



Reunião Anual da
ASSOCIAÇÃO PORTUGUESA DE CLASSIFICAÇÃO
E ANÁLISE DE DADOS (CLAD)

LIVRO DE PROGRAMA E RESUMOS

ESTBarreiro/IPS, 10 e 11 de abril de 2015

Ficha Técnica

Presidente das Jornadas

José Dias (Presidente da CLAD)

Secretário das Jornadas

Otília Dias (ESTBarreiro/IPS)

Comissão Organizadora

Ana Sousa Ferreira (Universidade de Lisboa)

Anabela Marques (ESTBarreiro/IPS)

Clara Carlos (ESTBarreiro/IPS)

Otília Dias (ESTBarreiro/IPS)

Raquel Barreira (ESTBarreiro/IPS)

Telma Guerra Santos (ESTBarreiro/IPS)

Título: XXII Jornadas de Classificação e Análise de Dados
(JOCLAD 2015).
Livro de Resumos.

Produzido: Instituto Nacional de Estatística

Editores: Ana Sousa Ferreira, Anabela Marques, Clara Carlos,
Otília Dias, Raquel Barreira, Telma Guerra Santos

ISBN: 978-989-98955-1-5

Prefácio

Iniciarei este Prefácio com uma retrospectiva histórica da Associação Portuguesa de Classificação e Análise de Dados (CLAD). A CLAD nasceu em 1994, em primeiro lugar, da vontade de um grupo de (futuros) associados, decididamente manifestada durante as I Jornadas de Classificação e Análise de Dados, as JOCLAD'93, então organizadas pelo Laboratório de Estatística e Análise de Dados da Faculdade de Psicologia e Ciências da Educação da Universidade de Lisboa, Helena Bacelar Nicolau, Ana Sousa Ferreira e Otilia Dias e que teve uma forte participação de colegas da Société Francophone de Classification (SFC). A grande receptividade e apoio vinda da SFC e de outras sociedades estrangeiras congêneres, especialmente a Britânica e a Norte-Americana, animou este grupo, do qual tive a honra de fazer parte, e a que se juntou Fernando Costa Nicolau, a trabalhar arduamente no sentido de, não só termos a nossa associação rapidamente legalizada, mas também estarmos em condições de solicitar a adesão à International Federation of Classification Societies (IFCS), o que se efetivou ainda antes da CLAD completar um ano de idade.

Ao longo destes anos, uma parte importante da atividade nacional e internacional da CLAD enquadrou-se no âmbito de projetos de intercâmbio apoiados inicialmente pela Embaixada de França, o British Council, a Fundação Luso-Americana para o Desenvolvimento e a JNICT (atual FCT). A participação estrangeira nas JOCLAD era constante e contava sempre com vários congressistas de reconhecido mérito científico internacional, vindos da Alemanha, Bélgica, França, Reino Unido e USA. As conferências convidadas e as comunicações livres, sobre temas de investigação e desenvolvimento de metodologias, eram de grande qualidade e interesse e as aplicações, incluídas nas comunicações ou apresentadas em *posters*, abarcavam campos francamente diversificados e estavam presentes através da participação de alguns “curiosos”/”estudiosos” muito interessados nesses novos métodos de Análise de Dados no âmbito da sua vida profissional. Entretanto, a atividade científica em Portugal sofreu alterações e a investigação passou a ser menos apoiada, pelo que a atuação da CLAD começou a centrar-se mais nos diferentes grupos de investigação com atividade científica relevante nas áreas de atuação da CLAD, sediados em instituições de ensino superior português, e as apresentações de aplicações por parte de utilizadores “estranhos” à comunidade científica foram diminuindo.

Os anos foram-se passando: entraram novos sócios, os órgãos sociais foram-se alterando com a participação de sócios mais novos, mas as Jornadas Científicas da CLAD têm tido lugar anualmente, sem qualquer interrupção, com mais ou menos participantes.

No ano passado, nas XXI Jornadas de Classificação e Análise de Dados, a CLAD comemorou os 20 anos de vida, e ao fazer esta retrospectiva lembrei-me de um artigo que publiquei num dos primeiros Boletins Informativos da CLAD e que pela sua atualidade aqui deixo:

“Uma das leis do movimento, conhecida por primeira lei de Newton, diz-nos que um objeto em descanso permanecerá em descanso e um objeto em movimento continuará em movimento com uma velocidade constante, isto é, a velocidade constante em linha reta, a não ser que sofra uma força exterior. Esta lei de Newton também é conhecida por lei da inércia, e é assim que me lembro dela quando a estudei no liceu. Inércia é a propriedade da matéria relacionada com a tendência de um objeto permanecer em repouso ou em movimento uniforme. Não sei porquê mas nunca mais me esqueci destas ideias sobre inércia e gosto de aplicá-las a situações que não são da física. A CLAD é um corpo em movimento, e se queremos que o seu movimento não seja apenas retilíneo e a velocidade constante precisamos de lhe aplicar novas e estimulantes forças, até porque a massa da CLAD é cada vez maior e isso mede a sua inércia. Maior inércia, mais força é necessário para a mover. Aqui o paralelo com as leis da física começa a escapar-me, não sei se as forças que precisamos de aplicar são externas ou internas. Precisamos das duas, certamente. Precisamos de impulsionar a CLAD de várias maneiras porque os desafios, solicitações, ideias e projetos são muitos. As forças da CLAD são o trabalho coletivo dos seus sócios. Hoje, porque a CLAD cresceu é cada vez mais necessária essa ação nos vários campos de intervenção e de trabalho da nossa associação. Mas hoje, ser sócio da CLAD é também ter o privilégio de poder participar em vários tipos de trabalho coletivo verdadeiramente aliciantes e criativos. Do Boletim à Revista com Teses de Mestrado e Doutoramento, passando pelos Cursos e Seminários, Reflexões sobre Estatística e Análise de Dados, há toda uma diversidade de tarefas e formas de intervir que esperam por cada um de nós. E hoje já não há distâncias, e por isso as inércias individuais também podem ser menores. É por isso que continuamos a procurar fazer todos os esforços para podermos fazer dela mais do que um veículo de comunicação, pois ela poderá ser um grande e diversificado espaço de trabalho coletivo.”
Ontem, como hoje nada me parece mais atual para continuar a dar vida a esta Associação.

A Comissão Organizadora agradece aos professores convidados, a todos os autores e moderadores de sessões, aos membros da Comissão Científica, aos colegas que procederam à revisão dos trabalhos que constam deste livro e aos participantes em geral. Uma nota particular ao Professor Gilles Celeux, INRIA, parceiro da CLAD desde o seu início e à Professora Cristina Rocha da Faculdade de Ciências da Universidade de Lisboa que lecionam os dois minicursos destas Jornadas, bem como ao Professor Ruy Ribeiro com quem contactámos no

Los Alamos National Laboratory, EUA e que aceitou prontamente ser orador convidado de uma sessão plenária na área de que é responsável na Faculdade de Medicina da Universidade de Lisboa, Bioestatística, e à Professora Fernanda Figueiredo da Faculdade de Economia da Universidade do Porto oradora convidada da sessão plenária na área de Controlo de Qualidade. Agradecemos ainda ao Instituto Nacional de Estatística e ao Banco de Portugal que, mais uma vez, para além de patrocinarem este evento científico se aliam através dos seus corpos técnicos, organizando algumas das Sessões Temáticas que constam do Programa.

Por último, desejamos agradecer a todas as entidades que, direta ou indiretamente, apoiaram ou patrocinaram estas Jornadas.

O nosso obrigada a todos!

Barreiro, Abril de 2015

P'la Comissão Organizadora JOCLAD 2015

Otília Dias

ORGANIZAÇÃO

Presidente das Jornadas

José Dias (Presidente da CLAD)

Secretário das Jornadas

Otília Dias (ESTBarreiro/IPS)

Comissão Organizadora

Ana Sousa Ferreira (Universidade de Lisboa)

Anabela Marques (ESTBarreiro/IPS)

Clara Carlos (ESTBarreiro/IPS)

Otília Dias (ESTBarreiro/IPS)

Raquel Barreira (ESTBarreiro/IPS)

Telma Guerra Santos (ESTBarreiro/IPS)

Comissão Científica

Adelaide Figueiredo (Universidade do Porto)

A. Manuela Gonçalves (Universidade do Minho)

Anabela Marques (ESTBarreiro/IPS)

Ana Sousa Ferreira (Universidade de Lisboa)

Ana Lorga da Silva (Universidade Lusófona)

Carlos Ferreira (Universidade de Aveiro)

Carlos Soares (Universidade do Porto)

Catarina Marques (Instituto Universitário de Lisboa)

Clara Carlos (ESTBarreiro/IPS)

Conceição Amado (Universidade de Lisboa)

Fátima Salgueiro (Instituto Universitário de Lisboa)

Fernanda Sousa (Universidade do Porto)

Fernando Nicolau (Universidade Nova de Lisboa)

Helena Bacelar-Nicolau (Universidade de Lisboa)

Irene Oliveira (Universidade de Trás-os-Montes e Alto Douro)

Isabel Silva Magalhães (Universidade do Porto)

Jorge Cadima (Universidade de Lisboa)

José Gonçalves Dias (Instituto Universitário de Lisboa)

José Palma (ESTSetúbal/IPS)

Manuela Neves (Universidade de Lisboa)

Margarida Cardoso (Instituto Universitário de Lisboa)
Luís Miguel Grilo (Instituto Politécnico de Tomar)
Paula Brito (Universidade do Porto)
Paula Vicente (Instituto Universitário de Lisboa)
Paulo Infante (Universidade de Évora)
Pedro Campos (Universidade do Porto)
Pedro Duarte Silva (Universidade Católica Portuguesa)
Susana Faria (Universidade do Minho)
Victor Lobo (Universidade Nova de Lisboa)

APOIOS



PROGRAMA

1º DIA, SEXTA-FEIRA, 10 DE ABRIL DE 2015

9:30 Minicurso

Sala 0.02 Minicurso Presidente: José Dias	Sala 0.07 Minicurso Presidente: Otília Dias
Gilles Celeux (INRIA Saclay-Île-de-France, França) Model-based Clustering - presentation of the Mixmod software	Cristina Rocha (Faculdade de Ciências, UL, Portugal) Análise de Dados nas Ciências da Saúde.

10:30 Pausa para café

11:00 Minicurso (*continuação*)

12:45 Almoço

12:00 Registo e entrega de documentação

14:00 Sessão de Abertura das Jornadas – Anf. 0.01

**Diretor da ESTBarreiro/IPS, Secretário da Comissão Organizadora,
Presidente da CLAD, Presidente da CMBarreiro, Presidente do IPS**

14:30 Sessão Plenária I – Anf. 0.01

Ruy Ribeiro (Los Alamos National Laboratory, EUA e FMUL, Portugal)

**Fitting viral dynamic models to longitudinal data using a population based
approach**

Moderador: Helena B. Nicolau

15:30 Sessão Temática I – Instituto Nacional de Estatística

	Anf. 0.01 Sessão INE –Desafios nas Estatísticas Oficiais IV Moderador: Carlos Marcelo
15:30	Cristina Neves Estatísticas do Comércio Internacional: como se obtêm os dados finais?
15:45	Sónia Torres O cálculo das estimativas mensais de emprego e desemprego divulgadas em 2014
16:00	Eduarda Góis, Esperança Figueiredo A riqueza e o rendimento das famílias em Portugal
16:15	Eduarda Góis Taxas de pobreza ancoradas no tempo

A. Manuela Gonçalves, Marco Costa, Joana Silva

A metodologia *Bootstrap* associada ao método de Holt-Winters na previsão de séries temporais

Paulo Infante, Gonçalo Jacinto, Anabela Afonso, José Conde, Teresa Engana, Jorge Vicente, Pedro Silva, Ana Carla Coelho, Paula Costa

Alguns benefícios do envelhecimento ativo: um caso de estudo

Anabela Afonso, Maria Filomena Mendes, Paulo Infante

Alguns motivos para indivíduos entre os 40 e os 49 anos quererem ter filhos

Áurea Sousa, Maria da Graça Batista, Ana Carolina Arruda

Análise das perceções sobre a liderança num organismo público dos Açores com recurso a métodos quantitativos

Ana Lorga da Silva, Artur Parreira

Análise de um Teste de Motivação: Comparação em Países Lusófonos

Carla Santos, Célia Nunes, Cristina Dias, João Tiago Mexia

Building up complex COBS

O. Silva, S. N. Caldeira, M. Mendes, S. P. Botelho, M. J. Martins

Caracterização da praxe académica com recurso à análise de dados

José Guerreiro, Carla Torre, Zilda Mendes

Consumo Distrital de Antidiabéticos em Portugal – Exemplo Prático da Análise Classificatória

Ricardo Jorge Correia, José Gonçalves Dias, Mário Sérgio Teixeira

Determinants of firm performance: a structural equation approach

A. Manuela Gonçalves, Rui Gomes

Estudo do Papel Mediador de Variáveis Psicológicas na Relação entre a Intenção e a Prática de Exercício Físico

Cristina Dias, João Tiago Mexia, Carla Santos, Célia Nunes

Models of stochastic symmetric matrices with truncation

Susana Faria, A. Rui Gomes

O Papel Mediador da Avaliação Cognitiva na Relação entre Stress e Saúde Psicológica

Isabel Preto, Mário Freitas, Luís Dourado, Michael McCall

Participação de crianças e adolescentes no planeamento das suas localidades

Ana Sousa Ferreira, Rosário Lima

Perceção de barreiras na carreira e diversidade de género: Uma abordagem multivariada numa amostra de trabalhadores do sector energético

17:00 Sessões Paralelas I

	Anf. 0.01 Classificação, Regressão e Processamento de Imagem Moderador: Manuela Neves	Sala 0.02 Análise de Dados em Educação Moderador: Anabela Marques	Sala 0.07 Análise de Dados em Ciências Empresariais e Sociais Moderador: Isabel Silva Magalhães
17:00	Eunice Carrasquinha, Conceição Amado, Ana M. Pires A new cleaning method for image restoration using robust confidence intervals	Margarida Vilela, M. Rosário Oliveira Escolas Secundárias atípicas no panorama nacional	Margarida Vilela, Conceição Amado, Beatriz Sousa, Jéssica Vieira, Luís Medeiros Previsão de Churners: Quem? E Quando?
17:20	Pedro Maia, António Abreu Real Time Activity Recognition for Cycling with Smartphone	Susana Faria, Conceição Portela Avaliação do desempenho dos alunos Portugueses em Matemática – PISA 2009	Lídia Patrícia Tomé, Maria Filomena Mendes Multidimensional relationship: education and fertility
17:40	Conceição Castro, Fernanda A. Ferreira A hedonic price approach for hotels in Porto	Otília Dias, Raquel Pereira, Boguslawa Sardinha Papel da ação social no Ensino Superior Politécnico: expetativas futuras dos estudantes bolseiros com recurso à Análise de Multivariada de Dados	Filipe Ribeiro, Maria Filomena Mendes A mortality study under different selected frameworks

18:00 Sessão Temática II – Banco de Portugal

	Anf. 0.01 Economia e Finanças Moderador: S. Filipa Lima
18:00	Patrícia Antunes, Rita Pisco Casting a light on shadow banking activity in Portugal
18:20	Rodrigo Batista, Ricardo Colaço Financing of NFC – A comparison with other economic indicators

18:40 Assinatura do Protocolo com a Associação Portuguesa de Demografia

18:50 Reunião da Assembleia Geral da CLAD – Anf. 0.01

19:30 Jantar das Jornadas ou IGNITE (19:00) – ESTBarreiro/IPS

2º DIA, SÁBADO, 11 DE ABRIL DE 2015

9:45 Sessões Paralelas II

	<p>Sala 0.02</p> <p>Aplicação de Métodos de Regressão</p> <p>Moderador: Paulo Infante</p>	<p>Anf. 0.01</p> <p>Métodos de Classificação</p> <p>Moderador: Susana Faria</p>	<p>Sala 0.07</p> <p>Análise de Dados em Economia, Gestão e Ensino</p> <p>Moderador: Adelaide Figueiredo</p>
9:45	<p>Andreia Maciel, Maria Filomena Mendes, Paulo Infante, Anabela Afonso, Rita Freitas</p> <p>Um perfil de quem não tem nem quer ter filhos em Portugal</p>	<p>André Maia, Tiago Cunha, Carlos Soares, Pedro Henriques Abreu</p> <p>TweetProfiles@Sapo: relating Spatio-Temporal patterns on Twitter with Sapo content</p>	<p>Paula Vicente, Catarina Marques, Elizabeth Reis</p> <p>Are callbacks a waste of time in mCATI surveys?</p>
10:05	<p>Carla Henriques, Joaquim Antunes, Sónia Duarte</p> <p>Recurso a produtos dietéticos e prática de exercício físico – Quem? Porquê? Como?</p>	<p>Maria José Amorim, Margarida Cardoso</p> <p>Distribuições de índices de concordância entre agrupamentos</p>	<p>José Gonçalves Dias</p> <p>Uncertainty analysis in conjoint studies</p>
10:25	<p>Ana Cristina Bico Matos, Carla Henriques, Bernardo Faria, Manuel Pestana</p> <p>Estudo de biomarcadores de uma doença renal</p>	<p>Fernanda Sousa</p> <p>Distribuições de Coeficientes de Comparação de Dendrogramas</p>	<p>João Martins, Luís Merca</p> <p>Data Warehouse de suporte à melhoria contínua do ensino na Escola Naval</p>

10:45 Sessão Plenária II – Anf. 0.01

Fernanda Figueiredo (FEP e CEAUL, Portugal)

Controlo Estatístico de Processos: Reamostragem em Planos de Aceitação

e

Cartas de Controlo

Moderador: Fernanda Sousa

11:45 Pausa para café

12:00 Sessão Temática III – Importância da Estatística na Educação

	Anf. 0.01 <u>Importância da Estatística na Educação</u> Moderador: Ana Sousa Ferreira
12:00	Luísa Canto e Castro Loura Desafios Estatísticos nas Estatísticas da Educação
12:20	Olívia Sousa A Importância da Estatística na Educação
12:40	Isabel Flores Modelling Efficiency in Education

13:00 Almoço

14:00 Sessão Temática IV – Cartas de Controlo de Qualidade

	Anf. 0.01 <u>Cartas de Controlo de Qualidade</u> Moderador: Fernanda Figueiredo
14:00	Luís Grilo Control charts for individual measurements of peroxide index.
14:20	Paulo Infante Monitorização de tempos de procedimentos académicos
14:40	Adelaide Figueiredo Monitorização da variabilidade de um processo skew-normal multivariado

