúrgico e a resposta do paciente, é justificada pela variabilidade farmacocinética e farmacodinâmica entre pacientes. Uma dose que é excessiva para um paciente talvez não seja suficiente para um outro. Por esta razão, e sem surpresa, nenhum regime único de concentração de fármaco ou de combinação de fármacos se aplica a todos os pacientes.

O objetivo deste trabalho é obter uma previsão robusta em tempo real da dose em estado estacionário ou dose de manutenção individualizada (u_{ss}) que conduza e fixe o nível do bloqueio neuromuscular, NMB, no valor desejado. Para atingir este fim, é considerada a seguinte abordagem: primeiro encontrar um conjunto de atributos que caracterizem o comportamento do indivíduo e depois usar esses atributos para conduzir a análise preditiva individualizada com base no princípio de aprendizagem, a partir de dados por meio de técnicas estatísticas.

Este estudo de caso centra-se no relaxante muscular *atracurium* e na sua influência no NMB. O sistema de bloqueio neuromuscular induzido pelo *atracurium* é um sistema dinâmico. Sistemas dinâmicos aparecem frequentemente em biomedicina e, algumas vezes, não são identificáveis em tempo real ou são mal identificados, como neste caso. No presente contexto, a análise espectral de Walsh-Fourier — que é uma ferramenta poderosa para a análise exploratória de séries temporais — e a regressão robusta são conjugadas com sucesso para levarem a um conjunto de preditores informativo que contribuem diretamente

para prever a dose individual de *atracurium* a administrar. Tendo caracterizado o individuo, duas famílias conhecidas de métodos de classificação supervisionada são consideradas: aprendizagem baseada em casos, *Instance Based*, e árvores de decisão, *Decision Trees*. Todos os classificadores levam a previsões que apresentam erros relativos inferiores a 10% em cerca de 99% dos casos. Este mesmo estudo pode facilmente estender-se a outros protocolos de administração do fármaco, a outros fármacos anestésicos ou mesmo a outros sistemas fisiológicos.

Sobrevivência relativa do cancro colo-rectal e do estômago no Sul de Portugal

(Comunicação)

Ricardo SÃO JOÃO – *ESTGS-IPS & CEAUL*

ricardo.sjoao@esg.ipsantarem.pt

Ana Papoila – *FCM-UNL & CEAUL*

Ana Miranda – *ROR-Sul*

Resumo

A sobrevivência é um conceito de importância ímpar em vários domínios e, em particular, na oncologia. Neste contexto, a sobrevivência vem expressa por uma taxa (*survival rate*) que indica a percentagem de indivíduos vivos durante um dado período de tempo após lhes ter sido diagnosticado cancro. Todavia, a morte de um doente com cancro poderá ficar a dever-se ao próprio cancro ou a outras causas. Se a causa específica de morte não for conhecida, a taxa de sobrevivência deverá ser designada por taxa de sobrevivência global (*overall survival rate*), muito utilizada no prognóstico das neoplasias.

Uma questão pertinente e habitual no contexto clínico, reside em saber qual a contribuição da neoplasia em si para a sobrevivência do doente. Como a causa específica de morte é frequentemente desconhecida ou imprecisa, remete-se a questão anterior para a sobrevivência relativa (*relative survival*).

A principal vantagem da sobrevivência relativa é não depender da causa específica de morte, razão essa que motivou a sua utilização no presente estudo. A sobrevivência relativa pode ser considerada uma medida que expressa o excesso de mortalidade/risco num grupo de indivíduos com uma determinada doença. O seu cálculo baseia-se na taxa de risco global λ (*overall hazard rate*) e na taxa de risco esperada λ^* (*expected hazard rate*), sendo esta calculada com base nas tábuas de mortalidade (*life tables*) de uma população com características análogas (p.e. idade, sexo, região, ...) ao grupo em análise.

A equação (9) permite determinar a taxa de mortalidade por causa específica de cancro ν (*excess hazard*) através da relação:

$$\lambda = \lambda^* + \nu \tag{9}$$

Como a causa de morte é desconhecida, não é possível determinar a proporção

de óbitos que se devem à doença em análise, justificando-se *per si* a adopção da função de sobrevivência relativa $r(t)$ (*cumulative relative survival function*), definida por:

$$r(t) = \frac{S_o(t)}{S_p(t)} \quad (10)$$

onde $S_o(t)$ representa a sobrevivência observada no grupo em análise e $S_p(t)$ representa a sobrevivência esperada determinada com base nas tábuas de mortalidade. Da equação anterior, resulta que $r(t)$ só admite valores positivos com particular interesse para valores inferiores à unidade, na situação em que as neoplasias apresentam uma taxa de sobrevivência inferior à do grupo populacional de referência, livre de doença. Os modelos de sobrevivência relativa são modelos de regressão e, neste estudo, foram utilizadas três abordagens: modelos aditivos, multiplicativos e transformados.

Com base no registo de cancro de base populacional da Região Sul de Portugal (ROR Sul) o presente estudo tem como objectivo avaliar a sobrevivência relativa de 11859 doentes diagnosticados com cancro do estômago e 21575 doentes diagnosticados com cancro colo-rectal num período de *follow-up* de cinco anos. Os resultados evidenciaram diferenças significativas na sobrevivência relativa quanto ao tipo de neoplasia, grupo etário, género, tempo de *follow-up* e região geográfica inserida no ROR-Sul.

Agradecimentos

Este estudo foi financiado por Fundos Nacionais através da FCT-Fundação para a Ciência e Tecnologia, no âmbito do projecto PEst-E/MAT/UI0006/2011.

Referências

- [1] Dickman,P.W., Sloggett, A., Hills, M., Hakulinen,T. (2003). Regression models for relative survival. *Statistics in Medicine* 23, 51–64.
- [2] Holleczeck,B., Brenner, H. (2013). Model based period analysis of absolute and relative survival with R: Data preparation, model fitting and derivation of survival estimates. *Computer methods and programs in biomedicine* 110, 192–202.

Altura do Geopotencial através da Análise em Componentes Independentes para Séries Temporais Estendidas

(Comunicação)

Fernando SEBASTIÃO – *Departamento de Matemática, Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Leiria e CM-UTAD*

fsebast@ipleiria.pt

Irene Oliveira – *Departamento de Matemática, Universidade de Trás-os-Montes e Alto Douro e CM-UTAD*

Resumo

A Análise Espectral Singular Multicanal (MSSA) (Golyandina *et al.*, 2001) é uma técnica que incide na análise de dados multivariados estruturados sob a forma de séries temporais. Esta técnica corresponde a uma Análise em Componentes Principais (ACP) (Jolliffe, 2002) da matriz estendida (ou dos desfasamentos) das séries temporais iniciais, também conhecida na área do clima por Análise de Funções Ortogonais Empíricas Estendidas (EEOF) (von Storch e Zwiers, 1999; Hannachi *et al.*, 2007).

A técnica da Análise em Componentes Independentes (ACI) (Hyvärinen *et al.*, 2001) é introduzida para explorar a matriz estendida das séries temporais como uma alternativa ao método clássico MSSA. A ACI é frequentemente usada em áreas como o processamento de sinais e de imagens, em indicadores biomédicos, em telecomunicações, em aplicações financeiras entre outras. No entanto, na área da climatologia surgem ainda poucas referências à sua aplicação tais como Aires *et al.* (2000), Basak *et al.* (2004), Hannachi *et al.* (2009) ou Sebastião e Oliveira (2013).

Apresenta-se uma aplicação da ACI estendida a um conjunto de dados meteorológicos, constituído por observações ao longo de 50 anos, de janeiro de 1958 a dezembro de 2007, correspondentes aos valores médios mensais da altura do geopotencial a 500 mbar, registados no setor do Hemisfério Norte delimitado por (20žN - 80žN) e (0žE - 357.5žE). Os dados distribuem-se segundo uma grelha espacial uniforme de 2.5ž ao longo da latitude e da longitude,

perfazendo um total de 25 por 144 pontos.

O objetivo deste trabalho consiste em avaliar a potencialidade da ACI estendida na interpretação de fenômenos climatológicos com características de propagação temporal e espacial, através da elaboração de mapas de clima e da obtenção de resultados de comparação entre esta e as outras técnicas mencionadas.

Referências

- [1] Aires, F., Chédin, A. e Nadal, J-P. (2000) *Independent component analysis of multivariate time series: Application to the tropical SST variability*, Journal of Geophysical Research. 105(D13), 17,437–17,455.
- [2] Basak, J., Sudarshan, A., Trivedi, D. e Santhanam M. S. (2004) *Weather data mining using independent component analysis*, Journal of Machine Learning Research. 5, 239–253.
- [3] Golyandina, N. E., Nekruktin, V. V. e Zhigljavsky, A. A. (2001) *Analysis of Time Series Structure: SSA and Related Techniques*. Chapman and Hall, Boca Raton.
- [4] Hannachi, A., Jolliffe, I. T. e Stephenson, D. B. (2007) *Empirical orthogonal functions and related techniques in atmospheric science: a review*, International Journal of Climatology. 27, 1119–1152.
- [5] Hannachi, A., Unkel, S., Trendafilov, N. T. e Jolliffe, I. T. (2009) *Independent component analysis of climate data: a new look at EOF rotation*, Journal of Climate. 22, 2797–2812.
- [6] Hyvärinen, A., Karhunen, J. e Oja, E. (2001) *Independent Component Analysis*. John Wiley and Sons, Inc., U.S.A..
- [7] Jolliffe, I. T. (2002) *Principal Component Analysis (2nd Edition)*. Springer-Verlag, New York.
- [8] Sebastião, F. e Oliveira, I. (2013) *Independent component analysis for extended time series in climate data*, Advances in Regression, Survival Analysis, Extreme Values, Markov Processes and Other Statistical Applications. Studies in Theoretical and Applied Statistics. 427–436.
- [9] von Storch, H. e Zwiers, F. W. (1999) *Statistical Analysis in Climate Research*. Springer-Verlag, New York.

Modelos GARCH de valores inteiros associados a leis infinitamente divisíveis

(Comunicação)

Filipa SILVA – *CMUC, Departamento de Matemática, Universidade de Coimbra*

mat0504@mat.uc.pt

Nazaré Mendes-Lopes – *CMUC, Departamento de Matemática, Universidade de Coimbra*

Esmeralda Gonçalves – *CMUC, Departamento de Matemática, Universidade de Coimbra*

Resumo

A necessidade de oferecer respostas adequadas à modelação estocástica de séries temporais de valores inteiros não negativos, tem sido realçada por vários autores tanto mais que tal tipo de séries surge com frequência e de forma natural em diversos contextos e áreas científicas tais como a medicina, a economia, o turismo ou a meteorologia.

Neste sentido, têm surgido na literatura vários modelos para séries temporais de valores inteiros não negativos, entre os quais se destacam os modelos INGARCH (INteger-valued GARCH), propostos em 2006 por Ferland, Latour e Oraichi [2], claramente inspirados nos GARCH clássicos de Bollerslev [1] mas com distribuição condicional de Poisson.

Com o objetivo de alargar este tipo de processos, particularmente no que diz respeito à família de distribuições condicionais, vamos no presente trabalho introduzir e analisar uma nova classe de modelos de valores inteiros com evolução análoga para a média condicional mas em que a família de leis condicionais associada é a das distribuições infinitamente divisíveis com suporte em \mathbb{N}_0 .

Para além do referido modelo de Poisson, esta classe geral inclui outras famílias já estudadas na literatura, de que salientamos os modelos GP-INGARCH em que a lei condicional é de Poisson Generalizada [3]. Novos casos particulares serão apresentados neste trabalho. Além disso, será desenvolvida a análise probabilista daquela classe geral de processos no que diz respeito à estacionaridade e ergodicidade, uniformizando e alargando significativamente os resultados do mesmo tipo estabelecidos em [2] e [3].

Referências

- [1] Bollerslev, T., Engle, R.F., Nelson, D.B. (1994) ARCH models, in Engle and McFadden, ed.: Handbook of Econometrics, vol.4, Elsevier Science, Amsterdam.
- [2] Ferland, R., Latour, A. e Oraichi, D. (2006) *Integer-valued GARCH process*, Journal of Time Series Analysis 27, 923–942.
- [3] Zhu, Fk. (2012) *Modelling overdispersed or underdispersed count data with generalized Poisson integer-valued GARCH models*, Journal of Mathematical Analysis and Applications, 389 1, 58–71.

Como modelar os efeitos espaço-temporais na proporção de área ardida em fogos florestais?

(Comunicação)

Giovani SILVA – *CEAUL & Instituto Superior Técnico*

gsilva@math.ist.utl.pt

Paulo Soares – *CEAUL & Instituto Superior Técnico*

Resumo

O impacto dos fogos florestais pode ser estudado através da modelação da proporção de área ardida dos fogos tendo em conta efeitos espaço-temporais possivelmente não lineares. Esta modelação por vezes faz-se através de uma transformação da proporção na reta real com vista ao uso de modelos de regressão gaussiana para a variável resposta (Marques et al., 2009). Todavia, outras abordagens podem ser mais vantajosas, e.g., os modelos de regressão beta, onde os parâmetros de regressão podem ser interpretados em termos da média da variável resposta ou de uma razão de chances, quando a função de ligação *logit* é considerada (Ferrari e Cribari-Neto, 2004).

Os objetivos deste trabalho são i) comparar alternativas na modelação da proporção de área ardida com base em dados de incêndios florestais em Portugal continental, ii) investigar efeitos espaço-temporais possivelmente não lineares, iii) avaliar e interpretar modelos relevantes na caracterização dos fogos florestais, e.g., estimar quantidades de interesse relacionadas com a variação espacial não observadas entre as regiões em estudo usando modelos auto-regressivos condicionais (Besag et al., 1991).

Neste trabalho, os modelos espaço-temporais para proporção aqui analisados enquadram-se também na classe de modelos hierárquicos bayesianos (Amaral-Turkman et al., 2011). Além disso, perante um maior número de regiões e de anos observados torna-se necessário adotar uma mistura de distribuições para modelar conjuntamente a proporção de área ardida e o significativo número de casos em que não ocorrem quaisquer incêndios. Por fim, as inferências sobre os parâmetros dos modelos espaço-temporais para proporção foram obtidas a partir de métodos de Monte Carlo via cadeia de Markov (MCMC).

Agradecimentos

Este trabalho foi parcialmente financiado pelo projeto FCT Pest-OE/MAT/UI0006/2011.

Referências

- [1] Amaral-Turkman, M.A., Turkman, K.F., Le Page, Y., Pereira J.M. (2011) *Hierarchical space-time models for fire ignition and percentage of land burned by wildfires*, Environmental and Ecological Statistics, 18, 601–617.
- [2] Besag, J., York, J.C., Mollié, A. (1991) *Bayesian image restoration, with two applications in spatial statistics (with discussion)*, Annals of the Institute of Statistical Mathematics, 43, 1–59.
- [3] Ferrari, S.L.P., Cribari-Neto, F. (2004) *Beta regression for modeling rates and proportions*, Journal of Applied Statistics, 31, 799–815.
- [4] Marques, S., Borges, J.G., Garcia-Gonzalo, J., Moreira, F., Carreiras, J.M.B., Oliveira, M.M., Cantarinha, A., Botequim, B., Pereira, J.M.C. (2009) *Characterization of wildfires in Portugal*, European Journal of Forest Research, 130, 775–784.

Estimação da taxa de prevalência na presença de problemas de má classificação utilizando testes compostos

(Comunicação)

Ricardo SOUSA – *Escola Superior de Tecnologia da Saúde de Lisboa do Instituto Politécnico de Lisboa, CEAUL — Centro de Estatística e Aplicações da Universidade de Lisboa*

ricardo.sousa@estesl.ipl.pt

Rui Santos – *Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Leiria, CEAUL — Centro de Estatística e Aplicações da Universidade de Lisboa*

João Paulo Martins – *Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Leiria, CEAUL — Centro de Estatística e Aplicações da Universidade de Lisboa*

Resumo

A utilização de testes compostos surge na Estatística com o trabalho de Dorfman (1943), com o objetivo de determinar a dimensão ótima para cada grupo em função da taxa de prevalência, no sentido de minimizar o número esperado de testes necessários para a identificação de todos os indivíduos infetados. Todavia, o recurso a análises conjuntas não visa unicamente a classificação de indivíduos, podendo ser igualmente útil na estimação da taxa de prevalência. Para este fim não é necessária a posterior realização de testes individuais (uma vez que não se pretende identificar os indivíduos infetados, mas somente estimar a taxa de prevalência da doença, tendo ainda a vantagem de manter o anonimato dos indivíduos infetados). Deste modo, os estimadores obtidos pela aplicação de testes compostos têm, sob determinadas condições, melhor comportamento que os estimadores tradicionais baseados em testes individuais, cf. Sobel e Elashoff (1975), Loyer (1983) e Garner *et al.* (1989). O enviesamento, a eficiência e a robustez destes estimadores foram examinados em diversos artigos, tais como os de Chen e Swallow (1990), Lancaster e Keller-McNulty (1998) ou Hung e Swallow (1999). Assim sendo, os estimadores baseados em

testes conjuntos (para taxas de prevalência baixas) permitem não só a obtenção de ganhos monetários (diminuindo o número de testes efetuados) mas igualmente a obtenção de estimativas mais precisas em comparação com as obtidas com base em testes individuais.

Porém, na maioria das aplicações dos testes compostos é suposto que os testes classificam sempre corretamente os indivíduos e, quando não o fazem, consideram que as probabilidades associadas aos erros de classificação são iguais nos testes individuais e nos testes conjuntos (uma simplificação pouco realista uma vez que a utilização de testes conjuntos implica o aumento da probabilidade de ocorrência de má classificação, cf. Santos, Pestana e Martins, 2013).

Deste modo, o principal objetivo deste trabalho é comparar, via simulação, a performance de diferentes metodologias para estimar a taxa de prevalência, utilizando testes individuais e compostos, em situações onde podem ocorrer erros de classificação.

Trabalho financiado por Fundos Nacionais através da Fundação para a Ciência e a Tecnologia, no âmbito do projeto PEst-OE/MAT/UI0006/2011.

Referências

- [1] Chen, C.L., Swallow, W.H. (1990). *Using group testing to estimate a proportion, and to test the binomial model*, Biometrics 46, 1035-1046.
- [2] Dorfman, R. (1943). *The detection of defective members of large populations*, Ann. Math. Statist. 14, 436-440.
- [3] Garner, F.C., Stapanian, M.A., Yfantis, E.A., Williams, L.R. (1989). *Probability estimation with sample compositing techniques*, Journal of Official Statistics 5, 365-374.
- [4] Hung, M., Swallow, W.H. (1999). *Robustness of group testing in the estimation of proportions*, Biometrics 55, 231-237.
- [5] Lancaster, V.A., Keller-McNulty, S. (1998). *A review of composite sampling methods*, Journal of the American Statistical Association 93, 1216-1230.
- [6] Loyer, M.W. (1983). *Bad probability, good statistics, and group testing for binomial estimation*, The American Statistician 37, 57-59.

- [7] Santos, R., Pestana, D. and Martins, J.P. (2013). *Extensions of Dorfman's Theory*, P.E. Oliveira et al. eds., *Recent Developments in Modeling and Applications in Statistics, Studies in Theoretical and Applied Statistics, Selected Papers of the Statistical Societies*, Springer, 179–189.
- [8] Sobel, K.M., Elashoff, R.M. (1975). *Group testing with a new goal, estimation*, *Biometrika* 62, 181-193.

Identificação de variáveis ambientais relevantes para o Site Index do *Eucalyptus Globulus*

(Comunicação)

Manuela SOUTO DE MIRANDA – CIDMA e DMat, Universidade de Aveiro
manuela.souto@ua.pt

Margarida Mendes Silva – RAIZ, grupo Portucel Soporcel

Conceição Amado – CEMAT e IST, Universidade de Lisboa

Resumo

Para avaliar a adequação de um local em relação a um investimento florestal, usam-se variáveis biométricas das árvores e povoamentos, ou índices calculados a partir de observações já existentes no local. O *Site Index* é o indicador mais utilizado em gestão florestal para quantificar a capacidade produtiva de um povoamento; este indicador, descrito em Skovsgaard and Vanclay (2008), baseia-se na altura das árvores dominantes do povoamento a uma idade de referência, que no caso do *Eucalyptus Globulus* é de 10 anos. Quando não existem observações num local pretendido, é importante saber como usar as variáveis ambientais que influenciam as variáveis biométricas determinantes para o cálculo da produtividade.

Neste trabalho apresenta-se um estudo do *Site Index* efetuado para a espécie *Eucalyptus Globulus*, em povoamentos comerciais puros, no território nacional. As variáveis ambientais consideradas foram selecionadas tendo em conta a sua contribuição para a disponibilidade de água às plantas e a facilidade de acesso e recolha de observações, contemplando variáveis climáticas, variáveis relacionadas com o solo e variáveis topográficas.

A metodologia seguida incluiu análise multivariada convencional e robusta. Após uma Análise em Componentes Principais identificaram-se dois fatores limitantes do crescimento das árvores, relacionados com o clima e com o solo. A abordagem robusta das Componentes Principais mostrou-se superior à convencional ao evidenciar a diferente contribuição das variáveis ambientais e melhorou a deteção de observações atípicas. A análise em Componentes Principais ainda apontou para a potencial existência de dois grupos de observações. Essa possibilidade foi investigada através de uma Análise de Agrupamentos,

que na versão robusta se baseou no método de agrupamento aparado proposto em García-Escudero *et al.* (2008). Tanto o método convencional como o método robusto conduziram a idênticas soluções, resultando em dois grupos de povoamentos. Os grupos foram posteriormente justificados pela localização geográfica, correspondendo a duas situações climáticas distintas no território nacional.

Referências

- [1] García-Escudero, L., Gordaliza, A., Matrán, C. and Mayo-Isar, A. (2008) *A general trimming approach to robust cluster analysis*, *Annals of Statistics* 36, 1324-1345.
- [2] Skovsgaard, J.P. and Vanclay, J.K. (2008) *Forest site productivity: a review of the evolution of dendrometric concepts for even-aged stands*, *Forestry* 81, 13-21.

Cálculo do tamanho da amostra para estimar a sensibilidade e a especificidade

(Comunicação)

Ana SUBTIL – *CEMAT e Faculdade de Ciências da Universidade de Lisboa, Universidade de Lisboa*

asubtil@ihmt.unl.pt

M. Rosário de Oliveira – *CEMAT e Departamento de Matemática, Instituto Superior Técnico, Universidade de Lisboa*

Luzia Gonçalves – *CEAUL e Unidade de Saúde Pública Internacional e Bioestatística, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa*

Resumo

O cálculo do tamanho da amostra para estimar uma proporção é um tópico muito relevante no contexto biomédico. Uma prática corrente, é recorrer a condições sobre intervalos de confiança (IC) para uma proporção para determinar o tamanho da amostra. Tradicionalmente, escolhe-se o método de Wald para construir este IC, pela sua simplicidade e facilidade de cálculo. No entanto, a vasta investigação em torno deste tópico tem vindo a destacar as limitações do método de Wald e a discutir métodos alternativos, fornecendo recomendações quanto à sua adequação a diferentes situações práticas [1,2]. Paralelamente, o cálculo do tamanho da amostra baseado em IC, tem igualmente sido alvo de renovada atenção pelos investigadores [3]. Os trabalhos desenvolvidos baseiam-se numa amostra de dimensão fixa, onde se conta o número de entidades que têm a característica de interesse (cuja probabilidade de ocorrência é a proporção em estudo). Na prática, é frequente seleccionar um conjunto de n indivíduos aos quais se aplicam um (ou mais) teste(s) de diagnóstico, com o objectivo de estimar a(s) sensibilidade(s) (Se, probabilidade de o teste identificar correctamente os indivíduos com a infecção ou doença) e especificidade(s) (Sp, probabilidade de o teste identificar correctamente os indivíduos que não sofrem dessa infecção ou doença) desse(s) teste(s). Neste contexto, as estimativas convencionais destas quantidades baseiam-se no conhecimento do número de indivíduos com a infecção ou doença, que por sua

vez é uma variável aleatória e não um valor fixo, como é considerado nos diversos estudos para proporções binomiais. Desta forma, surge a necessidade de redefinir os critérios usados na avaliação do desempenho de IC para probabilidades condicionais, onde Se e Sp são um caso particular, já que os estudos anteriores não são necessariamente transponíveis para esta situação.

No presente trabalho, usamos as definições modificadas da probabilidade de cobertura e do valor esperado do comprimento do IC aleatório para determinar o tamanho da amostra adequado à estimação da Se e Sp por um método exacto, contudo mais dispendioso computacionalmente, e por um método aproximado. Ao calcular o tamanho da amostra, o investigador tem de optar por estabelecer condições referentes à Se ou Sp de um dos testes de diagnóstico em estudo. Neste trabalho, discute-se o impacto que adoptar uma dimensão amostral respeitando critérios relativamente a uma das medidas poderá ter no desempenho dos IC das restantes Se e Sp .

Referências

- [1] Brown, L.D., Cai, T.T., Dasgupta, A. (2002) *Confidence intervals for a Binomial proportion and asymptotic expansions*, The Annals of Statistics 30, 160–201.
- [2] Pires A.M., Amado C. (2008) *Interval estimates for a Binomial proportion: comparison of twenty methods*, REVSTAT 6, 165–197.
- [3] Gonçalves, L., Oliveira, M.R., Pascoal, C., Pires, A.M. (2012) *Sample size for estimating a binomial proportion: comparison of different methods*, Journal of Applied Statistics 39, 2453–2473.

Cartas de controlo para a média e desvio padrão com amostragem dupla e instantes de amostragem predefinidos

(Comunicação)

Renata TEIXEIRA – *Doutoramento em Matemática, CIMA-UE, Universidade de Évora*

renatateixeira@netcabo.pt

Paulo Infante – *CIMA-UE e DMAT, ECT da Universidade de Évora*

Fernanda Figueiredo – *CEAUL e Faculdade de Economia da Universidade do Porto*

Resumo

Diversos estudos mostram que podemos melhorar o desempenho estatístico das cartas de controlo variando a dimensão amostral, os instantes de recolha das amostras e os limites de controlo ao longo do monitorização do processo. Estas cartas podem ser classificadas, do ponto de vista da sua implementação, em duas categorias: cartas cujos parâmetros são fixos, mas não necessariamente constantes, sendo definidos no início do controlo do processo; e cartas em que pelo menos um dos parâmetros varia em tempo real em função da estatística amostral. As primeiras são conhecidas por cartas com parâmetros predefinidos e as últimas por cartas de controlo adaptativas ou dinâmicas. Vários têm sido os métodos apresentados e esta temática continua a merecer grande interesse por parte dos investigadores podendo, por exemplo, citar-se [4] e [5] relativamente a controlo por variáveis e [1] relativamente a controlo por atributos.

Neste trabalho, recuperamos o método DSPSI (*Double Sampling with Predetermined Sampling Intervals*), apresentado inicialmente para uma carta de controlo para a média em [3], que combina o procedimento PSI (*Predetermined Sampling Intervals*) estudado e analisado em [6] e o método de amostragem dupla (DS - *Double Sampling*) apresentado em [2]: no início da monitorização do processo os instantes de amostragem são calendarizados de acordo

com o tempo de vida do sistema, de modo que a taxa cumulativa de risco entre dois quaisquer instantes de amostragem consecutivos seja constante, e a dimensão das amostras a analisar, em cada instante predefinido, é definida de acordo com o método de amostragem dupla: em cada instante de amostragem recolhem-se duas amostras de tamanho n_1 e n_2 ; se a inspeção da primeira amostra de tamanho n_1 não permitir concluir sob o estado do processo, inspeccionam-se as restantes n_2 unidades e considera-se a informação recolhida do conjunto das duas amostras para decidir sobre o estado do processo.

Neste sentido, aplicamos este procedimento ao uso simultâneo de cartas para a média e para o desvio padrão, considerando as distribuições de Weibull e Burr XII para o tempo de vida do sistema, de modo a analisar o desempenho do método para diferentes tipos de taxa de risco. Realiza-se, ainda, um estudo comparativo com diversos esquemas adaptativos de amostragem que têm sido apresentados na literatura.

Na análise do desempenho estatístico, utilizamos as medidas AATS (*Adjusted Average Time to Signal*) e ANI (*Average Number of Inspected Items*) e apresentamos algumas linhas gerais de implementação do método. Para todas as distribuições e esquemas considerados, este método apresenta globalmente melhor desempenho, sendo tanto mais eficaz quanto maior a diversidade de alterações a que o processo está sujeito. Por outro lado, em geral a sua eficácia aumenta com a diminuição do número de amostras analisadas quando o processo está sob controlo e para sistemas com taxas de risco acentuadamente crescentes.

Referências

- [1] Araújo Rodrigues, A., Epprecht, E. K., Magalhães, M. S. (2011) *Double Sampling Control Charts for Attributes*, Journal of Applied Statistics 38, 87–112.
- [2] Daudin, J. J. (1992) *Double Sampling X-Charts*, Journal of Quality Technology 24, 78–87.
- [3] Infante, P., Rosmaninho, E. (2007) *A Combined Double Sampling and Predetermined Sampling Intervals*, Bulletin of the International Statistical Institute, vol. LXII.

- [4] Khoo, M.B., Lee, H.C., Wu, Z. Chen, C.H., Castagliola, P. (2010) *A Syhthetic Double Sampling Control Chart for the Process Mean*, IIE Transactions 43, 23–38.
- [5] Lee, P.H., Chang, C.Y., Tornng, C.C. (2012) *A Design of s Control Charts with a Combined Double Sampling and Variable Sampling Interval Scheme*, Communications in Statistics - Theory and Methods 41, 153–165.
- [6] Rodrigues Dias, J., Infante, P. (2008) *Control Charts with Predetermined Sampling Intervals*, International Journal of Quality & Reliability Management 25, 423–435.

Coeficiente correlacional de concordância e coeficiente de concordância *top-down*: algumas recomendações suscitadas por aplicações na área do Desporto

(Comunicação)

Júlia TELES – *Secção de Métodos Matemáticos e CIPER, Faculdade de Motricidade Humana, Universidade Técnica de Lisboa*

jteles@fmh.utl.pt

Resumo

Quem trabalha na área da Estatística Aplicada ainda se depara com a necessidade de encontrar soluções para questões que lhe são colocadas com alguma rapidez, porque “o trabalho já está pronto, falta apenas a análise estatística”, como diz frequentemente o nosso interlocutor! O coeficiente correlacional de concordância e o coeficiente de concordância *top-down* foram utilizados para dar resposta a algumas dessas questões.

A avaliação da fiabilidade entre juízes, ou da reprodutibilidade de uma experiência, método ou instrumento, é uma temática que desperta o interesse de investigadores de diversas áreas, nomeadamente do Desporto e da Saúde (e.g., Atkinson and Nevill [1]). O coeficiente correlacional de concordância (Lin [5,6]) é um índice utilizado na avaliação da concordância entre pares de observações de natureza contínua. Este coeficiente pode ser escrito como o produto de duas componentes, que quantificam a precisão e a exatidão, a qual, por sua vez, pode ser expressa à custa de deslocamentos de escala e de localização. Deste modo, em situações de falta de acordo, é fácil identificar a sua origem e propor eventuais ações corretivas. Outra vantagem deste coeficiente é a sua intuitiva generalização para mais de dois avaliadores (e.g., Barnhart et al. [2]). No entanto, é preciso ter cuidado na utilização deste índice, pois quando a variabilidade entre indivíduos é substancialmente superior à variabilidade entre avaliações, este coeficiente sobrestima o valor da concordância. Esta sobrestimação ocorre também com outras medidas de concordância, nomeadamente com uma das mais utilizadas, o coeficiente de correlação intraclasse (e.g., McGraw e Wong [7]).

Se o objetivo for a avaliação do acordo entre diversas amostras de *ranks*,

o coeficiente de concordância de Kendall (Kendall e Babington-Smith [4]) é a estatística a que a maioria dos investigadores recorre. No entanto, em situações em que o objectivo é a avaliação do acordo entre as observações de ordem inferior ou de ordem superior, o coeficiente de concordância *top-down* (Iman e Conover [3]), que utiliza os scores de Savage (Savage [8]), revela-se uma opção mais adequada do que o coeficiente de concordância de Kendall (Teles [9]).

Nesta comunicação são apresentados diversas aplicações, na área do Desporto, destes coeficientes de concordância para mais de duas amostras. Estes exemplos e os resultados de estudos de simulação possibilitam a apresentação de algumas recomendações para a utilização destas duas medidas de acordo.

Referências

- [1] Atkinson, G., Nevill, A.M. (1998) *Statistical Methods For Assessing Measurement Error (Reliability) in Variables Relevant to Sports Medicine*, Sports Medicine 26(4), 217–238.
- [2] Barnhart, H.X., Haber, M., Song, J.L. (2002) *Overall concordance correlation coefficient for evaluating agreement among multiple observers*, Biometrics 58, 1020–1027.
- [3] Iman, R.L., Conover, W.J. (1987) *A measure of top-down correlation*, Technometrics 29, 351–357.
- [4] Kendall, M.G., Babington-Smith, B. (1939) *The problem of m rankings*, Annals of Mathematical Statistics 10, 275–287.
- [5] Lin, L.I.-K. (1989) *A concordance correlation coefficient to evaluate reproducibility*, Biometrics 45, 255–268.
- [6] Lin, L.I.-K. (1992) *Assay validation using the concordance correlation coefficient*, Biometrics 48, 599–604.
- [7] McGraw, K.O., Wong, S.P. (1996) *Forming inferences about some intra-class correlation coefficients*, Psychological Methods 1, 30–46.
- [8] Savage, I.R. (1956) *Contributions to the theory of rank order statistics — the two-sample case*, Biometrics 10, 275–287.
- [9] Teles, J. (2012) *Concordance coefficients to measure the agreement among several sets of ranks*, Journal of Applied Statistics 39, 1749–1764.

Atracção do máximo de um modelo INARMA(1,1) na classe de Anderson

(Comunicação)

Maria da Graça TEMIDO – *CMUC, Dep. de Matemática, Fac. de Ciências e Tecnologia, Universidade de Coimbra*

mgtm@mat.uc.pt

Sandra Dias – *CM-UTAD, Dep. de Matemática, Universidade de Trás-os-Montes e Alto Douro*

Resumo

Para funções de distribuição discretas (f.d's) F que, como a Binomial Negativa, têm limite superior do suporte infinito ($w_F = \infty$) e satisfazem

$$\lim_n \frac{1 - F(n-1)}{1 - F(n)} = r > 1, \quad (11)$$

o máximo de n variáveis aleatórias (v.a's) independentes e identicamente distribuídas (i.i.d.) com f.d. F não tem distribuição limite não degenerada, sob qualquer normalização, na classe max-estável. Contudo, Anderson [1] provou que uma f.d. discreta, com $w_F = \infty$, verifica (11) se e só se existe uma sucessão real $\{b_n\}$ tal que

$$\exp(-r^{-x-1}) \leq \liminf_n F^n(x + b_n) \leq \limsup_n F^n(x + b_n) \leq \exp(-r^{-x}), \quad (12)$$

$x \in \mathbb{R}$.

A classe de f.d's satisfazendo (11) é designada classe de Anderson e denotada por $\mathcal{C}_A(r)$.

Por outro lado, Temido [6] provou que para f.d's discretas F com $w_F = \infty$, existe uma sucessão inteira crescente $\{k_n\}$ que satisfaz $\lim_n k_{n+1}/k_n = r \geq 1$, e uma sucessão real $\{b_n\}$ tal que $F^{k_n}(x + b_n) \equiv P(M_{k_n} \leq x + b_n) \rightarrow_n G(x) := \exp(-\eta r^{-[x]})$, $x \in \mathbb{R}$, para algum $\eta > 0$, se e só se F verifica (11), i.e., pertence a $\mathcal{C}_A(r)$. Esta f.d. $G(x) := \exp(-\eta r^{-[x]})$ é designada Gumbel discreta e é uma lei max-semiestável (ver Pancheva [4]).

Em McKenzie [3] é introduzido o processo estacionário INARMA(1,1) definido por

$$X_n = \beta \star Z_n + V_n W_{n-1}, \text{ com } W_n = \alpha \star W_{n-1} + U_n Z_n,$$

onde \star denota o operador aleatório Binomial, $\{Z_n\}$, $\{U_n\}$ e $\{V_n\}$ são sucessões de v.a's i.i.d., $\{U_n\}$ e $\{V_n\}$ têm ambas distribuição de Bernoulli (com parâmetros α e β , respetivamente) e W_0 é independente de todas as outras variáveis. McCormick e Park [3] consideraram a distribuição Binomial Negativa para $\{Z_n\}$ e $\{W_n\}$ e estabeleceram o limite (12). Neste trabalho assumimos que $\{Z_n\}$ e $\{W_n\}$ têm margens em $\mathcal{C}_A(r)$, em particular $1 - F_Z(n) \sim An^\xi(1+\lambda)^{-n}$, $n \rightarrow +\infty$ ($r = 1 + \lambda$), $A, \xi, \lambda > 0$, que satisfaz (11). Para a sucessão $\{X_n\}$ validamos o comportamento de independência assintótica e de dependência local que passamos a recordar.

A sucessão estacionária $\{X_n\}$ satisfaz a condição de independência assintótica $D_{k_n}(u_n)$ se, para quaisquer inteiros $1 \leq i_1 < \dots < i_p < j_1 < \dots < j_q \leq k_n$ tais que $j_1 - i_p > \ell_n$, se tem

$$\begin{aligned} & \left| P\left(\bigcap_{s=1}^p \{X_{i_s} \leq u_n\}, \bigcap_{m=1}^q \{X_{j_m} \leq u_n\}\right) - \right. \\ & \left. P\left(\bigcap_{s=1}^p \{X_{i_s} \leq u_n\}\right)P\left(\bigcap_{m=1}^q \{X_{j_m} \leq u_n\}\right) \right| \leq \alpha_{n, \ell_n}, \end{aligned}$$

onde $\lim_n \alpha_{n, \ell_n} = 0$ para alguma sucessão $\ell_n = o_n(k_n)$. Temido e Canto e Castro [5] provaram que, para este tipo de sucessões estacionárias, a distribuição limite de M_{k_n} , sob normalização linear, é max-semiestável, caso exista. A condição de dependência local $D'_{k_n}(u_n)$, (introduzida em [5]), verifica-se se existir uma sucessão de inteiros positivos $\{s_n\}$ tal que $k_n/s_n \rightarrow +\infty$, $s_n \alpha_{n, \ell_n} \rightarrow 0$ e $\lim_n k_n \sum_{t=2}^{\lfloor k_n/s_n \rfloor} P(X_t > u_n, X_1 > u_n) = 0$. Estas condições são adaptações das conhecidas condições de Leadbetter adequadas ao contexto de max-semiestabilidade. Como é esperado, sob $D_{k_n}(u_n)$ e $D'_{k_n}(u_n)$, o máximo M_{k_n} e $F_X^{k_n}$ têm a mesma distribuição limite.

Neste trabalho mostramos que a sucessão estacionária $\{X_n\}$ verifica $D_{k_n}(u_n)$ e $D'_{k_n}(u_n)$. Consequentemente o máximo M_{k_n} é atraído para a distribuição Gumbel discreta.

Referências

- [1] Anderson, C.W. (1970) *Extreme value theory for a class of discrete distribution with applications to some stochastic processes*, Journal of Applied Probability. 7, 99–113.
- [2] McCormick, W., Park, Y. (1992) *Asymptotic analysis of extremes from autoregressive negative binomial processes*, Journal of Applied Probability. 29, 904–920.
- [3] McKenzie, E. (1986) *Auto regressive-moving-average processes with negative binomial and geometric marginal distribution*, Advances in Applied Probability. 18, 679–705.
- [4] Pancheva, E. (1992) *Multivariate max-semistable distributions*, Theory of Probability and its Applications. 18, 679–705.
- [5] Temido, M.G., Canto e Castro, L. (2003) *Max-semistable laws in extremes of stationary random sequences*, Theory of Probability and its Applications. 47, 365–374.
- [6] Temido, M.G. (2002) *Domínios de atracção de funções de distribuição discretas*, Novos Rumos em Estatística, Editores: Carvalho, L., Brilhante, F., Rosado, F., Edições SPE, Lisboa. 415–426.

Agradecimentos: As autoras agradecem o apoio da FCT - projetos PEst-OE/MAT/UI4080/2011 e PEst-C/MAT/UI0324/2011 e do Programa Europeu COMPETE.

Combinando testes BHEP para escolhas extremas e não-extremas do parâmetro de suavização

(Comunicação)

Carlos TENREIRO – *CMUC, Departamento de Matemática, Universidade de Coimbra*

tenreiro@mat.uc.pt

Resumo

Sendo X_1, \dots, X_n, \dots uma sucessão de cópias independentes dum vector d -dimensional e absolutamente contínuo X , com densidade de probabilidade f , desconhecida, o problema clássico do teste dum hipótese multivariada de normalidade é o de, com base em X_1, \dots, X_n , testar a hipótese

$$H_0 : f \in \mathcal{N}_d,$$

contra uma hipótese alternativa geral, onde \mathcal{N}_d é a família das densidades de probabilidade normais sobre \mathbb{R}^d . Diversos procedimentos para testar H_0 são propostos na literatura (ver a bibliografia incluída em Tenreiro [6]). Uma família de tais procedimentos é constituída pelos testes BHEP (Baringhaus–Henze–Epps–Pulley) baseados na distância L_2 ponderada entre a função característica amostral associada aos resíduos standardizados $Y_j = S_n^{-1/2}(X_j - \bar{X}_n)$, $j = 1, \dots, n$ (\bar{X}_n e S_n são a média e a matriz de covariância amostrais), definida por

$$\Psi_n(t) = \frac{1}{n} \sum_{j=1}^n \exp(it'Y_j), \quad t \in \mathbb{R}^d,$$

e a função característica Φ da distribuição normal standard em \mathbb{R}^d com densidade

$$\phi(x) = (2\pi)^{-d/2} \exp(-x'x/2), \quad x \in \mathbb{R}^d.$$

A função de peso que consideramos é dada por

$$t \rightarrow |\Phi_h(t)|^2 = \exp(-h^2 t't),$$

onde Φ_h é a função característica de $\phi_h(\cdot) = \phi(\cdot/h)/h^d$ e h é um número real estritamente positivo que deve ser escolhido pelo utilizador. Assim, o teste BHEP é baseado na estatística

$$B(h) = n \int |\Psi_n(t) - \Phi(t)|^2 |\Phi_h(t)|^2 dt = (2\pi)^d \frac{1}{n} \sum_{i,j=1}^n Q(Y_i, Y_j; h),$$

com

$$Q(u, v; h) = \phi_{(2h^2)^{1/2}}(u - v) - \phi_{(1+2h^2)^{1/2}}(u) - \phi_{(1+2h^2)^{1/2}}(v) + \phi_{(2+2h^2)^{1/2}}(0),$$

para $u, v \in \mathbb{R}^d$. A simplicidade da expressão anterior para $B(h)$, justifica a escolha da função de peso considerada. Contrariamente à quase totalidade dos procedimentos de teste existentes, os testes baseados em $B(h)$ são universalmente convergentes, para qualquer escolha do parâmetro de suavização $0 < h < \infty$.

Em Tenreiro [5] propõe-se a utilização dum método considerado em Fromont e Laurent [1], que pode ser visto como um melhoramento do método clássico de Bonferroni, para apresentar um novo teste de normalidade multivariada que combina os testes clássicos de Mardia [3] com os procedimentos de teste $B(h_S)$ e $B(h_L)$ considerados em Tenreiro [4], mostrando o novo teste um bom desempenho para um vasto conjunto de distribuições alternativas quando comparado com os mais recomendados testes numa hipótese multivariada de normalidade. No entanto, outras combinações são possíveis. No presente trabalho estudamos uma dessas combinações que envolve as estatísticas $B(h)$ para escolhas extremas e não-extremas do parâmetro de suavização h . O teste múltiplo que propomos combina assim os testes $B(h_S)$ e $B(h_L)$ com os testes de normalidade considerados em Henze [2], cujas estatísticas de teste se podem obter da estatística $B(h)$ fazendo $h \rightarrow 0$ e $h \rightarrow \infty$.

Referências

- [1] Fromont, M., Laurent, B. (2006) *Adaptive goodness-of-fit tests in a density model*, Ann. Statist. 34, 680–720.
- [2] Henze, N. (1997) *Extreme smoothing and testing for multivariate normality*, Statis. Probab. Lett. 35, 203–213.
- [3] Mardia, K.V. (1970) *Measures of multivariate skewness and kurtosis with applications*, Biometrika 57, 519–530.

- [4] Tenreiro, C. (2009) *On the choice of the smoothing parameter for the BHEP goodness-of-fit test*, *Comput. Statist. Data Anal.* 53, 1038–1053.
- [5] Tenreiro, C. (2011) *An affine invariant multiple test procedure for assessing multivariate normality*, *Comput. Statist. Data Anal.* 55, 1980–1992.
- [6] Tenreiro, C. (2013) *Combinando testes de Mardia e BHEP na avaliação duma hipótese multivariada de normalidade.*, *Boletim da SPE*, 2013 (Primavera), 15–21.

Some comparisons between sequential order statistics

(Comunicação)

Nuria TORRADO – *CMUC and Departamento de Matemática, Universidade de Coimbra*

nuria.torrado@mat.uc.pt

Resumo

Sequential order statistics provide an unified approach to a variety of models of ordered random variables such as order statistics, type II censored order statistics, k -th record values and k_n records from nonidentical distributions, among others. Sequential order statistics (SOS) are proposed in Kamps [3]. Cramer and Kamps [2] gave an equivalent definition to that given by Kamps [3] of sequential order statistics as follows:

Definition: Let F_1, \dots, F_n be distributions functions with $F_1^{-1}(1) \leq \dots \leq F_n^{-1}(1)$ and let V_1, \dots, V_n be independent random variables with $V_r \sim \text{Beta}(n - r + 1, 1)$, $1 \leq r \leq n$. Then, the random variables

$$X_r^* = F_r^{-1}(X^{(r)}),$$

with $X^{(r)} = 1 - V_r \bar{F}_r(X_{r-1}^*)$, $1 \leq r \leq n$, $X_{(0)}^* = -\infty$, are called sequential order statistics based on $\{F_1, \dots, F_n\}$.

The marginal distribution functions $F_{*,1}, \dots, F_{*,n}$ of SOS $X_{1,n}^*, \dots, X_{n,n}^*$ based on $\{F_1, \dots, F_n\}$ are given by:

$$F_{*,1}(t) = 1 - (1 - F_1(t))^n,$$

$$F_{*,r}(t) = \begin{cases} F_{*,r-1}(t) - \int_{-\infty}^t \left(\frac{\bar{F}_r(t)}{\bar{F}_r(z)} \right)^{n-r+1} dF_{*,r-1}(z) & \text{if } F_r(t) < 1, \\ 1 & \text{if } F_r(t) = 1. \end{cases}$$

Some stochastic properties and ageing notions of sequential order statistics have been studied recently in Burkschat and Navarro [1], Navarro and Burkschat [4], Torrado et al. [5] and Zhuang and Hu [6]. In this talk, a short review

of these references is given, and also, I will discuss recent results (and some ongoing work) obtained in collaboration with M. Burkschat.

Acknowledgements

This work was supported by the Portuguese Government through the Fundação para a Ciência e Tecnologia (FCT) under the grant SFRH/BPD/91832/2012.

Referências

- [1] Burkschat, M., Navarro, J. (2011) *Aging properties of sequential order statistics*, Probab. Eng. Inform. Sci. 25, 1–19.
- [2] Cramer, E., Kamps, U. (2003) *Marginal distributions of sequential and generalized order statistics*, Metrika 58, 293–310.
- [3] Kamps, U. (1995) *A concept of generalized order statistics*, J. Statist. Plann. Inference 48, 1–23.
- [4] Navarro, J., Burkschat, M. (2011) *Coherent systems based on sequential order statistics*, Naval Research Logistics 58, 123–135.
- [5] Torrado, N., Lillo, R.E., Wiper, M.P. (2012) *Sequential Order Statistics: Ageing and Stochastic Orderings*, Methodol. Comput. Appl. Probab. 14, 579–596.
- [6] Zhuang, W., Hu, T. (2007) *Multivariate stochastic comparisons of sequential order statistics*, Probab. Eng. Inform. Sci. 21, 47–66.

Estimação nos modelos bivariados de médias móveis de valor inteiro

(Comunicação)

Cristina TORRES – *Instituto Politécnico do Porto - ISCAP e Universidade do Porto*

cmptorres@gmail.com

Isabel Silva – *Faculdade de Engenharia, Universidade do Porto*

Maria Eduarda Silva – *Faculdade de Economia, Universidade do Porto e CIDMA*

Resumo

Nas últimas décadas, o estudo das séries temporais de contagem univariadas tem sido objecto de interesse crescente na literatura. Uma das classes de modelos mais populares é a dos modelos auto-regressivos e médias móveis de valor inteiro não-negativo, INARMA, obtida através da substituição da multiplicação por um operador aleatório, chamado *thinning*, nos modelos ARMA convencionais. Os modelos INAR para séries de contagem univariadas têm sido amplamente estudados na literatura no que diz respeito quer às suas propriedades probabilísticas quer à inferência estatística. Por outro lado, a classe alargada de modelos de médias móveis para séries de valor inteiro, INMA, em que as operações *thinning* em instantes diferentes são dependentes entre si, não tem sido objecto de tanta atenção. De facto os modelos INMA não são Markovianos pelo que a inferência estatística apresenta dificuldades adicionais.

Actualmente o interesse na análise de séries temporais de contagem centra-se em modelos e métodos para séries multivariadas. Neste trabalho, consideram-se os modelos INMA bivariados propostos por Torres *et al.* (2012). Nesta classe alargada de modelos bivariados INMA, a estrutura de dependência entre as duas séries temporais é introduzida pela dependência entre os dois processos de chegada enquanto que a dependência em cada série é definida pela dependência de operações *thinning* em instantes diferentes. Consideram-se estimadores baseados em momentos: método dos momentos (MM), método dos momentos generalizados (GMM) e método dos momentos eficiente (EMM), assim como estimadores baseados na função geradora de probabilidades.

Referências

- [1] Torres, C., Silva, I., Silva, M.E. (2012). *Modelos bivariados de médias móveis de valor inteiro*. XX Congresso Anual da Sociedade Portuguesa de Estatística, 27-29 de Setembro, Porto, Portugal.

Invariância Estrutural da Qualidade de Vida de Pessoas com Doenças Neurológicas e com Doenças Metabólicas: Análise Multigrupo

(Comunicação)

Estela VILHENA – *IPCA Barcelos; ICBAS e ISPUP - Universidade do Porto*
evilhena@ipca.pt

J. Pais Ribeiro – *FPCE-Universidade do Porto, UIPES*

L. Pedro – *ESTeSL IP, Lisboa*

I. Silva – *Universidade Fernando Pessoa, Porto*

R.F. Meneses – *Universidade Fernando Pessoa, Porto*

H. Cardoso – *UMIB/ICBAS e Hospital Santo António/CHP, Portugal*

A. Martins da Silva – *UMIB/ICBAS e Hospital Santo António/CHP, Portugal*

D. Mendonça – *ICBAS, ISPUP, Universidade do Porto*

Resumo

A qualidade de vida (QdV) é um constructo composto por três fatores (bem-estar geral, saúde física e saúde mental) que contribuiu para o bem-estar de um indivíduo e para o ajustamento a uma determinada doença.

A Análise Multigrupo foi utilizada, usando Modelos de Equações Estruturais, com o objetivo de avaliar a invariância estrutural da QdV de indivíduos portadores de doenças neurológicas e de doenças metabólicas. Este tipo de análise pretende avaliar se a estrutura do modelo de medida e do modelo estrutural é invariante em diferentes grupos, com características diferentes.

Numa primeira fase, os modelos foram ajustados individualmente para cada um dos grupos com base no Lagrange Multiple Test. Os resultados encontrados indicaram um bom ajustamento dos modelos: $\chi^2_{(16)}=63.0139$, $p<0.001$, CFI=0.97, RMSEA=0.08; $\chi^2_{(16)}=46.6181$, $p<0.001$, CFI=0.98 e RMSEA=0.074, para o grupo dos doentes neurológicos e metabólicos, respetivamente. Aplicou-se Análise Fatorial Confirmatória Multigrupo (para os grupos independentes), com restrições de igualdade entre os grupos, de

forma testar a invariância dos modelos de medida e estrutural, para a estrutura da QdV. Os valores dos índices de ajustamento encontrados sugerem que os modelos se ajustam bem aos dados (modelo de medida, $\chi^2_{(37)}=98.3414$, $p<0.001$, CFI=0.972, RMSEA=0.055; modelo estrutural, $\chi^2_{(40)}=106.3413$, $p<0.001$, CFI=0.97, RMSEA=0.055).

Concluiu-se que a estrutura da QdV está bem representada pelos três fatores e que existe invariância estrutural entre os grupos em estudo.

Referências

- [1] Byrne, B. M. (2006). Structural Equation Modeling With EQS: Basic Concepts, Applications and Programming (Second Edition ed.). Lawrence Erlbaum Associates, Publishers.
- [2] Tabachni, B. and Fidell, L. (1996). Using Multivariate Statistics (Third edition). New York: HarperCollins College Publishers.
- [3] Kline, R. B. (2011). Principles and Practice of Structural Equation Modeling, (Third Edition ed.). New York: The Guilford Press.

Posters



Generalized Estimating Equations and Mixed Effects Modeling Approaches in Capture-Recapture Closed Population Models

(Poster)

Md. Abdus Salam AKANDA – *Research Center in Mathematics and Applications (CIMA-U.E.), Department of Mathematics, University of Évora, 7000-671 Évora, Portugal and Department of Statistics, Biostatistics & Informatics, University of Dhaka, Bangladesh.*

akanda_du@yahoo.com

Russell Alpizar-Jara – *Research Center in Mathematics and Applications (CIMA-U.E.), Department of Mathematics, University of Évora, 7000-671 Évora, Portugal.*

Abstract

Estimation of animal population parameters is an important issue in ecological statistics. Many estimation methods have been developed for the capture-recapture closed population models (Otis et al., 1978; Amstrup et al., 2005). Development of capture-recapture models dealing with individual heterogeneity in capture probabilities has been one of the most challenging tasks. Failure to account for such heterogeneity has long been known to cause substantial bias in population estimates (Otis et al., 1978; Hwang and Huggins, 2005). We are interested in estimating closed population size using model M_h , where only heterogeneity is considered, there is no time effect or behavioral response to capture, and the capture probabilities depend on covariates. The use of covariates (or auxiliary variables), if available, has been proposed as an alternative way to partially cope with the problem of heterogeneous capture probabilities (Pollock et al., 1984; Huggins, 1989; Alho, 1990). In the analysis of a capture-recapture data, Hwang and Huggins (2005) and Zhang (2012) examined the effect of heterogeneity on the estimation of population size by solving estimating equations. In this work we use Generalized Estimating Equations approaches (GEE) to account

for individual observed heterogeneity, modeling capture probabilities as a function of individual observed covariates, and to account for a correlation structure among capture occasions (Liang and Zeger, 1986). A mixed effects modeling approach may also be used to model individual observed and unobserved heterogeneity in capture-recapture experiments motivating the use of generalized linear mixed models (GLMM) (Pinheiro and Bates, 2000). Some authors have previously introduced the use of GLMM (logit models with normal random effects) (e.g., Coull and Agresti, 1999; Stoklosa et al., 2011) in capture-recapture literature. An advantage of using GLMM for the estimation of capture probabilities is to accommodate not only the heterogeneity of observed individual characteristics, but also the uncertainty that the observed individual characteristics cannot explain. A real example of house mice (*Mus musculus*) data (Otis et al., 1978) is used for illustrative purposes and for comparison to currently used methodology. Conditional arguments are used to obtain a Horvitz-Thompson-like estimator for estimating population size. A simulation study is also conducted to show the performance of the estimation procedure and for comparison within and between methodologies.