15:00 Sessões Paralelas III

	Anf. 0.01 <u>Modelos Longitudinais</u> Moderador: Paula Vicente	Sala 0.02 <u>Análise de Dados em Controlo de Qualidade</u> Moderador: Luís Grilo
15:00	Isabel Silva, Maria Eduarda Silva, Cristina Torres Diagnostic tools for bivariate time series of counts	C. Saraiva, I. Oliveira, C. Martins, J. Ventanas, C. García A seleção de compostos voláteis como indicadores da qualidade sensorial da carne fresca
15:20	Paula C. R. Vicente, Maria de Fátima Salgueiro O efeito da existência de omissões nas estimativas de um modelo com trajetória latente linear: Um estudo de simulação	Dora Carinhas, Paulo Infante, Pilar Pestana Controlo estatístico da qualidade em ensaios interlaboratoriais – caso particular de nutrientes em águas salinas

15:40 Sessão Plenária III – Anf. 0.01

Gilles Celeux (Université Paris-Sud e INRIA, França)

Mixture model selection criteria focussing on a clustering purpose

Moderador: Ana Sousa Ferreira

16:40 Sessão de Encerramento das Jornadas –Anf. 0.01

17:00 Passeio com Moscatel de Honra

RESUMOS

Índice

Model-based clustering - presentation of the mixmod software	9
Análise de Dados nas Ciências da Saúde	11
Fitting viral dynamic models to longitudinal data using a population based approach	15
Controlo Estatístico de Processos: Reamostragem em Planos de Aceitação e Cartas de Controlo	17
Mixture model selection criteria focusing on a clustering purpose.....	19
Estatísticas do Comércio Internacional: como se obtêm os dados finais?	23
O cálculo das estimativas mensais de emprego e desemprego divulgadas em 2014	25
A riqueza e o rendimento das famílias em Portugal.....	27
Taxas de pobreza ancoradas no tempo	29
Casting a light on shadow banking activity in Portugal	31
Financing of NFC – A comparison with other economic indicators	33
Desafios Estatísticos nas Estatísticas da Educação	35
A importância da Estatística na Educação	37
Modelling Efficiency in Education	39
Control charts for individual measurements of peroxide index.....	41
Monitorização de tempos de procedimentos académicos.....	43
Monitorização da variabilidade de um processo skew-normal multivariado ..	45
A new cleaning method for image restoration using robust confidence intervals	49
Real Time Activity Recognition for Cycling with Smartphone.....	51
A hedonic price approach for hotels in Porto	53
Escolas Secundárias atípicas no panorama nacional	55

Avaliação do desempenho dos alunos Portugueses em Matemática – PISA 2009	57
Papel da ação social no Ensino Superior Politécnico: expetativas futuras dos estudantes bolseiros com recurso à Análise de Multivariada de Dados	59
Previsão de <i>Churners</i>: Quem? E Quando?	61
Multidimensional relationship: education and fertility	63
A mortality study under different selected frameworks	65
Um perfil de quem não tem nem quer ter filhos em Portugal*	67
Recurso a produtos dietéticos e prática de exercício físico – Quem? Porquê? Como?	69
Estudo de biomarcadores de uma doença renal	71
TweeProfiles@Sapo: relating Spatio-Temporal patterns on Twitter with Sapo content	73
Distribuições de índices de concordância entre agrupamentos	75
Distribuições de Coeficientes de Comparação de Dendrogramas	77
Are callbacks a waste of time in mCATI surveys?	79
Data Warehouse de suporte à melhoria contínua do ensino na Escola Naval ...83	
Diagnostic tools for bivariate time series of counts	85
O efeito da existência de omissões nas estimativas de um modelo com trajetória latente linear: Um estudo de simulação	87
A seleção de compostos voláteis como indicadores da qualidade sensorial da carne fresca	89
Controlo estatístico da qualidade em ensaios interlaboratoriais – caso particular de nutrientes em águas salinas	91
A metodologia <i>Bootstrap</i> associada ao método de Holt-Winters na previsão de séries temporais	95
Alguns benefícios do envelhecimento ativo: um caso de estudo	97

Alguns motivos para indivíduos entre os 40 e os 49 anos quererem ter filhos*	99
Análise das perceções sobre a liderança num organismo público dos Açores com recurso a métodos quantitativos	101
Análise de um Teste de Motivação: Comparação em Países Lusófonos.....	103
Building up complex COBS	105
Caracterização da praxe académica com recurso à análise de dados.....	107
Consumo Distrital de Antidiabéticos em Portugal – Exemplo Prático da Análise Classificatória	109
Determinants of firm performance: a structural equation approach.....	111
Estudo do Papel Mediador de Variáveis Psicológicas na Relação entre a Intenção e a Prática de Exercício Físico	113
Models of stochastic symmetric matrices with truncation	115
O Papel Mediador da Avaliação Cognitiva na Relação entre Stress e Saúde Psicológica.....	117
Participação de crianças e adolescentes no planeamento das suas localidades	119
Perceção de barreiras na carreira e diversidade de género: Uma abordagem multivariada numa amostra de trabalhadores do sector energético	121
Índice de Autores.....	123

MINICURSOS

Minicurso – 6ª feira, 10 de abril, Sala 0.02 (9h30)

Model-based clustering - presentation of the mixmod software

Gilles Celeux¹

¹Inria Saclay-Île-de-France, gilles.celeux@inria.fr

In this mini course, I will present model-based clustering methods:

I will present Gaussian mixtures models for continuous data sets and the multivariate multinomial distributions for categorical data sets.

I will present EM-like algorithms to estimate the parameters of the numerous possible mixture models through maximum likelihood methodology. The important problems of spurious local maximizers and the dependence of the algorithms over their starting positions will be discussed.

I will consider the model selection issue and the choice of the number of cluster according to various model selection criteria.

Extensions as clustering of mixed data, variables selection, alternative mixture models will be presented.

If enough time, Bayesian inference for mixtures as the block clustering model will be introduced.

Along the lecture, the main features of the software mixmod and its extension Rmixmod in R will be presented and used to illustrate the models and methods introduced.

References

Biernacki, C., Celeux, G., Govaert, G. & Langrognet, F. (2006). Model-based cluster and discriminant analysis with the mixmod software. *Computational Statistics and Data Analysis*, 51, 587-600.

Fraley, C., Raftery, A. E. (2002). Model-Based Clustering, Discriminant Analysis, and Density Estimation. *J. Am. Statistical Association*, 97, 611-631.

Lebret, R., Iovleff, S., Langrognet, F., Biernacki, C., Celeux, G. & Govaert, G. (2015) Rmixmod: The R package of the Model-Based Unsupervised, Supervised and Semi-supervised Classification Mixmod Library. *Journal of Statistical Software*. (In Press).

McLachlan, G. J., Peel, D. (2000). *Finite Mixture Models*. Wiley, New York.

Minicurso – 6ª feira, 10 de abril, Sala 0.07 (9h30)

Análise de Dados nas Ciências da Saúde

Cristina Rocha¹

¹Departamento de Estatística e Investigação Operacional, Faculdade de Ciências, Universidade de Lisboa, cmrocha@fc.ul.pt

As Ciências da Saúde constituem uma das áreas científicas mais interessantes de aplicação da Estatística. De facto, como a incerteza é inerente a muitas situações que ocorrem em Saúde, a Estatística tem um papel fundamental na utilização adequada da informação recolhida, revelando-se um guia indispensável na tomada de decisões nesta área.

Neste mini-curso serão apresentados conceitos e métodos de inferência estatística que são fundamentais para a realização de estudos no âmbito das ciências biomédicas. O curso será de cariz essencialmente prático, sendo os métodos introduzidos com recurso a exemplos provenientes desta área. O software estatístico R será utilizado para ilustrar a aplicação dos métodos estatísticos, através da análise de dados que foram obtidos em diferentes situações.

SESSÕES PLENÁRIAS

Sessão Plenária I – 6ª Feira, 10 de abril, Anf. 0.01 (14h30)

Fitting viral dynamic models to longitudinal data using a population based approach

Ruy M. Ribeiro¹

¹*Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM, USA and Laboratório de Biomatemática, Faculdade de Medicina da Universidade de Lisboa, Portugal, ruyribeiro@medicina.ulisboa.pt*

Summary

Viral dynamic analyses of HIV-infected patients under treatment have shed light into many details and rates of the viral infection process. Here we will present this approach, with specific examples of analyzing longitudinal viral load data for subjects starting antiretroviral treatment. We will discuss the relevance of statistical models (mixed-effects) together with a mechanistic interpretation of the infection process to shed light on the biology of HIV.

Keyword: Drug treatment, HIV, Mixed-effects, Viral dynamics.

Introduction

Viral infectious diseases represent a huge burden for public health, with important implications for mortality, morbidity and health care costs [1]. Quantitative studies of viral load kinetics under antiviral treatment have allowed the definition of viral turnover and infected cell turnover in many different infections. Analyses of human infections as diverse as human immunodeficiency virus (HIV) [2; 3] or hepatitis C virus (HCV) [4] have demonstrated how quickly circulating virus is produced and cleared, even in long-term chronic infections, when the infection seems to be in clinical latency and the patient remains (mostly) asymptomatic.

Results

The standard model of HIV infection includes target cells, T , infected cells, I and free virus, V [2; 3]. Cells are infected at rate βTV , proportional to the availability of target cells and free virus, generating productively infected cells, I , which die at rate δ . In turn, virus, V , is produced by cells I at rate p per cell and is cleared from the circulation at rate c per virion. These infection processes are modified by the effect of drug treatment, which depending on the mode of action can affect infectivity (β), production of virus (p), or other processes. This model can be written as a system of ordinary differential equations, which can be solved under appropriate assumptions. We will present the development of these models and analyze a specific application to HIV-infected individuals treated with a combination of antiretrovirals. To fit the model to the data, we used a mixed-effects approach, with subject as the random effect. The covariates and the random effect covariance structure were chosen based on likelihood ratio tests. The objective was to estimate the parameters governing the dynamics of viral infection, and the efficacy of the drugs in reducing infection.

In Figure 1, we present the model prediction for the decline in viral load under therapy. Adjusting these curves to the data, we can estimate under various circumstances the death rate of infected cells, the clearance rate of free virus, and the efficacy of the drugs in blocking specific steps of the viral lifecycle. Using mixed-effects models to analyze this longitudinal data is a natural and powerful approach that allows simultaneous fitting of all the data, and analyses of important covariate factors with influence on the parameters estimated.

This approach showed that during clinical latency of HIV the viral turnover is huge. The half-life of free virus was estimated at ~45 minutes, implying total production and clearance of over 10^{10} virions per day. The half-life of productively infected cells was estimated at ~1 day, implying that about 10^7 cells are infected and killed every day by the virus [3]. These estimates have important clinical implications. For example, given the mutation rate of the virus, one understands the large diversity of the viral quasi-species and why drug resistance leads to treatment failure, or why it is so difficult to find an effective vaccine for this virus. One clinical corollary is that HIV must be treated with combination therapy (multiple drugs) [5].

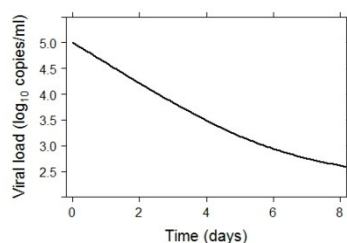


Figure 1. Predicted plasma HIV RNA decay under therapy

Conclusions

Modeling the dynamics of viral infections is a new field [6]. Matching these mechanistic models with mixed-effects fitting of data from individuals under treatment (and indeed other cases) has proven to be very successful in providing biological and clinical insight into the lifecycle of these viruses.

References

1. FAUCI, A. S., & MORENS, D. M. (2012) The perpetual challenge of infectious diseases. *N Engl J Med*, 366, 454-461.
2. PERELSON, A. S. (2002) Modelling viral and immune system dynamics. *Nat Rev Immunol*, 2, 28-36.
3. PERELSON, A. S., & RIBEIRO, R. M. (2013) Modeling the within-host dynamics of HIV infection. *BMC Biology*, 11, 96.
4. LAYDEN, T. J., LAYDEN, J. E., RIBEIRO, R. M., & PERELSON, A. S. (2003) Mathematical modeling of viral kinetics: a tool to understand and optimize therapy. *Clin Liver Dis*, 7, 163-178.
5. COFFIN, J. M. (1995) HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. *Science*, 267, 483-489.
6. NOWAK, M. A., & MAY, R. M. (2000). *Virus dynamics: Mathematical principles of immunology and virology*. Oxford: Oxford University Press.

Sessão Plenária II – sábado, 11 de abril, Anf. 0.01 (10h45)

Controlo Estatístico de Processos: Reamostragem em Planos de Aceitação e Cartas de Controlo

Fernanda Figueiredo¹

¹Faculdade de Economia da Universidade do Porto, e CEAUL, Universidade de Lisboa, otilia@fep.up.pt.

Sumário

Os planos de amostragem e as cartas de controlo são as ferramentas mais utilizadas em controlo estatístico da qualidade para decidir sobre a aceitação ou rejeição de lotes de matéria-prima ou de produto final, e monitorizar processos produtivos ou de serviços, respetivamente. Na literatura podemos encontrar diferentes tipos de planos de amostragem e de cartas de controlo, com aplicações nas mais diversas áreas, e bastante eficientes caso os pressupostos subjacentes à sua implementação não sejam violados. No entanto, quando trabalhamos com problemas reais os dados que temos de analisar dificilmente cumprem os pressupostos necessários para garantir a eficiência esperada para estes procedimentos, e a utilização de técnicas de reamostragem como o *bootstrap* poderão ajudar a contornar algumas dificuldades que surgem na sua implementação.

Palavras-chave: *Bootstrap*, Cartas de controlo, Controlo Estatístico da Qualidade, Planos de amostragem, Simulação de Monte Carlo.

Introdução

Os planos de amostragem para atributos, onde cada item para inspeção é classificado apenas em defeituoso ou não defeituoso, são os mais usuais, mas em muitas situações práticas temos de considerar planos de amostragem para variáveis, nos quais as características de qualidade são medidas em escala numérica (veja-se, por exemplo, Montgomery, 2009 e Gomes *et al.*, 2010). Em geral, um plano para variáveis requer a recolha de amostras de dimensão menor do que um plano para atributos face a um mesmo nível de proteção, mas os custos unitários de amostragem são em geral mais elevados. Mas a principal desvantagem dos planos para variáveis reside no facto da distribuição da característica de qualidade ter de ser conhecida, ou então estimada. Nos casos mais frequentes, admitimos que a característica de qualidade em estudo é Gaussiana, mas na prática podemos ter outros modelos conhecidos ou desconhecidos, para os quais a avaliação analítica dos planos de amostragem é difícil ou mesmo impossível. Note-se ainda que muitas vezes em contexto real apenas temos acesso a medições da característica de qualidade associadas a um conjunto histórico de itens previamente inspecionados, sendo usualmente difícil modelar esses dados. Como podemos então avaliar e comparar o desempenho de planos de amostragem específicos?

As cartas de controlo foram introduzidas por Shewhart em 1924 para monitorizar um processo de produção, mas atualmente o seu domínio de aplicação tem sido alargado a diversas áreas, incluindo os serviços. Para a implementação de uma carta de controlo os valores nominais do processo precisam de ser conhecidos, ou então necessitam de ser estimados. Na prática, a distribuição dos dados do processo e os seus parâmetros são em geral desconhecidos, sendo estimados a partir de uma amostra de referência constituída por m subgrupos de dimensão n . O desempenho da carta de controlo implementada com limites estimados só será semelhante ao da carta correspondente implementada com limites exatos se considerarmos um elevado

número de subgrupos iniciais na estimação (veja-se, por exemplo, Jensen *et al.*, 2006, Chen, 1997, e Castagliola & Figueiredo, 2013), o que é uma limitação do ponto de vista prático.

Metodologia e Resultados

Na indústria alimentar é muito importante controlar a presença de certas substâncias químicas na matéria-prima, que irão afetar a qualidade do produto final. A partir de um conjunto de dados relativos a medições efetuadas em amostras de material retirado de lotes de dimensão elevada, comparamos diferentes planos de amostragem utilizando a metodologia *bootstrap* (Efron & Tibshirani, 1993) e simulações de Monte Carlo (Figueiredo *et al.*, 2014). Num contexto de cartas de controlo, analisamos o desempenho de uma carta de médias implementada com limites estimados, e apresentamos os benefícios de utilizar a metodologia *bootstrap* para obter uma amostra de referência maior para a estimação dos valores nominais do processo (Figueiredo & Gomes, 2015).

Conclusão

Os casos analisados mostram que a utilização da metodologia *bootstrap* permite a avaliação de planos de amostragem complexos, mesmo no caso em que a taxa de amostragem é excessivamente reduzida e a característica de qualidade é difícil de modelar. Esta metodologia fornece também alguns contributos para a implementação de cartas de controlo com limites estimados, quando dispomos apenas de uma amostra de referência de dimensão reduzida para a estimação desses limites.

Agradecimentos

Este trabalho é financiado por Fundos Nacionais através da FCT – Fundação para a Ciência e a Tecnologia no âmbito do projecto PEst-OE/MAT/UI0006/2014.

Referências

- CASTAGLIOLA, P. & FIGUEIREDO, F. (2013) The median chart with estimated parameters. *European Journal of Industrial Engineering*, 7:5, 594-614.
- CHEN, G. (1997) The mean and standard deviation of the run length distribution of \bar{X} charts when control limits are estimated. *Statistica Sinica*, 7:3, 789-798.
- EFRON, B. & TIBSHIRANI, R. J. (1993) *An Introduction to the Bootstrap*. Chapman and Hall.
- FIGUEIREDO, F., FIGUEIREDO, A. & GOMES, M. I. (2014) Comparison of sampling plans by variables using bootstrap and Monte Carlo simulations. *International Conference of Computational Methods in Sciences and Engineering 2014, AIP Conference Proceedings* 1618, 535-538.
- FIGUEIREDO, F. & GOMES, M. I. (2015) *Control charts implemented on the basis of a bootstrap reference sample*. SMTDA Book 2 of selected papers (aceite para publicação).
- GOMES, M. I., FIGUEIREDO, F. & BARÃO, M. I. (2010) *Controlo Estatístico da Qualidade*, 2ª edição, Edições I.N.E.
- 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. *Journal of Quality Technology*, 38:4, 349-364.
- MONTGOMERY, D. C. (2009) *Introduction to Statistical Quality Control: a Modern Introduction*, 6th edition, John Wiley & Sons.

Sessão Plenária III – Sábado, 11 de abril, Anf. 0.01 (15h40)

Mixture model selection criteria focusing on a clustering purpose

Gilles Celeux,
INRIA Saclay-Île-de-France, gilles.celeux@inria.fr

In the framework of model-based clustering, I will review model selection criteria for choosing the number of clusters. I will pay special attention to the Integrated Completed Likelihood (ICL) of a mixture model. Different ways to approximate ICL with a full Bayesian view or with a maximum likelihood view will be presented and discussed. The position of ICL with respect to BIC criterion will be discussed too and the differences between these two criteria will be highlighted. Finally, some extensions of ICL, useful in applied contexts will be presented.

References

- BAUDRY, J.-P. (2015). Estimation and model selection for model-based clustering with the conditional classification likelihood. (Submitted).
- BAUDRY, J.-P., CARDOSO, M., CELEUX, G., AMORIM, M. J. & SOUSA FERREIRA, A. (2014). Enhancing the selection of a model-based clustering with external categorical variables. *Advances in Data Analysis and Classification*, on line.
- BAUDRY, J.-P., RAFTERY, A., CELEUX, G., LO, K. & GOTTARDO, R. (2010). Combining mixture components for clustering. *Journal of Computational and Graphical Statistics*, 19, 332-353.
- BERTOLETTI, M., FRIEL, N. & RASTELLI, R. (2014). Choosing the number of components in a finite mixture model using an exact Integrated Completed Likelihood criteria. *arXiv preprint arXiv:1411.4257* [stat.CO].
- BIERNACKI, C., CELEUX, G. & GOVAERT, G. (2000). Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22, 719-725.
- BIERNACKI, C., CELEUX, G. & GOVAERT, G. (2010) Exact and Monte Carlo calculations of integrated likelihoods for the latent class model. *Journal of Statistical Planning Inference*, 140, 2991-3002.
- CELEUX, G., SOROMENHO, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification*, 13, 195-212.

SESSÕES TEMÁTICAS

Sessão Temática I – Sessão INE – 6ª feira, 10 de abril, Anf. 0.01 (15h30)

Estatísticas do Comércio Internacional: como se obtêm os dados finais?

Cristina Neves¹

¹*Instituto Nacional de Estatística, cristina.neves@ine.pt*

Sumário

As estatísticas do Comércio Internacional de bens, produzidas pelo Instituto Nacional de Estatística (INE), resultam de um processo integrado de produção estatística, que envolve vários procedimentos de recolha, análise de informação e controlo de qualidade, que permitem a obtenção dos dados finais que mensalmente são disponibilizados ao público em geral e comparáveis ao nível dos restantes países da União Europeia.

Palavras-chave: Comércio, Estatísticas, Exportações, Importações, Internacional.

Introdução

As estatísticas do Comércio Internacional de bens são produzidas e divulgadas mensalmente pelo INE, e resultam da agregação de informação sobre transações Intra-UE de bens (entre Portugal e os restantes Estados-Membros da União Europeia – UE) e transações Extra-UE de bens (entre Portugal e os Países Terceiros).

Em 1 de janeiro de 1993, a entrada em vigor do Mercado Único determinou a supressão da maior parte das formalidades e dos controlos aduaneiros, ligados às trocas de bens entre os Estados-Membros da UE. Desde então, as estatísticas do Comércio Intra-UE de bens passaram a ser compiladas com base em informação recolhida diretamente junto das empresas, através do Sistema Intrastat. As transações Extra-UE continuam a ser compiladas com base na informação administrativa proveniente das alfândegas.

Com vista à redução da carga estatística, apenas as empresas com importações/exportações anuais superiores a determinados limiares são inquiridas no âmbito do Sistema Intrastat, sendo as transações abaixo desses limiares estimadas pelo INE, com vista à obtenção dos valores globais de transações de bens.

Resultados

No âmbito do processo de compilação das estatísticas do Comércio Internacional de bens, existe um compromisso de qualidade com os processos e produtos estatísticos, pelo que têm vindo a ser implementados pelo INE procedimentos sistemáticos de controlo de qualidade, através de um sistema integrado de produção estatística, envolvendo as suas várias fases.

Para o efeito contribuiu a implementação, pelo INE, do SIGINQ – Sistema Global de Gestão de Inquéritos, que consiste num sistema modular de gestão de todos os processos e procedimentos inerentes a uma operação estatística, desde a gestão de amostra, recolha e registo de informação, análise e validação, até à divulgação dos resultados.

Recolha de informação: a recolha de informação das transações Intra-UE de bens via Sistema Intrastat, é feita maioritariamente através de recolha eletrónica, via *Webform* através do *WebInq* (inquéritos eletrónicos na Web) do INE, consistindo também no primeiro momento de análise e controlo de qualidade, através da implementação de regras de validação automáticas.

Análise de microdados: com vista à melhoria da qualidade da informação de base, foram implementados modelos automáticos de deteção de *outliers*, através de um modelo de análise de *scores*, que permite identificar situações de potencial erro, com efetivo impacto nos resultados a divulgar, permitindo assim uma melhor gestão de recursos e um enfoque nas questões de maior importância.

Utilização de outras fontes de informação: no processo de produção das estatísticas do Comércio Internacional são utilizadas fontes administrativas e outras fontes internas ao INE, na atualização da amostra do Sistema Intrastat, na análise de microdados, na análise de valores globais e para a obtenção de estimativas de não resposta e de transações abaixo do limiar.

Controlo de qualidade: para além das validações de conteúdo efetuadas quer na recolha de dados quer na análise de microdados, a utilização das outras fontes de informação tem também como objetivo a implementação de procedimentos de controlo de qualidade. Adicionalmente são produzidos internamente relatórios de qualidade e de monitorização de revisões, com vista à avaliação da qualidade global dos resultados apurados.

Divulgação de resultados: na disponibilização de informação ao público em geral, para além dos resultados mensais divulgados, é dada importância à meta-informação associada que permite ao utilizador uma correta leitura dos dados publicados. No cumprimento do Código de Conduta para as Estatísticas Europeias, procura-se ainda disponibilizar aos utilizadores informação adicional que permita avaliar a relevância, precisão e fiabilidade da informação disponibilizada, como sejam os pressupostos inerentes à divulgação de dados revistos (política de revisões), quebras de série e análise de séries temporais.

Conclusões

A compilação das estatísticas do Comércio Internacional resulta de um método integrado, que permite a junção, de forma harmonizada e coerente, das várias fases do processo de produção estatística, com vista à obtenção dos dados finais.

A obtenção de informação de qualidade é um objetivo fundamental que se pretende sempre alcançar, de modo a que a informação estatística produzida seja precisa e fiável, mas também oportuna e relevante para os seus utilizadores.

Sessão Temática I – Sessão INE – 6ª feira, 10 de abril, Anf. 0.01 (15h45)

O cálculo das estimativas mensais de emprego e desemprego divulgadas em 2014

Sónia Torres¹

¹*Instituto Nacional de Estatística, sonia.torres@ine.pt*

Sumário

Nesta apresentação, faz-se uma descrição da metodologia adotada pelo INE para o cálculo das estimativas mensais de emprego e desemprego obtidas a partir do Inquérito ao Emprego, cuja divulgação foi iniciada em 2014 – médias móveis de três meses (trimestres móveis) centradas, integrando uma projeção para o terceiro mês do último trimestre móvel. A projeção referida é realizada através da utilização de modelos de alisamento exponencial em espaço de estados. A especificação dos modelos é escolhida, de forma automática, no *software* R (função ETS: *Error, Trend, Seasonality*).

Palavras-chave: Estimativas mensais, *ETS*, Inquérito ao emprego, Modelos de alisamento exponencial em espaço de estados, Trimestres móveis.

Em novembro de 2014, o INE iniciou a divulgação mensal de estimativas do Inquérito ao Emprego, em complemento da publicação das estimativas trimestrais habituais.

A metodologia selecionada para o cálculo destas estimativas – médias móveis de três meses (trimestres móveis) centradas, integrando uma projeção para o terceiro mês do último trimestre móvel – foi reconhecida pelo Eurostat, que aceitou publicar as estimativas assim obtidas da população desempregada e da taxa de desemprego para Portugal nos seus *News Release* mensais, em substituição das estimativas obtidas até então a partir de um modelo de desagregação temporal de séries temporais, mediante a utilização de informação mensal auxiliar (no caso, o desemprego registado do IFEP), designada por metodologia Chow-Lin.

Em 2013 e 2014, o INE testou várias metodologias alternativas à metodologia Chow-Lin para a produção mensal das estimativas pretendidas tendo como fonte o Inquérito ao Emprego, de entre as quais se salientam o estimador de médias móveis de três meses, o estimador de um mês isolado, um conjunto de estimadores combinados e a metodologia Chow-Lin com utilização de informação auxiliar mensal proveniente do Inquérito ao Emprego. Estas metodologias foram avaliadas/comparadas segundo um conjunto de critérios/propriedades que concorrem para a definição da qualidade das estimativas: atualidade/oportunidade e possibilidade de manutenção dos prazos de divulgação mensal atual do Eurostat, precisão das estimativas, volatilidade das estimativas, consistência com as estimativas trimestrais e magnitude das revisões.

Dos estudos conduzidos, foi possível identificar a metodologia das médias móveis de três meses como a que garantia globalmente a maior qualidade das estimativas. Dada a existência de um objetivo adicional sobre a manutenção da atualidade/oportunidade na divulgação de estimativas mensais pelo Eurostat e a inexistência de dados recolhidos para o terceiro mês do último trimestre móvel centrado, houve necessidade de se proceder a uma projeção para esse

terceiro mês (estimativa considerada provisória e revista no mês seguinte, quando a recolha já está completa).

Nesta apresentação, descrevem-se o processo de cálculo das estimativas dos trimestres móveis publicadas mensalmente e o processo de seleção da metodologia de projeção do terceiro mês do último trimestre móvel.

O processo de cálculo das estimativas dos trimestres móveis estrutura-se essencialmente em cinco etapas: a informação utilizada; o cálculo dos ponderadores (calibração); as projeções realizadas; o cálculo das médias móveis; o ajustamento de sazonalidade.

A projeção referida é realizada através da utilização de modelos de alisamento exponencial em espaço de estados. Esta metodologia foi selecionada entre um conjunto de metodologias alternativas, nomeadamente modelos SARIMA e modelos de previsão *ad hoc*, tendo sido a que permitiu obter menores níveis de revisão na passagem das estimativas provisórias a definitivas. Para a escolha automática da especificação dos modelos de previsão selecionados (com base nos critérios *Bayesian Information Criterion* e *Akaike Information Criterion*), é utilizada a função ETS (*Error, Trend, Seasonality*) do *package Forecast* do *software R*, desenvolvida por Robert Hyndman (2008).

Referências

HYNDMAN, R. J., KOEHLER, A. B., SNYDER, R. D. & GROSE, S. (2002) A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, 18(3), 439–454.

HYNDMAN, R. J. & KHANDAKAR, Y. (2008a) Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software*, 27(3).

HYNDMAN, R. J., KOEHLER, A. B., ORD, J. K. & SNYDER, R. D. (2008b) *Forecasting with Exponential Smoothing: The State Space Approach*, Springer-Verlag.

TORRES, S. (2008) Taxas de desemprego mensais – Estimativas para Portugal. *Estatísticas do Emprego – 2º trimestre de 2008*, 38-44.

TORRES, S. (2010) Taxas de desemprego mensais – Estimativas para Portugal – Parte II. *Estatísticas do Emprego – 4º trimestre de 2010*, 53-60.

Sessão Temática I – Sessão INE – 6ª feira, 10 de abril, Anf. 0.01 (16h00)

A riqueza e o rendimento das famílias em Portugal

Eduarda Góis¹, Esperança Figueiredo²

¹Instituto Nacional de Estatística, eduarda.gois@ine.pt;

²Instituto Nacional de Estatística, esperanca.figueiredo@ine.pt

Sumário

Os resultados estatísticos disponibilizados no âmbito do Inquérito à Situação Financeira das Famílias (ISFF) permitem caracterizar a distribuição da riqueza líquida das famílias, nomeadamente a sua concentração e a correlação com o rendimento monetário das famílias. Em Portugal, a análise dos dados permite concluir sobre a existência de uma forte assimetria na distribuição da riqueza líquida e associação entre níveis de riqueza e de rendimento mais elevados.

Palavras-chave: Decis, Rendimento monetário bruto, Riqueza líquida.

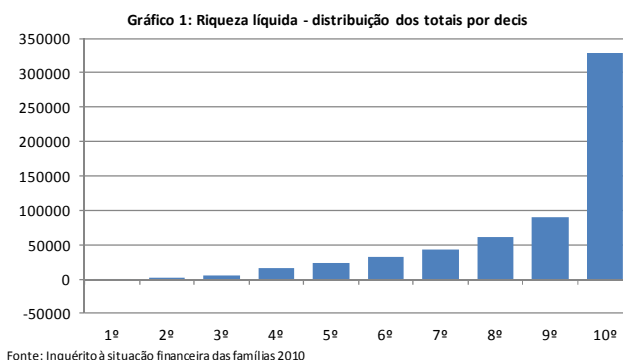
Introdução

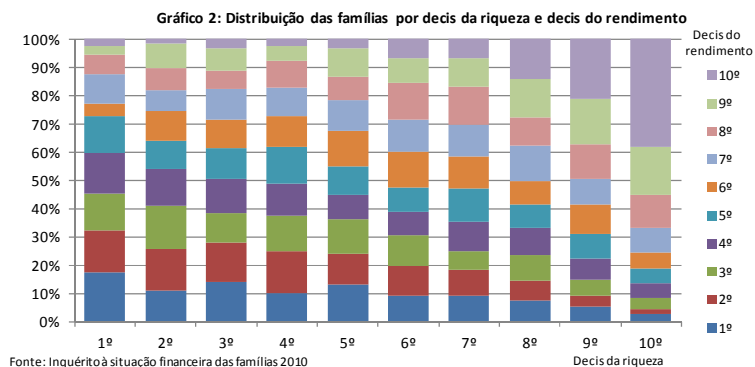
Em 2010 foi lançada a primeira recolha de dados no âmbito do projeto *Household Finance and Consumption Survey*, implementado em todos os Estados-Membros da área do Euro. Em Portugal, a realização do inquérito (ISFF) é da responsabilidade do Banco de Portugal e do Instituto Nacional de Estatística.

Os conceitos de referência para a análise que se apresenta são a riqueza líquida – diferença entre o valor total da riqueza e o valor total da dívida à data da recolha dos dados – e o rendimento monetário bruto – rendimento monetário obtido pelos agregados, de carácter regular e não regular, proveniente do trabalho, de outros rendimentos privados, das pensões e outras transferências sociais, antes da dedução dos impostos devidos e das contribuições obrigatórias para a proteção social. A riqueza inclui ativos reais e ativos financeiros. As classes de rendimento e de riqueza líquida baseiam-se nos decis da distribuição destas variáveis estimadas para a população.

Resultados

A distribuição da riqueza líquida em Portugal caracterizava-se em 2010 por uma forte assimetria, em que 10% das famílias detinham mais de metade da riqueza líquida, e o primeiro decil da distribuição se caracterizava por valores de riqueza líquida negativos.



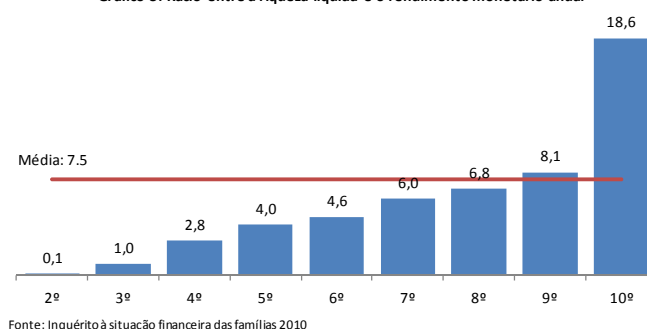


A análise dos valores da riqueza em simultâneo com a distribuição dos rendimentos monetários permite concluir pela existência de uma relação entre detentores de maior riqueza e de rendimentos mais elevados, todavia não sendo linear, o que parece indicar a influência de

outros determinantes para a ordenação das duas variáveis para muitas das famílias observadas.

A análise do rácio entre a riqueza líquida e o rendimento monetário anual permite uma aproximação ao número de anos em que, em média, as famílias conseguiriam manter o nível de recursos monetários idêntico ao de 2010 apenas recorrendo à riqueza líquida.

Gráfico 3: Rácio entre a riqueza líquida e o rendimento monetário anual



Conclusões

Os resultados do ISFF permitem concluir que a assimetria na distribuição da riqueza líquida é bastante superior à verificada na distribuição do rendimento monetário. A distância entre as famílias do decil de riqueza mais elevado para as restantes regista uma dimensão considerável, mesmo quando comparada com o decil imediatamente inferior (a riqueza dos 10% mais ricos é 4 vezes superior à dos 10% imediatamente inferiores).

Por outro lado, Portugal é um país com uma riqueza líquida inferior à média da área do Euro mas também com um rendimento monetário inferior à média, pelo que é pertinente analisar a importância da riqueza para fazer face a períodos de eventual perda do rendimento habitual.

Referências

INSTITUTO NACIONAL DE ESTATÍSTICA (2012) Inquérito à Situação Financeira das Famílias 2010. Destaque à Comunicação social, INE e Banco de Portugal.

BANCO CENTRAL EUROPEU (2013) The Eurosystem Household Finance and Consumption Survey – Results from the first wave. BCE.

FARINHA RODRIGUES, C., ANDRADE, I. (2013) Growing Inequalities and their Impacts in Portugal. GINI Country Reports, Amsterdam Institute for Advanced Labour Studies.

Sessão Temática I – Sessão INE – 6ª feira, 10 de abril, Anf. 0.01 (16h15)

Taxas de pobreza ancoradas no tempo

Eduarda Góis¹

¹*Instituto Nacional de Estatística, eduarda.gois@ine.pt*

Sumário

Os indicadores de referência para avaliação da pobreza na União Europeia e na OCDE, baseados no cálculo de linhas de pobreza relativas nacionais, nem sempre são claros para os cidadãos, que antecipam uma evolução do risco de pobreza em linha com a alteração global dos rendimentos, contrariamente ao verificado recentemente em Portugal.

Será necessário equacionar a aproximação entre o indicador de pobreza e a forma como os cidadãos percecionam a evolução das condições de vida, nomeadamente através de medidas de pobreza absolutas, ou, num futuro próximo, através do uso crescente de medidas de pobreza ancoradas no tempo.

Palavras-chave: Alterações na distribuição do rendimento, Linhas de pobreza ancoradas no tempo, Linhas de pobreza relativas.

Introdução

Os indicadores baseados no cálculo de linhas de pobreza relativas permitem ter em conta as diferenças entre os níveis de rendimento entre países e as alterações da distribuição do rendimento ao longo do tempo. As linhas de pobreza relativas correspondem a uma percentagem específica da mediana da distribuição do rendimento disponível, ou seja, do centro de gravidade da distribuição que assim constitui um valor chave para a análise da pobreza. Na União Europeia, o risco de pobreza é a percentagem de pessoas que vivem com um rendimento disponível equivalente inferior à linha de pobreza relativa (60% da mediana), adotada como limiar de pobreza.

Esta medida nem sempre é clara para os utilizadores, para quem é exetável que um aumento geral dos ganhos e benefícios resulte na redução da taxa de pobreza (ou que uma redução geral resulte num aumento do risco de pobreza), mas esta relação pode não ocorrer com uma medida de pobreza baseada numa linha relativa: o resultado dependerá do equilíbrio entre as alterações individuais, impacto sobre a mediana do rendimento e dimensão das alterações entre pobres, menos pobres e não pobres.

Resultados

A comparação de dados do Inq. Condições de Vida e Rendimento das Famílias com as Contas Nacionais não evidencia uma associação linear negativa entre as taxas de variação anual do risco de pobreza e as taxas de variação anual do rendimento disponível *per capita*.