References

- [1] Alho, J.M. (1990) *Adjusting for nonresponse bias using logistic regression*, *Biometrika* 77(3), 617–624.
- [2] Amstrup, S. C., McDonald, T. L. and Manly, B. F. J. (eds.). (2005). *Handbook of Capture-Recapture Analysis*, Princeton University Press: Princeton, New Jersey.
- [3] Breslow, N.E. and Clayton, D.G. (1993) *Approximate Inference in Generalized Linear Mixed Models*, *Journal of the American Statistical Association* 88(421), 9–25.
- [4] Coull, B.A. and Agresti, A. (1999) *The use of mixed logit models to reflect heterogeneity in capture-recapture studies*, *Biometrics* 55, 294–301.
- [5] Huggins, R. M. (1989) *On the statistical analysis of capture experiments*, *Biometrika* 76, 133–140.

- [6] Hwang, W. H. and Huggins, R. M. (2005) *An examination of the effect of heterogeneity on the estimation of population size using capture-recapture data*, *Biometrika* 92, 229–233.
- [7] Liang, K. and Zeger, S. L. (1986) *Longitudinal data analysis using generalized linear models*, *Biometrika* 73(1), 13–22.
- [8] Otis, D. L., Burnham, K. P., White, G. C. and Anderson, D. R. (1978) *Statistical Inference From Capture Data on Closed Animal Populations*, *Wildlife Monograph* 62, Washington, DC: The Wildlife Society.
- [9] Pinheiro, J. and Bates, D. M. (2000) *Mixed-Effects Models in S and S-PLUS*, *Statistics and Computing Series*, Springer-Verlag, New York, NY.
- [10] Pollock, K., Hines, J. and Nichols, J. (1984) *The use of auxiliary variables in capture-recapture and removal experiments*, *Biometrics* 40, 329–340.
- [11] Stoklosa, J., Hwang, W., Wu, S. and Huggins, R. (2011) *Heterogeneous capture-recapture models with covariates: A partial likelihood approach for closed populations*, *Biometrics* 67, 1659–1665.
- [12] Zhang, S. (2012) *A GEE approach for estimating size of hard-to-reach population by using capture recapture data*, *Statistics* 46(2), 175–183.

Modelos de Cura Aplicados ao Cancro da Mama

(Poster)

Carina ALVES – Núcleo de Gestão de Doentes e Estatística, SESARAM

anacarina.alves@gmail.com

Ana Maria Abreu – CCCEE e CCM, Universidade da Madeira

Resumo

O desenvolvimento da medicina, nomeadamente no que diz respeito à cura de doenças do foro oncológico, tem sido um forte motivo para a aplicação de modelos de cura, em detrimento ou como complemento dos modelos mais tradicionais da análise de sobrevivência. Assim, o foco passou a estar não só na estimação do tempo de vida dos indivíduos (onde se inclui o efeito das covariáveis relevantes) mas também na probabilidade de cura dos pacientes.

Neste âmbito, considerou-se uma base de dados referente a 833 mulheres diagnosticadas com cancro da mama, do tipo não inflamatório, entre 1998 e 2005, na Região Autónoma da Madeira.

Através da estimativa de Kaplan-Meier da função de sobrevivência, aos 14 anos de *follow-up*, a sobrevida estimada para o total de casos situou-se nos 63,8%. Verificou-se a existência de um maior risco de morte em mulheres na faixa etária dos 50 a 59 anos. Comprovou-se que o estadiamento tem um papel preponderante em relação ao prognóstico, sendo que, quanto mais avançado o estadio pior o prognóstico.

O modelo de cura foi aplicado apenas ao grupo de mulheres no estadio IV, pois só neste caso se admitiu que o tempo de *follow-up* era suficiente (Yu *et al.*, 2004): a mediana do tempo de vida foi de 1 ano e 6 meses e o tempo mínimo de *follow-up* foi de 5 anos. Com o modelo de cura baseado na distribuição de Chen (Chen, Z., 2000), obteve-se uma taxa de cura de 7,4%.

Referências

- [1] Chen, Z. (2000) *A new two-parameter lifetime distribution with bathtub shape or increasing failure rate function*, Statistics and Probability Letters. 49, 155–161.
- [2] Yu, B., Tiwari, R.C., Cronin, K.A., Feuer, E.J. (2004) *Cure fraction from the mixture cure models for grouped survival data*, Statistics in Medicine. 23, 1733–1747.

Monte Carlo na Avaliação de Opções Americanas

(Poster)

Maria de Lurdes BABO – *Instituto Politécnico do Porto - ISCAP e Universidade do Porto*

lbabo@iscap.ipp.pt

Nelson Areal – *Escola de Economia e Gestão da Universidade do Minho*

Maria Eduarda Silva – *Faculdade de Economia, Universidade do Porto e CIDMA*

Resumo

A determinação do preço justo de um contrato de opções, trouxe enormes desafios a diversos ramos da Matemática procurando desenvolver e aperfeiçoar modelos e métodos que melhor representem os comportamentos dos mercados financeiros.

A avaliação de opções americanas apresenta uma dificuldade acrescida, uma vez que é necessário determinar uma estratégia óptima ao seu exercício antecipado, já que estas opções podem ser exercidas em qualquer momento até à sua maturidade. Investigações recentes mostram que metodologias baseadas em técnicas de simulação podem ser usadas com sucesso neste tipo de opções (Duan and Simonato (2001), Longstaff and Schwartz (2001), Stentoft (2005)). Neste trabalho, usamos métodos de Monte Carlo para avaliar opções Americanas, recorrendo à abordagem sugerida por Longstaff and Schwartz (2001), combinando modelos GARCH para o subjacente com *Filtered Historical Simulation* (Barone, Engle and Mancini (2008)).

Referências

- [1] Duan, J., Simonato, J., (2001) *American option pricing under GARCH by a Markov chain approximation*, Journal of Economic Dynamics and Control 25, 1689–1718.
- [2] Longstaff, F., Schwartz, E. (2001) *Valuing American options by simulation: A simple least-squares approach*, Review of Financial Studies 14, 113–147.

- [3] Stentoft, L. (2005) *Pricing American options when the underlying asset follows GARCH processes*, Journal of Empirical Finance 12, 576–611.
- [4] Barone-Adesi, G., Engle, R., Mancini, L. (2008) *Valuing American options by simulation: A simple least-squares approach*, Review of Financial Studies 21, 1223–1258.

A Tuberculose Pulmonar em Crianças e Adolescentes em Portugal Continental

(Poster)

Marília BETTENCOURT SILVA – *Escola Nacional de Saúde Pública, Universidade Nova de Lisboa*

mariliabettencourtsilva@gmail.com

Patrícia A. Filipe – *Escola de Ciência e Tecnologia, Universidade de Évora, CIMA/UE*

Dulce Gomes – *Escola de Ciência e Tecnologia, Universidade de Évora, CIMA/UE*

Cristiana Areias – *Escola Nacional de Saúde Pública, Universidade Nova de Lisboa*

Carla Nunes – *Escola Nacional de Saúde Pública, Universidade Nova de Lisboa, CIESP e CMDT*

Resumo

A Tuberculose Pulmonar (TBP) nas crianças e adolescentes até aos 18 anos de idade representa o foco deste estudo. As faixas etárias mais jovens são de particular interesse uma vez que representam o caso em que a transmissão da infeção foi recente. Em 2011 estimou-se, pela primeira vez a nível mundial, meio milhão de casos e 64000 mortes em crianças (<15 anos) devido à tuberculose (TB)[1]. Alguns autores referem que a incidência da TBP nas crianças ainda está mal documentada e sugere a dificuldade de um diagnóstico de TBP. Com o intuito de conhecer melhor a transmissão de TBP nesta faixa etária e contribuir para o controlo da mesma, estudou-se o número de casos de crianças notificadas com TBP e a demora no diagnóstico (tempo decorrido desde o início dos sintomas até ao diagnóstico da doença), entre 2000 e 2010 em Portugal continental. A base de dados utilizada foi disponibilizada pelo Sistema Nacional de Vigilância da Tuberculose (SVIG-TB). Realizamos um estudo descritivo e analítico do número de casos de TBP em relação às variáveis sexo, idade, distrito, coinfeção com VIH, número de tratamentos anteriores, motivo do diagnóstico e demora no diagnóstico. Foram utilizadas técnicas de estatística clássica para caracterizar os casos notificados com

TBP com idades entre os 0 e os 18 anos e técnicas de análise de sobrevivência para caracterizar a demora no diagnóstico e modelar o risco de ocorrência do mesmo. Do total de 28612 casos notificados com TBP no período em estudo, 985 (3,4%) casos são de crianças e adolescentes até aos 18 anos de idade (dos quais 49% têm idades inferiores a 15 anos). Os casos em estudos são maioritariamente masculinos correspondendo a 52,2%, a média de idades é de 11,9 anos ($dp=6,1$). O distrito com maior número de crianças com TBP é o do Porto com 31,6% dos casos seguido de Lisboa com 26%. Apenas 1,8% das crianças apresenta coinfeção com o VIH no momento do diagnóstico. A mediana global da demora no diagnóstico da TBP é de 57 dias[2], no caso particular das crianças com menos de 15 anos é de 41 dias e para o grupo etário dos 15 aos 18 anos é de 45 dias. O teste log-rank apenas revelou diferenças significativas nas curvas de sobrevivência para o número de tratamentos anteriores e para o motivo do diagnóstico ($p \leq 0,043$). As possíveis razões da demora no diagnóstico devem ser objeto de um estudo continuado para que se possa adequar as medidas de saúde pública necessárias à sua diminuição. O grupo estudado apresenta maior incidência nas crianças com idades entre 15 e 18 anos, consistente com outros estudos, este facto pode estar associado à entrada ativa na comunidade[3].

Referências

- [1] WHO. STOPTB Partnership - Global Tuberculosis Report 2012. WHO: Geneva. 2012.
- [2] Gomes, D., Filipe, P.A., Nunes, C. e de Sousa, B. (2013) *Penalized spline smoothing for delay in Pulmonary Tuberculosis diagnosis*, Proceedings of the 28th International Workshop on Statistical Modelling, (Muggeo VMR, Capursi V, Boscaïno G, Lovison G, editors), vol.2, 621-25.
- [3] Gonçalves, J., Cerqueira, A., Machado, C., Carvalho, F., Cruz S., Gonçalves, A. e Martins, S. (2012) *Tuberculose em idade pediátrica: características, incidência e distribuição geográfica (2000-2010)*, Acta Médica Pediátrica, 43(3), 104-10.

Estudo da dinâmica populacional da gamba e da faneca

(Poster)

Nuno Miguel BRITES – *Instituto Português do Mar e da Atmosfera, I.P. & Centro de Investigação em Matemática e Aplicações, Universidade de Évora, Portugal*

nbrites@ipma.pt

Manuela Azevedo – *Instituto Português do Mar e da Atmosfera, I.P., Portugal*

Cristina Silva – *Instituto Português do Mar e da Atmosfera, I.P., Portugal*

Resumo

A gamba (*Parapenaeus longirostris*) e a faneca (*Trisopterus luscus*) são duas espécies importantes na economia nacional, encontrando-se entre as dez principais espécies no *ranking* de desembarques em valor em Portugal continental [1]. Dada a importância destas espécies, é necessário modelar e avaliar as suas dinâmicas populacionais de modo a caracterizar o seu estado de exploração e prever potenciais de captura.

Os actuais modelos de avaliação de recursos pesqueiros podem categorizar-se em modelos analíticos [2], modelos de produção [2,3] e modelos para dados limitados [4]. Os modelos analíticos consideram a evolução das coortes ao longo do tempo, estimando a dimensão da população por idades. Os modelos de produção analisam a evolução da biomassa ao longo dos anos, estimando a capacidade de sustento e a taxa intrínseca de crescimento da biomassa. No caso dos modelos para dados limitados, a estimação é restrita a valores de capturas sustentáveis, uma vez que estes modelos apenas requerem dados de captura e algumas suposições sobre a variação da biomassa.

Os modelos serão aplicados a dados de captura e índices de abundância da faneca e da gamba, no período compreendido entre 1980-2010. Para cada modelo estimam-se parâmetros populacionais, caracteriza-se o estado de exploração e projectam-se capturas a curto prazo. Na discussão é dado enfoque aos aspectos relacionados com a selecção do modelo que, para cada recurso, melhor descreve a sua dinâmica.

Referências

- [1] Azevedo, M., et al. (2012) *Avaliação Inicial do Estado Ambiental das Populações de Peixes e Moluscos Explorados Comercialmente na ZEE Continental Portuguesa: Descritor 3 da Diretiva Quadro da Estratégia Marinha DQEM in* www.dgrm.min-agricultura.pt.
- [2] Haddon, M. (2001) *Modelling and Quantitative Methods in Fisheries*. Chapman & Hall.
- [3] Prager, M. (1994) *A nonequilibrium surplus-production model*, Fisheries Bulletin. 92(2), 372–389.
- [4] MacCall, A. (2009) *Depletion-corrected average catch: a simple formula for estimating sustainable yields in data-poor situations*, Fisheries Research. 110, 331–341.

Estudo correlacional entre a idade cronológica, a maturação esquelética da mão e punho e a maturação esquelética das vértebras cervicais, numa população portuguesa

(Poster)

Lima S.—*Universidade de Coimbra, Faculdade de Medicina, Medicina Dentária*

Alves S.—*Universidade de Coimbra, Faculdade de Medicina, Medicina Dentária*

Teixeira V.—*Universidade de Coimbra, Faculdade de Medicina, Medicina Dentária*

CANOVA F.—*Instituto Politécnico de Coimbra, ISEC, DFM*
fcanova@isec.pt

Maló L.—*Universidade de Coimbra, Faculdade de Medicina, Medicina Dentária*

Resumo

O objetivo deste estudo é estabelecer a correlação entre a idade cronológica, a maturação esquelética da mão e punho e a maturação esquelética das vértebras cervicais numa amostra infantojuvenil portuguesa.

Materiais e Métodos: Foram selecionadas de forma aleatória 285 radiografias da mão e punho e telerradiografias de perfil da face de crianças portuguesas de ambos os sexos, com idades compreendidas entre os 7 e os 16 anos, no período circumpubertal. A maturação esquelética das vértebras cervicais foi avaliada através do método desenvolvido por Baccetti e colaboradores, enquanto que a maturação esquelética da mão e punho foi avaliada através do método desenvolvido por Grave e Brown. Os resultados obtidos pela aplicação de ambos os métodos foram correlacionados com a idade cronológica através do coeficiente de correlação de Spearman.

Resultados: A maturação esquelética das vértebras cervicais e a idade óssea da mão e punho apresentaram uma correlação positiva forte ($p < 0.01$, r_s masculino = 0.806, r_s feminino = 0.803). Verificou-se também uma

correlação positiva forte entre a maturação das vértebras cervicais e idade cronológica ($p < 0.01$, r_s masculino = 0.778; r_s feminino = 0.788) e entre a maturação da mão e punho e idade cronológica ($p < 0.01$, r_s masculino = 0.820; r_s feminino = 0.847).

Conclusões: Com base nesta amostra há evidência estatística para afirmar que na população portuguesa existe uma correlação significativa entre a maturação esquelética determinada a partir das vértebras cervicais e a maturação esquelética determinada a partir da mão e punho. Estas variáveis estão fortemente correlacionadas, apresentando a correlação o sentido positivo. Deste modo, é justificável a utilização do método de determinação da maturação esquelética através das vértebras cervicais na prática ortodôntica, no tratamento de crianças portuguesas, evitando assim exposição a radiação adicional. De forma idêntica, a forte correlação entre a idade cronológica e os estados de maturação esquelética da mão e punho e entre a idade cronológica e os estados de maturação esquelética das vértebras cervicais sugere que, dentro de certos limites, a idade cronológica pode ser utilizada para determinar o estado de maturação esquelética de um indivíduo.

Referências

- [1] Grave KC., Brown T. (1976) *Skeletal ossification and the adolescent growth spurt*, Am J Orthod and Dentofacial Orthop. 69 (6): 611-619
- [2] Franchi L, Baccetti T, McNamara JA.(2000) *Mandibular growth as related to cervical vertebral maturation and body height*, Am J Orthod and Dentofacial Orthop. 118: 335-340
- [3] Baccetti T., Franchi L., McNamara, JA.(2002) *An improved version of the cervical vertebral maturation (CVM) method for the assessment of mandibular growth*, Angle Orthod. 72: 316-323
- [4] Baccetti T., Franchi L., De Toffol L., Ghiozzi B., Cozza P.(2006) *The diagnostic performance of chronologic age in the assessment of skeletal maturity*, Prog Orthod. 7(2): 176-188
- [5] Santiago RC., Costa LF de M., Vitral RWF., Fraga MR., Bolognese AM., Maia LC.(2012) *Cervical vertebral maturation as a biologic indicator of skeletal maturity. A systematic review*, Angle Orthod. 82 (6): 1123-1131

Estudo comparativo da maturação esquelética de crianças portuguesas com a de crianças de outras nacionalidades, através do Método TW3

(Poster)

Teixeira V.—*Universidade de Coimbra, Faculdade de Medicina, Medicina Dentária*

Alves S.—*Universidade de Coimbra, Faculdade de Medicina, Medicina Dentária*

Lima S.—*Universidade de Coimbra, Faculdade de Medicina, Medicina Dentária*

CANOVA F.—*Instituto Politécnico de Coimbra, ISEC, DFM*
fcanova@isec.pt

Maló L.—*Universidade de Coimbra, Faculdade de Medicina, Medicina Dentária*

Resumo

O objetivo deste estudo é determinar a maturidade esquelética rádio-cúbito-ossos curtos (RUS) numa amostra portuguesa e comparar as suas características com as de populações congéneres de outras nacionalidades.

Materiais e Métodos: Foram selecionadas de forma aleatória 480 radiografias da mão e punho de crianças portuguesas de ambos os sexos, com idades compreendidas entre os 7 e os 15 anos. Foi utilizado o método Tanner-Whitehouse 3 (TW3) na análise das radiografias, com vista à determinação da maturação esquelética.

Foram elaborados gráficos dos percentis 50 dos "RUS scores" das crianças portuguesas que posteriormente foram comparados com os correspondentes das crianças chinesas, belgas e japonesas, por sexo. As curvas suavizadas foram obtidas recorrendo ao programa MATLAB, pela aplicação do método dos mínimos quadrados na regressão polinomial de grau 3.

Resultados: Com base na nossa amostra há evidência estatística para afirmar que na população portuguesa existem diferenças significativas entre a idade esquelética e a idade cronológica em ambos os sexos. Tais resultados

foram confirmados pela utilização de testes t-de *Student* para comparação de médias ($p < 0.01$). De um modo geral, podemos afirmar que os rapazes apresentaram, um atraso na idade esquelética em relação à idade cronológica, enquanto que as raparigas na faixa etária dos 10-14 anos (período sobreponível ao surto pubertário) apresentaram idade esquelética mais avançada do que a idade cronológica. A utilização dos percentis 50 para os "RUS scores" mostrou que as raparigas portuguesas maturam mais cedo do que as belgas e as asiáticas, enquanto que os rapazes portugueses, apesar de terem um percentil 50 sobreponível ao dos rapazes chineses e japoneses, maturam mais cedo do que os belgas.

Conclusões: As raparigas portuguesas maturam mais cedo do que as chinesas, japonesas e belgas, enquanto que os rapazes portugueses apesar de terem uma maturação esquelética semelhante à dos chineses e japoneses, maturam mais cedo do que os rapazes belgas.

Referências

- [1] Beunen G., Lefevre J., Ostryn M., Renson R., Simons J., Van Gerven D.(1990) *Skeletal maturity in Belgian youths assessed by the Tanner-Whitehouse method (TW2)* , Ann Hum Biol. 17(5): 355-376
- [2] Freitas D., Maia J., Beunen G., Lefevre J., Claessens A., Marques A., Rodrigues A., Silva C., Crespo M., Thomis M., Sousa A., Malina R.(2004) *Skeletal maturity and socio-economic status in Portuguese children and youths: the Madeira growth study* , Ann Hum Biol. 31(4): 408-420
- [3] Ashizawa K., Asami T., Anzo M., Matsuo N., Matsuoka H., Murata M., et al. (1996) *Standard RUS skeletal maturation of Tokyo children* , Ann Hum Biol. 23(6): 457-469
- [4] Zhang SY., Liu LJ., Wu ZL., Liu G., Ma ZG., Shen XZ., et al. (2008) *The diagnostic performance of chronologic age in the assessment of skeletal maturity* , Ann Hum biol. 35(3): 349-354

Controlo estatístico de qualidade de nutrientes em águas salinas

(Poster)

Dora CARINHAS – *Mestranda em Modelação Estatística e Análise de Dados (Universidade de Évora) e Instituto Hidrográfico da Marinha*

dora.carinhas@hidrografico.pt

Paulo Infante – *CIMA-UE e DMAT, ECT da Universidade de Évora*

Maria do Pilar Silva – *Instituto Hidrográfico da Marinha*

Resumo

A agricultura intensiva, os esgotos e os resíduos provenientes dos grandes centros urbanos são a causa do excesso de nutrientes nos ecossistemas aquáticos. Os níveis de nutrientes determinam a quantidade de microrganismos que podem crescer em qualquer sistema aquático. Assim, se a população de microrganismos se modificar repentinamente o ecossistema inteiro pode transformar-se num caos.

O Instituto Hidrográfico da Marinha, através da Divisão de Química e Poluição do Meio Marinho, promove e realiza estudos e trabalhos destinados a ampliar o conhecimento da química da água do mar e do meio marinho, na costa, estuários e outras áreas de interesse nacional.

Devido à grande concorrência no mercado nacional e internacional, os laboratórios procuram a garantia dos seus ensaios com um mínimo custo possível. Neste contexto, a aplicação de ferramentas estatísticas que permitam aumentar a fiabilidade dos resultados laboratoriais assume um papel muito importante.

Este trabalho tem como principal objetivo modelar e analisar dados de ensaios laboratoriais, realizados no Instituto Hidrográfico da Marinha, relativos a nutrientes em águas salinas, usando, em particular, técnicas de controlo estatístico de qualidade aplicadas no âmbito da acreditação segundo a norma NP EN ISO/IEC 17025 [4]. Trata-se de uma norma que descreve os requisitos gerais para o reconhecimento da competência de um laboratório em realizar ensaios e/ou calibrações, incluindo amostragem, que é utilizada pelos labora-

tórios para desenvolver os seus sistemas de gestão para a qualidade, e para as atividades administrativas e técnicas (veja-se, por exemplo [1]). De acordo com esta norma, o laboratório deve ter procedimentos de controlo da qualidade para monitorizar a validade dos ensaios e calibrações realizadas. A análise destes dados permitirá que sejam tomadas ações de correção planeadas sempre que não sejam satisfeitos os critérios predefinidos.

A norma NP EN ISO/IEC 17025 refere a necessidade dos laboratórios participarem em Ensaios de Comparação Interlaboratorial (ECI) como garantia da fiabilidade dos resultados. A participação em ECI permite evidenciar a competência de um laboratório através da avaliação do seu desempenho, reforçando assim a confiança nas técnicas desenvolvidas. Os ECI não só contribuem para a avaliação do desempenho dos laboratórios, como também são a ferramenta de controlo da qualidade utilizada, no Instituto Hidrográfico, para monitorizar a validação de métodos e deteção de erros sistemáticos.

Neste contexto, realizamos estudos de repetibilidade e reprodutibilidade, evidenciando a mais valia da sua introdução tendo em vista os objetivos definidos pelo laboratório. A este propósito veja-se [2] onde se pode ver uma aplicação deste estudos na determinação de água nas amostras de óleo mineral.

Aplicam-se, também, outras técnicas estatísticas usuais no tratamento de informações oriundas de procedimentos laboratoriais [3], tais como os testes de Grubbs (definido como sendo um teste para avaliar a variabilidade entre laboratórios), Dixon (utilizado para detetar a presença de outliers) e Cochran (definido como sendo um teste para estudar a variabilidade interna de um laboratório).

As cartas de controlo constituem um método simples e económico para monitorizar os resultados obtidos em laboratório. São imensas as suas aplicações em diversas áreas fora da indústria. Neste trabalho são aplicadas a sistemas de medição, sendo dado um especial destaque à formalização de linhas gerais para uma melhor operacionalização das cartas já existentes e para a implementação de outras cartas de controlo, como as EWMA, procurando definir a carta mais apropriada para diferentes variáveis e contextos. Por fim, mostramos como a análise de capacidade do processo em cumprir as especificações pode ser uma ferramenta muito útil ao laboratório.

Referências

- [1] Abdel-Fatah Hesham, T. M. (2010) *ISO/IEC 17025 Accreditation: Between the desired gains and the reality*, Journal of Quality Assurance 13, 21–27.
- [2] Chui, Q. S. H., Antonoff, H. B., Olivieri, J. C. (2002) *Utilização de índices r e R obtidos de programas interlaboratoriais para o controle de precisão do método analítico: Determinação de água por Karl Fischer*, Quim. Nova 25, 657–659.
- [3] International Standard Organization (1998) ISO 5725 - Accuracy (trueness and precision) of Measurement Methods and Results.
- [4] ISO/IEC 17025 (2005) Requisitos Gerais de Competência para laboratórios de ensaio e calibração, Instituto Português da Qualidade.
- [5] Mullins, E. (2003) Statistics for the quality control chemistry laboratory, The Royal Society of Chemistry.