Com o cálculo de linhas de risco de pobreza ancoradas, i.e. limiares de pobreza com valores que resultam da atualização da linha de pobreza relativa num determinado ano base (âncora) com base na evolução do Índice de Preços no Consumidor, obtém-se uma série que é independente dos efeitos da distribuição do rendimento no tempo.

Tabela 1: Cálculo das taxas de risco de pobreza ancoradas em 2004

Ano de referência do rendimento	Rendimento disponível per capita		Limiar de pobreza (€)	Taxa de pobreza relativa	Alteração IPC	Limiar de pobreza ancorado a 2004		Taxa de pobreza ancorada em 2004
	€	Variação				€	Variação	
2004	9 776		4 317	19,4%	2,4%	4 317		19,4%
2005	10 125	3,6%	4 386	18,5%	2,3%	4 520	2,3%	19,8%
2006	10 422	2,9%	4 544	18,1%	3,1%	4 661	3,1%	19,1%
2007	10 914	4,7%	4 886	18,5%	2,5%	4 775	2,4%	17,7%
2008	11 368	4,2%	4 969	17,9%	2,6%	4 899	2,6%	17,2%
2009	11 325	-0,4%	5 207	17,9%	-0,8%	4 858	-0,8%	15,2%
2010	11 674	3,1%	5 046	18,0%	1,4%	4 926	1,4%	16,4%
2011	11 543	-1,1%	4 994	17,9%	3,7%	5 106	3,7%	18,7%

Fonte: INE - Contas económicas anuais das famílias, ICOR - Inquérito às Condições de Vida e Rendimento (EU-SILC), Índice de Preços no Consumidor (IPC).

Conclusões

A utilização das linhas de pobreza relativas é coerente com a noção de que ser pobre é em grande medida uma condição relativa, que depende da condição dos restantes residentes no país (ou região) em estudo, e permite que este aspeto seja automaticamente reconhecido no cálculo dos indicadores de pobreza quando toma o centro de gravidade da distribuição do rendimento como base de referência. Por outro lado, não é consistente com a constatação de que abaixo de um determinado nível de rendimento as pessoas não conseguem ter acesso a um conjunto mínimo de bens e serviços que assegura viver com dignidade.

Em Portugal, o rendimento disponível das famílias reduziu-se em 2011 e 2012, afetando principalmente os que se situam no centro da distribuição dado que as políticas implementadas preservaram, pelo menos parcialmente, os mais pobres. Como a linha de pobreza relativa tende a variar em linha com o rendimento monetário disponível, por construção, o limiar de pobreza reduziu-se e um número considerável de montantes de rendimento antes potenciadores da pobreza deixou de ser classificado como tal, observando-se em 2011 um risco de pobreza (17,9%) igual ao valor mínimo observado na série e em 2012 um risco (18,7%) próximo do valor médio da série.

O que se propõe é a utilização crescente de linhas de pobreza ancorada enquanto complemento da linha de pobreza relativa, obtendo-se desta forma uma série que é independente dos efeitos da distribuição do rendimento no tempo, apenas dependente da âncora e da trajetória dos preços, tendo como vantagem a possibilidade de monitorizar a evolução da pobreza a partir de um ano base. A escolha da âncora implica algum cuidado e atualização, nomeadamente quando ocorrem fortes alterações na estrutura dos rendimentos.

Referências

FARINHA RODRIGUES, C. (2007) *Distribuição do rendimento, desigualdade e pobreza: Portugal nos anos 90*, Coimbra, Portugal, Edições Almedina S.A..

GÓIS, E. (2013) How clear are relative poverty measures to the common public? *Seminar "The way forward in poverty measurement"*. UNECE, Working paper 13.

Sessão Temática II – Sessão Banco de Portugal – 6ª feira, 10 de abril, Anf. 0.01 (18h00)

Casting a light on shadow banking activity in Portugal

Patrícia Antunes¹, Rita Pisco²

¹ Banco de Portugal, paantunes@bportugal.pt;

² Banco de Portugal, rmpisco@bportugal.pt

Abstract

In line with most recent debates on financial stability, the present study analyses the shadow banking activity in Portugal using the available statistical information. Empirical data shows a significant growth of activity from 2001 until 2010, with an increase of 30% to 63% as a percentage of GDP mainly driven by securitisation operations; at the end of the 3rd quarter of 2014 it amounted to 46% of GDP. Nonetheless, these values are smaller than the ones registered in other economic and financial jurisdictions.

Keywords: Credit intermediation, Securitisation, Shadow banking.

Introduction

Over recent years, in the aftermath of the financial crisis the shadow banking activity became a focus of international analysis as a source of financial stability concerns, leading the G20 leaders, in November 2010, to request the Financial Stability Board (FSB) to develop an assessment on shadow banking, in order to prompt policy initiatives to enhance its monitoring and regulation. According to the FSB, in broader terms, shadow banking system can be described as credit intermediation involving entities and activities outside the regular banking system. The entities engaged in shadow banking are financial intermediaries that ultimately conduct maturity transformation (long term assets financed by short term liabilities) within the process of credit intermediation (and/or transformation). From a financial stability perspective, their high interconnectedness with the banking system heightens the leverage and the procyclicality of the financial system underlying potential risks.

The European Central Bank published, in 2012, a paper with the first analysis on the size and the structure of shadow banking within the euro area, having observed that the total amount of shadow banking activity in the 2nd quarter of 2011 was of about 11 trillion euro (112% of euro area GDP). In the present study, the universe of entities that were considered as relevant were those identified in Banco de Portugal's Financial Stability Report - May 2014, namely: investment funds (including money market funds), financial vehicles corporations (FVC) engaged in securitisation and other financial intermediaries (OFI).

Analysis and results

The aim of this study is to analyse the shadow banking activity in Portugal, using the published statistical data, comparing its evolution against the euro area as a whole. The inexistence of a single statistical database for this analysis, led to the combination of different

statistical outputs, namely Monetary and Financial Statistics and National Financial Accounts, which made the quest for results more challenging.

Since 2000, when analysing the total assets of the relevant shadow banking subsectors (excluding MMF which only amounted about 2.5 billion euro in the end of 2014), it is possible to observe a positive growth rate until the end of 2007, reaching a peak of 30% in 2003, mostly due to the development of FVC activity. In fact, having begun their activity in Portugal in 2001, FVC recorded an impressive growth until the end of 2011, with total assets reaching a maximum of 62 billion euro, more than quadrupling its size since 2003.

As a whole, total assets of shadow banking system, now also including MMF, amount to nearly 80 billion euro (46% of GDP) in the 3rd quarter of 2014 comparing with 444 billion euro (255% of GDP) of banks. More than 90% of the securitised assets, as at the 3rd quarter of 2014, were loans originally granted by the banking system, demonstrating the very strong interconnectedness between the banking sector and the shadow banking system. Also for investment funds, excluding MMF, it is possible to observe that banks hold 33% of the total issued investment fund units, at the end of 2014. In the case of OFI, credit intermediation has a relevant roll, since about 75% (in the 3rd quarter of 2014) of total assets correspond to loans and advances. As consequence of this interconnectedness, since 2003 the other intermediaries have contributed to the grossing-up of bank's liabilities, reflecting mainly securitisation operations.

Conclusions

Although the credit intermediation activity in Portugal is largely undertaken by the banks, the shadow banking system increased its relevance until 2010, mainly driven by securitisation. Even so, data seem to indicate that the shadow banking activity in Portugal is not as relevant as in the euro area.

Disclaimer

The usual disclaimer applies.

References

BAKK-SIMON, K., BORGIOLO, S., GIRON, C., HEMPELL, H., MADDALONI, A., RECINE, F., ROSATI, S., "Shadow banking in the euro area an overview" <http://www.ecb.europa.eu/pub/pdf/scpops/ecbocp133.pdf>, (accessed 10 March 2015)

Financial Stability Board, "Strengthening Oversight and Regulation of Shadow Banking" https://www.financialstabilityboard.org/publications/r_130829c.pdf, (accessed 10 March 2015)

Banco de Portugal, Financial Stability Report - May 2014, <http://www.bportugal.pt/en-US/EstabilidadeFinanceira/Publicacoes/RelatoriodeEstabilidadeFinanceira/REFAnteriores/Pages/Edicoesanteriores.aspx>, (accessed 10 March 2015)

Sessão Temática II – Sessão Banco de Portugal – 6ª feira, 10 de abril, Anf. 0.01 (18h20)

Financing of NFC – A comparison with other economic indicators

Batista, Rodrigo¹, Colaço, Ricardo²

¹Banco de Portugal, rsbatista@bportugal.pt;

²Banco de Portugal, rcsilva@bportugal.pt

Abstract

From December 2009 until the end of 2014, loans granted by the resident financial system to Portuguese non-financial corporations (NFC) have shrunk by almost 24%. In this paper we present an analysis of the NFC loans' evolution and compare it with related economic indicators. We further break it down by economic activity and geographical region, and observe notorious heterogeneity in the behaviour of credit and other variables across the different regions and activities.

Keywords: Central bank statistics, Credit, Economic Indicators, Gross Value Added.

Introduction

According to Leão *et al.* (2014), the financing of non-financial corporations (NFC) is a key factor for their performance and the restrictions imposed since the beginning of the Portuguese Economic and Financial Assistance Programme have been one of the main barriers to the competitiveness of the Portuguese NFC. Since the end of 2009 loans granted by the resident financial system to Portuguese NFC have decreased by almost 24% while the economic climate indicator has been negative since November 2010, exhibiting a positive relation with the loans' year on year rate of change (Chart 1).

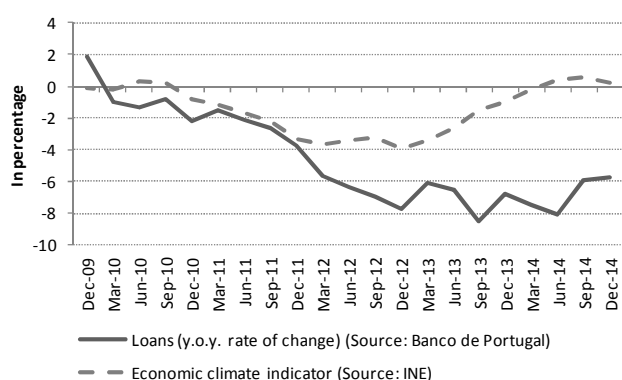


Chart 1: Loans granted to NFC and economic climate indicator

In this paper, by jointly analysing the information publicly available at Banco de Portugal and other public databases, we present a comparison between the evolution of loans granted to NFC and other economic indicators, breaking down by economic activity (according to NACE) and geographical region, revealing different relations across regions and sectors of economic activity.

Results and Conclusions

The joint analysis of Banco de Portugal statistics, based on the Central Credit Register's data, and data available at INE's website, allows a comparison between the gross value added (GVA) by NFC and the loans granted to NFC at a NUTS III region level. Comparing both variables at their 2012 levels, it is observable a close relation between the share of GVA each region adds and its respective share of loans granted to NFC. Excluding the NUTS III regions of *Grande Lisboa* and *Grande Porto* from the analysis, due to their relative size difference to other regions, the close relation is still observable as seen in Chart 2.

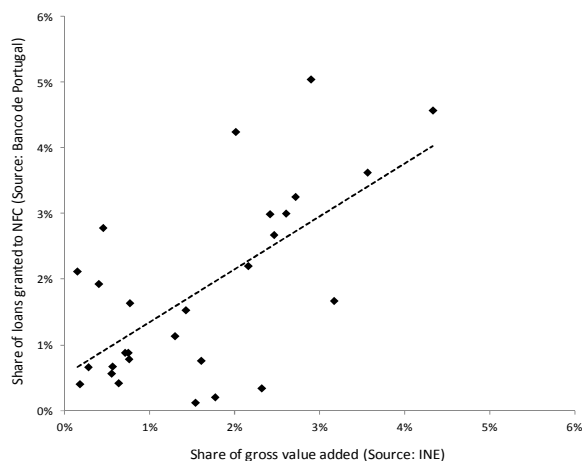


Chart 2: Share of loans granted to NFC and share of GVA – NUTS III

The same sources of data also allow a comparison of the same variables at a sector of economic activity level. Analysing the rate of change between 2009 and 2013, one can observe that sectors of activity with higher rates of change in loans granted also experienced higher rates of change in their GVA.

The analysis of statistics published by Banco de Portugal, within different scopes, also allows for interesting analysis. Comparing the total debt of private NFC vis-à-vis the financial sector and the returns on total assets (ROA) of private NFC by sector of activity, between December 2009 and September 2014, it is possible to see that the loans' year on year rate of change and changes in ROA follow distinct patterns across sectors, although not allowing for a clear relation to be established. This result provides evidence that, despite the loans' reduction, there were some sectors able to increase their ROA.

Disclaimer

The usual disclaimer applies.

References

LEÃO, J., MARTINS, A. & GONÇALVES, J. (2014) Crescimento Económico e Financiamento da Economia Portuguesa. *Boletim Mensal da Economia Portuguesa*, 11|2014, 55-69.

**Sessão Temática III – Importância da Estatística na Educação – sábado, 11 de abril,
Anf. 0.01 (12h00)**

Desafios Estatísticos nas Estatísticas da Educação

Luísa Canto e Castro Loura¹

¹*Direção Geral de Estatísticas da Educação e Ciência e CEAUL, luisa.castro@dgeec.mec.pt;*

Sumário

Esta apresentação tem por objetivo dar a conhecer os processos de recolha dos dados dos alunos que frequentam as nossas escolas, universidades e institutos politécnicos, descrever as características desses dados em termos de qualidade e diversidade e apresentar algumas das análises que a DGEEC tem vindo a desenvolver como forma de apoio à decisão e, também, de divulgação junto do grande público. Dar-se-á destaque aos modelos de previsão do número de alunos, aos métodos de comparação dos resultados das escolas em anos sucessivos e aos indicadores de valor esperado, de valor acrescentado e de alinhamento entre classificações internas e classificações de exame.

Palavras-chave: Análise de dados, Estatísticas da educação, Recolha de dados,

Introdução

A Direção-Geral de Estatísticas da Educação e Ciência é o organismo do Ministério da Educação e Ciência que tem a responsabilidade pelo processo de recolha, validação e apuramento dos dados sobre o sistema educativo e de formação. Com o avanço das tecnologias, todo este processo se tem tornado mais célere e eficiente libertando recursos para se procederem a algumas análises de suporte ao planeamento e à decisão. Com impacto no planeamento das ofertas é de referir o desenvolvimento de um modelo de previsão do número de alunos no ensino básico e secundário ([1]), modelo esse que contempla e avalia o impacto do alargamento da escolaridade obrigatória e que foi alicerçado em dados censitários e nos dados mais recentes da distribuição etária em cada ano de escolaridade e em cada oferta formativa. Outra área onde se revelou importante proceder a uma análise a nível central, foi a da avaliação dos resultados das escolas tomando por referência os exames nacionais: avaliação dos resultados tendo em conta variáveis de contexto ([2]), análise da progressão entre o 9.º ano e o 12.º ano e análise do alinhamento entre classificações internas e classificações de exame ([3]), avaliação da melhoria em anos sucessivos ([4]), análise da evolução do percentil médio das notas de entradas nas licenciaturas e mestrados ([5]).

Resultados

De entre os estudos e análises acima referidos, apenas no respeito ao modelo de previsão de alunos e ao de análise da melhoria de resultados escolares em anos sucessivos é que há lugar a uma apreciação global em termos de resultados. Todos os restantes dão informação de referência para cada escola, uma vez que permitem situá-la por comparação com todas as restantes, permitindo assim que estas concentrem esforços nas áreas em que considerem haver necessidade e possibilidade de melhoria.

Os principais resultados da aplicação do modelo de previsão do número de alunos para Portugal continental revelam uma tendência geral de decréscimo do número de alunos nos três ciclos do ensino básico (decrécimo de cerca de 44 mil alunos, até 2018/19, no 1.º ciclo, representando 11% do total registado em 2012/13, decréscimo de cerca de 19 mil alunos no 2.º ciclo – 8%, e decréscimo de cerca de 12 mil alunos no 3.º ciclo – 3%). Nos próximos cinco anos prevê-se que esse decréscimo seja mais sentido no 1.º ano do 1.º ciclo. Ao nível do ensino secundário, e para as modalidades regular e artístico, prevê-se um aumento de 3% (cerca de 5 mil alunos a mais em 2018/19 do que em 2012/13). Já para as modalidades de cariz profissionalizante prevê-se um aumento de 5% (cerca de 8 mil alunos a mais).

Quanto à análise da melhoria de resultados das escolas nos exames nacionais em anos sucessivos, por decorrer desta a atribuição de crédito horário aos agrupamentos de escolas, importa aqui reportar o número total dos que superaram os critérios para atribuição desse crédito – 68 de entre o total de 811.

Agradecimentos: Esta apresentação destaca algumas das análises desenvolvidas desde 2012 na DGEEC e, por terem envolvido as diversas equipas de especialistas e técnicos, opto por agradecer a todos através da referência a quem teve papel de coordenação: João Baptista, Nuno Rodrigues, Joaquim Santos, Carlos Malaca e Joana Duarte.

Referências

Links:

[1] – Página WEB da DGEEC - <http://www.dgeec.mec.pt/np4/PREVISOES/>

[2] – Página WEB da DGEEC - <http://www.dgeec.mec.pt/np4/120/>

[3] – Portal INFOESCOLAS - <http://www.infoescolas.pt>

[4] – Página WEB da DGEEC - <http://www.dgeec.mec.pt/np4/253/>

[3] – Portal INFOCURSOS - <http://www.infocursos.pt>

Sessão Temática III – Importância da Estatística na Educação – sábado, 11 de abril,
Anf. 0.01 (12h20)

A importância da Estatística na Educação

Olívia Sousa

Professora do Ensino Básico, em destacamento no IAVE, oliviasousa6@gmail.com

Sumário

A Estatística assume um papel de extrema relevância na Educação, quer enquanto disciplina escolar quer como ferramenta no domínio da avaliação das aprendizagens, proporcionando métodos e técnicas de análise. Após uma breve referência à integração da Estatística nos currículos escolares, será abordada a pertinência de alguns dos seus procedimentos na análise de resultados obtidos em testes de conhecimento.

Palavras-chave: Educação estatística, Índice de dificuldade, Índice de discriminação, Psicometria, TCT e TRI.

Introdução

O papel da Estatística na Educação pode considerar-se duplo, quer enquanto disciplina escolar, que contribui para a formação dos alunos, quer como ferramenta transdisciplinar, que proporciona métodos e técnicas de análise aplicáveis a qualquer domínio do saber.

O ensino da Estatística inicialmente restringiu-se ao ensino superior. Posteriormente alargou-se ao secundário e, na década de noventa, foi integrado nos programas de Matemática de todos os níveis de escolaridade. A inclusão do ensino da estatística em todos os níveis de escolaridade nasceu de um movimento de contestação aos conteúdos e métodos de ensino então reduzidos ao cálculo de rotinas e procedimentos complexos. Em alternativa, este movimento propunha o enfoque no papel formativo da análise de dados na educação das crianças, desde o ensino elementar (Holmes, 2000; Sheaffer, 2000).

Estas mudanças conduziram à valorização de outros conteúdos estatísticos que, apesar de menos exigentes em pressupostos matemáticos, constituem importantes ferramentas de análise em numerosos domínios, nomeadamente, na análise de resultados de testes de conhecimento.

Resultados

Os testes de conhecimento são instrumentos de recolha de dados frequentemente utilizados em educação. Um teste é um conjunto de itens que constituem uma *amostra* representativa de um *domínio* de conhecimentos. A partir da análise e da classificação das respostas dos alunos aos itens de um teste fazem-se *inferências* acerca dos seus níveis de competência no domínio que o teste avalia (Madaus, Haney e Kreitzer, 2000).

Usualmente, a análise dos dados de um teste é feita no âmbito de um de dois quadros teóricos: a teoria clássica dos testes (TCT) ou a teoria da resposta aos itens (TRI). Propõem

ambos diferentes modelos (aditivo e probabilístico) e unidades de análise (pontuação no teste e padrão de respostas aos itens) respetivamente (Embretson & Reise, 2000).

Apesar das diferenças, é possível, em qualquer dos modelos, proceder a uma análise de resultados, recorrendo aos parâmetros psicométricos dos itens de um teste. Nesta comunicação, será apresentada uma tabela com uma amostra de dados obtidos num teste, que serão utilizados para ilustrar o modo como se podem calcular e interpretar os índices de dificuldade e de discriminação dos itens.

O índice de dificuldade traduz a taxa de sucesso num item e o índice de discriminação a diferença de sucesso entre os alunos que obtiveram as melhores e as piores classificações no teste. O índice de discriminação é um indicador da correlação entre o sucesso no item e o sucesso no teste global (Kelley, Ebel & Linacre, 2002). A distribuição das classificações por item pode também fornecer informações pertinentes para a análise dos resultados.

Conclusões

A análise dos parâmetros dos itens e da distribuição das classificações das respostas permite a identificação de erros que podem comprometer a validade do próprio teste: erros na construção dos itens, e.g., instruções de realização incompletas ou ambíguas, opções de resposta não plausíveis em itens de seleção; erros na construção do teste, e.g., itens interdependentes, distribuição desequilibrada do número de itens por nível de dificuldade ou por nível cognitivo.

Referências

- HOLMES, P. (2000) What sort of statistics should be taught in schools: And why? In LOUREIRO, C., OLIVEIRA, F. & BRUNHEIRA, L. (Orgs.) *Ensino e aprendizagem da estatística* (pp. 49-56). Lisboa: SPE, APM, DE e EIO da FCUL.
- KELLEY, T., EBEL, R. & LINACRE, J. (2002) Item discrimination indices. *Rasch Measurement Transactions*, 16(3), 883-4, www.rasch.org (acedido em 21 de maio de 2012).
- MADAUS, G., HANEY, W. & KREITZER, A. (2000) The role of testing in evaluations. In D. STUFFLEBEAM, G. MADAUS & T. KELLAGHAN (Eds.), *Evaluation models: Viewpoints on Educational and human services evaluation* (2nd ed., pp. 113-125). NY: Kluwer Academic Publishers.
- SHEAFTER, R. (2000) Statistics for a new century. In BURKE, M. J. & CURCIO, F. R. (Orgs) *Learning mathematics for a new century* (pp. 158-173). Reston, VA: NCTM.
- EMBRETSON, S. E. & REISE, S. P. (2000) *Item Response Theory for Psychologists*. NJ, London: Lawrence Erlbaum Associates, Publishers.

**Sessão Temática III – Importância da Estatística na Educação – sábado, 11 de abril,
Anf. 0.01 (12h40)**

Modelling Efficiency in Education

Isabel Flores¹

¹*ISCTE, icfvs@iscte-iul.pt;*

Abstract

An analysis of aggregate expenditure in education was compared to PISA's scores as a mean of understanding how money link to performance. We analysed global education expenditure, teachers' pay, class size and number of teaching hours, infrastructure and teaching resources investment in European countries. Four groups of efficiency /effectiveness are proposed based on the combination of expense versus results.

Palavras-chave: Education Effectiveness, Efficiency; PISA Scores; Public Expenditure.

Introduction

This paper has the objective of throwing some light on the important subject of education efficiency. Measuring both input and output in social investment is always delicate and far from being straightforward, though it is vital to a well-oiled system.

In a world with finite resources, knowing the price of educating citizens, and how to better use the investment to maximize the outcome is a central question to politicians, managers, parents and every taxpayer. A more efficient system does not need to be feared as disinvestment in education and worst service for citizens. Welch (Welch, 1998) argues that historically “efficiency” means blind cuts to education. Politicians, economists, school managers cannot turn a blind eye on where to spend the money. More money in the system is not, always, a meaning for quality. Efficiency is about choosing the most efficient mix, and deciding in where to invest (Woessmann & Schuetz, 2006).

Results

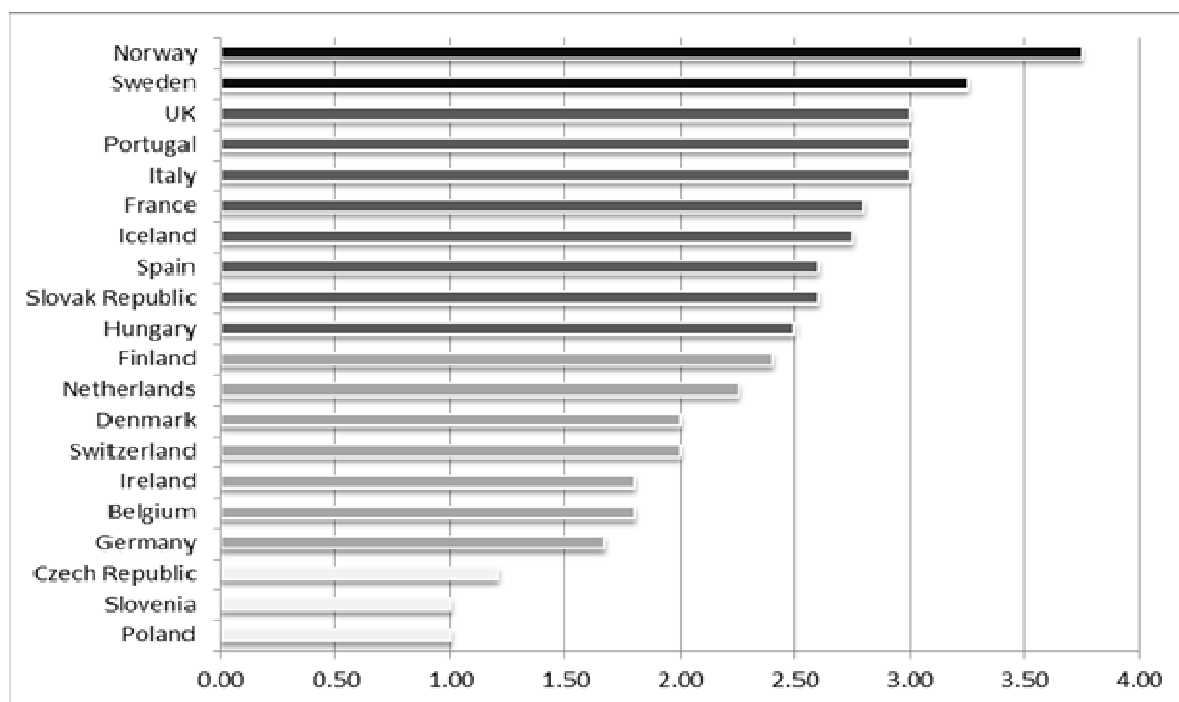
Can schools deliver a better outcome for the same investment? How much will it cost to increase the outcome? Where and how are the countries spending their budgets?

To answer the above questions we gathered aggregate information on the quantity and quality of money into the system (OECD, 2013) and compared those variables to PISA Scores.

A rank in efficiency is proposed. This scale was created by attributing points to each country according to their partial efficiency in the considered variables – Education Expenditure, Teachers pay, teaching time/class sizes and infrastructure. Efficiency (good results at a low cost) scored 1; Efficacy scored 2; results that weren't so good but with low expense scored 3 and finally poor results and lot of investment scored 4.

Conclusion

In terms of global efficiency, it is possible to clearly identify four groups, where groups with a smaller score represent the most efficient.



:1 - Aggregate Efficiency

Poland, Slovenia and Czech Republic are the most efficient countries, meaning PISA scores above average, while expense is kept at a low level. The second group, labelled as effective, is constituted by Germany, Belgium, Ireland, Switzerland, Denmark, Netherlands and Finland. South and East poorer European countries integrate the third group, where the UK emerges as the surprising exception.. Finally, we have a group of just two countries who are very rich and spend much money on education though results are not reaching the desired levels – Norway and Sweden.

References

OECD. (2013). *Education at a Glance 2013*, Paris, OECD Publishing.

WELCH, A. R. (1998) The Cult of Efficiency in Education: Comparative reflections on the reality and the rhetoric. *Comparative Education*, 34(2), 157–175.

WOESSMANN, L. & SCHUETZ, G. (2006) *Efficiency and Equity in European Education and Training Systems*, Brussels, European Commission Education and Culture.

**Sessão Temática IV – Cartas de Controlo de Qualidade – sábado, 11 de abril,
Anf. 0.01 (14h00)**

Control charts for individual measurements of peroxide index

Luís M. Grilo¹, Helena L. Grilo², Isabel M. Nogueira³

¹ *Unidade Departamental de Matemática e Física do Instituto Politécnico de Tomar, e Centro de Matemática e Aplicações (CMA), FCT, UNL, lgrilo@ipt.pt;*

² *Centro de Sondagens e Estudos Estatísticos, Instituto Politécnico de Tomar, helenagrilo56@gmail.com;*

³ *Unidade Departamental de Engenharia, Instituto Politécnico de Tomar, inogueira@ipt.pt.*

Summary

The peroxide index is an important variable measured in order to assess the quality of olive oil. However, the normality assumption of its distribution fails the test. The control limits of individual and moving range charts for peroxide index of virgin olive oil are obtained considering the average and also the median moving range. A non-parametric method based on empirical quantiles, quite robust against deviations from normality, is also applied with the bootstrap procedure to estimate control limits for individual charts. The conclusions obtained with these approaches are relatively different.

Key-words: Non-normality, Non-parametric, Empirical quantiles, Bootstrap, Robustness.

Introduction

Olive oil is a food product with growing importance in health due to its nutritional richness and therapeutic properties. The variable “peroxide index (mEq O₂/Kg) of virgin olive oil” (*PI*), which can be affected by several factors (such as climate, soil and latitude), measures its initial oxidation and subsequent deterioration to assess the quality as well as to characterize the type of oil (e.g., virgin, extra virgin). Control charts have been used effectively for years to monitor processes and to detect abnormal behaviors so, in this study, we obtain individual (*X*) and moving range (*mR*) Shewhart charts in order to analyse central location and to monitor eventual changes in the spread, respectively. The empirical distribution of *PI* is skewed and the normality assumption is violated, which may seriously affect the performance of the corresponding control charts, in what regards the false alarm rate and the time required to detect the occurred changes (Jensen *et al.*, 2006, Borrer *et al.*, 1999, Stoumbos & Reynolds, 2000). According to Wheeler (1995, 2009) these type of charts are a robust tool and work well under non-normality. Despite the controversy, we obtain and compare *X* and *mR* charts for *PI* computing the control limits based on the average and on the median of *mR* (Wheeler, 2010, Grilo & Grilo, 2015). Some simulation studies show the excellent performance of the empirical quantile control chart, which is a special case of the bootstrap control chart (Willemain & Runger, 1996), for a broad range of (non)normal distributions (Vermaat *et al.*, 2003). Thus, with individual observations stem from a process which is statistically in-control, we also estimate robust control limits based on empirical quantiles to make the *X* control chart more sensitive to persistent assignable causes.

Results

Based on a sample size of $n = 32$ individual measurements for PI (obtained in laboratory) of one olive oil supplier of a Portuguese brand leader, we obtain and compare the statistical performance of control limits for X charts considering different approaches. The control limits for the X chart based on empirical quantiles, more sensitive to small shifts in the mean, are calculated with 1000 bootstrap samples (obtained with replacement from the original data).

Conclusions

Although all the original measurements of PI are within the legal specification limits for human consumption, this case study is considered very interesting by the members of the enterprise, once it allows to evaluate the variable PI itself and to compare its behaviour among other brand suppliers. The approaches used here are easy to compute and help to judge whether or not a process is statistically in-control. The control limits estimated with non-parametric method are sufficiently accurate.

Acknowledgments

This work was partially supported by the Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) through the project UID/MAT/00297/2013 (Centro de Matemática e Aplicações).

References

- BORROR, C. M., MONTGOMERY, D. C. & RUNGER, G. C. (1999) Robustness of the EWMA control chart to non-normality. *Journal of Quality Technology*, 31 (3), 309-316.
- GRILO, L. M. & GRILO, H. L. (2015) Individual and moving range control charts in the production of olive oil. *12th International Conference of Numerical Analysis and Applied Mathematics (ICNAAM 2014)*, AIP (American Institute of Physics) Conference Proceedings (in print).
- 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. *Journal of Quality Technology*, 38 (4), 349-364.
- STOUMBOS, Z. G. & REYNOLDS, M. R. Jr. (2000) Robustness to non-normality and autocorrelation of individuals control charts. *Journal of Statistical Computation and Simulation*, 66, 145-187.
- VERMAAT, M. B., DOES, R. J. M. M. & KLAASSEN, C. A. J. (2003) A comparison of Shewhart individuals control charts based on normal, non-parametric, and extreme-value theory. *Quality and Reliability Engineering International*, 19 (4), 337-353.
- WHEELER, D. J. (2010). Individuals Charts Done Right and Wrong, *Quality Digest*, <http://www.qualitydigest.com>, (acedido em 02 de março 2015).
- WHEELER, D. J. (2009). Do you have leptokurtophobia?, *Quality Digest*, <http://www.qualitydigest.com>, (acedido em 02 de março 2015).
- WHEELER, D. J. (1995) *Advanced Topics in Statistical Process Control*, SPC Press, Knoxville, Tennessee.
- WILLEMAIN, T. R. & RUNGER, G. C. (1996) Designing control charts using an empirical reference distribution. *Journal of Quality Technology*, 28 (1), 31-38.

**Sessão Temática IV – Cartas de Controlo de Qualidade – sábado, 11 de abril,
Anf. 0.01 (14h20)**

Monitorização de tempos de procedimentos académicos

Paulo Infante¹, Rita Freitas², Fernanda Otilia Figueiredo³, Gonçalo Jacinto¹, João Dias⁴,

¹*CIMA-UE e ECT, Universidade de Évora, pinfante@uevora.pt / gjcj@uevora.pt*

²*MMEAD, Universidade de Évora, ritabf8@gmail.com*

³*CEAUL e FEUP, otília@fep.up.pt*

⁴*Serviços Académicos, Universidade de Évora, jpsd@uevora.pt*

Sumário

Os tempos associados a procedimentos académicos assumem uma grande importância no bom funcionamento de uma Universidade e na imagem que projeta para o exterior. Neste trabalho discutimos como implementar cartas de controlo para monitorizar esses tempos, alguns dos quais têm distribuições que se afastam muito da normal. Mostramos também como a implementação dessas cartas pode ajudar a identificar causas responsáveis pelo aumento ou diminuição desses tempos. Por fim, para os casos sob controlo estatístico analisamos a capacidade do serviço em cumprir as especificações.

Palavras-chave: Cartas de Controlo, Controlo da Qualidade, Índice de Capacidade do Processo.

Introdução

Depois da entrega de uma dissertação de mestrado, o aluno deixa de poder controlar o tempo que ainda lhe falta até obter o grau. A dissertação demorará algum tempo até chegar ao Conselho Científico, depois decorrerá algum tempo até à formação do júri e em seguida decorrerá algum tempo até o júri reunir e emitir o despacho liminar. Em caso de parecer positivo, decorrerá algum tempo até à realização das provas. Após a discussão, o aluno terá que esperar algum tempo até emissão de certificado de conclusão do curso ou do diploma.

Monitorizar estes tempos através de uma carta de controlo adequada (Gomes *et al.*, 2010) possibilita a avaliação e controlo do desempenho do processo ao longo do tempo, permitindo a distinção entre a variação natural do processo e outras formas de variação, que levam à identificação das causas responsáveis e, conseqüentemente, à implementação de ações que permitem reduzir a variabilidade e/ou trazer o processo para os valores alvo pretendidos. Por outro lado, todos estes tempos têm limites de especificação definidos pelos regulamentos internos da Universidade e conseqüentemente podem ser alvo de estudos de análise de capacidade do processo.

Com base numa amostra constituída por 1095 alunos diplomados que ingressaram na Universidade de Évora entre o ano letivo 2007/2008 e o ano letivo 2012/2013, seguidos até Julho de 2013, implementaram-se cartas de controlo retrospectivas para monitorização destes tempos, mas que poderão ficar desde logo implementadas no serviço, numa perspetiva de controlo em tempo real. Como alguns destes tempos 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 se elaboraram outras cartas de controlo baseadas na distribuição *skew-normal* de acordo com o procedimento descrito em Figueiredo & Gomes (2013).

Resultados

Uma análise exploratória dos tempos dos vários procedimentos permitiu identificar valores excessivamente elevados, pelo que em vários casos se optou por truncar a amostra, usando limites que se entenderam razoáveis em função do conhecimento que se tinha sobre o processo e das suas especificações. Dada a complexidade de interpretação das cartas associadas a cada um dos tempos monitorizados, nos casos em que o processo estava fora de controlo estatístico foram elaboradas cartas de controlo estratificando por diferentes fontes de variação: ano de conclusão do Mestrado, Escola a que pertence o Curso e Curso de Mestrado. Para alguns foi possível identificar comportamentos do processo distintos em dois ou mais períodos temporais, pelo que também se implementou uma carta para cada período, tendo sido investigadas as causas para tais comportamentos.

Vários foram os casos em que a distribuição *skew-normal* se ajustou aos dados. Para alguns destes casos, além das cartas de controlo X-SN (Figueiredo & Gomes, 2013) foram elaboradas também cartas de controlo EWMA, dada a sua robustez, verificando-se comportamentos distintos do processo e assim atestar a importância da distribuição.

No estudo de capacidade do processo, para alguns tempos o processo mostrou-se incapaz, tendo-se obtido índices de capacidade inferiores a 1. O melhor resultado para os serviços foi obtido para o tempo de emissão de um diploma de grau no ano 2010 (Limite Superior Especificação = 30 dias; $\hat{\mu} = 8,2$ dias, $\widehat{Cpku} = 1,68$).