Um modelo populacional estocástico para uma espécie vulnerável de tubarão pelágico, o tubarão-raposo-olhudo (*Alopias superciliosus*) no Oceano Atlântico

(Poster)

Rui COELHO – *Instituto Português do Mar e da Atmosfera (IPMA, I.P.) e Centro de Ciências do Mar do Algarve (CCMAR)*

rpcoelho@ipma.pt

Joana Fernandez-Carvalho – *Instituto Português do Mar e da Atmosfera (IPMA, I.P.) e Centro de Ciências do Mar do Algarve (CCMAR)*

Miguel N. Santos – *Instituto Português do Mar e da Atmosfera (IPMA, I.P.)*

Karim Erzini – *Centro de Ciências do Mar do Algarve (CCMAR)*

Resumo

No Oceano Atlântico os tubarões pelágicos são normalmente capturados como espécies acessórias de pescarias oceânicas, incluindo a pesca de palangre de superfície dirigida a espadartes e atuns. A pescaria Portuguesa de palangre de superfície opera numa vasta área do Atlântico, e apesar de ter como espécie alvo sobretudo o espadarte (*Xiphias gladius*), os tubarões oceânicos constituem uma componente importante das capturas [1]. Neste trabalho desenvolveu-se um modelo de dinâmica populacional para uma espécie de tubarão pelágico com um ciclo biológico particularmente vulnerável, o tubarão raposo-olhudo (*Alopias superciliosus*), sendo a análise realizada com matrizes de Leslie estruturadas por idade [2]. O tubarão raposo-olhudo é uma espécie com distribuição global em águas tropicais e temperadas, e apesar de ser capturado frequentemente nesta pescaria permanece como uma das espécies menos estudada, designadamente no Oceano Atlântico. Devido a incertezas relativamente aos vários parâmetros biológicos usados na análise populacional, nomeadamente em relação a parâmetros de sobrevivência e de reprodução, foi feita uma simulação de Monte Carlo de modo a incorporar estocasticidade aos parâmetros. A estimativa de crescimento populacional obtida com es-

tas simulações ($\lambda = 1.008$; Intervalo de Confiança a 95% = [0.9968, 1.0189]) sugere que esta espécie de tubarão pelágico possui um potencial de crescimento muito baixo, e que corresponde a um dos valores de λ mais baixos conhecidos para todos os tubarões [3]. Foram estimadas as elasticidades para as matrizes, tendo-se verificado que os maiores valores correspondem às sobrevivências das idades correspondentes à fase juvenil, enquanto as elasticidades para as sobrevivências dos adultos, assim como para os parâmetros reprodutivos, são menores. Este padrão ao nível das elasticidades parece ser característico das espécies de tubarões que possuem taxas de crescimento mais lentas, em que a proteção dos juvenis parece ser a melhor estratégia ao nível de conservação das populações. Por outro lado, nas espécies de crescimento mais rápido e com maiores taxas de fecundidade, as elasticidades parecem estar distribuídas de forma mais homogênea entre as idades dos juvenis e dos adultos. Este trabalho apresenta nova e importante informação sobre a dinâmica populacional do tubarão raposo-olhudo no Oceano Atlântico, e vem corroborar a informação já disponível para esta espécie que parece ser uma das de tubarões mais vulneráveis à pesca. Os resultados deste trabalho permitem agora uma melhor compreensão da dinâmica populacional da espécie, e poderão ser utilizados para apoio à sua melhor gestão e conservação.

Referências

- [1] Coelho, R., Fernandez-Carvalho, J., Lino, P.G., Santos, M.N. (2012) *An overview of the hooking mortality of elasmobranchs caught in a swordfish pelagic longline fishery in the Atlantic Ocean*, Aquatic Living Resources 25, 311–319.
- [2] Caswell, H. (2001) *Matrix Population Models: Construction, Analysis, and Interpretation* 2nd Ed, Sinauer Associates.
- [3] Cortes, E. (2002) *Incorporating uncertainty into demographic modeling: application to shark populations and their conservation*, Conservation Biology 16, 1048–1062.

Estimador para a média do estado de um modelo em espaço de estados baseado na propagação do viés dos preditores do filtro de Kalman

(Poster)

Marco COSTA – *ESTGA – Universidade de Aveiro e CMAF – Universidade de Lisboa*

marco@ua.pt

Magda MONTEIRO – *ESTGA e CIDMA – Universidade de Aveiro*

Resumo

Neste trabalho pretende-se estudar o impacto da substituição da verdadeira média do estado de um modelo em espaço de estados nos preditores do filtro de Kalman. O efeito da utilização de uma estimativa da média é discutido para modelos invariantes e para modelos com coeficientes variáveis, embora constantes. Os resultados obtidos permitem propor um novo estimador para a média cujo erro quadrático médio é comparado com o estimador de máxima verosimilhança num conjunto de cenários e para erros gaussianos.

O estudo do erro quadrático médio dos preditores do filtro de Kalman associado a um modelo em espaço de estados com parâmetros estimados tem sido abordado por alguns autores (Rodríguez e Ruiz, 2012; Ansley e Kohn, 1986; Durbin e Koopman, 2000; Quenneville e Singh, 2000; Pfeiffermann e Tiller, 2005). Monteiro e Costa (2012) apresentam expressões analíticas para o viés induzido nos preditores $\hat{\beta}_{t|t-1}$ e $\hat{\beta}_{t|t}$ do filtro de Kalman com a média de β_t estimada por $\hat{\mu}$ e os restantes parâmetros do modelo em espaço de estados multivariado são conhecidos.

Considere-se o modelo em espaço de estados univariado

$$\begin{aligned} Y_t &= h_t \beta_t + e_t \\ \beta_t &= \mu + \phi(\beta_{t-1} - \mu) + \epsilon_t \end{aligned}$$

onde Y_t é a variável observável, β_t o estado não observável com média μ , ϕ o parâmetro autoregressivo tal que $|\phi| < 1$, e_t e ϵ_t os erros ruídos branco não

correlacionados com variâncias σ_ϵ^2 e σ_δ^2 , respetivamente.

Monteiro e Costa (2012) mostraram que tomando uma estimativa de μ , $\hat{\mu} = \mu + \delta$, os preditores do filtro de Kalman $\hat{\beta}_{t|t-1}^*$ e $\hat{\beta}_{t|t}^*$ deixam de ser centrados e os viés podem ser obtidos através de expressões analíticas diretas ou pelas equações recursivas $\text{viés}(\hat{\beta}_{t|t-1}) = (1 - \phi)\delta + \phi\text{viés}(\hat{\beta}_{t-1|t-1})$ e $\text{viés}(\hat{\beta}_{t|t}) = (1 - k_t h_t)\text{viés}(\hat{\beta}_{t|t-1})$ onde k_t é o *ganho* do filtro de Kalman e $\text{viés}(\hat{\beta}_{1|0}) = \delta$.

Aplicando recursivamente as expressões anteriores para os viés é possível estimar δ através do método dos mínimos quadrados $\hat{\delta}$, obtendo-se um novo estimador para a média dado por $\hat{\hat{\mu}} = \hat{\mu} - \hat{\delta}$. Este estimador é comparado com o estimador de máxima verosimilhança através do erro quadrático médio e das respetivas distribuições empíricas.

Referências

- [1] Ansley, C.F., Kohn, R. (1986) *Prediction mean squared error for state space models with estimated parameters*, Biometrika. 73, 467–473.
- [2] Durbin, J., Koopman, S. (2000) *Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives. With comments*, Journal of the Royal Statistical Society Series B. 62, 3–56.
- [3] Monteiro, M., Costa, M. (2012) *A note on prediction bias for state space models with estimated parameters*, ICNAAM, In AIP Conference Proceedings, Greece. 1479, 2090–2093.
- [4] Pfeffermann, D., Tiller, R. (2005) *Bootstrap approximation to prediction MSE for state space-models with estimated parameters*, Journal of Time Series Analysis. 21, 219–236.
- [5] Quenneville, B., Singh, A.C. (2000) *Bayesian prediction mean squared error for state space models with estimated parameters*, Journal of Time Series Analysis. 26, 893–916.
- [6] Rodríguez, A., Ruiz, E. (2012) *Bootstrap prediction mean squared errors of unobserved states based on the Kalman filter with estimated parameters*, Computational Statistics and Data Analysis. 56, 62–74.

Volatility of Inflation Rate in Mozambique: A Variance Decomposition

(Poster)

Bruno DAMÁSIO – *ISEG - Departamento de Matemática*

bdamasio@iseg.utl.pt

João Nicolau – *ISEG - Departamento de Matemática*

Carlos Barros – *ISEG - Departamento de Economia*

Resumo

This paper investigates the volatility of inflation in Mozambique from December 1995 to October 2012 and proposes a new procedure to decompose inflation volatility according to the various CPI components: a Triangular GARCH model. We show that monsoon weather has an intense effect on the inflation and inflation volatility and that the food, housing, water, electricity, gas and other fuels and transportation are the major contributors to the variability of average inflation. The paper is directly comparable with Bekaert et al. (2005) and Christiansen (2007), but our model is more general in that it does not assume that causality goes only in one direction. In our case, there is not a causal ordering of the (inflation component) variables, as implicitly defined in Bekaert et al. (2005) and Christiansen (2007). In fact, we allow the inflation components to interact with each other. However, for obvious reasons, we do assume that inflation is caused, in Granger sense, by all inflation components. Based on these findings we propose some recommendations for economic policy.

Referências

- [1] Bekaert, G., Harvey, C. R. and Ng, A. (2005) *Market integration and contagion*, *Journal of Business*: 78, 39–69.
- [2] Christiansen, C. (2007) *Volatility-Spillover Effects in European Bond Markets.*, *European Financial Management*: 13(5), 923–948.

MRSA e Autocarros Públicos

(Poster)

Fernanda DIAMANTINO – *DEIO e CEAUL, Faculdade de Ciências da Universidade de Lisboa*

mfdiamantino@fc.ul.pt

Marta Aires-de-Sousa – *Escola Superior de Saúde da Cruz Vermelha Portuguesa (ESSCVP), Lisboa*

Resumo

A infeção por *Staphylococcus aureus* resistente à meticilina (MRSA) tem vindo a revelar-se um problema em termos de saúde pública [1]. Durante muito tempo as infeções causadas por MRSA eram exclusivamente adquiridas em ambiente hospitalar. Mais recentemente, foram aparecendo novos reservatórios que explicam a propagação cada vez maior na comunidade: os autocarros públicos constituem um destes reservatórios.

Com este trabalho procura-se averiguar a associação entre a presença de MRSA e as rotas de autocarros que passam, no mínimo, por um hospital. Investiga-se também se a sazonalidade que a infeção por MRSA apresenta em outros países, referida na literatura [2], se verifica também ao nível da contaminação dos autocarros em Portugal.

Agradecimentos

Este trabalho foi financiado por Fundos Nacionais através da FCT - Fundação para a Ciência e a Tecnologia no âmbito dos projetos PEst-OE/MAT/UI0006/2011 e PEst-OE/EQB/LA0004/2011.

Referências

- [1] Aires-de-Sousa, M., Correia, B., de Lencastre H. (2008) *Changing patterns in frequency of recovery of five methicillin-resistant Staphylococcus aureus clones in Portuguese hospitals: surveillance over a 16-year period.*, J. Clin. Microbiol. 46, 2912–2917.
- [2] Mermel, L.A., Machan, J.T., Parenteau, S. (2011) *Seasonality of MRSA Infections*, PLoS One 6: e17925.

Regressão Logística como ferramenta de Melhoria da Qualidade dos Cuidados de Saúde: Óbitos Potencialmente Preveníveis no Hospital Professor Doutor Fernando Fonseca, EPE

(Poster)

Fernando FARIA – *Direcção da Qualidade, Hospital Professor Doutor Fernando Fonseca, EPE*

fernando.a.faria@hff.min-saude.pt

Paulo Infante – *CIMA-UE e DMAT, ECT da Universidade de Évora*

Resumo

A utilização da mortalidade nos hospitais tem sido largamente usada para aferir a qualidade dos cuidados prestados ([6]). Em [2] é referido que a taxa de mortalidade bruta é um mau indicador, pois não tem em conta a complexidade dos pacientes, sendo as taxas ajustadas ao risco através de modelos estatísticos mais informativas e mais comparáveis. Em [1], sugerem centrar o foco nas mortes preveníveis.

Este trabalho tem como objetivo determinar os óbitos potencialmente preveníveis através da construção de um modelo de regressão logística. O modelo permitirá estimar as probabilidades de morte dos indivíduos, para posteriormente serem efectuadas auditorias aos óbitos com Probabilidade de ocorrência inferior a 10%.

Utilizou-se a base de dados dos Grupos de Diagnóstico Homogéneos (GDH) do Hospital Fernando Fonseca, relativa ao ano de 2012. Os Grupos de Diagnósticos Homogéneos (GDH) fazem parte de um sistema de classificação, criado nos anos 70 nos Estados Unidos, que procura definir os episódios hospitalares baseando-se nos diagnósticos e serviços prestados ao doente com o propósito de melhorar o financiamento dos hospitais ([3]). São usados em Portugal desde a divulgação da circular normativa 1/89, do Gabinete do Secretário de Estado da Saúde. Da base de 2012 foram excluídos os seguintes

casos: episódios de ambulatório, doentes saídos das unidades de cuidados intensivos, doentes saídos dos serviços do Departamento de Pediatria, doentes saídos da Obstetrícia e Urgência Obstétrica, doentes com GDH de doenças mentais, óbitos ocorridos na urgência com internamento inferior a 36 horas e episódios com GDH de caso social. Deste modo, a amostra ficou constituída por 23659 indivíduos.

Para o modelo foram consideradas como possíveis preditores, a idade, sexo, presença/ausência em UCI ao longo do internamento (UCI), GDH sem complicações/com complicações/com complicações major (Complicações), duração do internamento (em dias), índice de Case mix de episódio e peso relativo do episódio.

A modelação foi feita tal como preconizada em [5], tendo sido realizada uma minuciosa análise de resíduos. Através do teste de Hosmer e Lemeshow foi possível concluir que o modelo tem um bom ajustamento aos dados (valor $p=0,89$). O modelo também revelou uma excelente capacidade discriminativa, com uma área abaixo da curva ROC igual a 0,90. Este modelo foi validado internamente por bootstrap e será alvo de uma validação externa com os dados do primeiro semestre de 2013.

Podemos concluir que uma maior idade, ter estado em UCI, ser Homem e ter GDH com complicações são factores de risco para a ocorrência de óbito. O efeito da idade depende do sexo e do tipo de complicações, que o efeito das complicações depende da idade e de ter ou não estado na UCI e o efeito do sexo apenas depende da idade.

Com este modelo conseguiremos estimar as probabilidades de morte com alto grau de confiança para deste modo proceder à auditoria dos casos com probabilidade inferior a 10%, contribuindo para o processo de melhoria contínua do serviço.

Uma vez que existem estudos que apontam para 840-40000 mortes potencialmente preveníveis por ano em Inglaterra ([4]) e outros que afirma a que 4,1% das mortes hospitalares possam resultar de algum evento adverso prevenível ([7]), e dado que se trata de tentar evitar a morte de seres humanos, é de extrema importância o desenvolvimento de ferramentas cada vez mais robustas de modo a mitigar o máximo possível a ocorrência destes eventos.

Referências

- [1] Dubois , R.W., Brook, R.H. (1988) *Preventable deaths: Who, How Often, and Why ?* , Annals of Internal Medicine 109, 582–589.
- [2] Dubois , R.W., Brook, R.H., Rogers, W.H. (1987) *Adjusted Hospital Death Rates: a potential Screen for Quality of Medical Care* , American Journal of Public Health 77, 1161–1166.
- [3] Fetter, R.B., Freeman, J.L. (1983) *Diagnosis Related Groups: Product Line Management within Hospitals* , Academy of Management Review 11, 41–54.
- [4] Hogan, H.; Healey, F., Neale, G.; Thomson, R., Vincent, C, Black, N. (2012) *reventable deaths due to problems in care in English acute hospitals : a retrospective case record review study* , BMJ Quality and Safety 21, 737–745.
- [5] Hosmer, D.W., Lemeshow, S. L. (2004) *Applied Logistic Regression*, John Wiley & Sons.
- [6] Travassos, C., Noronha, J.C., Martins, M. (1999) *Mortalidade Hospitalar como indicador de qualidade: uma revisão* , Ciência & Saúde Colectiva 4, 367–381.
- [7] Zegers, M., de Bruijne, M. C, Wagner, C., Hoonhout, L.H.F., Waaijman, R., *et. al* (2009) *Adverse events and potentially preventable deaths in Dutch hospitals: results of a retrospective patient record review study* , Quality & Safety in Health Care 18, 297–302.

Sobreposição em Misturas de Regressões Lineares

(Poster)

Susana FARIA – *Departamento de Matemática e Aplicações. CMAT, Centro de Matemática. Universidade do Minho, Portugal*

sfaria@math.uminho.pt

Gilda Soromenho – *Instituto de Educação, Universidade de Lisboa*

Resumo

Actualmente temos assistido a um crescente interesse pelos modelos de mistura finita uma vez que se têm revelado ser o modelo mais adequado em diversos campos de aplicação. Estes modelos são muito utilizados para modelar a heterogeneidade não observada populacional.

Para analisar a heterogeneidade dos dados, várias medidas de entropia têm sido amplamente estudadas. Em modelos de mistura, critérios de entropia podem ser aplicados para estudar a sobreposição das componentes da mistura (Leisch, 2004).

Neste trabalho começamos por estudar uma medida de entropia em modelos de mistura de regressões lineares de duas componentes baseada na distância simétrica de Kullback-Leibler (Frühwirth-Schnatter, 2006). Esta medida permite investigar a sobreposição das componentes de um modelo de mistura.

Usando este critério, um estudo de simulação foi realizado para investigar o desempenho do algoritmo EM na determinação das estimativas de máxima verosimilhança dos parâmetros de modelos de mistura de regressões lineares de duas componentes. Neste estudo, diferentes configurações das verdadeiras retas de regressão componentes da mistura foram analisadas para diferentes níveis de sobreposição.

A principal conclusão deste trabalho é que a precisão das estimativas dos parâmetros de modelos de mistura de regressões lineares de duas componentes não depende da configuração das verdadeiras retas de regressão componentes da mistura mas do grau de sobreposição destas componentes.

Agradecimentos

Este trabalho é financiado por Fundos FEDER através do Programa Operacional Factores de Competitividade - COMPETE e por Fundos Nacionais através da FCT - Fundação para a Ciência e a Tecnologia no âmbito do projecto PEst-C/MAT/UI0013/2011

Referências

- [1] Cavanaugh, J. E., (1999). *A large sample model selection criterion based on Kullback symmetric divergence*, Statistics and Probability Letters 44, 333–344.
- [2] Frühwirth-Schnatter, S., (2006) *Finite Mixture and Markov Switching Models*, Springer, Heidelberg.
- [3] Hafidi, B. and Mkhadri, A., (2010) *The Kullback information criterion for mixture regression models*, Statistics and Probability Letters, 80, 807–815.
- [4] Leisch, F., (2004). *Exploring the Structure of Mixture Model Components*, In J. Antoch (ed.) *Compstat 2004, Proceedings in Computational Statistics*, 1405–1412. Physika Verlag, Heidelberg, Germany.
- [5] Naik, P.A., Shi, P., Tsai, C.L., (2007) *Extending the akaike information criterion to mixture regression models*, J. Amer. Statist. Assoc. 102, 244–254.
- [6] Park, S., Rao, M., Shin, D.W., (2012) *On cumulative residual Kullback-Leibler information*, Statistics and Probability Letters 82, 2025–2032.

Estimação não paramétrica do coeficiente de dependência de cauda: uma abordagem heurística

(Poster)

Marta FERREIRA – *Centro de Matemática da Universidade do Minho/DMA*
msferreira@math.uminho.pt

Sérgio Silva – *Universidade do Minho*

Resumo

O coeficiente de dependência na cauda, usualmente denotado TDC, é uma medida de dependência extremal bivariada, descrevendo, grosso modo, a probabilidade de uma variável exceder um valor elevado, dado que a outra também excede. Este conceito é de grande utilidade em aplicações na área das finanças (e.g., Poon et al. 2003, Frahm et. al 2006 e Schmidt and Stadtmüller 2006), na área ambiental (e.g., Capéraà et. al, 1997 e Aghakouchak et. al, 2010), entre outras, pelo que, a sua estimação assume um papel importante. Este trabalho tem por objetivo a estimação não paramétrica desta medida. Mais precisamente, e através de simulação, estuda-se um procedimento de estimação heurístico proposto em Frahm et. al (2006) para um vasto leque de casos.

Referências

- [1] Aghakouchak, A., Ciach, G., Habib, E. (2010) *Estimation of tail dependence coefficient in rainfall accumulation fields*, *Advances in Water Resources* 33, 1142–1149.
- [2] Capéraà, P., Fougères, A.L., Genest, C. (1997) *A nonparametric estimation procedure for bivariate extreme value copulas*, *Biometrika* 84, 567–577.
- [3] Frahm, G., Junker, M., Schmidt, R. (2005) *Estimating the tail-dependence coefficient: properties and pitfalls*, *Insurance Math. Econom.* 37(1), 80–100.
- [4] Poon, S.H., Rockinger, M., Tawn, J. (2003) *Modelling extreme-value dependence in international stock markets*, *Statist. Sinica* 13, 929–953.
- [5] Schmidt, R., Stadtmüller, U. (2006) *Nonparametric estimation of tail dependence*, *Scand. J. Statist.* 33, 307–335.

Aninhamento generalizado e aninhamento ortogonal em modelos mistos

(Poster)

Sandra Maria Bargão Saraiva FERREIRA – *Departamento de Matemática e Centro de Matemática, Universidade da Beira Interior*

sandraf@ubi.pt

Dário Ferreira – *Departamento de Matemática e Centro de Matemática, Universidade da Beira Interior*

Célia Nunes – *Departamento de Matemática e Centro de Matemática, Universidade da Beira Interior*

João T. Mexia – *Departamento de Matemática e CMA, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Monte da Caparica*

Resumo

Neste trabalho vamos considerar modelos do tipo $Y = \sum_{i=1}^w X_i \beta_i$, com β_0 fixo e os β_1, \dots, β_w aleatórios, independentes, com vectores médios nulos e matrizes de covariâncias $\sigma_1^2 I_{c_1}, \dots, \sigma_w^2 I_{c_w}$. Diz-se que o aninhamento é generalizado se tivermos $\Omega_i = R(X_i), i = 0, \dots, w$, com $\Omega_0 \subset \dots \subset \Omega_w = R^n$, onde \subset indica inclusão estrita e $R(A)$ representa o espaço imagem da matriz A . Por outro lado, o aninhamento será ortogonal se a matriz de projecção ortogonal sobre Ω_0 e as matrizes $M_i = X_i X_i', i = 1, \dots, w$, comutarem. Pretendemos com este trabalho apresentar detalhadamente as condições que originam estes dois tipos de aninhamento.

Referências

- [1] Ferreira, S.S. (2006) *Inferência para modelos ortogonais com segregação*, PhD Thesis, UBI.
- [2] Mexia, J.T. . *Best linear unbiased estimates, duality of F tests and the Scheffé multiple comparison method in the presence of controlled heteroscedasticity*. Comput. Statist. Data Anal. 10, **3**, 271–281, 1990.
- [3] Seely, J. (1971). *Quadratic subspaces and completeness*. The Annals of Mathematical Statistics. 42, 2, 710–721.

[4] Zmysłony, R. . *A characterization of best linear unbiased estimators in the general linear model*, Mathematical Statistics and Probability Theory, Proc. Sixth Internat. Conf., Wisla, Lecture Notes in Statist., **2**, Springer, New York-Berlin, 365–373, 1978.

Metodologia STATIS aplicada à monitorização de processos

(Poster)

Adelaide Maria FIGUEIREDO – *Faculdade de Economia e LIAAD-INESC TEC, Universidade do Porto*

adelaide@fep.up.pt

Fernanda Otília Figueiredo – *Faculdade de Economia, Universidade do Porto, e CEAUL*

Resumo

A avaliação da qualidade global de um produto ou de um serviço em situações reais depende de mais do que uma característica de qualidade, pelo que o desenvolvimento de cartas de controlo para dados multivariados é crucial. A utilização de cartas de controlo univariadas para monitorizar cada uma das características de qualidade separadamente é obviamente inapropriada, sendo usual proceder-se à monitorização de dados multivariados através de uma carta de controlo univariada cuja estatística permite monitorizar simultaneamente o vector médio do processo e a matriz de covariâncias, ou considerando simultaneamente duas cartas de controlo univariadas, uma para monitorizar o vector médio do processo e a outra para monitorizar a matriz de covariâncias. A maior parte destes esquemas de controlo são baseados na estatística de Hotelling e são implementados sob a hipótese de dados normais multivariados. Atendendo a que temos várias variáveis em jogo para definir a qualidade de um produto/serviço, sempre que um esquema/carta de controlo emite um sinal, é de todo importante tentar averiguar quais as variáveis do processo responsáveis pela emissão de sinal, e também identificar quais os parâmetros do processo que sofreram alterações. Neste trabalho iremos aplicar a metodologia STATIS (Structuration des Tableaux à Trois Indices de la Statistique) na monitorização de dados multivariados, e averiguar qual o seu contributo para uma melhor identificação das causas associadas à emissão de um sinal de fora de controlo. O STATIS (Lavit et al., 1994) é um método de análise exploratória a três dimensões baseado na Análise em Componentes Principais, que nos permite analisar vários quadros de indivíduos descritos por variáveis

quantitativas avaliadas em diferentes instantes de tempo ou circunstâncias. Em Gourvénec et al. (2005) e Figueiredo et al. (2012) são apresentadas duas aplicações interessantes da metodologia STATIS a conjuntos de dados reais.

Referências

- [1] Figueiredo, A., Figueiredo, F., Monteiro, N., Straume, O. (2012) *Restructuring in privatised firms: a Statis approach*, Structural Change and Economic Dynamics 23, 108-116.
- [2] Gourvénec S., Stanimirova I., Saby C.A. (2005) *Monitoring batch process with the STATIS approach*, J. Chemometrics 19, 288-300.
- [3] Lavit, C., Escoufier, Y., Sabatier, R., Traissac, P. (1994) *The ACT (Statis method)*, Computational Statistics & Data Analysis 18, 97-119.

Distúrbios Respiratórios do Sono: Estudo de Prevalência e Estimação de Preditores com recurso a Modelo de Regressão Logística Categorial

(Poster)

João Paulo FIGUEIREDO – *Escola Superior de Tecnologia da Saúde de Coimbra - Departamento das Ciências Complementares*

jpfigueiredo@estescoimbra.pt

Paulo Caseiro – *Escola Superior de Tecnologia da Saúde de Coimbra - Departamento de Cardiopneumologia*

Cláudia Gil – *Escola Superior de Tecnologia da Saúde de Coimbra - Técnica Superior de Cardiopneumologia*

Jorge Conde – *Escola Superior de Tecnologia da Saúde de Coimbra - Departamento de Cardiopneumologia*

Resumo

Os Distúrbios Respiratórios do Sono (DRS) são cada vez mais uma das preocupações da atualidade devido à sua maior expressão na sociedade e que conduzem a uma degradação da qualidade de vida do indivíduo. Os principais sinais e sintomas associados a DRS são a roncopatia, paragens respiratórias e Sonolência Diurna Excessiva (SDE). As consequências desses distúrbios envolvem sonolência excessiva, risco de acidentes de trabalho e viação, além de déficits cognitivos e doenças cardiovasculares. Também a Síndrome de Apneia Obstrutiva do Sono (SAOS) é a segunda perturbação do sono mais comum tratando-se de uma doença crónica, progressiva, incapacitante, com alta mortalidade e morbidade cardiovascular[1]. A manifestação desta síndrome está associada a episódios recorrentes de obstrução parcial ou total das vias aéreas superiores (VAS) durante o sono [2][3]. Neste sentido quando falamos de SDE esta não é mais do que uma manifestação da presença de SAOS e outros sintomas associados [4][5]. Objetivos de estudo: Avaliar a prevalência de sinais e sintomas associados a DRS, na população adulta portuguesa e os determinantes que mais concorrem para os DRS. Estudo

Observacional, tipo Secional. O período de recolha de dados: Outubro de 2011 a Março de 2012. A população-alvo de estudo: residentes em Portugal Continental, de ambos os sexos. A amostragem não probabilística e propagação Geométrica. Critérios de inclusão: idade ≥ 30 anos e não ter sido diagnosticado DRS. A recolha de dados: inquérito de morbilidade. Foram aplicadas estatísticas descritivas e testes de hipóteses: Quiquadrado da Independência, Exato de Fisher e Regressão Logística (Categorial)[6][7]. Software estatístico: IBM SPSS Statistics e MedCalc.