Conclusões

A monitorização dos tempos de procedimentos pode constituir uma mais valia ao permitir identificar causas de variação que, depois de eliminadas, permitem tornar o processo mais capaz em cumprir o especificado pelos regulamentos, reduzindo o tempo de espera e, conseqüentemente, melhorando a qualidade do serviço prestado.

Agradecimentos: Aos Serviços Académicos e aos Serviços de Informática da Universidade de Évora pela cedência dos dados e pelo apoio incondicional que nos deram ao longo do trabalho.

Referências

- FIGUEIREDO, F. & GOMES M. I. (2013) The Skew-Normal Distribution. *REVSTAT - Statistical Journal*, 11 (1), 83-104.
- GOMES, M. I., FIGUEIREDO, F. & BARÃO, M. I. (2010) *Controlo Estatístico da Qualidade*, 2ª ed., Edições SPE.

Sessão Temática IV – Cartas de Controlo de Qualidade – sábado, 11 de abril,
Anf. 0.01 (14h40)

Monitorização da variabilidade de um processo skew-normal multivariado

Adelaide Figueiredo¹, Fernanda Figueiredo²

¹Faculdade de Economia da Universidade do Porto e LIAAD-INESC Porto,
adelaide@fep.up.pt;

²Faculdade de Economia da Universidade do Porto, e CEAUL, Universidade de Lisboa,
otilia@fep.up.pt.

Sumário

A distribuição skew-normal multivariada é muito versátil para modelar dados assimétricos em aplicações práticas, e, em particular, para modelar simultaneamente várias características de qualidade em Controlo Estatístico de Processos. Neste estudo iremos considerar uma carta de controlo baseada na metodologia *Statis* para monitorizar a matriz de covariâncias de um processo skew-normal multivariado, e estudaremos a *performance* desta carta considerando vários processos skew-normal multivariados com características diferentes.

Palavras-chave: Carta de controlo, Controlo Estatístico da Qualidade, Distribuição skew-normal multivariada, Simulação de Monte Carlo, *Statis*.

Introdução

Em Controlo Estatístico de Processos é usualmente necessário monitorizar simultaneamente várias características de qualidade. Frequentemente estas características de qualidade estão correlacionadas, pelo que as técnicas multivariadas de controlo da qualidade são mais apropriadas do que as técnicas univariadas, as quais permitem apenas monitorizar separadamente cada uma das características.

As cartas de controlo são as ferramentas mais utilizadas em Controlo Estatístico de Processos para monitorizar processos industriais e de serviços, e foram introduzidas por Shewhart em 1924 nos Laboratórios *Bell*. As cartas de controlo ajudam-nos a decidir se o processo que está a ser monitorizado está sob controlo ou fora de controlo. Quando uma carta de controlo emite um sinal de que o processo está fora de controlo, que pode ser eventualmente um falso alarme, é importante averiguar quais as causas que são responsáveis pela emissão desse sinal, de modo a tomar-se as ações necessárias para que o processo volte a ficar sob controlo.

Neste estudo iremos apresentar algumas propriedades da distribuição skew-normal multivariada estudada por Azzalini & Dalla Valle (1996) e Azzalini & Capitanio (1999), e em seguida iremos propor um esquema de controlo baseado no método *Statis* para monitorizar a variabilidade de um processo skew-normal multivariado. Por fim, analisaremos a *performance* destes esquemas de controlo analisando situações distribucionais distintas.

Metodologia e Resultados

Várias cartas de controlo têm sido propostas na literatura para monitorizar o vetor médio e/ou a matriz de covariâncias de um processo multivariado (ver por exemplo, Hotelling, 1947,

Alt, 1985 e Chen *et al.*, 2005). Figueiredo & Figueiredo (2014) propuseram um esquema de controlo baseado na metodologia *Statis* para controlar a variabilidade de um processo normal multivariado. A metodologia *Statis* foi introduzida por L'Hermier des Plantes (1976), e mais tarde desenvolvida por Lavit (1988) e Lavit *et al.* (1994).

Seguindo uma metodologia análoga à utilizada no trabalho de Figueiredo & Figueiredo (2014), iremos propor uma carta de controlo baseada no coeficiente RV de similaridade entre matrizes (Escoufier, 1973) para controlar a variabilidade de um processo skew-normal multivariado. Apresentaremos resultados de simulação sobre a eficiência da carta, medida através do Average Run Length (ARL), calculado para vários processos skew-normal bivariados e multivariados.

Conclusão

Os casos analisados mostram que a carta proposta permite detetar facilmente alterações nas correlações entre as variáveis num processo skew-normal multivariado, constituindo assim uma ferramenta bastante útil numa grande variedade de aplicações industriais.

Agradecimentos

Este trabalho é financiado por Fundos FEDER através do Programa Operacional Fatores de Competitividade – COMPETE e por Fundos Nacionais através da FCT – Fundação para a Ciência e a Tecnologia no âmbito dos projectos FCOMP-01-0124-FEDER-037281 e PEst-OE/MAT/UI0006/2014.

Referências

- ALT, F. B. (1985) Multivariate quality control. IN KOTZ, S. and JOHNSON, N. L. (Eds.) *Encyclopedia of Statistical Sciences*, Wiley, New York.
- AZZALINI, A. & CAPITANIO, A. (1999) Statistical applications of the multivariate skew normal distribution. *J. R. Statist. Soc. B*, 61, part 3, 579-602.
- AZZALINI, A. & DALLA VALLE, A. (1996) The multivariate skew-normal distribution. *Biometrika*, 83 (4), 715-726.
- CHEN, G., CHENG, S. W. & XIE, H. (2005) A new multivariate control chart for monitoring both location and dispersion. *Communications in Statistics: Simulation and Computation*, 34, 203-217.
- ESCOUFIER, Y. (1973) Le traitement des variables vectorielles. *Biometrics*, 29, 751-760.
- FIGUEIREDO, A. & FIGUEIREDO, F. (2014) Monitoring the variability of a multivariate normal process using STATIS. IN GILLI, M., GONZALEZ-RODRIGUEZ, G. & NIETO-REYES, A. (Eds.) *Proceedings of COMPSTAT 2014*, 443-450.
- HOTELLING, H. (1947) Multivariate quality control, illustrated by the air testing of sample bombsights. IN EISENHART, C., HASTAY, M.W. and WILLIS, W.A. (Eds.) *Techniques of Statistical Analysis*. McGraw Hill, New York, 111-184.
- LAVIT, C. (1988) *Analyse Conjointe de Tableaux Quantitatives*, Collection Méthodes+Programmes, Masson.
- LAVIT, C., ESCOUFIER, Y., SABATIER, R. & TRAISSAC, P. (1994) The ACT (Statis method). *Computational Statistics and Data Analysis*, 18, 97-119.
- L'HERMIER DES PLANTES, H. (1976) *Structuration des Tableaux a Trois Indices de la Statistique*. Thèse de 3^{ème} cycle. Université de Montpellier II.

SESSÕES PARALELAS

Sessão Paralela I – Classificação, Regressão e Processamento de Imagem – 6ª feira, 10 de abril, Anf. 0.01 (17h00)

A new cleaning method for image restoration using robust confidence intervals

Eunice Carrasquinha¹, Conceição Amado², Ana M. Pires³.

¹*CEMAT – Instituto Superior Técnico, eunice.trigueirao@tecnico.ulisboa.pt;*

²*CEMAT – Instituto Superior Técnico, conceicao.amado@tecnico.ulisboa.pt;*

³*CEMAT – Instituto Superior Técnico, apires@math.ist.utl.pt;*

Abstract

A particular issue of signal processing is image processing, often the improvement of images is the key. This work will focus on image restoration, in particular in image denoising, where the aim is to obtain an image with the least possible noise. There are in literature many methods for image denoising and it is usually assumed that the noise follows a Normal distribution. However, in real data, this could not be true. For this reason, a new image denoising method based on robust bootstrap confidence intervals is proposed, considering that the noise distribution does not need to follow a Normal distribution, but a distribution belonging in the neighbourhood of a contaminated Normal distribution.

Key-words: Bootstrap confidence intervals, Image denoising, Image processing, Robust statistics.

Introduction

Signal processing is involved in many fields of science, and the signals can be represented in many forms. Image processing deals with signals in the form of images, such as photographs, video frames, medical images, and so on. Some of the frequently used types of images are: binary, gray-scale and colour images. The images that are going to be used for application examples in this work are gray-scale images.

Image restoration is one of the main areas of image processing and its goal is to recover the images from degraded observations. In this work we are going to study the problem of distorted images, or most common blurred images. Blurring occurs when the intensity of a pixel is related to the intensity in a larger neighbourhood of the corresponding section. A blurred image is a mathematical model, which contains a blur image and random noise. Nevertheless, blurring is a reversible operation, and it is often called image deconvolution or deblurring. In image deblurring an approximation of the original image is recovered by using a variety of methods.

There are in literature many examples of methods associated to image deblurring. Recently for spatially invariant blurring models Nagy & O'Leary (2002) and Taroudaky (2011) applied image reconstruction algorithms to subimages and construct confidence intervals for pixel values.

In this work, we propose a first step in the process of image restoration, the denoising step, based on the computation of confidence intervals where the distribution of the random

noise, which is always present in blurring images, is not necessary Normal as described in Nagy & O'Leary (2002) and Taroudaky (2011). The method used to compute the confidence intervals was the nonparametric bootstrap, since the sampling distribution of the statistical function is unknown. The classical bootstrap is affected by atypical observations (increasing this issue due to the resample with replacement) and therefore an appropriate choice is the use of robust estimators in the resampling process. The robust bootstrap is based on the trimmed-mean and the Winsorized-mean robust location estimators (Amado, 2003).

Results

A new method based on robust confidence intervals is proposed for cleaning an image, in which the removal of the noise is done using robust confidence intervals, and where other types of noise, besides the additive Normal noise were considered. In order to accomplish this, we had simulated some blurred images with additive noise, where the noise it was generated from Normal and non-Normal distributions. Over these images, the confidence intervals were computed based on robust bootstrap. The results show, regardless of the distribution used to model the noise, the lower and upper bounds are a good approximation of the blurred image, without the influence of noise. Notice, that the proposed technique, is applied and valid to deblurred images that are contaminated with outliers.

Acknowledgments: This work has been supported by Fundação para a Ciência e Tecnologia (FCT, Portugal).

References

- AMADO, M. C. (2003). *Bootstrap Robusto com base na função influência*. PhD thesis. Universidade Técnica de Lisboa, Instituto Superior Técnico.
- NAGY, J. G. & O'LEARY, D. P. (2002). Image restorations through subimages and confidence intervals. *Electronic Transactions on Numerical Analysis*, 13, 22-37.
- TAROUDAKI, V. (2011, Spring Semester) Image deblurring-computation of confidence intervals. Technical Report, University of Maryland.

Sessão Paralela I – Classificação, Regressão e Processamento de Imagem – 6ª feira, 10 de abril, Anf. 0.01 (17h20)

Real Time Activity Recognition for Cycling with Smartphone

Pedro Maia¹, António Abreu²

¹ Student at EST Setúbal/IPS, jpedro.maia@gmail.com;

² EST Setúbal/IPS, antonio.abreu@estsetubal.ips.pt;

Abstract

Studies in activity recognition focus on precision. Computation algorithms for 3% or 4% boost in accuracy are acceptable for computers but will decrease battery time on a smartphone. In this work we present a simple but efficient classifier app for cycling. The application distinguishes three cycling activities: not pedaling (NP), pedaling (PD) and sprinting (SP). We achieved 91.82% accuracy with a two node decision tree on data collected at 5Hz frequency with a non overlapping sliding window.

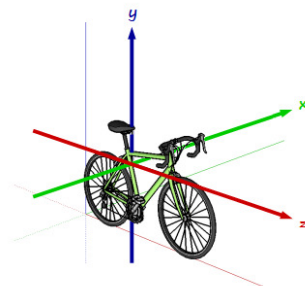
Keywords: Activity recognition; Cycling, Real time, Smartphone.

Introduction

Popular bike apps like Runtastic Mountain Bike GPS, Runtastic Road Bike Tracker, Strava Running and Cycling GPS rely on GPS sensor for every information, e.g. caloric expenditure is table based on velocity (calculated from GPS data) which is prone to error. In downhill trips the effort is less than in uphill trips. In addition to collect GPS data, our cycling activity recognition app gives useful information that can also be used for better caloric consumption estimates.

Design

A smartphone was mounted on the seat tube with the screen pointing to the bike's front. All three accelerometer channels (X, Y and Z) were sampled at approximately 20ms rate in different courses (hill and plane). A total of 177 minutes of raw data resulted in 158 minutes of useful segmented data, of which 18.6%, 67.2% and 14.2% correspond to NP, PD and SP, respectively. Two other training sets were created by downsampling the data. For each training set (20ms, 60ms and 200ms sample rate) we extracted the standard deviation from a 32 sample window. We used Gini's diversity index decision tree algorithm for every possible combination of channels. The decision trees were then optimized and pruned with a 10 fold cross validation.



**Fig. 1: Smartphone mounted on bike
Results**

Fig. 2: Smartphone axis on bike

In Table 1 we present the accuracy and the average F-score (1) of the three activities, for every decision tree implemented.

$$F\text{-score} = 2 \cdot \frac{\textit{precision} \cdot \textit{recall}}{\textit{precision} + \textit{recall}} \quad (1)$$

Table 1: 10 fold cross validation pruned tree

Android sensor delay	GAME		UI		NORMAL	
	rate 20ms, freq 50Hz		rate 60ms, freq 16.6Hz		rate 200ms, freq 5Hz	
Channel	Accuracy	F-Score	Accuracy	F-Score	Accuracy	F-Score
(X)	0.914839	0.893300	0.921153	0.903841	0.918291	0.903858
(Y)	0.816912	0.713656	0.830592	0.737528	0.824588	0.730897
(Z)	0.843647	0.782696	0.855823	0.802962	0.837331	0.775042
(X, Y)	0.934815	0.920701	0.938950	0.928445	0.948276	0.940107
(X, Z)	0.914839	0.893300	0.927686	0.908144	0.925787	0.908188
(Y, Z)	0.928807	0.912877	0.941428	0.929575	0.935532	0.923381
(X, Y, Z)	0.961475	0.954236	0.954945	0.946779	0.967766	0.962258

Conclusions

We demonstrated successfully with this work, that it's possible to create a simple activity classifier for cycling, using only one channel (X) and to our knowledge with the lowest sampling rate (200ms, 5Hz) used in a activity recognition study. The classification is useful by itself but it can also be used for better caloric assessment on app trackers such as Strava.

Decision trees with features from all channels provide better accuracy, but increase algorithm complexity, which in turn increases computation time, leading to quicker battery exhaustion. Energy efficiency study of our solution remains to be made, but it's our belief that simple and energy efficient solutions will be preferred by smartphone users, and most academic studies don't take energy efficiency into account, because they focus solely on precision.

References

- ZHAN, A., CHANG, M., CHEN, Y., & TERZIS, A. (2012) Accurate Caloric Expenditure of Bicyclists using Cellphones. ESKICIOGLU, R., CAMPBELL, A. & LANGENDOEN, K. (Eds.) *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*. ACM New York.
- YEO, J., LIM, Y., RYU, M., & YANG, Y. (2012) An Automatic Recognition of Bicycle Riding State by Using a Smartphone. CHANG, C., CAYTILES, R. & GELOGO, Y. (Eds.) *Proceedings of the International Conference, SoftTech 2012*. Science and Engineering Research Support Society.

Sessão Paralela I – Análise de Dados em Educação – 6ª feira, 10 de abril,
Sala 0.02 (17h00)

Escolas Secundárias atípicas no panorama nacional

Margarida Vilela¹, M. Rosário Oliveira²

¹ Instituto Superior Técnico, Universidade de Lisboa, margarida.azeitona@tecnico.ulisboa.pt;

² CEMAT e Departamento de Matemática, Instituto Superior Técnico, Universidade de Lisboa, rosario.oliveira@tecnico.ulisboa.pt;

Sumário

Desde 2001, o Ministério da Educação e Ciência divulga os resultados nos exames nacionais e os diversos órgãos de comunicação social analisam esses dados e produzem *rankings* de escolas que geram polémica entre especialistas e responsáveis educativos.

Neste trabalho usam-se as abordagens clássicas e robustas de análise em componentes principais para ordenar as escolas secundárias tendo por base os resultados divulgados relativamente ao ano letivo 2012/2013 (*vide* JNE, 2013). Inclui-se também uma caracterização das instituições de ensino atípicas identificadas, de acordo com o tipo (pública ou privada) e a sua localização geográfica.

Palavras-chave: Componentes principais, Ensino secundário, Métodos robustos, Observações atípicas.

Introdução

A publicação dos *rankings* das escolas foi introduzida na rotina do ano escolar em 2001. Desde então, cada órgão de comunicação social estabelece critérios próprios (por exemplo, disciplinas consideradas, mínimo de provas por disciplina, *etc.*) e efetua a sua ordenação de escolas baseada nos resultados dos exames nacionais.

Neste trabalho pretende-se ordenar as escolas de acordo com o desempenho dos seus alunos do ensino secundário, do perfil de Ciências e Tecnologias, uma vez que este é preferencialmente o perfil dos alunos que ingressam no Instituto Superior Técnico. Cada escola é caracterizada de acordo com notas dos seus alunos nos exames nacionais em 5 disciplinas: Matemática A, Português, Biologia e Geologia, Geometria Descritiva, Física e Química A e ainda baseadas nas notas atribuídas pelos professores no final do ano letivo nessas mesmas disciplinas, denominadas classificações internas de frequência (CIF).

Para eliminar potenciais distorções, consideraram-se apenas escolas com pelo menos 10 alunos aprovados em cada uma das disciplinas em estudo. A conjugação destes critérios permitiu obter um conjunto formado por 225 instituições de ensino, caracterizadas pelas medianas das notas de exame e de CIF, em cinco disciplinas da área de Ciências e Tecnologias, que serão usadas para estabelecer um *ranking* de escolas, baseado na análise em componentes principais (ACP).

A construção das componentes principais pode ser muito afetada pela existência de um pequeno conjunto de instituições com características muito distintas da globalidade das suas congéneres, o que tornará este indicador menos interessante para o fim em vista. Inspirado nas dificuldades e soluções frequentemente encontradas na estimação das componentes principais

baseadas em amostras com observações atípicas, usam-se métodos robustos de construção de componentes principais (Croux *et al*, 2007; Croux-Ruiz-Gazen, 2005; Filzmoser & Todorov, 2011; Hubert *et al*, 2005; Locantore *et al*, 1999).

Para além de acomodar convenientemente as instituições de ensino atípicas neste processo, é igualmente relevante identificar e perceber a razão da sua “atipicidade” no panorama global do ensino em Portugal. Por esta razão, a ACP robusta, em combinação com a distância de Mahalanobis robusta, foram usadas para identificar estas instituições (*vide* Pascoal, 2014).

Discussão e Resultados

A análise preliminar dos dados permitiu verificar que as CIF são, por norma, superiores às classificações de exame. Note que as CIF das várias disciplinas são sempre superiores a 100, uma vez que só nessas condições os alunos são admitidos a exame (*vide* Sousa e Vilela, 2013 e Vilela, 2015).

Os resultados principais obtidos para a ordenação das escolas estão refletidos na deteção de instituições atípicas baseada na ACP robusta. Estas instituições dividem-se em três grupos. O primeiro é formado por escolas com fraco desempenho na globalidade das avaliações consideradas e é constituído essencialmente por escolas públicas. O segundo e terceiro grupos são formados por escolas com bons resultados globais e estão localizadas maioritariamente na região Norte e Lisboa e Vale do Tejo. O segundo grupo é constituído por escolas privadas com os resultados mais elevados e o terceiro grupo para além de escolas privadas contém igualmente escolas públicas e os seus resultados são bons mas, em geral, mais baixos que os do grupo dois.

Referências

- CROUX, C., FILZMOSER, P., & OLIVEIRA, M. R. (2007). Algorithms for Projection - Pursuit robust principal component analysis. *Chemometr. Intell. Lab.*, 87(2),218–225.
- CROUX, C. AND RUIZ-GAZEN, A. (2005). High breakdown estimators for principal components: the projection-pursuit approach revisited. *J. Multivariate Anal.*, 95(1),206–226.
- FILZMOSER, P. AND TODOROV, V. (2011). Review of robust multivariate statistical methods in high dimension. *Anal. Chim. Acta*, 705(1–2),2 – 14.
- HUBERT, M., ROUSSEEUW, P., AND VANDEN BRANDEN, K. (2005). ROBPCA: A new approach to robust principal component analysis. *Technometrics*, 47(1),64–79.
- JNE (2013). DGE - Bases de Dados - Arquivo 1998 - 2013 - Júri Nacional de Exames, <http://www.dge.mec.pt/jurinacionalexames/data/jurinacionalexames/estatisticas/enes2013.zip>, (acedido em 30 de Novembro 2013)
- LOCANTORE, N., MARRON, J.S., SIMPSON, D.G., TRIPOLI, N., ZHANG, J.T. AND COHEN, K.L. (1999). Robust principal component analysis for functional data. 8(1),1–73.
- PASCOAL, C. (2014). *Contributions to Variable Selection and Robust Anomaly Detection in Telecommunications*. Tese de Doutoramento. Técnico Lisboa, UL, Lisboa.
- SOUSA, B. AND VILELA, M. (2013). Projecto em Análise Multivariada. *Technical Report*, Técnico Lisboa.
- VILELA, M. (2015). Projecto em Tópicos Avançados de Análise Multivariada. *Technical Report*, Técnico Lisboa.

Sessão Paralela I – Análise de Dados em Educação – 6ª feira, 10 de abril,
Sala 0.02 (17h20)

Avaliação do desempenho dos alunos Portugueses em Matemática – PISA 2009

Susana Faria¹, Conceição Portela²

¹ CMAT – Centro de Matemática, Departamento de Matemática e Aplicações, Universidade do Minho, sfaria@math.uminho.pt

² CEGE - Centro de Estudos em Gestão e Economia, Universidade Católica Portuguesa do Porto, Portugal, csilva@porto.ucp.pt

Sumário

Neste trabalho, aplicando um modelo multinível de regressão quantílica, analisam-se os resultados dos alunos portugueses relativamente à literacia em Matemática com base nos dados disponibilizados pelo PISA 2009. O estudo pretende identificar os principais factores que podem determinar os resultados dos alunos portugueses e perceber em que medida estes fatores influenciam os resultados dos alunos em diferentes quantis de desempenho. Pretende ainda estudar a contribuição das escolas portuguesas nos resultados dos seus alunos de diferentes quantis de desempenho.

Palavras-chave: Desempenho dos alunos, Modelo multinível de regressão quantílica, PISA 2009

Introdução

O PISA – *Programme International for Students Assessment* desenvolvido pela Organização para a Cooperação e Desenvolvimento Económico (OCDE) visa avaliar, em ciclos de 3 anos, a capacidade dos estudantes de 15 anos para utilizarem os conhecimentos adquiridos em situações que ocorrem no mundo real. Este programa avalia o desempenho dos alunos em três domínios: leitura, matemática e ciências.

Neste estudo pretende-se analisar os resultados dos alunos portugueses no domínio da Matemática com base nos dados do PISA 2009 disponibilizados pela OCDE.

Dada a estrutura hierárquica dos dados (os alunos estão organizados em turmas e as turmas em escolas), aplicámos modelos de regressão multinível que, na sua construção, incorporam as diferentes hierarquias observacionais dos dados (Raundenbush & Bryk, 2002). Como estávamos interessados em identificar os principais fatores que influenciam os resultados dos alunos portugueses com diferentes níveis de desempenho, ajustámos um modelo multinível de regressão quantílica (Geraci & Bottai, 2014). Inicialmente, estimou-se um modelo multinível de regressão quantílica apenas com variáveis explicativas ao nível do aluno para identificar os principais fatores que podem determinar os resultados dos alunos portugueses. De seguida, estimou-se um modelo de regressão linear usando como variável resposta o efeito das escolas para analisar os fatores que justificam o impacto das escolas nos resultados dos alunos.

Resultados

Este estudo evidenciou que o género tem um impacto positivo (os rapazes apresentam melhores resultados) e aproximadamente constante para os vários quantis de desempenho. A condição de imigrante do aluno leva a que este obtenha menores resultados, sendo esta condição mais penalizadora para alunos nos quantis de desempenho mais baixos. O grau frequentado pelo aluno tem um grande impacto nos resultados sendo este impacto positivo para os alunos que frequentam o secundário e negativo para os alunos que frequentam o ensino profissional.

Relativamente aos fatores familiares, as expectativas da família e do aluno desempenham um papel muito importante no sucesso dos alunos bem como o seu nível socioeconómico. O impacto do nível socioeconómico é positivo, mas aproximadamente constante para os vários quantis de desempenho, evidenciando que estudantes em ambientes socioeconómicos mais favorecidos apresentam resultados a matemática superiores independentemente da sua posição na distribuição dos resultados. O apoio da família nos trabalhos de casa, bem como o recurso dos alunos a explicações foram variáveis que apresentaram sinal contrário ao esperado (impacto negativo nos resultados) devido à endogeneidade associada a estas variáveis. No que diz respeito a variáveis relativas à aprendizagem, a repetição de ano aparece com um grande impacto negativo nos resultados dos alunos repetentes. As estratégias de aprendizagem revelaram-se variáveis interessantes e maioritariamente com impacto positivo nos resultados (e.g. estratégias de elaboração, de controle, de sumarização), mas as estratégias de memorização aparecem com um impacto negativo nos resultados dos alunos, sendo estas estratégias particularmente penalizadoras para os alunos que se encontram nos níveis de desempenho mais baixo. Este estudo mostrou que as escolas portuguesas desempenham um papel importante na explicação do sucesso dos alunos em matemática. O maior impacto das escolas verificou-se para os alunos do quantil 0.5 onde cerca de 10% da variabilidade encontrada nos resultados dos alunos pode ser explicada pela escola que frequentam. No entanto, apesar da escola que o aluno frequenta ter impacto nos seus resultados, apenas as variáveis localização da escola e dimensão (número total de alunos) explicam o efeito da escola. Este estudo permitiu ainda mostrar que há grandes diferenças no desempenho dos alunos a matemática nas escolas portuguesas, existindo escolas em que os alunos têm um desempenho superior ao desempenho esperado em todos os quantis de desempenho e noutras escolas os alunos têm um desempenho inferior ao desempenho esperado em todos os quantis de desempenho.

Agradecimentos: S. Faria foi financiada pelo Centro de Matemática da Univ. do Minho com Fundos Nacionais através da FCT no âmbito do projeto PEst-OE/MAT/UI0013/2014.

Referências

- GERACI, M. & BOTTAI, M. (2014) Linear quantile mixed models. *Statistics and Computing* 24 (3), 461-479.
- RAUDENBUSH, S.W. & BRYK, A.S. (2002) *Hierarchical Linear Models* (Second Edition), Thousand Oaks, Sage Publications.

Sessão Paralela I – Análise de Dados em Educação – 6ª feira, 10 de abril,
Sala 0.02 (17h40)

Papel da ação social no Ensino Superior Politécnico: expectativas futuras dos estudantes bolseiros com recurso à Análise de Multivariada de Dados

Otília Dias¹, Raquel Pereira², Boguslawa Sardinha³

¹ ESTBarreiro/IPS, otilia.dias@estbarreiro.ips.pt;

² ESCE/IPS, raquel.pereira@esce.ips.pt;

³ ESCE/IPS, boguslawa.sardinha@esce.ips.pt;

Sumário

O objetivo geral deste trabalho é o de analisar o impacto dos apoios sociais no acesso e frequência do Ensino Superior Politécnico Público (ESPP) por parte de estudantes oriundos de famílias economicamente carenciadas. Serão apresentados os resultados de um inquérito por questionário, realizado aos estudantes bolseiros do ESPP no ano letivo 2013/2014, no sentido de perceber qual o papel da ação social neste tipo de ensino, na perspetiva dos estudantes apoiados. Para além da caracterização do público-alvo dos apoios sociais, analisaremos a importância que os estudantes bolseiros atribuem aos apoios sociais, bem como as suas aspirações profissionais e expectativas futuras.

Palavras-chave: Ação social, Análise Ordinal, Análise Multivariada, Ensino superior politécnico.

Introdução

O objetivo deste estudo é apresentar a ação social no ESPP como uma das parceiras que poderá influenciar a igualdade de oportunidade de frequência neste tipo de ensino aos estudantes com níveis de rendimento mais baixos e surge no âmbito de um projeto financiado pelo Fundo Social Europeu (FSE) tendo sido proposto pelas Escolas Superiores de Ciências Empresariais (ESCE) e de Tecnologia do Barreiro (ESTBarreiro) do Instituto Politécnico de Setúbal (IPS) iniciado em janeiro de 2014.

As dificuldades financeiras que Portugal tem enfrentado, influenciaram drasticamente o prosseguimento de estudos de muitos estudantes portugueses. Compete aos Serviços de Ação Social (SAS) proporcionar as melhores condições de estudo aos estudantes, sobretudo aos carenciados, através da disponibilização de um conjunto de apoios diretos e indiretos.

Resultados

Foi realizado um inquérito com 44 questões e enviado aos estudantes bolseiros dos 14 ISPP, tendo-se obtido 708 respostas oriundas do Instituto Politécnico de Bragança, do Instituto Politécnico de Cávado e Ave, do Instituto Politécnico de Setúbal e do Instituto Politécnico de Viseu. Neste questionário, de natureza confidencial, para além da caracterização pessoal e

familiar também foram inquiridas questões referentes aos apoios sociais e às aspirações profissionais e expectativas dos estudantes bolseiros.

No trabalho que aqui apresentamos, a análise dos dados foi realizada com o *software* IBM SPSS 22.0, começando por uma breve caracterização da amostra, seguindo-se a análise das relações entre as variáveis que nos pareciam interessantes, após conversa com a administradora do SAS IPS, aplicando os coeficientes de correlação adequados. O estudo efetuado por recurso à análise multivariada apenas incidiu sobre a variável 44: “ Em que medida considera que a obtenção de um curso superior o vai ajudar em cada um dos seguintes aspetos?” Esta questão foi respondida considerando a resposta a 12 itens, numa escala que vai de (1) *Nada influente* a (5) *Muitíssimo influente*.

Sabendo que a Análise em Componentes Principais (ACP) é uma técnica que pode ser aplicada apenas a variáveis quantitativas, e que neste estudo temos 12 variáveis (V44A, V44B,..., V44L) qualitativas medidas numa escala ordinal, aplicámos uma variante desta conhecida técnica – a Análise Ordinal (AO).

Conclusões

Os resultados obtidos na análise multivariada têm uma interpretação coerente, que podemos resumir como sendo: A primeira componente relaciona-se com as motivações sociais, a segunda com as motivações pessoais, e por fim a terceira componente inclui as motivações de autoestima relacionadas com a realização pessoal e profissional que os jovens estudantes esperam vir a alcançar com a obtenção de um curso superior.

Agradecimentos

- Este projeto foi financiado pelo Programa Operacional de Assistência Técnica do Fundo Social Europeu, no âmbito do QREN - POAFSE 000755402014 - Estratégias de Ação Social no Ensino Superior.
- Aos Serviços de Ação Social dos Institutos Politécnicos Públicos que colaboraram no questionário.

Referências

CERDEIRA, M. (2008) *O Financiamento do Ensino Superior Português – a partilha de custos*, Tese de Doutoramento em Ciências da Educação, Faculdade de Psicologia e de Ciências da Educação, Universidade de Lisboa.

DIAS, O. SARDINHA, B. S e PEREIRA, R. (2014). *Ação Social no Ensino Superior: Um caso de estudo no Instituto Politécnico de Setúbal*, no Eixo Temático *O Financiamento e a expansão do Ensino Superior*, FORGES, Universidade Agostinho Neto e Universidade Mandume Ya Ndemufayo.

LEBART, L.; MORINEAU, A.; PIRON, M. (1995): *Statistique Exploratoire Multidimensionnelle*, Dunod Ed., Paris.

Sessão Paralela I – Análise de Dados em Ciências Empresariais e Sociais – 6ª feira, 10 de abril, Sala 0.07 (17h00)

Previsão de *Churners*: Quem? E Quando?

Margarida Vilela¹, Conceição Amado¹, Beatriz Sousa¹, Jéssica Vieira¹, Luís Medeiros²,
¹ Instituto Superior Técnico e CEMAT, Universidade de Lisboa,
ana.d.sousa@tecnico.ulisboa.pt, conceicao.amado@tecnico.ulisboa.pt,
jessica.vieira@tecnico.ulisboa.pt, margarida.azeitona@tecnico.ulisboa.pt;
² Colours' Company, luis.medeiros@colours.pt

Sumário

A importância do *Churn Management* prende-se com o facto de ser menos dispendioso reter clientes do que angariar novos. A maioria dos métodos de *Data Mining* aplicados a este problema tem permitido a criação de modelos preditivos bem-sucedidos, no entanto estes dificilmente identificam o momento exato em que um cliente irá abandonar a empresa.

Assim, o objetivo deste trabalho é a aplicação de modelos de análise de sobrevivência usuais e modelos de fragilidade à previsão de *churners* de uma empresa de cartões de combustível, e a sua posterior comparação com alguns métodos de *Data Mining*.

Palavras-chave: Análise de sobrevivência, *Churn*, *Data mining*, Modelos de fragilidade paramétricos.

Introdução

No âmbito do *Churn Management*, há uma preocupação constante em identificar clientes em risco de abandonar a empresa, para que se possa agir no sentido de os reter. Importa referir que, neste contexto, um *churner* é um cliente que inicia um período de 7 semanas sem consumir no mês em estudo ou no seguinte.

O desenvolvimento de modelos preditivos baseados em métodos de *Data Mining*, em geral, conduz a bons resultados. Contudo, a análise de sobrevivência surge como uma metodologia mais natural quando o foco da análise é não só a identificação de possíveis desistentes, mas acima de tudo a localização dessa ocorrência no tempo.

O conjunto de dados trabalhado foi cedido pela Colours' Company e continha registos relativos a clientes portadores de cartões de abastecimento de combustível. Neste tipo de estudos é usual considerar 3 tipos principais de variáveis: comportamentos, perceções e informações demográficas do cliente. No entanto, neste estudo as covariáveis utilizadas focam-se essencialmente no comportamento dos clientes e nos seus hábitos de consumo. A satisfação do cliente foi, por diversas vezes, identificada como sendo um dos fatores preponderantes no processo de retenção, mas variáveis que medem o nível de satisfação dos clientes são subjetivas e difíceis de obter. Assim, é inevitável a existência de covariáveis omissas no modelo e torna-se aconselhável incluir efeitos aleatórios que modelem essas covariáveis ocultas. Neste contexto surgem os modelos de fragilidade, um tipo particular de modelos de sobrevivência, no qual se atribui um efeito (fragilidade) a cada cliente ou um efeito comum a um grupo de clientes (PORTELA, 2010, pp. 93-94).

A seleção das variáveis mais apropriadas influencia diretamente o desempenho dos modelos de classificação, pelo que o primeiro passo do processo de modelação foi a análise e posterior reformulação das variáveis a considerar.

Resultados

Uma análise preliminar dos dados permitiu detetar que os *churners* são a classe minoritária, correspondem a menos de 1% das instâncias e a maioria dos clientes nunca registou períodos sem consumo ou este foi um evento muito esporádico.

As variáveis consideradas podem ser agrupadas em dois blocos:

1. Cartões: proporções de cartões utilizados em transações (face ao número total).
2. Consumos
 - rácios de consumos médios e de consumos medianos;
 - variáveis referentes a máximos e mínimos;
 - tendências dos mínimos (decréscimos de consumo);
 - número de decréscimos registados nas últimas 6 semanas;
 - número de semanas sem consumo;
 - número de mínimos.

Na construção de modelos de fragilidade foram consideradas as seguintes distribuições *baseline*, Exponential, Weibull e Gompertz, combinadas com as distribuições de fragilidade, Gama, Gaussiana Inversa, Estável Positiva e Lognormal. A modelação dos dados foi efetuada recorrendo a alguns pacotes do software *R*, nomeadamente *parfm* e *flexsurv*.

Conclusões

Na generalidade, os modelos de fragilidade obtidos são precisos na identificação do instante em que o cliente desiste, quando sabemos que de facto o cliente irá abandonar a empresa, isto é, dado o seu processo de estimação de parâmetros, estes modelos não podem ser utilizados como modelos de classificação.

Adicionalmente, os resultados dos modelos de análise de sobrevivência que não incorporam termos de fragilidade foram comparados com os obtidos pelos modelos de Data Mining usuais (Árvores de decisão, *AdaBoost* e *Random Forest*) tendo-se verificado que ao nível da capacidade de classificação, estes últimos apresentam melhor desempenho.

Agradecimentos: Este trabalho foi realizado em colaboração com a Colours' Company e desde já gostaríamos de agradecer pela disponibilização dos dados analisados, bem como por toda a ajuda prestada.

Referências

- DUCHATEAU, L. & JANSSEN, P. (2010) *The Frailty Model*. New York, Springer.
- PORTELA, S. (2010) *Modelling partial customer churn in the Portuguese fixed telecommunications industry by using survival models*. Tese de doutoramento. ISCTE, IUL, Lisboa. Disponível em [www:<http://hdl.handle.net/10071/4416>](http://hdl.handle.net/10071/4416).
- LU, J.X. (2002) Predicting customer churn in the telecommunications industry: an application of survival analysis modeling using SAS, *SAS User Group International (SUGI27) Online Proceedings*, Paper No. 114-27, SAS Institute Inc., Cary, NC, USA.