Resultados: a amostra (426 indivíduos), 45,3% eram do sexo masculino e 54,7% eram do sexo feminino. Média de idades por sexo: homens $M=52$ anos $DP=12,74$ anos e mulheres $M=51$ anos $DP=12,46$ anos). IMC 46% dos participantes tinham excesso de peso e 15,5% eram obesos. Hábitos tabágicos: 13,1% fumavam, 58,22% ($n=248$) consumiam álcool e 14,3% recorre/recorreu a medicamentos para dormir. A prevalência de DRS foi de 25,59%. Com aplicação do modelo de regressão logística constatou-se:IMC apresentou um efeito sobre a variável dependente (DRS) (Wald=13,625($gl=1$); $p\text{-value}<0,001$ (OR=2,629). Com base no rácio de oportunidade, concluímos que a probabilidade de ter DRS foi exponencialmente superior em pessoas que revelaram um IMC alterado face às que apresentavam peso normal. Os preditores idade e sexo não apresentaram um efeito significativo sobre o Logit da probabilidade de ter DRS ($p\text{-value}>0,05$). Comportamentos de risco que predisseram a presença de DRS foram: fumar (Wald=3,942($gl=1$); $p\text{-value}<0,05$; OR=0,455) e consumo de álcool (Wald=6,965($gl=1$); $p\text{-value}<0,01$; OR=1,898). Segundo o rácio de oportunidade, a probabilidade de ter DRS foi significativamente superior em indivíduos consumidores de álcool face aos não consumidores. O preditor turnos não revelou um efeito significativo ($p\geq 0,05$). Sinais noturnos constatámos que: roncopatia (Wald=9,671($gl=1$); $p\text{-value}<0,01$; OR=2,499) e apneias (Wald=7,518($gl=1$); $p\text{-value}<0,01$; OR=2,163) foram os sinais que revelaram um efeito significativo à exceção da variável nictúria ($p\text{-value}>0,05$). Quanto às variáveis clínicas, a HTA foi o único preditor a apresentar um efeito sobre o Logit da probabilidade de ter DRS (Wald=10,171($gl=1$); $p\text{-value}<0,01$; OR=2,126). Veio a concluir-se que a probabilidade de ter DRS foi exponencialmente superior em pessoas que clinicamente eram hipertensos face às normotensas. Dados epidemiológicos revelam que os DRS apresentam alta incidência e prevalência, podendo

acometer qualquer faixa etária. Contudo, o pico de incidência está situado entre os 30 e os 60 anos de idade, com maior acometimento de homens obesos nesta faixa etária.

Referências

- [1] Burger, R., Di Ninno, Q., Caixeta, E. (2004) *A relação entre apnéia do sono, ronco e respiração oral*, Revista de Atualização Científica em Fonoaudiologia e Educação. 3, 266–271.
- [2] *Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research*. The Report of an American Academy of Sleep Medicine Task Force. (1999) *Sleep*. 22(5), 667–89.
- [3] Penzel, T., Moller, M., Becker, H., Knaack, L. Peter, J (2001) *Effect of sleep position and sleep stage on the collapsibility of the upper airways in patients with sleep apnea* *Sleep*. (4), 90–95.
- [4] Adão, P. (2007) *Síndrome da Apneia Obstrutiva do Sono: Fisiopatologia, Epidemiologia, Consequências, Diagnóstico e Tratamento*, Arquivos Médicos. 21(5/6), 159–73
- [5] Teixeira, F. (2006) *Distúrbios respiratórios obstrutivos do sono: síndrome de apneia-hipopneia obstrutiva do sono e síndrome de resistência das vias aéreas superiores* *Revista Portuguesa de Clínica Geral*. 613–623.
- [6] Tabachnick, BG., Fidell, LS. (2007) *Using Multivariate Statistics*, Fifth Edition, Pearson International Edition.
- [7] Rosner, B. (2006) *Fundamentals of Biostatistics*, Sixth Edition, International Student Edition, Thomson Brooks Cole.

Avaliação da Qualidade do Ar nas Cidades de Aveiro, Lisboa, Portimão e Porto:

Análise Retrospetiva

(Poster)

João Paulo FIGUEIREDO – *Escola Superior de Tecnologia da Saúde de Coimbra - Departamento de Ciências Complementares*

jpfigueiredo@estescoimbra.pt

Marta Pinto Vasconcelos – *Escola Superior de Tecnologia da Saúde de Coimbra - Departamento de Saúde Ambiental*

Inês Ribeiro – *Técnica Superior de Saúde Ambiental*

Ana Ferreira – *Escola Superior de Tecnologia da Saúde de Coimbra - Departamento de Saúde Ambiental*

Resumo

A qualidade do ar (QA) ao longo das últimas décadas tem sido objeto de estudo e de medidas legislativas face ao crescimento da atividade industrial e da população. Uma vez que as consequências da degradação da QA causam efeitos sobre a saúde humana e sobre os ecossistemas é necessário a aplicação de medidas de controlo da qualidade da mesma. A rede nacional de monitorização da qualidade do ar aplicada em todo o território português permite analisar as concentrações dos poluentes atmosféricos emitidos pelas várias fontes emissoras em Portugal [1][2][3]. O presente estudo visou analisar a evolução da QA nas cidades de Aveiro, Lisboa, Portimão e Porto no período de 2005 a 2011. Tipo de estudo Observacional e coorte retrospectiva. A recolha analítica teve como suporte o sítio da Internet QualAR (Agência Portuguesa do Ambiente). Os poluentes em estudo foram a PM10, o NO₂, o CO, e o O₃. As estações selecionadas para este estudo de evolução da qualidade do ar tiveram por base o mesmo tipo de ambiente e o mesmo foco de poluição. Todas as estações avaliadas caracterizam-se por estações de ambiente urbano e zonas de tráfego automóvel. Para a descrição dos parâmetros aplicaram-se medidas descritivas simples. Quanto aos testes de hipóteses: teste t-Student para a média populacional, General Linear Model: Repeated Measures e ANOVA a 1 fator [4][5]. A interpretação dos testes estatísticos foi realizada com base num nível

de confiança de 95% para um erro máximo até 5%. O tratamento estatístico dos dados adquiridos realizou-se com recurso ao software IBM SPSS Statistic versão 21.0. Quanto à distribuição de concentração de CO nas quatro estações em estudo em 2005 até 2011 num período de 8h a média de concentração desse poluente não ultrapassou o valor limite de proteção da saúde humana, (10 000m³), de forma significativa ($p < 0,0001$). Quanto à concentração média do poluente PM₁₀ (40m³) de 2005 a 2011 em cada estação de tráfego avaliado verificou-se que, no ano de 2007 e 2011 na estação de Aveiro a concentração ultrapassou o valor limite anual para proteção da saúde humana. A estação de Lisboa revelou o maior número de excedências face ao valor limite em todos os anos de estudo. Na estação de Portimão os anos de 2005, 2007 e 2011 apresentaram excedências ao valor de referência, sendo os anos de 2007 e 2011 que apresentaram concentrações mais elevadas de PM₁₀. Na estação do Porto, o ano de 2005 foi o único que apresentou excedência sobre o valor limite mas este foi muito próximo do valor de referência ($p < 0,001$). No que diz respeito ao poluente NO₂ a estação de Lisboa registou em todos os anos uma concentração acima do valor limite ($p < 0,001$), sendo estas concentrações muito elevadas face ao valor de referência (40m³). A estação de Aveiro, dos vários anos em estudo, não apresentou qualquer excedência ao valor limite e a sua concentração média foi relativamente baixa face ao valor de risco. Portimão também não apresentou uma concentração média anual deste poluente acima do valor limite/risco para a saúde humana. A estação localizada na cidade do Porto apresentou excedências ao valor limite em todo os anos em estudo, sendo maioritariamente estes valores registados um pouco acima do valor de referência. Verificou-se ainda, que os grandes centros urbanos, Lisboa e Porto foram os que apresentaram concentrações bastante superiores do valor limite anual de concentração de NO₂. Em relação ao poluente O₃ (valor limite: 120m³), foi necessário analisar os dados por períodos de 8h. A estação de monitorização do Porto foi a única a avaliar a concentração deste poluente. Considerando todos os períodos de 8H diários o ano de 2010 foi o mais crítico a nível de concentração média, uma vez que registou as concentrações mais elevadas apesar de serem inferiores ao valor de risco para a saúde humana. Verificou-se ainda que o período de 8H das 8H01 às 16H foi o que apresentou as maiores concentrações deste poluente.

Em geral, entre estações ao nível da concentração anual dos poluentes,

verificou-se que a estação de Lisboa foi a que apresentou maiores concentrações anuais dos poluentes, uma vez que esta estação está localizada numa das principais vias de tráfego de Lisboa, com congestionamentos frequentes e com um grande fluxo de veículos. As estações que registaram menor concentrações anuais foram as estações de Aveiro e Portimão, uma vez que são cidades com uma menor densidade populacional e de menor atividade antropogénica.

Referências

- [1] Lameiras, H., Póvoas, F. (2003) *Qualidade do Ar – Proteger a Saúde Humana.*, Comissão de Coordenação e Desenvolvimento Regional do Centro. Coimbra.
- [2] Agência Portuguesa do Ambiente (2008) *Evolução da qualidade do ar em Portugal entre 2001 e 2005 .*, Relatório. Amadora.
- [3] Baptista, C.E. (2008) Análise comparativa da qualidade do ar em Portugal, Dissertação. Universidade de Aveiro (Aveiro).
- [4] Cunha, G., Eiras, M., Teixeira, N. (2011) *Bioestatística e Qualidade na Saúde.* Lidel: Lisboa.
- [5] Maroco, J., Bispo, R. (2005) *Estatística aplicada às Ciências Sociais e Humanas.* Manuais Universitários, Climepsi Editores: Lisboa.

Porque duram tanto tempo algumas dissertações de Mestrado?

(Poster)

Rita FREITAS – *Mestrado em Modelação Estatística e Análise de Dados, Universidade de Évora*

ritabf8@gmail.com

Paulo Infante – *CIMA-UE e DMAT, ECT da Universidade de Évora*

Gonçalo Jacinto – *CIMA-UE e DMAT, ECT da Universidade de Évora*

Fernanda Figueiredo – *CEAUL e Faculdade de Economia da Universidade do Porto*

João Dias – *Serviços Académicos da Universidade de Évora*

Resumo

A informação disponível e gerada pelos Serviços Académicos de uma Universidade é um instrumento de elevada relevância no apoio a diversos serviços da universidade, bem como à Reitoria e ao seu gabinete de promoção da qualidade, com o objetivo de uma melhoria contínua de modo a atingir níveis de excelência. De entre as muitas variáveis que podem ser analisadas com esse objectivo, o tempo de duração dos cursos assume uma importância bastante relevante, pois tem impacto não só na projeção da Universidade para o exterior, mas também no próprio orçamento das Universidades. Para além disso, uma análise deste tipo de dados pode ser usada na reformulação de alguns cursos.

No caso particular dos Mestrados, uma variável fulcral que pode adiar a sua conclusão no tempo previsto é o tempo de duração das respetivas dissertações. Neste trabalho, com base dados recolhidos nos Serviços Académicos da Universidade de Évora, iremos analisar e modelar estatisticamente o tempo de duração de uma dissertação de Mestrado.

Começamos por modelar esta variável recorrendo à análise de sobrevivência [4], considerando preditores sociodemográficos e de cariz académico: idade de ingresso, sexo, tipologia e distrito da área de residência, naturalidade, nacionalidade, habilitações literárias dos pais, área de formação inicial, curso, área científica predominante do curso, modo de ingresso, tempo de conclusão da parte curricular e número de ECTS em creditações concebidas.

Com base num modelo de regressão logística [3], procuramos também traçar o perfil mais provável do aluno que necessita de pedir semestres adicionais para terminar a sua dissertação. Neste modelo, considera-se também a nota da dissertação, a média de conclusão do mestrado e o ano de conclusão como possíveis variáveis explicativas.

Depois da entrega da dissertação nos serviços académicos, outros trâmites administrativos podem atrasar ou acelerar o tempo até discussão pública e, conseqüentemente, influenciar significativamente o tempo de duração do curso. Neste trabalho serão também analisados alguns tempos parciais, como sejam o tempo de formação do júri, tempo de realização da primeira reunião do júri e tempo desde a primeira reunião até realização de provas públicas. Para a análise destes tempos parciais recorreremos fundamentalmente ao controlo estatístico de qualidade, medindo índices de capacidade do serviço e elaborando cartas de controlo de qualidade retrospectivas, mas que poderão ficar desde logo implementadas no serviço numa perspetiva de controlo em tempo real.

A implementação de cartas de controlo de qualidade, além de possibilitar a avaliação e controlo do desempenho de um processo ao longo do tempo, permite a distinção entre a variação natural do processo e outras formas de variação, que levam à identificação das causas responsáveis, possibilitando a implementação de ações que reduzem a variabilidade e/ou o ajustamento do serviço para valores alvo previamente especificados, melhorando a qualidade. Como alguns destes tempos parciais são muito assimétricos e com caudas pesadas, não é recomendável a aplicação de cartas baseadas na distribuição normal, apesar da sua simplicidade e popularidade, pelo que aplicaremos cartas de controlo desenvolvidas em [2], as quais são baseadas na distribuição normal assimétrica ([1]).

Referências

- [1] Azzalini, A. (2005) *The Skew-Normal Distribution and Related Multivariate Families*, Scandinavian Journal of Statistics 32, 159–188.
- [2] Figueiredo, F., Ivette Gomes, M. (2013) *The Skew-Normal Distribution in SPC*, REVSTAT - Statistical Journal 11, 83–104.
- [3] Hosmer, D.W.; Lemeshow, S. L. (2004) *Applied Logistic Regression*, John Wiley & Sons.
- [4] Rocha, C.; Papoila, A. L. (2009) *Análise de Sobrevivência*, Edições SPE.

Evolução da Tuberculose Pulmonar em Portugal Continental, 2000 a 2010

(Poster)

Dulce GOMES – *CIMA, Universidade de Évora*

dmog@uevora.pt

Ana L. Brás – *CIMA, Universidade de Évora*

Patrícia Filipe – *CIMA, Universidade de Évora*

Carla Nunes – *CMDT, Escola Nacional de Saúde Pública, Universidade Nova de Lisboa*

Resumo

A seguir ao VIH/AIDS, a Tuberculose é a segunda doença infecciosa que mais mata a nível mundial. Trata-se de um importante problema de saúde pública, apesar do número estimado de pessoas no mundo que desenvolve a doença ter vindo a decrescer lentamente de ano para ano [1]. A Tuberculose Pulmonar (PTB) representa a maioria dos casos notificados, sendo ainda a principal forma de transmissão da doença, pelo que o conhecimento da sua evolução temporal é essencial na monitorização da Tuberculose. Actualmente, existem vários métodos de análise temporal que permitem identificar variações ao longo do tempo e realizar inferências válidas, permitindo assim uma melhor compreensão de doenças importantes em saúde pública e de potenciais factores de risco [2]. Este trabalho pretende analisar a evolução temporal das taxas de incidência mensais de PTB em Portugal Continental e em áreas geográficas de alto/baixo risco de desenvolver doença, identificadas em estudos prévios. A importância de características demográficas, como o sexo e a classe etária, é também considerada.

Material e Métodos: Os dados analisados correspondem às taxas de incidência mensal de PTB em Portugal Continental, entre 2000 e 2010. Foram utilizadas duas fontes oficiais de dados: o número de casos de PTB notificados em Portugal Continental neste período, disponibilizado pelo Programa Nacional de Controlo de Tuberculose; e a população em risco, estimada pelo Instituto Nacional de Estatística (INE). As taxas de incidência

mensais foram analisadas ao nível do Continente e das áreas geográficas de elevado risco (zonas metropolitanas do Porto e de Lisboa) e baixo risco (restante país) de PTB já identificadas em estudos prévios [3]. Os dados foram ainda divididos segundo características demográficas específicas, como o sexo e o grupo etário (0-4, 5-14, 15-24, 25-34, 35-44, 45-54, 55-64 e >65). O estudo da tendência e sazonalidade foi realizado através de um método de decomposição (STL) da série temporal, e a modelação e previsão teve por base o ajustamento de modelos Integrados Auto-Regressivos e Médias Móveis Sazonais (SARIMA). Toda a análise foi realizada com o *software* “R”.

Resultados e discussão: Ao longo de 132 meses foram notificados 28615 casos de PTB em Portugal Continental. A taxa de incidência média por mês foi de 2,17 casos por 100 000 habitantes. Observou-se em geral uma tendência decrescente e constante da taxa de incidência mensal de PTB, e verificou-se um comportamento similar tanto em homens como em mulheres. Ambas as áreas metropolitanas do Porto e de Lisboa apresentaram uma tendência decrescente, apesar do declínio ser mais marcado no Porto. Nas classes etárias acima dos 15 anos de idade a tendência foi decrescente, contudo crianças dos 0 aos 14 anos de idade apresentaram uma tendência aproximadamente constante ao longo do período em estudo. Estes resultados são consistentes com um declínio da endemia em Portugal, ao apontarem para uma diminuição e/ou estabilização da transmissão recente de PTB em crianças e concomitante diminuição de casos notificados em adultos. Os modelos SARIMA com melhor ajustamento às séries do Continente e das áreas metropolitanas do Porto e Lisboa prevêm a continuação do declínio das taxas de incidência e revelam a presença de sazonalidade. A decomposição STL permitiu observar o perfil da componente sazonal ao longo do ano, verificando-se um acentuado decréscimo da taxa de incidência em Novembro e Dezembro.

Conclusões: Ao longo de 11 anos, Portugal Continental apresentou um declínio no número de notificações de PTB. A utilização de modelos matemáticos permite prever a continuação desta diminuição, caso se verifique o pressuposto de que as condições existentes no período em estudo se mantêm no futuro. Quanto à sazonalidade encontrada, vários factores

poderão estar implicados na diminuição de notificações no Inverno aqui evidenciada. Uma possibilidade é a presença de outras infecções respiratórias no Inverno associada aos sintomas inespecíficos da PTB, os quais podem dificultar o diagnóstico expedito e correcto de Tuberculose. Mais estudos serão necessários para determinar quais os factores explicativos da evolução temporal de PTB em Portugal.

Agradecimentos: Trabalho realizado no âmbito do projecto de investigação PTDC/SAU-SAP/116950/2010 e financiado pela FCT/MCTES.

Referências

- [1] Global Tuberculosis Report 2012. [cited 2013 Jul 2]. (<http://www.who.int/mediacentre/factsheets/fs104/en/index.html>)
- [2] Zeger, SL, Irizarry, R., Peng, RD. (2012). *On Time Series Analysis of Public Health and Biomedical Data*, Annu. Rev. Public Health; 27(1):57-79.
- [3] Nunes, C., Briz, T., Gomes, D., Filipe, P.A. (2011). *Pulmonary Tuberculosis and HIV/AIDS: joint space-time clustering under an epidemiological perspective*, Proceedings of the Spatial Data Methods for Environmental and Ecological Processes, 2nd Edition. Barbara Cafarelli (Ed.), Foggia e Gargano.

Análise da satisfação da procura no ensino superior em Portugal: um índice alternativo

(Poster)

A. Manuela GONÇALVES – *DMA - Departamento de Matemática e Aplicações. CMAT - Centro de Matemática. Universidade do Minho, Portugal*

mneves@math.uminho.pt

Raquel Oliveira – *CMAT - Centro de Matemática. Universidade do Minho, Portugal*

Rosa M. Vasconcelos – *Departamento de Engenharia Têxtil. 2C2T - Centro de Ciência e Tecnologia Têxtil. Universidade do Minho, Portugal*

Resumo

Neste estudo apresenta-se um índice de satisfação da procura alternativo aos cursos do ensino superior com o objectivo de se efectuar uma análise comparativa com o índice de satisfação da procura, apresentado pelas entidades reguladoras do ensino superior português. Para o efeito são considerados os dados disponíveis referentes ao concurso nacional, 1 fase, dos cursos de engenharia para o ano de 2010. A aplicação de metodologias multivariadas permitiram identificar e detectar *clusters* que caracterizam o índice de satisfação da procura agora proposto, permitindo, desta forma, proceder à sua comparação com a caracterização e identificação de *clusters* já realizada, em trabalho anterior, para o índice de satisfação da procura.

Referências

[1] <http://www.dges.mctes.pt/DGES/pt/Estudantes/Acesso> (acedido em 08 de Março 2013).

[2] <http://www.dados.gov.pt/PT/CatalogoDados/Dados.aspx?name=ClassNacionaldeareasdeeducacaoeformacao> (acedido em 08 de Março 2013).

[3] OECD, Organization for Economic Co-operation and Development. (2006) Reviews of National policies for education: Tertiary education in Portugal. Examiner's Report. Available at <http://www.dges.mctes.pt/NR/rdonlyres/8B016D34-DAAB-4B50-ADBB-25AE105AEE88/2564/Backgroundreport.pdf> (acedido em 08 de Março 2013).

[4] Oliveira, R., Gonçalves, A.M., Vasconcelos, R.M. (2013) *An empirical study on the index of satisfaction of student allocation in the Portuguese undergraduate engineering courses*, Proceedings of 15th ASMDA 2013 International Conference, Barcelona, Spain, 25-28 June.

Propriedade de Taylor e curtose em processos auto-regressivos

(Poster)

Esmeralda GONÇALVES – *Departamento de Matemática, Universidade de Coimbra, e CMUC*

Cristina M. Martins – *Departamento de Matemática, Universidade de Coimbra*

Nazaré Mendes-Lopes – *Departamento de Matemática, Universidade de Coimbra, e CMUC*

Resumo

O efeito de Taylor é uma característica presente em séries temporais de natureza diversa. Este facto estilizado foi, pela primeira vez, detetado por Taylor (1986) ao analisar os retornos de algumas séries financeiras. Taylor constatou que as autocorrelações das observações em módulo eram sistematicamente superiores às correspondentes autocorrelações das observações ao quadrado. A propriedade teórica correspondente denomina-se propriedade de Taylor. A presença desta propriedade numa determinada classe de modelos para séries temporais poderá contribuir para a seleção de um modelo mais adequado para descrever a dinâmica da série de interesse. No entanto, este estudo exige o conhecimento de momentos de X de ordem superior a 2 o que o torna teoricamente elaborado. Esta propriedade tem sido sobretudo analisada em modelos com características claras de leptocurtose como os condicionalmente heteroscedásticos, conforme pode ser encontrado nos trabalhos de He et Terasvirta (1999), Gonçalves, Leite e Mendes-Lopes (2009), Haas (2009), dedicados ao estudo desta propriedade na classe geral dos modelos GTARCH, tendo os resultados obtidos revelado uma relação forte entre a presença desta propriedade e valores elevados da curtose do modelo. Esta leitura é sobretudo coerente com a leptocurtose assinalada em séries temporais, em particular de natureza financeira como as que Taylor analisou, em que o referido facto estilizado está presente. Tanto quanto é do nosso conhecimento, não há qualquer informação sobre a presença desta propriedade nos modelos lineares o que, tendo em conta a importância destes na modelação geral de dados temporais, nos pa-

rece necessário ultrapassar. Com o objetivo de dar uma contribuição a um tal estudo, vamos neste trabalho avaliar a presença da propriedade de Taylor em processos autorregressivos não-negativos de ordem 1. A relação entre a validade desta propriedade e a curtose do modelo será analisada no caso geral e relacionada com a curtose do processo de erro. Concluiremos com aplicações deste estudo, e conseqüente discussão, a modelos autorregressivos associados a processos de erro com curtoses significativamente diferentes, sendo notória a importância deste parâmetro na presença no modelo da dita propriedade de Taylor.

Referências

- [1] Gonçalves, E., Leite, J., Mendes-Lopes, N. (2009) *A mathematical approach to detect the Taylor property in TARARCH processes*, Statistics and Probability Letters. 72, 602–610.
- [2] Haas, M. (2009) *Persistence in volatility, conditional kurtosis, and the Taylor property in absolute value GARCH processes*, Statistics and Probability Letters. 79, 1674–1683.
- [3] He, C., Teräsvirta, T. (1999) *Properties of moments of a family of GARCH processes*, Journal of Econometrics. 92, 173–192.
- [4] Taylor, S.J. (1986) *Modelling Financial Time Series*, John Wiley & Sons.

Seleção de variáveis na presença de valores omissos: uma aplicação na modelação do Índice de Massa Corporal (IMC) nos imigrantes africanos e brasileiros residentes em Lisboa e Setúbal *

(Poster)

Beatriz GOULÃO – *FCUL*

beatriz.goulao@gmail.com

Patrícia de Zea Bermudez – *FCUL e CEAUL* ¹

Valeska Andreozzi – *CEAUL* ¹

Resumo

Os dados omissos são comuns em estudos epidemiológicos [1]. Os métodos usados por diversos softwares para tratar este tipo de problema (por exemplo, a rejeição total dos registos com observações omissas nalguma das variáveis ? análise dos casos completos (CC)) não são satisfatórios. O procedimento habitualmente usado consiste em analisar apenas os casos completos (CC) [2]. De facto, se os valores omissos diferirem significativamente dos valores observados, então, não considerar os dados incompletos, poderá enviesar os resultados do estudo.

Existem na literatura diversas técnicas para tratar dados omissos, nomeadamente a substituição dos mesmos por um único valor (imputação simples) ou por vários (imputação múltipla). Esta investigação pretende avaliar o impacto de técnicas para tratamento de valores omissos na escolha de variáveis em modelos de regressão, cuja variável resposta é o IMC.

A amostra é formada por 1980 imigrantes brasileiros e africanos a viver em Portugal. Os dados foram recolhidos no âmbito do estudo de Saúde dos imigrantes, realizado em 2007. Este estudo pretendia investigar factores associados ao IMC. Elaboraram-se dois cenários de dados omissos:

1. tratamento de uma covariável com baixa percentagem de dados omissos;
2. simulação da existência de uma covariável com elevada percentagem de dados omissos e que está fortemente associada ao IMC.

Os resultados deste trabalho sugerem que a existência de uma baixa percentagem de dados omissos, numa variável explicativa pouco associada com a variável resposta, parece ter poucas implicações nos resultados finais, independentemente da técnica escolhida para lidar com os dados omissos. No caso de percentagens superiores de dados omissos, a análise CC é claramente inferior às técnicas de imputação.

Referências

- [1] Molenberghs G, Kenward MG. (2007) *Missing Data in Clinical Studies*, Wiley.
- [2] Rubin D. (1987) *Multiple Imputation for Nonresponse in Surveys*, Wiley.

¹ Este trabalho foi financiado pelos projectos da FCT: PEst-OE/MAT/UI0006/2011 e PTDC/MAT/118335/2010

* A recolha dos dados usados neste trabalho foi financiada pelo Alto Comissariado da Saúde. Agradecemos à equipa do IMP responsável pelo projeto, aos entrevistadores, assim como aos participantes voluntários.

O método *file-drawer* e o viés da publicação — um estudo de simulação comparativo para diferentes métodos de combinação de valores de prova p

(Poster)

Catarina GOUVEIA – DEIO, FCUL da Universidade de Lisboa

catarina.gouveia@gmail.com

M. Fátima Brilhante – DM, Universidade dos Açores and CEAUL

Fernando Sequeira – DEIO, FCUL, Universidade de Lisboa and CEAUL

Resumo

Em meta-análise os valores de prova p (p -values) desempenham um papel fulcral na combinação dos resultados de testes independentes. Todos os métodos de combinação de testes conhecidos se baseiam no facto de que os valores de prova p observados formam uma amostra de uma população uniforme padrão. Porém, quando se efetuam tais testes combinados, é provável que as amostras de valores de prova p usadas sejam enviesadas, muito por culpa do fenómeno vulgarmente conhecido como viés da publicação. Este fenómeno resulta do facto das sínteses meta-analíticas se basearem quase exclusivamente em estudos publicados, que maioritariamente produziram resultados estatisticamente significativos, i.e. apresentam valores de prova p inferiores a 0.05. Estudos não publicados, e por isso desconhecidos do público, não possuem geralmente esta propriedade "desejada", muito embora possam existir nesse grupo estudos que sejam bastante informativos.

Idealmente, as sínteses meta-analíticas devem basear-se em estudos publicados e não publicados relevantes sobre o mesmo assunto. Mas quando tal não for possível, o método *file-drawer* (estudos na gaveta) pode ser usado para lidar com o possível problema do viés da publicação. Basicamente, este método procura dar uma ideia sobre o número de estudos desconhecidos, que se fossem também considerados na análise, levariam à alteração da decisão de rejeitar a hipótese nula global H_0 para um teste combinado, para a decisão

de não rejeitar H_0 . (Note-se que a decisão de rejeitar H_0 é bastante provável quando o viés da publicação está presente.) Se este número for pequeno, então há razões para acreditar que existem estudos desconhecidos que seriam importantes considerar na síntese.

Partindo de amostras de valores de prova p de dimensão $n = 4, 5, 6, 10, 20, 30, 50, 100$, e que conduzem à rejeição de H_0 para diferentes métodos de combinação de testes, é determinado por simulação o número médio de valores de prova p adicionais necessário para mudar a decisão de rejeitar para a de não rejeitar H_0 . Para a geração das amostras consideram-se duas situações: a restrição de todos os valores de prova p serem inferiores a 0.05 e a não restrição. Com este estudo pretende-se avaliar o impacto do viés da publicação em cada método considerado.

Referências

- [1] Hartung, J., Knapp, G. and Sinha, B.K. (2008). *Statistical Meta-Analysis with Applications*. Wiley, New Jersey.
- [2] Kulinskaya, E., Morgenthaler, S., and Staudte, R. G. (2008). *Meta Analysis. A Guide to Calibrating and Combining Statistical Evidence*. Wiley, Chichester.
- [3] Pestana, D. (2011). Combining p -values. In M. Lovric (ed.) *International Encyclopedia of Statistical Science*, Springer-Verlag, pp. 1145–1147.

Aplicação do método POT a valores extremos de precipitação na Ilha da Madeira

(Poster)

Délia GOUVEIA-REIS – *Universidade da Madeira, Centro de Estatística e Aplicações da Universidade de Lisboa, Centro de Investigação de Montanha*
delia@uma.pt

Luiz Guerreiro Lopes – *Universidade da Madeira, Instituto de Ciências Agrárias e Ambientais Mediterrânicas, Centro de Investigação de Montanha*

Sandra Mendonça – *Universidade da Madeira, Centro de Estatística e Aplicações da Universidade de Lisboa*

Resumo

O estudo de valores extremos por meio da abordagem de Gumbel ou método dos blocos apresenta a desvantagem de considerar apenas um máximo por bloco, o que pode limitar esse estudo. Tal limitação é atenuada pela abordagem POT (*peaks over threshold*), uma vez que todas as observações que excedem um certo nível (*threshold*), u , são integradas na análise. É neste contexto que são aqui explorados valores diários de precipitação na Ilha da Madeira referentes ao período de 1950 a 1980. Os dados analisados, fornecidos pelo Departamento de Hidráulica e Tecnologias Energéticas do Laboratório Regional de Engenharia Civil, são provenientes de estações e postos udométricos mantidos pela antiga Junta Geral do Distrito Autónomo do Funchal, localizadas nas vertentes norte e sul da ilha. A Madeira apresenta diferentes regiões relativamente aos extremos de precipitação, condicionadas pela sua complexa orografia e pelo pronunciado desnível das suas vertentes. Assim, neste trabalho, pretende-se averiguar a influência de tais factores na escolha do limiar u e no ajustamento dos excessos de nível por modelos apropriados, tanto numa perspectiva paramétrica quanto semi-paramétrica.

Estudo sobre a certificação das empresas

(Poster)

Marisa Jesus HENRIQUES – *Departamento de Matemática, Universidade de Aveiro*

mjh@ua.pt

Nélia Silva – *Departamento de Matemática, Universidade de Aveiro e CIDMA*

Filomena Frazão – *Empresa Multidados, Aveiro*

Resumo

A certificação é um tema fundamental e atual no crescimento sustentado das empresas. Esta qualificação deixou de ser apenas uma vantagem competitiva para passar a ser um critério selecionador de empresas diferenciando-as das concorrentes. O conceito de melhoria contínua transmite para o mercado a imagem de empresas capazes de satisfazer e superar as exigências dos clientes, direcionando todos os colaboradores para esse objetivo em comum (ver Pinto e Iolanda [5]).

Neste trabalho, apresenta-se um estudo, realizado a nível nacional, sobre a certificação das empresas. Este estudo resultou de um estágio curricular, na empresa MultiDados no âmbito do Mestrado em Matemática e Aplicações. O objetivo era saber quais os setores de atividade que recorriam mais à certificação, quais foram as dificuldades sentidas no processo de certificação e posteriormente quais as vantagens e desvantagem sentidas após a certificação. Foram ainda estudados os motivos pelos quais alguns setores não aderiram à certificação.

Para obter as conclusões construiu-se um questionário (ver Hill e Hill [2] e Gomes [1]) feito a 383 empresas escolhidas aleatoriamente de uma base de dados (ver Vicente e outros [6]). No estudo estatístico recorreu-se a uma análise de dados categorizados efetuada com recurso ao SPSS (ver Paulino e Singer [3] e Pestana e Gagueiro [4]).

Referências

- [1] Gomes, P. (1998) Tópicos de Sondagens, Sociedade Portuguesa de Estatística.
- [2] Hill, A., Hill M.M. (2008) Investigação por questionário, Edições Silabo.

- [3] Paulino, C.D., Singer, J.M. (2006) Análise de dados categorizados, Edgard Blucher, São Paulo.
- [4] Pestana, N.H., Gagueiro, J.N.(2009) Análise Categórica, Árvores de Decisão e Análise de Conteúdo, Edições Lidel.
- [5] Pinto, A., Iolanda, S. (2009) Sistema de gestão e Qualidade, Edições Sílabo.
- [6] Vicente, P., Reis, E., Ferrão, F. (2001) Sondagens- A amostragem como factor decisivo de qualidade, Edições Sílabo.

Avaliação Estatística de Alguns Benefícios do Envelhecimento Ativo

(Poster)

Paulo INFANTE – *CIMA-UE e ECT da Universidade de Évora*
pinfante@uevora.pt

Anabela Afonso – *CIMA-UE e ECT da Universidade de Évora*
aafonso@uevora.pt

Gonçalo Jacinto – *CIMA-UE e ECT da Universidade de Évora*
gjcj@uevora.pt

Teresa Engana – *Divisão de Desporto da Câmara Municipal de Évora*
eresangana@cm-evora.pt

Resumo

A prática de atividade física para além de aumentar o bem-estar e a qualidade de vida (Brown et al., 2003) é eficaz na prevenção de várias doenças coronárias, diabetes de tipo 2, cancro, hipertensão, obesidade, depressão e osteoporose, e morte prematura (Warburton et al., 2006). Em 2012, a Organização Mundial da Saúde (OMS) reconheceu a importância da atividade desportiva para a saúde física, mental e psicológica (WHO, 2012).

Segundo os dados dos Censos 2011, cerca de 18% da população do Concelho de Évora tem 65 ou mais anos, sendo que por cada 100 jovens com menos de 15 anos existem 140 pessoas idosas com pelo menos 65 anos. Com vista a gerir com eficácia os assuntos relacionados com a longevidade, a Câmara Municipal de Évora através dos serviços competentes, em parceria com diversas instituições do Concelho, que compreendem associações de reformados, associações de moradores, centros de dia, centros de convívio e centros sociais, tem em curso um programa denominado Seniores Ativos. Esta iniciativa visa a promoção da atividade física e estilos de vida saudáveis através da mobilização da população sedentária com pelo menos 65 anos de idade, de modo a integrar a atividade física em rotinas diárias. Os exercícios são realizados em grupo, na piscina municipal, em espaços verdes ao ar livre e em espaços cobertos, orientados por professores de educação física, com

duração média de uma hora e regularidade de uma vez por semana. Este programa permite, assim, a passagem da resposta apática para a atitude de envelhecimento ativo.

No decurso do Ano Europeu do Envelhecimento Ativo e da Solidariedade entre Gerações, a Câmara Municipal de Évora em parceria com a Universidade de Évora (através do CIMA-UE), iniciou no final de 2012 um estudo sobre os benefícios físicos, psicológicos e sociais da atividade física do programa Seniores Ativos. Para este estudo foi considerada uma população alvo com 55 ou mais anos, para a qual se construíram 4 grupos de participantes: i) grupo padrão constituído por alguns membros dos Seniores Ativos, mantendo a frequência atual; ii) grupo constituído por membros dos Seniores Ativos que se voluntariaram para realizar a atividade três vezes por semana; iii) grupo virgem constituído por novos membros que iniciaram a atividade no início do estudo, com uma regularidade igual à do grupo padrão; iv) grupo de controlo constituído por pessoas que não fazem atividade física e ao qual deixarão de pertencer caso resolvam iniciar atividade física regular. Os participantes em cada um dos grupos foram submetidos no início e no final desta fase do estudo a uma bateria de testes para avaliar os parâmetros físicos associados à aptidão funcional (força, flexibilidade, resistência, velocidade, agilidade e equilíbrio) e foram avaliados relativamente à pressão arterial, colesterol, glicose e IMC. Também responderam a um questionário que focou, entre outras, questões relativas à sua ocupação, estado funcional e historial clínico respeitante à toma de medicamentos para hipertensão, doenças cardíacas, colesterol, diabetes, distúrbios da tiróide, osteoporose e doença pulmonar. Foi ainda recolhida uma estimativa do custo mensal com medicamentos.

Neste trabalho apresentamos os resultados preliminares deste estudo. Começamos por efetuar uma caracterização dos quatro grupos e realizamos comparações, procurando retirar algumas conclusões sobre as diferenças observadas nos níveis de aptidão física e fundamentalmente nos custos/benefícios deste tipo de atividade para a saúde. Na análise dos dados iremos recorrer a medidas de associação, testes paramétricos e não paramétricos e a modelos lineares generalizados para modelar o custo/benefício associado à prática da atividade física.

Referências

- [1] Brown, D.W., *et al.* (2003) Associations between recommended levels of physical activity and health-related quality of life. Findings from the 2001 Behavioral Risk Factor Surveillance System (BRFSS) survey. *Preventive Medicine*, 37, 520-528.
- [2] Warburton, D., Nicol, C., Bredin, S. (2006) *Health benefits of physical activity: The evidence*, *Canadian Medical Association Journal*, 174, 801–809.
- [3] WHO (2012). Global recommendations on physical activity for health. World Health Organization. Acedido a 8 de julho de 2013. Disponível em http://whqlibdoc.who.int/publications/2010/9789241599979_eng.pdf

Segmentação de séries de variabilidade extraídas do eletrocardiograma

(Poster)

Argentina LEITE – *Escola de Ciências e Tecnologia, Universidade de Trás-os-Montes e Alto Douro e CM-UTAD*

tinucha@utad.pt

Rute Almeida – *Faculdade de Ciências, Universidade do Porto e CMUP; CIBER-BBN, Espanha*

Ana Paula Rocha – *Faculdade de Ciências, Universidade do Porto e CMUP*

Maria Eduarda Silva – *Faculdade de Economia, Universidade do Porto e CIDMA*

Resumo

Em muitas situações práticas as séries a analisar são longas e apresentam características não estacionárias com alterações na média, na variância ou mesmo na estrutura de dependência. A detecção dos pontos de mudança associados a quebras estruturais é importante quando se tem em vista a modelação da série. Uma abordagem comum à modelação de séries longas não estacionárias é a sua segmentação em registos mais curtos aproximadamente estacionários, aos quais se adequam modelos estacionários. Diz-se então que se tem séries localmente estacionárias e modelos estacionários por bocados.

Neste trabalho considera-se a segmentação de séries longas de variabilidade cardio-respiratória. Estas séries são extraídas de forma automática de electrocardiogramas de alta resolução e exprimem a variabilidade batimento-a-batimento da frequência cardíaca e da respiração, assim como outras relacionadas, tais como a variabilidade da duração da actividade ventricular, a dispersão da repolarização cardíaca, que foram associadas a risco cardíaco aumentado em diversas patologias. Estes registos apresentam vários pontos de mudança associados às diversas situações da vida diária como mudanças de posição corporal, actividade física e estado de alerta. Usam-se diversas abordagens para a segmentação das séries baseadas em modelos da família ARFIMA-GARCH que permitem a descrição conjunta de estruturas de correlação de termo curto e termo longo, bem como de características heterocedásticas.

Heterogeneidade Não Observada na Exposição ao Risco no Local de Trabalho: Uma Aplicação dos Modelos de Mistura Factorial a Dados Europeus

(Poster)

Catarina MARQUES – *Instituto Universitário de Lisboa (ISCTE-IUL),
Business Research Unit*
cmam@iscte.pt

José Gonçalves Dias – *Instituto Universitário de Lisboa (ISCTE-IUL),
Business Research Unit*

Resumo

É actualmente reconhecido que a exposição a factores de risco físico no local de trabalho pode influenciar negativamente a saúde e o bem-estar dos trabalhadores. A Saúde e Segurança no trabalho é uma das principais competências da União Europeia (UE) e uma das principais áreas da política social europeia. Enfrentar riscos novos e melhorar o acompanhamento do progresso são objectivos importantes das estratégias comunitárias em matéria de saúde e segurança no trabalho (2007-2012). Neste sentido, a UE tem regulado directivas e definido padrões que visam a criação de um ambiente saudável e seguro para os trabalhadores. Adicionalmente, as políticas de saúde e segurança no trabalho de cada país europeu envolvem medidas sobre as condições de trabalho, a carga dos empregadores. Contudo, contrariamente ao esperado, os níveis de exposição a riscos físicos no local de trabalho não têm diminuído muito a nível europeu desde o primeiro inquérito realizado sobre as condições de trabalho na Europa em 1991 [2].

O objectivo deste estudo é identificar grupos de trabalhadores que estejam expostos aos mesmos tipos de risco. Para tal, é aplicado um modelo de mistura factorial [1, 3, 4, 5], para identificar a heterogeneidade não observada na exposição a factores de risco no local de trabalho nos 27 países da UE. Estes modelos de análise factorial assumem que a população é heterogénea

incorporando uma variável latente discreta em modelos de estrutura de covariâncias. São usados os dados obtidos no 5º Inquérito Europeu sobre as Condições de Trabalho (EWCS) realizado em 2010, que fornece informação acerca da qualidade do trabalho e aspectos de emprego a nível europeu. Constituem a amostra 35372 trabalhadores dos 27 países da UE. Os resultados mostram que existem grupos distintos na população activa que estão expostos a diferentes perfis de risco e, conseqüentemente, poderão ter necessidades de informação diferentes. Usando esta abordagem poderão definir-se estratégias dirigidas a cada grupo, optimizando dessa forma recursos, e tornando mais efectivas as directrizes tanto a nível comunitário quanto nacional.

Referências

- [1] Arminger, G., Stein, P. (1997) *Finite mixtures of covariance structure models with regressors*, Sociological Methods & Research 26, 148–182.
- [2] EUROFOND (2012) Fifth European Working Conditions Survey. Publications Office of the European Union, Luxembourg.
- [3] Jedidi, K., Ramaswamy, V., DeSarbo, W. S., Wedel, M. (1996) *On estimating finite mixtures of multivariate regression and simultaneous equation models*, Structural Equation Modeling 3, 266–289.
- [4] Jedidi, K., Jagpal, H. S., DeSarbo, W. S. (1997) *Finite-fixture structural equation models for response-based segmentation and unobserved heterogeneity*, Marketing Science 16(1), 39–59.
- [5] Lubke, G. H., Muthén, B. O. (2005) *Investigating population heterogeneity with factor mixture models*, Psychological Methods 10, 21–39.

Modelo bayesiano de equações simultâneas para a estimação dos parâmetros da área basal e da mortalidade

(Poster)

Marco MARTO – *Universidade de Aveiro*

marcomarto@gmail.com

Isabel Pereira – *Universidade de Aveiro/CIDMA*

Margarida Tomé – *ISA/Centro de Estudos Florestais*

Resumo

Os modelos de crescimento florestais para a espécie *Eucalyptus Globulus* têm sido objeto de estudo e aperfeiçoamento pelo Centro de Estudos Florestais do Instituto Superior de Agronomia desde a década de 70 (Tomé *et al*[1] e [2]). Têm sido utilizados métodos de sistemas de equações simultâneas, nomeadamente o NSUR para possibilitar a estimação de equações não lineares nas variáveis biológicas e nos parâmetros estruturais das equações das variáveis biológicas dependentes a estimar. Pressupondo a existência de correlação contemporânea entre os resíduos das equações de G (área basal) e N (mortalidade) do modelo *Globulus 2.1*, estamos nas condições de aplicabilidade dos métodos referidos. Assim este estudo apresenta uma alternativa bayesiana para a estimação dos modelos de equações simultâneas clássicos, considerando uma metodologia NSUR (*Nonlinear Seemingly Unrelated Regression*) bayesiana com recurso a métodos de Monte Carlo baseados em cadeias de Markov, mais concretamente o algoritmo de Gibbs com o passo de Metropolis. desta forma generaliza-se o trabalho anteriormente feito por Marto, M. [3], comparando os resultados de estimação obtidos pelas abordagens clássica e bayesiana e analisando qual o melhor modelo preditivo. A validação do modelo é feita usando como medida de desvio a raiz quadrada do erro quadrático médio.

Referências

- [1] Tomé, M., Ribeiro, F. G. , Soares, P. (1998) *Silvipastoral systems in Portugal*, em Modelling the growth of tree plantations and agroforestry systems in south and east Africa (Pukkala, T. e Eerikainen, K. eds), Tiedonantoja Research Notes, 80, 23–33.
- [2] Tomé, M., Ribeiro, F. G., Soares, P. (2001) *O modelo Globulus 2.1*, Relatórios Técnico-Científicos do GIMREF n.º1/2001, Centro de Estudos Florestais, Instituto Superior de Agronomia, Lisboa.
- [3] Marto, M. (2011) *Estimação clássica e bayesiana de parâmetros do modelo Globulus 2.1*, Tese de Mestrado, Departamento de Matemática, Universidade de Aveiro.

The effect of counselling by healthcare professionals on breastfeeding success

(Poster)

Sílvia MENDONÇA – *DEIO-FCUL, Universidade de Lisboa*
silvia6m@gmail.com

Marília Antunes – *Centro de Estatística e Aplicações e DEIO-FCUL, Universidade de Lisboa*

Violeta Alarcão – *Instituto de Medicina Preventiva-FMUL, Universidade de Lisboa*

Rui Simões – *Instituto de Medicina Preventiva-FMUL, Universidade de Lisboa*

Resumo

This work aims to explain the role of BF counselling by healthcare professionals in breastfeeding (BF) success through a logistic regression analysis. The response variable *goal achievement* is binary, defined as the achievement of the intended BF duration.

There are 452 mothers that report a specific intended BF duration. Among these, most refer six (23.2%) and twelve months (45.4%) as planned duration. However, due to censored observations, the outcome can only be determined for 411, among which approximately one third (34.5%) achieved their goal. However, it is likely that the actual proportion of successes is higher than the one found, due to the differential loss of the mothers who breastfed longer.

The variables of interest in this model correspond to two questions where mothers were asked if they were ever counselled about BF benefits and techniques or difficulties. Both variables were categorized in three groups: was never counselled (*Never*), was first counselled before delivery (*Antenatal*) and was only counselled after delivery (*Postnatal*). The majority of mothers recall having been counselled at least once - 77% for benefits and 71% for techniques/difficulties. There are significant differences among both groups in terms of age, education level and previous BF experience.

The variables *maternal age, education, previous BF experience* and *country of origin* were considered as *a priori* confounders and were included indepen-

dently of statistical significance. Other biological variables were maintained in the model depending on the level of association with the response variable and the presence of confounding, measured as a change in effect measure [1]. Standardized deviance residuals revealed the existence of a non-linear relation between *maternal age* and the *logit*. This variable was transformed using regression polynomials of second and third degree and through cubic splines [2]. A cubic spline with three internal knots was chosen to correct for the non-linearity. An increase in maternal age leads to an increase in the adjusted values of the probability of success approximately 30 years of age, after which there is no change in the adjusted values.

Among the variables of interest, only counselling about BF benefits exhibited a borderline significant association with BF success. Prenatal counselling about benefits was associated with a 72% increase in the odds of success, compared to not having been counselled ($p=0.1165$).

The importance of counselling and support by healthcare professionals, at both prenatal and postnatal periods, has been confirmed in interventional studies. The main limitations of this analysis are the observational nature of the data and not knowing in detail the type of counselling that was given.

Acknowledgements: Data used in this work was collected as part of the Project 'Prevalência e Factores de Insucesso do Aleitamento Materno em Portugal' that was supported by the General Directorate of Health (DGS) and the Portuguese Society of Pediatrics. The authors thank the IMP research team and the voluntary participants. The research of Marília Antunes is partially funded by project Pest-OE/MAT/UI0006/2011.

Referências

[1] Greenland, S (1989) *Modeling and variable selection in epidemiologic analysis*, American Journal of Public Health 79, 340–349.

[2] Harrell, F (2001) *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*, Springer.

Diagnóstico em regressão binária

(Poster)

Isabel NATÁRIO – *CEAUL, Dep. de Matemática, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2829-516, Caparica, Portugal*
 icn@fct.unl.pt

Sílvia Shruballs – *CESUR, Instituto Superior Técnico - Technical University of Lisbon, Lisbon, Portugal*

Resumo

As aplicações do modelo de regressão binária são variadas e muito diversas, ocupando contudo um lugar muito destacado em estudos de biologia e medicina. Este modelo pressupõe que há uma variável aleatória binária de interesse, Y , cujo valor esperado, coincidente com sua probabilidade de sucesso $p = P(Y = 1)$, é modelado através de uma função de covariáveis x_1, \dots, x_p que se consideram influenciar essa resposta - preditor, η . Usualmente considera-se que este preditor é linear nas covariáveis, *i.e.*, $\eta = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$. A função que relaciona $E[Y|x_1, \dots, x_p] = p$ com o preditor designa-se por função de ligação, $g(p) = \eta$, e não é única. A estimação é feita por máxima verosimilhança [1]. É então imperativo proceder à análise de diagnóstico do modelo empregue, questionando se os pressupostos do modelo, incluindo a função de ligação escolhida, são razoáveis, face aos dados em mãos - avaliando assim a qualidade do ajuste [1], [2], [3], [4], [5], [6]. Contudo, parece haver grandes confusões e actuações incorrectas nestas análises, verificando-se muitas vezes erros grosseiros e inadmissíveis. Neste trabalho procura-se fazer um apanhado de boas práticas associadas a esta questão, apresentando um conjunto de procedimentos do que efectivamente pode e deve ser feito, indicando referências a software existentes que ajudem nesta análise. Ilustram-se os procedimentos num exemplo com dados reais de acidentes rodoviários com vítimas em Lisboa.

Agradecimentos

A ANSR e o LNEC forneceram parte da informação contida nos dados. Este trabalho foi parcialmente financiado por fundos nacionais através da Fundação Nacional para a Ciência e Tecnologia, Portugal - FCT, pelos projeto PEst-OE/MAT/UI0006/2011 e pelo projeto de investigação Spatial Analysis of Child Road Accidents (SACRA), PTDC/TRA/66161/2006

Referências

- [1] McCullagh, P., Nelder, J. (1989). Generalized Linear Models, 2nd edition, Chapman and Hall/CRC
- [2] Pregibon, D. (1980). Goodness of Link Tests for Generalized Linear Models. *Applied Statistics* 29, 15–24
- [3] Hosmer, D.W., Lemeshow, S. (1980). Goodness-of-fit tests for the multiple logistic regression model. *Communications in Statistics . Theory and Methods* 9, 1043–1068
- [4] Hosmer, D.W., Lemeshow, S. (2000). *Applied Logistic Regression, Second Edition*. Wiley Series in Probability and Mathematical Statistics, New York.
- [5] Rodríguez, G. (2007). Generalized Linear Models. Office of Population Research. Princeton University.
- [6] Ni, Y. (2010). *LDdiag* package manual. R-project

Barreiras à utilização dos serviços de saúde pelos imigrantes do Bangladesh a residir em Portugal

(Poster)

Carla OIKO *Unidade de Epidemiologia, Instituto de Medicina Preventiva, Faculdade de Medicina de Lisboa*

carla_oiko@yahoo.com.br

Violeta Alarcão *Unidade de Epidemiologia, Instituto de Medicina Preventiva, Faculdade de Medicina de Lisboa*

Rui Simões *Unidade de Epidemiologia, Instituto de Medicina Preventiva, Faculdade de Medicina de Lisboa*

Elisa Lopes *Unidade de Epidemiologia, Instituto de Medicina Preventiva, Faculdade de Medicina de Lisboa*

João Gomes *Departamento de Estatística e Investigação Operacional, Faculdade de Ciências de Lisboa*

Resumo

Com base em estudos anteriores, sabe-se que existem desigualdades na utilização de serviços de saúde entre as populações imigrantes a residir em Portugal e que a saúde física e mental dos migrantes tende a ser mais vulnerável.

Entre dezembro 2012 e março 2013, foi realizado um inquérito com o objetivo de caracterizar o estado de saúde e o acesso aos cuidados de saúde das comunidades imigrantes do Bangladesh, Índia e Paquistão. O objetivo específico deste trabalho foi identificar fatores associados à utilização dos serviços de saúde em Portugal entre os imigrantes adultos do Bangladesh. Foram recolhidos dados sociodemográficos e de saúde, autoavaliação da qualidade de vida e avaliação da saúde mental (MHI-5) através de um questionário estruturado aplicado presencialmente (taxa de adesão de 97%).

Os Bangladeshianos inquiridos ($n = 261$) tinham uma idade média de $32,2 \pm 7,3$ anos, 14,6% eram do sexo feminino, 51,7% solteiros e possuíam uma média de anos de escolaridade de $13,4 \pm 2,9$ anos. Estes residiam em Portugal há uma média de $2,6 \pm 3,5$ anos com 85,8% a residir há 5 ou menos anos. Entre

os participantes, 4,3% reportaram colesterol alto, 3,2% diabetes, 2,3% tensão arterial, e 89,3% referiram não possuir nenhuma das doenças referidas; 46,5% apresentaram algum sofrimento psicológico ($MHI \leq 52$). Do total da amostra, 46,4% já tinha recorrido a algum serviço de saúde em Portugal, através dos serviços de saúde dentária, consultas e/ou urgência num Centro de Saúde ou Hospital.

Recorrendo a modelos de regressão logística procurar-se-á identificar fatores que ajudem a explicar a utilização dos serviços de saúde em Portugal para esta comunidade, e nomeadamente testar a hipótese de uma associação entre o nível de sofrimento psicológico e uma maior utilização dos serviços de saúde.

Estes resultados serão também complementados com dados dos imigrantes da Índia e Paquistão, através do uso da mesma metodologia, para uma análise comparativa dos dados.

Referências

- [1] Dias, S (2013) Health status and preventative behaviors of immigrants by gender and origin: A Portuguese cross-sectional study. *Nursing and Health Sciences*.
- [2] Dias, S (2011) Healthcare-seeking patterns among immigrants in Portugal. *Health and Social Care in the Community* (2011) 19(5), 514–521.

Delineamentos Regressaionais Múltiplos, caso não equilibrado

(Poster)

Sandra OLIVEIRA – *Escola Superior de Ciências Empresariais do Instituto Politécnico de Setúbal*

sandra.oliveira@esce.ips.pt

Miguel Fonseca – *Centro de Matemática e Aplicações da Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa*

João Tiago Mexia – *Centro de Matemática e Aplicações da Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa*

Resumo

Quando a cada tratamento de um plano base corresponde uma regressão linear múltipla (modelo elementar) nas mesmas variáveis obtemos um Delineamento Regressional Múltiplo.

Neste trabalho o plano base é um factorial de base prima, caso completo. A estrutura algébrica do plano base assenta na teoria dos corpos de Galois e na utilização de álgebras de Jordan comutativas (AJC). É apresentada a forma como estes modelos estão associados a uma AJC o que permite obter estimadores UMVUE e derivar testes de hipóteses para os parâmetros do modelo.

Nesta exposição apresenta-se uma extensão dos modelos regressionais múltiplos no caso em que a matriz do modelo elementar pode não ser idêntica para todos os tratamentos do modelo base. Esta extensão assenta também na utilização de AJC e na aplicação do algoritmo definido para o plano base.

Referências

- [1] Covas, R. Mexia J.T., Zmyslony R., (2010). Lattices of Jordan algebras. *Linear Algebra and its Applications*, 2679-2690.
- [2] Jesus V., Mexia J.T., Fonseca M., Zmyslony R. (2009), Binary Operations and Canonical forms for factorial and related models. *Linear Algebra and its Applications*, Volume 430, 2781-2797.

- [3] Jesus V., Mexia J.T., Oliveira S., Binary operations on prime basis factorials. *Biometrical letters*, Vol. 46 (2009), No. 1, 1-14.
- [4] Mexia, J.T., Oliveira, S., (2006.) *Delimitamentos Regressionais com Base Factorial Prima*, Actas do XIV Congresso Anual da Sociedade Portuguesa de Estatística, Covilhã.
- [5] Moreira E.E., Ribeiro A.B., Mateus E.P., Mexia J.T., Ottosen L.M. (2005a) Regressional modeling of electro-dialytic removal of Cu, Cr and As from CCA timber waste: application to sawdust. *Wood Science and Technology* 39(4), 291-309.
- [6] Moreira E.E., Ribeiro A.B., Mateus E.P., Mexia J.T., Ottosen L.M. (2005b) Regressional modeling of electro-dialytic removal of Cu, Cr and As from CCA timber waste: application to wood chips. *Biometrical letters* 42(1), 11-25.
- [7] S. Ferreira, D. Ferreira, E. Moreira, J.T. Mexia (2010). Exact estimators for normal linear mixed models. *Far East Journal of Theoretical Statistics* 31(1): 33-48

Modelação e estimação de parâmetros de acontecimentos raros, em ambiente R

– as abordagens paramétrica e semi-paramétrica

(Poster)

Helena PENALVA – *Escola Superior de Ciências Empresariais/Instituto Politécnico de Setúbal*

helen.penalva@esce.ips.pt

Sandra Nunes – *Escola Superior de Ciências Empresariais / Instituto Politécnico de Setúbal e CMA*

Manuela Neves – *Instituto Superior de Agronomia/Universidade de Lisboa e CEAUL*

Resumo

Historicamente, as aplicações da Teoria de Valores Extremos iniciaram-se em duas principais áreas: a área ambiental, com o estudo dos níveis do mar, velocidade do vento, caudal dos rios, entre outros; e a área da fiabilidade. Actualmente, a Teoria de Valores Extremos tem surgido como uma das mais importantes áreas da Estatística utilizada em várias ciências aplicadas, tais como, a Biologia, a Geologia e risco sísmico, a Climatologia e o Ambiente, ver Coles (2001) e Beirlant *et al.* (2004). A Teoria de Valores Extremos pretende estimar e prever acontecimentos que poderão ser mais extremos do que aqueles que alguma vez já foram observados. Assim, a inferência estatística somente pode ser deduzida a partir das observações que são consideradas extremas num determinado contexto. O estudo do comportamento dos extremos de amostras de variáveis aleatórias pode ser feito usando duas abordagens: a paramétrica e a semi-paramétrica. Iremos tratar neste trabalho apenas a análise estatística de extremos univariados. Em contexto paramétrico, o primeiro modelo de distribuições do máximo é o chamado *Método dos Máximos Anuais*, *Método dos Máximos de Blocos* ou ainda *Método de Gumbel*, que considera os k valores máximos de sub-amostras de tamanho r , $n = k \times r$. Uma outra abordagem é a que considera a modelação da distribuição de excessos acima de certo limiar, conhecida por *Método POT (Peaks Over Thresholds)*, na qual se restringe

a nossa atenção às observações que excedem um certo nível elevado, u . A abordagem semi-paramétrica teve a sua origem na escola holandesa com os trabalhos de de Haan (1970), e neste caso apenas se admite que o modelo F subjacente aos dados da amostra verifica certas propriedades na cauda. Nestas distribuições, o parâmetro de forma, γ , designado por índice de cauda ou índice de valores extremos descreve o comportamento da cauda direita, $1 - F$, do modelo subjacente aos dados. A sua estimação precisa é muito importante e de enorme influência na estimação de outros parâmetros, tais como quantis elevados ou período de retorno de quantis elevados. Neste trabalho pretende-se abordar a utilização do software R, (R Development Core Team, 2012) na estimação de parâmetros de acontecimentos raros, aplicando as abordagens paramétrica e semi-paramétrica. Para isso iremos escolher dois conjuntos de dados: um já tratado na literatura e outro conjunto de dados relativos a níveis médios diários na estação hidrométrica do rio Fraga.

Agradecimentos: Investigação parcialmente suportada por Fundos nacionais através da FCT—Fundação para a Ciência e a Tecnologia, projecto PEst-OE/MAT/UI0006/2011 e PEst-OE/MAT7UI0297/2011(CMA/FCT/UNL).

Referências

- [1] Beirlant, J., Goegebeur, Y., Teugels, J. and Segers, J. (2004) *Statistics of Extremes: Theory and Applications*. Wiley, England.
- [2] Coles, S. (2001) *An Introduction to Statistical Modeling of extreme Values*. Springer-Verlag, London.
- [3] Haan, L.de (1970) *On regular variation and its Applications to the Week Convergence of Sample Extremes*. Mathematical Centre Tract 32, Amsterdam.
- [4] R Development Core Team (2012). R: A Language and Environment for Statistical Computing, R. Foundation for Statistical Computing, Vienna, Austria, URL <http://www.R-project.org/>. ISBN 3-900051-07-0.

Almost sure convergence for the maximum of nonstationary random fields

(Poster)

Luísa PEREIRA – *Universidade da Beira Interior*

lpereira@ubi.pt

Abstract

In recent years various authors discussed almost sure versions of distributional limit theorems. The first result on Almost Sure Central Limit Theorem (ASCLT) presented independently by Brosamler (1988), Schatte (1988) and Lacey and Philipp (1990) extended the classical central limit theorem to an almost sure version.

For an i.i.d. sequence $\{X_n\}_{n \in \mathbb{N}}$ with zero mean, unit variance and partial sum $S_k = \sum_{i=1}^k X_i$, $k \geq 1$, the simplest version of the ASCLT states that

$$\frac{1}{\log n} \sum_{k=1}^n \frac{1}{k} \mathbf{1}_{\{S_k \leq \sqrt{k}x\}} \rightarrow \Phi(x) \quad a.s.$$

for any fixed $x \in \mathbb{R}$, where *a.s.* means almost surely, $\mathbf{1}_A$ denotes the indicator function of the event A and $\Phi(x)$ is the standard normal distribution function.

Later on the ASCLTs for some other functions of random variables were studied. Namely, in Fahrner and Stadtmüller (1998), Cheng, Peng and Li (1998) and Berkes and Csáki (2001) the ASCLTs for the maximum of an i.i.d. random sequence were proved.

Let $\{X_n\}_{n \in \mathbb{N}}$ be an i.i.d. sequence, and let $M_k = \max_{1 \leq i \leq k} X_i$ denotes the partial maximum, $k \geq 1$. If there exist normalizing constants $a_k > 0$, $b_k \in \mathbb{R}$ and a nondegenerate distribution function $G(x)$ such that

$$P(M_n \leq a_n x + b_n) \rightarrow G(x),$$

then we have

$$\frac{1}{\log n} \sum_{k=1}^n \frac{1}{k} \mathbf{1}_{\{M_k \leq a_k x + b_k\}} \rightarrow G(x) \quad a.s.,$$

for any continuity point x of G . It is well known that $G(x)$ must be of the same type as the extreme value distribution $G(x) = \exp \left\{ -(1 + \gamma x)^{-\frac{1}{\gamma}} \right\}$, where γ is the so-called extreme value index.

On the other hand, the ASCLTs for the maximum of some dependent, stationary normal sequences were obtained by Csáki and Gonchigdanzan (2002), while the ASCLT for the maximum of some dependent, but not necessarily stationary sequences was established by Zouxiang and Nadarajah (2011) and Chen and Lin (2006). Some other works, which are also worthwhile to mention in this place are the papers of Fazekas and Rychlik (2003) and Choi (2010).

The aim of this paper is to prove a ASCLT for the maximum of some nonstationary random fields under some weak dependence conditions.

Acknowledgements

This research was partially supported by National Funds through FCT-Fundação para a Ciência e a Tecnologia, project PEst-OE/MAT/UI0212/2011, and by FEDER funding through COMPETE program and national funding through FCT within the research project PTDC/MAT/108575/2008

References

- [1] Berkes, I., Csáki, E. (2001). *A universal result in almost sure central limit theorem*, Stoch. Process. Appl., 94, 105-134.
- [2] Brosamler, G.A. (1988). *An almost sure everywhere central limit theorem*, Math. Proc. Camb. Phil. Soc., 104, 561-574.
- [3] Chen, S., Lin, Z. (2006). *Almost sure max-limits for nonstationary Gaussian sequence*, Statist. Probab. Lett., 76, 1175-1184.
- [4] Cheng, S., Peng, L., Qi, Y. (1998). *Almost sure convergence in extreme value theory*, Math. Nachr., 190, 43-50.
- [5] Choi, H. (2010). *Almost sure limit theorem for stationary Gaussian random fields*, Journal of the Korean Statistical Society, 39, 475-482.
- [6] Csáki, E., Gonchigdanzan, K. (2002). *Almost sure limit theorems for the maximum of stationary Gaussian sequences*, Statist. Probab. Lett., 58, 195-203.
- [7] Fahrner, I., Stadtmüller, U. (1998). *On almost sure max-limit theorems*, Statist. Probab. Lett., 37, 229-236.

- [8] Fazekas, I., Rychlic, Z. (2003). *Almost sure central limit theorems for random fields*, Math. Nachr., 259, 12-18.
- [9] Lacey, M.T., Philipp, W. (1990). *A note on the almost sure central limit theorem*, Statist. Probab. Lett., 9, 201-205.
- [10] Schatte, P. (1988) *On strong versions of the central limit theorem*, Math. Nachr., 137, 249-256.
- [11] Zuoxiang, P., Nadarajah, S. (2011) *Almost sure limit theorems for Gaussian sequences*, Theory Probab. Appl., 55, 361-367.

Análise longitudinal do número de consultas em doentes hipertensos

(Poster)

Bernardete PINHEIRO – *Centro de Investigação Sobre Economia Portuguesa (CISEP - ISEG/UTL)*

bpinheiro@cisep.iseg.utl.pt

Marília Antunes – *Centro de Estatística e Aplicações e DEIO-FCUL, Universidade de Lisboa*

Milene Fernandes – *Instituto de Medicina Preventiva, Faculdade de Medicina da Universidade de Lisboa*

Resumo

O estudo DIMATCH-HTA é um estudo observacional longitudinal em duas coortes (imigrantes e não imigrantes), desenhado com o objetivo de conhecer e avaliar os determinantes associados à adesão terapêutica e à mudança terapêutica no controlo da hipertensão arterial. Os dados foram recolhidos através de questionários administrados presencialmente no início do estudo, aos 6 e aos 12 meses e telefonicamente aos 3 e aos 9 meses após o início do estudo.

Um dos objectivos do estudo consiste em analisar os determinantes da utilização de cuidados de saúde entre os doentes hipertensos seguidos nos Cuidados de Saúde Primários, através da modelação dos dados recolhidos através dos questionários, usando modelos para dados longitudinais.

A variável indicadora do consumo de recursos foi o número de consultas médicas nos 3 meses anteriores a cada entrevista. Os valores de tensão arterial sistólica e diastólica foram recolhidos nas entrevistas presenciais aos 6 e 12 meses. Nos momentos 3 e 9, nos quais a variável não foi recolhida por técnicos de saúde por questões de delineamento do estudo, foram usados os valores auto-reportados pelo doente.

O número médio de consultas variou entre 0,50 (dp=0,85) e 0,16 (dp=0,50) aos 3 e 12 meses, respectivamente. As variáveis explicativas consideradas para a construção do modelo de regressão foram idade, sexo, etnia, estado civil,

rendimento, nível de ensino, seguro privado, subsistema de saúde, ocupação profissional, número de pessoas do agregado, número de comorbilidades, controlo de hipertensão, tempo desde diagnóstico de hipertensão e auto-percepção da doença. Foram estimados modelos marginais usando funções do package gee do R (versão 2.14.2) e foram testadas diferentes estruturas de dependência entre as medidas repetidas.

Os resultados preliminares dos modelos estimados indicaram que as variáveis número total de comorbilidades, o tempo desde diagnóstico de hipertensão, a percepção do doente relativamente à doença (controlado vs não controlado) e o tipo de unidade funcional (unidade de cuidados personalizados vs unidade de saúde familiar) eram estatisticamente significativas, independentemente das estruturas de correlação usadas. Além da consolidação dos resultados obtidos, no âmbito deste trabalho pretende-se ainda explorar modelos alternativos ao modelo de Poisson que permitam acomodar o excesso de zeros e a consequente sobredispersão que caracteriza estes dados de contagem.

Agradecimentos: O estudo DIMATCH-HTA foi co-financiado pela Fundação para Ciência e Tecnologia. Os autores agradecem à equipa de investigação da Unidade de Epidemiologia do Instituto de Medicina Preventiva, às USF /UCSP, médicos e participantes. O trabalho de Marília Antunes é parcialmente financiado pelo projecto Pest-OE/MAT/UI0006/2011.

Referências

- [1] Pan, W., Connett, J.E., 2002. *Selecting the Working Correlation Structure in Generalized Estimating Equations with. Application to the Lung Health Study*, Statistica Sinica 29, 12:475-490.
- [2] Spiensma AS, Hajos TR, de Boer MR, Heymans MW, Twisk JW, 2008. *A new approach to analyse longitudinal epidemiological data with an excess of zeros*, BMC Med Res Methodol. 2013;13:27
- [3] Winkelmann, R. (2008). *Econometric Analysis of Count Data*, Fifth edition, Berlin/New York, Springer.

A Escala de Coma FOUR na Pediatria

(Poster)

Alexandra PINTO – *Laboratório de Biomatemática, Faculdade de Medicina de Lisboa*

apinto@fm.ul.pt

Daniel Meireles – *Faculdade de Medicina de Lisboa*

Sofia Ferreira – *Faculdade de Medicina de Lisboa*

Francisco Abecasis – *Unidade de Cuidados Intensivos do Hospital de Santa Maria, Lisboa*

Resumo

O estado de consciência é um indicador importante de agravamento e melhoria durante a fase aguda após lesão cerebral. Consiste também num componente importante de prognóstico, uma vez que um estado de consciência alterado (“qualquer estado mental, induzido por vários agentes fisiológicos, psicológicos ou farmacológicos, que pode ser percebido subjectivamente pelo próprio indivíduo, ou de modo objectivo por um observador, como uma representação de um desvio na experiência subjectiva ou no funcionamento psicológico do sujeito, no estado de alerta e vigília, com base em determinados princípios e normas” [1]) pode evoluir para coma, estado vegetativo, recuperação lenta, mutismo acinético, síndrome locked-in ou morte, ou seja, estados de emergência médica. Deste modo, a validade e reprodutibilidade de ferramentas utilizadas na avaliação de alterações de estados de consciência são extremamente importantes, permitindo identificar a gravidade de cada doente e detectar o agravamento numa fase precoce.

De entre as várias escalas desenvolvidas com o intuito de: uniformizar a linguagem entre profissionais, agilizar a avaliação e objectivar a gravidade clínica, são utilizadas neste estudo, a Escala de Coma de Glasgow (GCS) e a Escala Full Outline of UnResponsiveness (FOUR). A GCS foi publicada oficialmente por Teasdale e Jennett em 1974, na revista *Lancet* como um apoio para a avaliação da profundidade e duração clínica da diminuição da consciência e do coma [2]. Esta escala baseia-se em três indicadores: 1) a abertura dos olhos, 2) a melhor resposta verbal e 3) a melhor resposta motora. A escala FOUR, inicialmente desenvolvida na clínica Mayo [3], foi testada por membros de Unidades de Cuidados Intensivos (UCI) e, posteriormente,

validada em testes realizados por enfermeiros especialistas de UCI, com maior ou menor grau de experiência profissional [4]. Esta escala é actualmente utilizada em vários centros de Neurotrauma e UCI por todo o Mundo. A Escala FOUR baseia-se em quatro componentes: 1) resposta ocular, 2) resposta motora, 3) reflexos de tronco cerebral e 4) respiração.

No presente estudo utilizou-se pela primeira vez a escala FOUR na Pediatria. Este estudo foi realizado de 1 de Maio de 2012 a 30 de Abril de 2013 e tem como objectivo comparar as escalas de coma GCS e FOUR, quando aplicadas em contexto pediátrico, por três grupos de avaliadores: Especialistas pediátricos, Internos de pediatria e Enfermeiros. Pretende-se também avaliar a concordância entre aplicadores e ainda, estudar o desempenho de ambas as escalas na previsão, quer da evolução clínica dos doentes, quer da mortalidade. Utilizaram-se neste estudo: Coeficiente Kappa de Cohen, Alfa de Cronbach, Coeficiente de Correlação de Spearman, Coeficiente de Contingência e curvas ROC. Este estudo veio confirmar vários resultados de trabalhos anteriores realizados na população adulta, nomeadamente, que a Escala FOUR apresenta maior valor preditivo na mortalidade intra-hospitalar e estado clínico à saída do hospital. Concluiu-se neste estudo que: 1) há interesse clínico na utilização da Escala FOUR para classificar os doentes e prever o seu prognóstico, 2) a Escala FOUR permite apurar e reunir um maior número de informações clínicas relevantes para os doentes pediátricos em estado crítico e 3) a Escala FOUR é válida para a população pediátrica.

Referências

- [1] Ludwig A. (1966) *Altered states of consciousness*, Arch Gen Psychiatry 15, 225–34.
- [2] Teasdale G., Jennett, B. (1974) *Assessment of coma and impaired consciousness. A practical scale*, Lancet 2, 81–4.
- [3] Wikdicks, E.F.M., Bamlet, W.R., Maramattom, B.V., Manno, E.M., McClelland, R.L. (2005) *Validation of a New Coma Scale: The FOUR Score*, Annals of Neurology 58, 585–93.
- [4] Wikdicks, E.F.M., Kokmen, E., O'Brien, P.C. (1998) *Measurement of impaired consciousness in the neurological intensive care unit: a new test*, J Neurol Neurosurg Psychiatry 64, 117–19.

Reconhecimento de Vogais usando Análise em Componentes Principais

(Poster)

Iola M.S. PINTO – *Área Departamental de Matemática, Instituto Superior de Engenharia de Lisboa, Rua Conselheiro Emídio Navarro 1, 1950-062 Lisboa, Portugal*

ipinto@adm.isel.pt

Carlos Meneses – *Área Departamental de Engenharia de Electrónica e Telecomunicações e de Computadores, Instituto Superior de Engenharia de Lisboa, Rua Conselheiro Emídio Navarro 1, 1950-062 Lisboa, Portugal*

André Duarte – *Área Departamental de Engenharia de Electrónica e Telecomunicações e de Computadores, Instituto Superior de Engenharia de Lisboa, Rua Conselheiro Emídio Navarro 1, 1950-062 Lisboa, Portugal*

Paulo Ruíz – *Área Departamental de Engenharia de Electrónica e Telecomunicações e de Computadores, Instituto Superior de Engenharia de Lisboa, Rua Conselheiro Emídio Navarro 1, 1950-062 Lisboa, Portugal*

Resumo

O reconhecimento de fala é o processo de transcrição de palavras isoladas para texto escrito, num computador ou telemóvel, de modo a facilitar a introdução de dados, sendo as aplicações praticamente ilimitadas. O reconhecimento de fala pode ser dependente ou independente do orador, sendo este último um processo menos robusto no sentido que, devido à maior variabilidade, verifica-se a existência de uma mais elevada incidência de erros.

As metodologias utilizadas na análise de sinais de fala permitem extrair as principais características do sinal convertendo-o num conjunto de parâmetros capazes de o descrever de um modo mais simplificado. Um dos parâmetros mais utilizado no reconhecimento de fala é o *MFCC*, *mel function cepstra coefficients*, [1], [2]. Para o reconhecimento de fala são utilizadas metodologias, como por exemplo, as que se baseiam em algoritmos como o *DTW*, *Dynamic time warping*, ou o *HMM*, *Hidden markov model*, [1], [2].

A análise em componentes principais (ACP) é um método estatístico multivariado com larga aplicação em diversas áreas do saber, sendo em par-

ticular reconhecida a sua importante função nas metodologias de análise e processamento de sinais [3], [4]. Este trabalho visa a implementação de um sistema de reconhecimento de vogais sustentadas independente do orador, recorrendo à utilização da técnica ACP, tendo como parâmetro a função MFCC de ordem 16. A utilização da análise em componentes principais permite reduzir a dimensionalidade dos dados, viabilizando a representação dos mesmos num espaço bi-dimensional ou tri-dimensional. A visualização dos dados no espaço das componentes principais permite, no caso concreto deste projeto, a identificação de grupos (*clusters*) que surgem naturalmente, indicando assim a existência de informação importante para o reconhecimento de vogais.

A amostra original de indivíduos é dividida em duas amostras de diferentes oradores: amostra de treino e a amostra de teste. Numa primeira fase procede-se à construção do modelo de treino com recurso à técnica ACP. Embora sejam encontrados cinco grupos correspondentes às cinco vogais a identificar, para cada grupo a aplicação da análise de *clusters*, com recurso ao algoritmo *K-means*, permite criar subclasses que permitem otimizar o processo de reconhecimento. Numa segunda fase procede-se à classificação das vogais da amostra de teste, com base numa regra de classificação que utiliza o critério do vizinho mais próximo com a distância euclidiana. Destaca-se como um dos resultados mais interessantes o valor de 1,5% de casos mal classificados, obtido a partir das três primeiras componentes principais e dois centróides por vogal.

Referências

- [1] Meneses, C.(2002) Processamento Digital de Fala, Sebenta de apoio, ISEL.
- [2] Rabiner, L.R., Schafer, R.W., (1978) Digital Processing of Speech Signals, Prentice-Hall, Signal Processing Series.
- [3] Everitt, B. S., Dunn, G. (2001) Applied Multivariate Data Analysis, Second edition, ARNOLD.
- [4] Johnson, R., Wichern, D.W. (2007) Applied Multivariate Statistical Analysis, 6^a, Pearson Education Inc, India.

Distribuições para modelos atuariais de risco

(Poster)

Celine QUEIRÓS – *Universidade do Minho*

celinequeiros@portugalmail.pt

Patrícia Gonçalves – *Universidade do Minho*

Irene Brito – *Universidade do Minho*

Resumo

A atividade seguradora, nas últimas décadas, está a ganhar terreno e a ocupar um lugar de destaque na sociedade. Quando um evento imprevisível ocorre solicitam-se os serviços de uma seguradora. Esta, através de um seguro transfere o risco de perdas financeiras em troca de um montante ou prémio. A probabilidade de ruína é uma preocupação constante para o negócio da atividade seguradora, ou seja, é a probabilidade de uma seguradora ficar sem reserva ou capital suficiente para pagar a indemnização de um determinado sinistro. Para a sobrevivência da mesma, é necessário avaliar o capital inicial investido e o cálculo do valor do prémio. O estudo da probabilidade de ruína de uma seguradora arruinar é feito usando o processo de reserva que se designa por $U(t), t \geq 0$. Apresenta-se assim o modelo clássico de Crámer-Lundberg:

$$U(t) = u + ct - S(t), t \geq 0,$$

onde $U(t)$ é designado de reserva de uma seguradora no instante $t \geq 0$, para fazer face a determinado risco, $u = U(0)$ representa o capital inicial e assume-se que os prémios são recebidos (por unidade de tempo) uma taxa constante que se denota por c . O processo $S(t) = \sum_{i=0}^{N(t)} X_i$, $t \geq 0$ representa o valor das indemnizações agregadas relativas ao intervalo de tempo $(0, t]$, onde $X_0 = 0$, para $i = 1, \dots, N(t)$, X_i representa o montante da i -ésima indemnização cujo sinistro ocorreu no intervalo de tempo $(0, t]$ e $N(t)$ representa o número de sinistros ocorridos no intervalo de tempo $(0, t]$. Este é um modelo probabilista, de tal forma que para $i = 1, \dots, N(t)$ X_i são variáveis aleatórias com certa distribuição e $N(t)$ é um processo estocástico de contagem. A escolha para a distribuição das variáveis X_i e do processo $N(t)$, dependem da escolha do tipo de seguros a considerar.

Diz-se que a ruína ocorreu, numa seguradora, se o processo de reserva é inferior a uma barreira de ruína, num certo instante ou instantes de tempo. Desta forma, define-se o primeiro instante de ocorrência de ruína, denotado por T , como $T = \min \{t, t \geq 0 \text{ e } U(t) < 0\}$

Em muitas situações não é possível obter o valor exacto da probabilidade de ruína, consegue-se apenas a sua quantificação de forma aproximada. Neste modelo, considera-se que $S(t)$ é um processo de Poisson composto, ou seja $N(t)$ é um processo de Poisson. Pelo Teorema fundamental do Risco, tem-se que para um processo de risco Poisson composto com capital inicial $u > 0$ e sendo R o coeficiente de ajustamento, a probabilidade de ruína é dada por

$$\psi(u) = \frac{e^{-Ru}}{E [e^{-RU(T)} | T < \infty]}.$$

Ressalva-se que no caso de acontecer a ruína, esta ocorre como consequência da ocorrência de uma indemnização. Este resultado propõe o cálculo da probabilidade de ruína sem especificar o tempo em que ela ocorre, mas em que número de indemnização.

Num modelo de risco em cálculo atuarial, é preciso especificar a distribuição para o número de indemnizações e para a severidade das indemnizações. Na prática, e quando se implementa o modelo, não há conhecimento da distribuição destas quantidades e tem-se a necessidade de enquadrar os dados da amostra ao modelo teórico. São seleccionadas algumas distribuições standard da Estatística, que usualmente são utilizadas para descrever o montante de indemnizações particulares para um tipo de carteira de seguro, com o objetivo de modelar os dados reais para diferentes carteiras de seguro. As distribuições apresentadas e caracterizadas com pormenor são a Gamma, Exponencial, Normal, Lognormal, Pareto, Burr, Weibull, Loggamma e mistura de duas distribuições Exponenciais. Para uma análise mais completa destas distribuições é determinada a função de excesso médio dada por $e(x) = E[X - x | X > x] = \frac{\int_x^{+\infty} (1 - F(u)) du}{1 - F(x)}$, $x > 0$, e sua representação gráfica.

Esta representa a esperança dos valores acima de x uma vez superado este valor, em contexto de seguradora interpreta-se como a sinistralidade esperada acima de x quando o número de sinistros é ilimitado. Se os valores de $e(x)$ apresentarem uma tendência crescente, significa que a cauda da distribuição

é mais pesada que a cauda da distribuição Exponencial, se pelo contrário os valores de $e(x)$ apresentarem uma tendência decrescente, significa que a cauda da distribuição é mais leve que a cauda da distribuição Exponencial, se os dados se distribuírem exponencialmente os valores de $e(x)$ apresentam uma forma linear.

Referências

- [1] Burneckı, K. et al. (2005) *Loss Distributions*, in *Statistical Tools for Finance and Insurance*, Springer.
- [2] Centeno, M. (2003). *Teoria de Risco na Actividade Seguradora* . Celta Editora.
- [3] Kaas, R. et al. (2009). *Modern Actuarial Risk Theory: using R* . Springer.
- [4] Reis, A. (2001). *Teoria da Ruína*. CEMAPRE-ISEG no 17/TA, Lisboa.
- [5] Reis, A. (2002). *O número de indemnizações até à ruína e recuperação*. CEMAPRE-ISEG, Lisboa.

Estimação robusta de equações simultâneas: o método dos momentos generalizados com informação limitada

(Poster)

Anabela ROCHA – *CIDMA e ISCA, Universidade de Aveiro*

anabela.rocha@ua.pt

Manuela Souto de Miranda – *CIDMA e DMat, Universidade de Aveiro*

João Branco – *CEMAT e IST, Universidade de Lisboa*

Resumo

Os modelos de equações simultâneas são modelos econométricos caracterizados por equações interdependentes, que generalizam os modelos de regressão multivariada na medida em que assumem heterocedasticidade e correlação entre erros e regressores. Tradicionalmente, no caso particular de existir homocedasticidade, os coeficientes são estimados pelo método dos mínimos quadrados em duas fases (*Two stages least squares*) ou pelo método dos mínimos quadrados em três fases (*Three stages least squares*). Quando há heterocedasticidade, o estimador determinado pelo método dos momentos generalizados (*GMM*), proposto por Hansen em 1982, veio a mostrar-se mais adequado às especificidades deste tipo de modelo.

Apesar do bom desempenho do estimador *GMM* perante os pressupostos do modelo, trabalhos mais recentes mostram que a investigação sobre sistemas de equações simultâneas continua ativa; veja-se por exemplo, Anderson *et al.* (2011), Iglesias and Phillips (2012), ou Youssef *et al.* (2013). Em particular, diversos autores têm focado a atenção em duas questões: conjugar as diferentes propostas com a estimação de informação limitada (ou seja, quando o processo de estimação é desenvolvido sucessivamente para cada equação); e/ou, por outro lado, encontrar estimadores robustos.

Uma versão robusta do estimador *GMM* foi desenvolvida em Rocha (2010). Posteriormente, um estudo de simulação apontou para a heterocedasticidade do modelo ser, de facto, pouco relevante na estimação robusta dos coeficientes com *GMM*. A questão é de grande importância, pois ignorar a heterocedastici-

dade sem prejuízos significativos permitiria reduzir fortemente a complexidade do processo de estimação e da própria aplicação do modelo. O presente trabalho alarga o estudo de simulação referido, investigando essa conjectura ao considerar novos cenários e procedendo à estimação de informação limitada, uma vez que desse modo se têm obtido melhores resultados em termos amostrais.

Referências

- [1] Anderson, T.W., Kunitomo, N. and Matsushita, Y. (2011) *On finite sample properties of alternative estimators of coefficients in a structural equation with many instruments*, Journal of Econometrics 165, 58-69.
- [2] Iglesias, E. and Phillips, G. (2012) *Improved instrumental variables estimation of simultaneous equations under conditionally heteroskedastic disturbances*, Journal of Applied Econometrics 27, 474-499.
- [3] Rocha, A. (2010) *Estimação robusta em modelos lineares de equações simultâneas*, Tese de Doutorado, Universidade de Aveiro.
- [4] Youssef, A.H., El-Sheikh, A.A., Mohammed, E.T.H. (2013) *New M-estimator objective function in simultaneous equations model (a comparative study)*, International Mathematical Forum 8, 1007-1022.

Sobreviver ao Cancro do Cólon e Reto

(Poster)

Mariana RODRIGUES – *Unidade de Investigação, SESARAM*

marianacr@gmail.com

Carina Alves – *Núcleo de Gestão de Doentes e Estatística, SESARAM*

Ana Maria Abreu – *CCCEE e CCM, Universidade da Madeira*

Resumo

De acordo com o Instituto Nacional de Estatística (INE), em 2008, registaram-se em Portugal 24 608 óbitos devido a tumores malignos, sendo a segunda maior causa de morte (23% do total de óbitos). No entanto, para a população entre os 45 e os 74 anos, a mortalidade por cancro passa para a primeira posição, com mais de 37% dos óbitos nessas idades. Em relação à incidência, o projeto GLOBOCAN revela que, no mesmo ano, em Portugal, foram registados 43 284 casos de cancro.

No que diz respeito ao cancro do cólon e reto, em 2008, ocorreram 6 952 novos casos e 3 691 óbitos em Portugal (fonte: GLOBOCAN), ou seja, cerca de 20 novos casos e 10 óbitos por dia! Além disso, segundo a mesma fonte, prevê-se que no ano de 2020, os valores ascendam a 8 178 novos casos e a 4 376 óbitos.

Esta realidade preocupante, constitui um forte motivo para os vários estudos que têm surgido sobre o assunto. Assim, o nosso contributo consistiu em fazer uma abordagem do ponto de vista da análise de sobrevivência, concretamente, através do estudo de uma base de dados de 800 indivíduos diagnosticados com cancro do cólon e reto, entre 2000 e 2009, na Região Autónoma da Madeira.

Sendo fundamental a deteção precoce desta doença, os resultados que obtivemos não são animadores pois dos estadios conhecidos 22% são do estadio IV. A covariável sexo não foi significativa no modelo de Cox que ajustámos. Nos tratamentos efetuados, a realização de cirurgia e de quimioterapia revelaram-se muito importantes para a sobrevivência, pois os doentes que não fizeram estes tratamentos apresentaram um risco de morte agravado em 63% e em 24%, respetivamente.

Aplicámos o modelo de cura ao estadio IV, pois só neste caso o tempo de *follow-up* foi suficiente (Yu *et al.*, 2004), sendo a mediana do tempo de vida de 267 dias e o tempo mínimo de *follow-up* de aproximadamente 3 anos e meio (1261 dias).

Considerámos as covariáveis quimioterapia e cirurgia na função de sobrevivência e a covariável quimioterapia no modelo de regressão logístico, tendo-se obtido uma taxa de cura de 3,6% para os indivíduos que não fizeram quimioterapia e de 4,4% para os que fizeram.

Referências

- [1] Yu, B., Tiwari, R.C., Cronin, K.A., Feuer, E.J. (2004) *Cure fraction from the mixture cure models for grouped survival data*, *Statistics in Medicine*. 23, 1733–1747.
- [2] GLOBOCAN (2008) *Cancer Incidence, Mortality and Prevalence Worldwide in 2008*, <http://globocan.iarc.fr/> (em 8/7/2013).

Medidas para avaliar a utilização de testes compostos

(Poster)

Rui SANTOS – *Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Leiria, CEAUL — Centro de Estatística e Aplicações da Universidade de Lisboa*

ruisantos@ipleiria.pt

João Paulo Martins – *Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Leiria, CEAUL — Centro de Estatística e Aplicações da Universidade de Lisboa*

Miguel Felgueiras – *Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Leiria, CEAUL — Centro de Estatística e Aplicações da Universidade de Lisboa, CIIC-IPL — Centro de Investigação em Informática e Comunicações do Instituto Politécnico de Leiria*

Resumo

A utilização de testes conjuntos (ou compostos) permite obter uma diminuição significativa no número de testes que é necessário realizar para a classificação de todos os indivíduos infetados em determinada população (recorrendo à metodologia de Dorfman (1943) e suas generalizações) ou para a obtenção de estimativas mais precisas para a taxa de prevalência de uma infeção (cf. Sobel e Elashoff, 1975). Por este motivo, os testes conjuntos são utilizados em análises clínicas, bem como no controlo de qualidade (por exemplo na amostragem para aceitação), com uma significativa poupança de recursos (cf. Boswell *et al.*, 1996). Contudo, a utilização de testes compostos pode igualmente implicar um aumento razoável da probabilidade de existência de erros de má classificação, que podem ser medidos pela especificidade e pela sensibilidade associadas ao processo selecionado (cf. Santos, Pestana e Martins, 2013). Neste trabalho, com base numa análise comparativa da performance dos testes compostos em algoritmos hierárquicos e algoritmos matriciais (cf. Kim e Hudgens, 2009), efetuada via simulação, são investigadas medidas que permitem inferir acerca da adequação da utilização de testes compostos. Deste modo, são identificadas medidas associadas à performance do teste individual e às distribuições que

caracterizam a substância em análise (de um indivíduo infetado e de um não infetado) de forma a permitir distinguir as aplicações para as quais a utilização de processos envolvendo testes compostos não acarreta o risco de uma elevada probabilidade de má classificação.

Referências

- [1] Boswell, M.T., Gore, S.D., Lovison, G., Patil, G.P. (1996). *Annotated bibliography of composite sampling, Part A: 1936-92*, Environ Ecol Stat 3, 1–50.
- [2] Dorfman, R. (1943). *The detection of defective members of large populations*, Ann. Math. Statist. 14, 436–440.
- [3] Kim, H., Hudgens, M. (2009). *Three-Dimensional Array-Based Group Testing Algorithms*, Biometrics 65, 903–910.
- [4] Santos, R., Pestana, D. and Martins, J.P. (2013). *Extensions of Dorfman's Theory*, P.E. Oliveira et al. eds., *Recent Developments in Modeling and Applications in Statistics, Studies in Theoretical and Applied Statistics, Selected Papers of the Statistical Societies*, Springer, 179–189.
- [5] Sobel, K.M., Elashoff, R.M. (1975). *Group testing with a new goal, estimation*, Biometrika 62, 181–193.

Agradecimentos

Trabalho financiado por Fundos Nacionais através da Fundação para a Ciência e a Tecnologia, no âmbito do projeto PEst-OE/MAT/UI0006/2011.

Indicadores estatísticos para avaliação do desempenho portuário

(Poster)

Catarina SILVA – *Universidade de Aveiro*

catarinarsilva@ua.pt

Isabel Pereira – *Departamento de Matemática e CIDMA, Universidade de Aveiro*

Luís Sousa – *Porto de Aveiro*

Resumo

O tratamento estatístico portuário pode tornar-se bastante complexo principalmente porque a atividade portuária é composta por termos e tarefas bastante específicas. O principal objetivo deste trabalho consiste em averiguar qual a dimensão da estatística no sistema portuário, quais as ferramentas estatísticas mais adequadas para tratar este tipo de dados e quais os indicadores que melhor permitem avaliar o desempenho da atividade portuária. Para tal são usados os dados do movimento portuário do porto da Figueira da Foz no ano 2012. As variáveis em estudo são a quantidade de mercadoria movimentada, o comprimento dos navios acostados no porto e a arqueação bruta dos mesmos. Realiza-se primeiramente uma proposta de alteração dos indicadores portuários, indicadores estes que são centrados na medição da quantidade de mercadoria movimentada e no número e características dos navios que atracam o porto. De seguida, usando testes não paramétricos compara-se o comportamento das variáveis nos diferentes trimestres do ano 2012. Por último, usa-se a análise de regressão para verificar quais os fatores que mais influenciam o desempenho da atividade portuária medido através da quantidade de mercadoria movimentada. O conjunto de dados é referente às características do porto da Figueira da Foz nos anos 2002 até 2012, e os fatores em estudo são: a entidade governadora (que varia entre Administração do Porto de Aveiro, S.A. e IPTM, IP - Instituto Portuário e dos Transportes Marítimos), a dimensão do porto, o indicador económico (PIB/capita), a classificação das gruas a nível de fiabilidade e generalização, o grau de investimento, a dimensão média dos navios, a taxa de unitização e a taxa de contentorização.

Estudo do Nível de Conforto na Interação de Contacto entre a Pele e Equipamentos Mecânicos

(Poster)

Filomena TEODORO – *Departamento de Matemática, Escola Superior de Tecnologia de Setúbal, Instituto Politécnico de Setúbal, Portugal e CEMAT, Instituto Superior Técnico, Universidade Técnica de Lisboa, Portugal*
maria.teodoro@estsetubal.ips.pt

Célio Figueiredo – *Departamento de Engenharia Mecânica, Escola Superior de Tecnologia de Setúbal, Instituto Politécnico de Setúbal, Portugal*

Paula Silva – *Departamento de Engenharia Mecânica, Escola Superior de Tecnologia de Setúbal, Instituto Politécnico de Setúbal, Portugal*

Resumo

Para um desenvolvimento adequado de equipamentos mecânicos é fundamental ter em consideração o conforto do utilizador [1]. A estimativa do contacto entre a pele e o equipamento permite-nos garantir níveis de conforto adequados. Deve ser feita a caracterização das propriedades de tecidos moles e nível de tolerância à pressão, deformação e força aplicadas, assim como da capacidade de construir equipamentos confortáveis [3]. Para a simulação do contacto foram obtidos alguns modelos não lineares [2]. Oitenta indivíduos jovens foram sujeitos a ensaios de compressão e descompressão considerando diferentes diâmetros de indentadores. Foi recolhida e tratada alguma informação sobre o estado físico e psicológico de cada indivíduo interveniente na experiência, aplicando técnicas de estatística multivariada.

Referências

- [1] Pons J.L. (2008), *Wearable Robots: Biomechatronic Exoskeletons*, Wiley Blackwell.
- [2] Seber, G. A. F., Wild, C. J. (1989), *Nonlinear Regression*, Wiley Blackwell.
- [3] Silva, P., Monteiro, A. , Bernardo, I., Claudio, R., Figueiredo-Pina, C. (2012), *Measuring Discomfort: From Pressure Pain Threshold To Soft Tissues Deformation*, *Journal of Biomechanics* 45 S1 S576.

Impact of value communication in decision making: an approach to the treatment of type 2 diabetes in Portugal

(Poster)

Daniel VIRIATO – *Novartis Farma-Produtos Farmacêuticos S.A., Portugal*
daniel.viriato@novartis.com

Marília Antunes – *Faculdade de Ciências da Universidade de Lisboa-
Departamento de Estatística e Investigação Operacional and CEAUL*

Abstract

The increasing budgetary constraints faced by healthcare payers have made the access decision increasingly complex and decentralized processes a growing trend. Thus, demonstration of new health technologies value has become critical to ensure patients' access to innovation.

The main goal of this study was to identify the most important value messages within decision making process in the treatment of Type 2 Diabetes Mellitus (T2DM) among General Practitioners (GP) coordinating Primary Care Centers.

For that purpose, regional meetings based on Health Economics (HE) training are taking place during 2013 with GPs, including the presentation of a cost-effectiveness analysis of vildagliptin vs sulphonilureas (SU) and a value communication tool that enables the estimation of meaningful impacts of T2DM.

Two surveys were developed to be applied at the beginning and at the end of training, in order to characterize the learning insights and to identify the factors considered in the decision process.

In this work, partial results based on a 51 GPs sample are presented. The positive impact of the training in HE concepts knowledge was statistically significant ($p=0.000$) with an 89.4% average rate of correct answers after training versus 10.59% before training.

More than 50% of GPs considered the increasing prevalence of T2DM, cost of macrovascular and microvascular complications and the impact of hypoglycaemias in patients' quality of life (QoL) and costs, as the most important factors in the decision process. The key decision factor to use vildagliptin was found to be the improvement of patients' QoL. The use of SU is explained due to its fast response, mainly in patients with Hb1Ac level above 7.5%.

Majority of GPs (87.5%) chose vildagliptin as the preferred therapeutic for T2DM patients not controlled with metformin.

This analysis demonstrated that both epidemiological and economic factors are important at local decision process. Clinically, patients' QoL is the most relevant achievement. Most GPs prefer vildagliptin to control T2DM, considering QoL improvements its greatest benefit.

Autores

- Abecasis, Francisco, 312
Abreu, Ana Maria, 236, 321
Afonso
 Anabela, 114, 290
 Lourdes Belchior, 92
Afreixo, Vera, 51
Aires-de-Sousa, Marta, 255
Akanda, Md. Abdus Salam, 235
Alarcão, Violeta, 297, 301
Aleixo, Sandra M., 112
Almeida
 João, 21
 Rute, 291
Alpizar-Jara, Russell, 114, 118, 235
Alves
 Carina, 236, 321
 Maria João, 26
 S., 244, 246
 Sandra Maria, 32
Amado, Conceição, 128, 210
Amaral Turkman
 Antónia, 195
 Antónia, 193
 M. A., 108
 Maria Antónia, 57
 Maria Antónia, 4
Amorim, Ana, 115
Andreozzi, Valeska, 282
Antunes
 Marília, 26, 193, 297, 310, 327
 Nelson, 86
Areal, Nelson, 238
Areias, Cristiana, 240
Athayde, Emilia, 116
Azevedo
 Cecília, 152, 169
 Manuela, 118, 242
Babo, Maria de Lurdes, 238
Bailey, Trevor C., 32
Banha, Rui, 143
Barros, Carlos, 254
Bastos, Carlos A. C., 51
Bergel, Agnieszka, 95
Bermudez, P. Zea, 20
Bernardino, Alexandre, 63
Bettencourt Silva, Marília, 240
Bianco, Ana M., 67
Boente, Graciela, 67
Bowman, Adrian, 7
Branco
 João A., 64
 João, 319
Brasil, Jorge, 159
Braumann, Carlos A., 8, 85, 123

Brilhante, M. Fátima, 284
 Brites
 Nuno, 85
 Nuno Miguel, 118, 242
 Brito, Irene, 317
 Brás, Ana L., 276
 Buescu, Jorge, 135

Cabral, Ricardo, 63
 Cadima, Jorge, 120
 Canova, F., 244, 246
 Canto e Castro, Luísa, 143
 Cardoso
 H., 229
 João Lourenço, 195
 Rui, 92
 Rui M.R., 99

Carinhas, Dora, 249
 Carita, Ana Isabel, 121
 Carlos, Clara, 85, 123
 Carmo, Manuel do, 126
 Carrasquinha, Eunice, 128
 Carvalho
 Alda, 130, 147
 Maria José, 132

Caseiro, Paulo, 268
 Casquilho, Miguel, 135
 Castro Ribeiro, Manuel, 46
 Catela Nunes, Luis, 103
 Clark, Taane, 17
 Coelho
 Rui, 138, 251

Conde, Jorge, 268
 Costa
 Marco, 253
 Maria da Conceição, 139

Costeira, João Paulo, 63
 Cotrim, Ricardo, 143
 Crujeiras, Rosa M., 39

Damáσιο
 Bruno, 140, 254
 Fábio, 130

de Haan, Laurens, 179
 de Zea Bermudez, Patrícia, 26, 108, 282
 Diamantino, Fernanda, 255
 Dias
 João, 273
 Maria João, 142
 Sandra, 220

Dionísio, João, 190
 Drakeley, Chris, 17
 Duarte
 André, 314
 Joana, 143

Duarte Silva, Pedro, 53

Economou, Theodoros, 32
 Egídio dos Reis
 Alfredo, 92
 Alfredo D., 95, 99

Engana, Teresa, 290
 Erzin, Karim, 251

Félix, Jorge, 21, 34
 Faria
 Fernando, 258
 Liliana, 146
 Susana, 146, 260

Felgueiras, Miguel, 323
 Fernandes
 Filipe Salvador, 147
 Milene, 310

Fernandez-Carvalho, Joana, 251
 Ferreira
 Ana, 271, 273
 Dário, 263
 Helena, 73
 Marta, 73, 261

- Sandra Maria Bargão Saraiva, 263
Sofia, 312
- Figueiredo
Adelaide Maria, 265
Célio, 325
Fernanda, 215, 273
Fernanda Otília, 149, 265
Jc Paulo, 268
João Paulo, 271
- Filipe
Patrícia, 85, 276
Patrícia A., 240
- Fonseca, Miguel, 150, 303
- Frazão, Filomena, 287
- Freitas
Adelaide, 132, 142
JorgeMilhazes, 75
Rita, 273
- García-Soidán, Pilar, 42
- Gil, Cláudia, 268
- Gomes
Dora, 44
Dulce, 240, 276
Fernando, 152
João, 301
M. Ivette, 155
Maria Ivette, 149
- Gonçalves
A. Manuela, 278
Elsa, 157
Esmeralda, 164, 203, 280
Luzia, 20, 213
Patrícia, 317
Rui, 158
- Gonçalves Dias, José, 293
- Goulão, Beatriz, 282
- Gouveia, Catarina, 284
- Gouveia-Reis, Délia, 285
- Grosso, Ana Rita, 55
- Guedes, Maria do Carmo, 159
- Guerra, Manuel, 87
- Guerreiro
Gracinda Rita, 92, 97
Guerreiro Lopes, Luiz, 285
- Henriques
Carla, 162
Marisa Jesus, 287
- Infante, Paulo, 126, 138, 215, 249, 258,
273, 290
- Jacinto, Gonçalo, 86, 273, 290
- Klein Tank, Albert, 179
- Kressner, Daniel, 167
- Leite
Argentina, 291
Joana, 164
- Lima, S., 244, 246
- Loja, Maria Amélia, 130
- Lopes, Elisa, 301
- Lourenço, Raphael, 166
- Macedo, Francisco, 167
- Maló, L., 244, 246
- Manjurano, Alphaxard, 17
- Marcelo, Carlos, 169
- Margalho, Luís, 171
- Marques, Catarina, 173, 293
- Martins
Antero, 157
Cristina M., 280
João Paulo, 208, 323
- Martins da Silva, A., 229
- Marto, Marco, 295
- Matos, Ana Cristina, 162

- Meireles, Daniel, 312
 Mendes
 Jorge, 126
 Jorge M., 29
 Mendes Silva, Margarida, 210
 Mendes-Lopes, Nazaré, 164, 203, 280
 Mendonça
 D., 229
 Jorge, 175
 Sandra, 285
 Sílvia, 297
 Teresa, 197
 Meneses
 Carlos, 314
 R.F., 229
 Menezes, Raquel, 42, 171, 186
 Mexia
 João T., 263
 João Tiago, 97, 150, 178, 303
 Miguens, Maria de Fátima, 97
 Milheiro de Oliveira, Paula, 192
 Minder, Christoph, 178
 Miranda
 Ana, 199
 João Luís de, 135
 Miranda Guedes, José, 130
 Molenberghs, Geert, 15
 Monteiro
 Ana Margarida, 105, 182
 Magda, 176, 253
 Moreira Freitas, Ana Cristina, 78
 Moreira, Elsa, 44, 150, 178
 Mota, Tiago, 190
 Natário, Isabel, 299
 Neves
 Cláudia, 179
 M. Manuela, 44
 Manuela, 305
 Maria Manuela, 181
 Nicolau, João, 140, 182, 254
 Niehus, Manfred, 190
 Nunes
 Carla, 240, 276
 Cláudia, 87
 Célia, 263
 Sandra, 305
 Oiko, Carla, 301
 Oliveira
 Carlos, 87
 Irene, 201
 M. Rosário, 20
 María, 39
 Raquel, 278
 Sandra, 303
 Oliveira, M. Rosário de, 213
 Osório, Hugo, 26
 Pacheco, António, 86, 167, 187
 Pais Ribeiro, J., 229
 Papoila, Ana, 199
 Pascoal, Rui, 105
 Paulino, Carlos Daniel, 15
 Pedro, L., 229
 Penalva, Helena, 305
 Pereira
 Isabel, 106, 139, 176, 295, 324
 Luísa, 308
 Maria João, 46
 Miguel, 55
 Paula, 183
 Piairol, Helena, 186
 Pimentel, Rita, 187
 Pina, Maria Fátima, 32
 Pinheiro, Bernardete, 310
 Pinto
 Alexandra, 312

- Constantino, 142
 Iola, 190
 Iola M.S., 314
 Pinto Vasconcelos, Marta, 271
 Pires, Ana M., 128
 Poletto, Frederico, 15
 Poncela, Pilar, 9
 Portela, Conceição, 146
 Prior, Ana Filipa, 192
 Proença, Maria da Conceição, 26
 Pérez, Jesús M, 114
 Queirós, Celine, 317
 Rabiais, Sara, 34
 Ramos
 Alexandra, 79
 Sandra, 193
 Rebelo, Maria Teresa, 26
 Ribeiro
 Conceição, 195
 Inês, 271
 Rocha
 Ana Paula, 291
 Anabela, 319
 Conceição, 197
 Rodríguez-Casal, Alberto, 39
 Rodríguez-Martínez, Eugenio V., 99
 Rodrigues
 Isabel M., 67
 João M. O. S., 51
 Mariana, 321
 Rosado, Fernando, 181
 Rubiños-López, Oscar, 42
 Ruiz, Esther, 9
 Ruíz, Paulo, 314
 São João, Ricardo, 199
 Sagitov, Serik, 88
 Sanfins
 Gabriel Rodrigues, 166
 Marco Aurélio, 166
 Santos
 Luís Ferreira dos, 162
 Miguel N., 138, 251
 Rui, 208, 323
 Scotto, Manuel, 106, 139, 176
 Sebastião, Fernando, 201
 Sepúlveda, Nuno, 17
 Sequeira, Fernando, 284
 Serra, Maria Conceição, 88
 Shruballs, Sílvia, 299
 Silva
 Catarina, 324
 Cristina, 242
 Filipa, 203
 Giovani, 205
 Isabel, 227
 Maria do Pilar, 249
 Maria Eduarda, 106, 159, 197, 227,
 229, 238, 291
 Maria João, 34
 Maria João, 21
 Nélia, 287
 Paula, 325
 Sérgio, 130, 261
 Silva-Fortes, Carina, 57
 Simões Coelho, Pedro, 29
 Simões, Rui, 297, 301
 Singer, Julio, 15
 Soares
 Amílcar, 46
 Paulo, 205
 Sobral, José, 147
 Sobreira, Sobreira, 103
 Soromenho, Paulo, 260
 Sousa
 Fernanda, 69
 Inês, 171

Inês, 186

Lisete, 55, 57

Luís, 324

Ricardo, 208

Souto de Miranda, Manuela, 210, 319

Subtil, Ana, 20, 213

Sá Carvalho, Marília, 32

Teixeira

Renata, 215

V., 244, 246

Teles, Júlia, 112, 217

Temido, Maria da Graça, 220

Tenreiro, Carlos, 223

Teodoro, Filomena, 325

Tomé, Margarida, 295

Torrado, Nuria, 225

Torres, Cristina, 227

Turkman

K. F., 108

Kamil Feridun, 183

Uña-Álvarez

J. de, 115

Jacobo de, 175

Vandewalle, Björn, 21, 34

Vasconcelos, Rosa M., 278

Vicente, Paula, 173

Vilhena, Estela, 229

Viriato, Daniel, 327

Weiß, Christian, 106