Sessão Paralela I – Análise de Dados em Ciências Empresariais e Sociais – 6ª feira, 10 de abril, Sala 0.07 (17h20)

Multidimensional relationship: education and fertility

Lídia Patrícia Tomé¹, Maria Filomena Mendes²

¹CIDEHUS – Universidade de Évora, lidiatome@uevora.pt;

² CIDEHUS – Universidade de Évora, mmendes@uevora.pt;

Summary

Influenced by the social and economic contexts and transformations, the educational level and the motherhood timing play a central role in shaping the life courses trajectory. E.g. in Portugal, the *explosion* of higher educated female resulted in the significant postponement of period fertility. We extend our analysis to the cohort fertility, (1) reconstructing life courses events for 5 European countries using IPUMS samples; (2) elaborating cohort life fertility tables to analyse the transition to motherhood; and (3) observing the impact of educational levels in the cohort fertility patterns.

Key words: Cohort reconstruction, Education, Fertility, Postponement.

Background and motivation

We assist in the European societies to a considerable improvement in the educational system and in the requalification of the labour market as result of higher educational levels. Thus, the relationship between educational attainment, and the transition to motherhood has been an important topic of discussion in the literature (e.g. Hoem *et al.* 2006; Lappegård & Rønsen 2005). Influenced by the social and economic contexts and transformations, the educational level and the motherhood timing play a central role in shaping the life courses trajectory. E.g. in countries, such as Italy, Portugal or Spain, the massification in the access to education in the 1960's and the *explosion* of female higher education enrolment seems to result in the significant postponement and decline of fertility. Once that is expected to observe different behaviours between different cohorts, the goals of the current research are, (1) to reconstruct life courses events for five European selected countries (Austria, Greece, Hungary, Portugal and Spain) using for that IPUMS samples; (2) elaborate cohort life fertility tables to analyse the transition to motherhood; and (3) observe the impact of educational levels in cohort fertility.

Data and methods

For the analysis of transition to motherhood we used fertility table methods (Chiang 1984). The analyses of survival curves are the result of data reconstruction of several cohorts by educational levels in the years of 1991 and 2001 using IPUMS database. The data used is the IPUMS census samples of 5% (Hungary, Portugal and Spain) and 10% (Austria and Greece) sample of the population in each one of the countries.

Preliminary Results – Portugal as an example in transition to motherhood

Figure 1 displays the estimated survival curves for the transition to motherhood for two groups of cohorts, 1950-1955 (women between 36-41 in 1991) and 1960-1965 (women between 36-41 in 2001). For the cohorts 1950-1955 (panel 1) we observe a decrease in the median age at first child (i.e., the age at which 50% of the respondents become mothers). For the cohorts 1960-1965 (panel 2) the exact opposite situation is observed, with the increase of the mean age. Between cohorts born in 1950-1955, the decline was from 27 (1950 cohort) to 24 (1955 cohort) years, while for the cohorts 1960-1965 the increase in the median age was smaller, from 25 in 1960 cohort to 26 years in 1965 cohort.

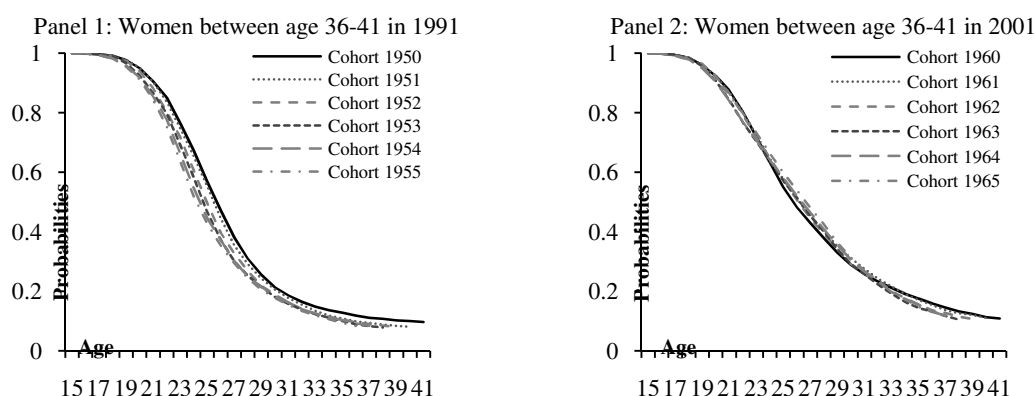


Figure 1: Transition to motherhood for women between age 36-41 in 1991 (1950 cohort) and 2001 (1960 cohort)
Source: Own elaboration; IPUMS

Preliminary exercises, not presented here revealed that median age by education level increased between the 1950 and 1960 cohorts and within the educational levels. The evolution of transition to motherhood presented similar behaviours between the different levels of education. Further research in the 1965 cohort, showed that Spain was the country with higher postponement while in 1955 Austria was the one with low postponement followed by a marked delay of fertility after age 26.

References

- CHIANG C. (1984) *The life table and its applications*, Robert e. Krieger publishing company.
- HOEM, J.M., NEYER, G. & ANDERSSON, G. (2006) Education and childlessness: The relationship between educational field, educational level, and childlessness among Swedish women born in 1955-59. *Demographic Research* 14 (15).
- LAPPEGÅRD, T. & RØNSEN, M. (2005) The multifaceted impact of education on entry into motherhood. *European Journal of Population*, 21.

Sessão Paralela I – Análise de Dados em Ciências Empresariais e Sociais – 6ª feira, 10 de abril, Sala 0.07 (17h40)

A mortality study under different selected frameworks

Filipe Ribeiro¹, Maria Filomena Mendes²

¹CIDEHUS - Universidade de Évora, fribeiro@uevora.pt;

²CIDEHUS - Universidade de Évora, mmendes@uevora.pt.

Abstract

The challenge to create a universal law that explains mortality over age resulted in different proposals obtaining very satisfactory results and growing complexity. Decision makers pursuit every day accurate information to take suitable and appropriate decisions to coping with countries' reality. With this study we intend to re-evaluate specific selected laws of mortality and calculate subsequent model-based life expectancies, contributing with enlightening knowledge to explain mortality trends.

Key words: Life expectancy, Maximum likelihood, Mortality laws.

Introduction

In Japan, a female born in 2012 could expect to live slightly more than 86 years, resulting from an average 0.5 yearly increase since 1947 (HMD, 2015). It is well known that this increase is only possible due to significant improvements in health and consequent reductions in mortality rates (Vaupel, 2010). Those rates have been widely studied attempting to generate a universal law of mortality. Gompertz (1825) proposed a simple parametric model $\mu = ae^{bx}$ to explain adult mortality, accounting for the level of mortality at the starting age a , and for the rate of mortality increase with age b . Later, Makeham (1860) add a constant to the model $\mu = ae^{bx} + c$ improving the results. The inclusion of this new parameter allows either to differentiate between background (c) and intrinsic (a) mortality. Despite the improvements brought by the inclusion of a third parameter in the model, mortality is still overestimated at oldest ages, leading to the suggestion of employing logistic models (Perks, 1932; Heligman and Pollard, 1980; Thatcher, 1999). Nevertheless, demographers view populations as a heterogeneous mixture of individuals that share the same baseline hazard of death, to which they are susceptible in a different and random way. Within this framework, by assumption (Gompertz, 1825; Vaupel *et al.*, 1979), the baseline hazard follows a Gompertz curve. Individual susceptibility can be reflected in different individual a 's (relative risk models), in different b 's (accelerated life models), or in both. However, recent evidence for a human mortality plateau (Gampe, 2010) speaks in favor of relative risks only. Letting a random variable Z , called frailty (Vaupel *et al.*, 1979), capture individual susceptibility, leads, very generically, to the widely used gamma-Gompertz frailty model.

Methodological approach

The attempt to create a universal law that explains mortality over age resulted in different proposals obtaining very satisfactory results and growing complexity. With this study we

intend to re-evaluate four of them accounting selected countries (the Gompertz's law (Gompertz, 1825); the Makeham's law (Makeham, 1860); the logistic approach developed by Thatcher (1999); and the so-called gamma-Gompertz model (Vaupel *et al.*, 1979)) and test its implications in estimating life expectancy. Under this framework, model estimation holds on the assumption that deaths follow a Poisson distribution (Brillinger, 1986).

Expected Results

From our study, we expect that, in a general way, the logistic models present better goodness of fit and as well that the Makeham's constant term also proves to be essential, likewise referenced in different studies (Horiuchi & Coale, 1990). Comparing the obtained outputs, we also expect that the gamma-Gompertz model and the gamma-Gompertz-Makeham will give an important insight to understand the distinctive patterns in the pace of mortality increase at the individual and population level providing distinctive and innovative outcomes.

References

- BRILLINGER, D.R. (1986) The natural variability of vital rates and associated statistics. *Biometrics*, 693-734.
- GAMPE, J. (2010) Human mortality beyond age 110. *In*: H. Maier, J.G.B. Jeune, J.M. Robine and J.W. Vaupel (eds), *Supercentenarians*, Heidelberg, 219-230.
- GOMPERTZ, B. (1825) On the nature of the function expressive of the law of human mortality, and on a mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London* 115, 513-585.
- HELIGMAN, L. & POLLARD, J.H. (1980) The age pattern of mortality. *Journal of the Institute of Actuaries* 107(1), 49-80.
- HORIUCHI, S. & COALE, A.J. (1990) Age patterns of mortality for older women: An analysis using the age-specific rate of mortality change with age. *Mathematical Populations Studies* 2(4), 245-267.
- Human Mortality Database (2014). University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org
- MAKEHAM, W.M. (1860) On the law of mortality. *Journal of the Institute of Actuaries* 13, 283-287.
- PERKS, W. (1932) On some experiments in the graduation of mortality statistics. *Journal of the Institute of Actuaries* 63, 12-40.
- THATCHER, A.R. (1999) The long-term pattern of adult mortality and the highest attained age (with discussion). *Journal of the Institute of Royal Statistical Society* 127(1), 5-43.
- VAUPEL, J.W. (2010) Biodemography of human aging. *Nature* 7288, 536-542.
- VAUPEL, J.W., MANTON, K. and STALLARD, E. (1979) The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* 16, 855-860.

Sessão Paralela II – Aplicação de Métodos de Regressão – sábado, 11 de abril,
Sala 0.02 (9h45)

Um perfil de quem não tem nem quer ter filhos em Portugal*

Andreia Maciel¹, Maria Filomena Mendes¹, Paulo Infante², Anabela Afonso², Rita Freitas³.

¹ CIDHEUS e ECS, Universidade de Évora, deiabarroso@hotmail.com / mmendes@uevora.pt;

² CIMA-UE e ECT, Universidade de Évora, pinfante@uevora.pt / aafonso@uevora.pt;

³ CIDHEUS e MMEAD, Universidade de Évora, ritabf8@gmail.com.

Sumário

Portugal regista atualmente um dos mais baixos níveis de fecundidade da Europa e do mundo. As consequências de um tal cenário são ainda mais preocupantes por nunca se terem registado tão poucos nascimentos e pelo sucessivo adiamento da entrada na maternidade e paternidade. Com base nos dados do Inquérito à Fecundidade realizado pelo INE em 2013, neste trabalho, recorrendo à regressão logística, identificamos algumas características de quem não quer ter filhos.

Palavras-chave: Amostra complexa, Demografia, Fecundidade, Regressão logística.

Introdução

No Inquérito à Fecundidade (IFEC), realizado no âmbito de um protocolo celebrado entre a Fundação Francisco Manuel dos Santos (FFMS) e o Instituto Nacional de Estatística (INE), os indivíduos foram interrogados sobre se tinham algum filho e, também, se ainda pretendiam ter ou ter mais filhos durante a sua vida reprodutiva. Aqueles que responderam que não tinham filhos e não pretendiam tê-los foram considerados como tendo tomado uma decisão definitiva, denominada de *childlessness* permanente. Por oposição, aqueles que não tendo filhos ainda pretendiam vir a tê-los foram considerados como estando numa situação não definitiva, designada por *childlessness* temporário.

À data do IFEC, estima-se que 38,5 % dos indivíduos ainda não tinham filhos biológicos. Excluindo aqueles que apesar de não terem filhos ainda não têm definida a sua decisão reprodutiva, 8,3 % dos residentes em Portugal esperam permanecer sem filhos no termo da sua vida reprodutiva, enquanto 29,2 % esperam ser *childlessness* apenas temporariamente.

Devido à situação de muito baixa fecundidade na sociedade portuguesa, torna-se pertinente identificar factores determinantes para a decisão definitiva de não ter filhos. Para tal recorreu-se a um modelo de regressão logística, pois a variável que nos interessa explicar é dicotómica (0 - *Childlessness* temporário; 1 - *Childlessness* permanente). Este modelo foi ajustado seguindo a estratégia definida em Hosmer e Lemeshow (2013). Atendendo a que estávamos em presença de uma amostra complexa estratificada e multi-etápica (INE, 2013), representativa ao nível da NUTS II para as mulheres e ao nível nacional para os homens, e a que não tínhamos toda a informação sobre o desenho da amostra, optámos por utilizar os pesos normalizados corrigidos para o efeito do desenho da amostra, i.e., deff (Osborne, 2011).

* Este resumo e as respetivas conclusões não podem ser citados.

Resultados

Para a decisão de permanecer sem filhos sobressai a importância da educação. Uma pessoa com instrução até ao básico tem quase o dobro das possibilidades de permanecer sem filhos do que uma com o ensino superior. Assim, investir na educação pode ser um fator potenciador da fecundidade futura.

Quer um número ideal de filhos inferior a dois, quer a convicção de que a realização pessoal não passa pelo projeto de parentalidade, potenciam a decisão de permanecer sem filhos durante todo o seu curso de vida.

O tempo dedicado ao mercado de trabalho condiciona a decisão apenas no caso dos homens: um homem que não trabalha a tempo inteiro tem mais possibilidades de permanecer sem filhos.

As diferenças entre homens e mulheres alteram-se com a idade: para os mais jovens, com menos de 23 anos, sem trabalho ou com trabalho a tempo parcial, as mulheres são menos suscetíveis de permanecerem sem filhos do que os homens; já a partir dos 35 anos, a relação inverte-se, sendo elas as que revelam maiores possibilidades de não querer ter filhos.

Por último, a experiência de conjugalidade é um elemento central para os planos reprodutivos dos residentes em Portugal. O facto de até aos 36 anos de idade não ter casado nem coabitado é factor potenciador para que homens e mulheres não tenham nem queiram vir a ter filhos.

Conclusões

O perfil mais provável de quem não tem nem deseja ter filhos é alguém do sexo feminino com mais de 30 anos e que trabalha 36 ou mais horas semanais, com ensino básico ou inferior, com um ideal de filhos inferior a 2, que nunca teve cônjuge ou companheiro e que discorda totalmente que a realização pessoal passa pela parentalidade.

Agradecimentos: Este trabalho insere-se no projeto “Determinantes da Fecundidade em Portugal”, financiado pela Fundação Francisco Manuel dos Santos e apresenta um extrato limitado dos resultados preliminares com vista à sua discussão inter-pares.

Referências

HOSMER D., LEMESHOW S., STURDIVANT, R. (2013) *Applied Logistic Regression*, 3rd Ed., New York, USA, Wiley.

INE (2013) *Inquérito à fecundidade*. Documento metodológico Versão 1.0, Lisboa, Instituto Nacional de Estatística.

OSBORNE, J. W. (2011) Best practices in using large, complex samples: the importance of using appropriate weights and design effect compensation. *Practical Assessment, Research & Evaluation*, 16, 1-7.

Sessão Paralela II – Aplicação de Métodos de Regressão – sábado, 11 de abril,
Sala 0.02 (10h05)

Recurso a produtos dietéticos e prática de exercício físico – Quem? Porquê? Como?

Carla Henriques¹, Joaquim Antunes², Sónia Duarte³

¹Escola Superior de Tecnologia e Gestão de Viseu e CMUC, carlahenriq@estv.ipv.pt;

²Escola Superior de Tecnologia e Gestão de Viseu e CI&DETS, jantunes@estv.ipv.pt;

³soniamvduarte@hotmail.com

Sumário

O presente trabalho tem como objetivo a análise do comportamento do consumidor de produtos dietéticos, bem como a caracterização de praticantes de exercício físico, relacionando estes dois recursos com a procura da imagem ideal. Pretendemos saber quais os fatores que influenciam a procura destes recursos e de que forma esta procura está diretamente relacionada com a maior ou menor satisfação com a imagem corporal e com a Ansiedade Físico-Social (AFS).

Palavras-chave: Análise Fatorial, Regressão Linear, Regressão Logística.

Introdução

Este estudo teve por base dados recolhidos a partir de um questionário *online*, a partir do qual se extraiu informação relativa a 540 residentes do distrito de Viseu. O inquérito foi difundido através de contactos de correio eletrónico e através de redes sociais. Pretendia-se investigar o que leva as pessoas à procura de produtos dietéticos e à prática de exercício físico regular, os dois recursos que constituem o principal objeto deste estudo, relacionando-os com a autoavaliação da própria imagem corporal. Dos 540 inquiridos 238 (44%) praticavam exercício físico com regularidade e 269 (49.8%) consumiam, ou já tinham consumido, produtos dietéticos. A motivação para a realização deste estudo reside na importância crescente conferida à imagem corporal pela sociedade, a qual incita à procura da “beleza ideal”. O mercado está atento a esta realidade, disponibilizando produtos e recorrendo a meios de divulgação que tentam ir ao encontro dos desejos e necessidades da sociedade. Com este trabalho espera-se dar um contributo para um melhor entendimento das motivações dos consumidores destes dois recursos. Neste trabalho utilizou-se uma escala que mede a ansiedade social relacionada com o físico - Escala de Ansiedade Físico-Social (EAFS) (Hart *et al.*, 1989). Para determinar as razões que levam os indivíduos a se exercitarem, recorreu-se a um conjunto de 24 itens traduzidos da Escala de Razões para se Exercitar desenvolvida por Silberstein *et al.* (1988).

Resultados

Através de uma análise bivariada e da construção de modelos de regressão logística, identificaram-se fatores relacionados com os recursos que são objeto deste estudo, produtos dietéticos e prática regular de exercício físico. Nos modelos de regressão logística a prática regular de exercício físico (sim/não) e o consumo de produtos dietéticos (sim/não) constituíram as variáveis dependentes. Como possíveis preditores destes dois recursos, analisaram-se a idade o género, a ansiedade físico-social (AFS), a insatisfação corporal, entre outros. A análise fatorial exploratória foi aplicada aos 24 itens de razões para a prática de exercício físico, permitindo encontrar seis fatores que explicam 73,4% da variância total. Os fatores encontrados são: Fitness/tonificação, Saúde, Boa Disposição, Atratividade, Controlo de peso, Diversão. A AFS foi avaliada através da pontuação total dos itens da escala, tendo-se também utilizado os três fatores obtidos por aplicação da análise fatorial, que definem dimensões específicas deste constructo latente (nível de ansiedade relativo à insegurança com o próprio corpo, nível de ansiedade relativo à exposição do próprio corpo e nível de ansiedade relativo ao desagrado com o seu corpo). A amostra permitiu identificar o género e a idade como fatores que estão relacionados com a prática de exercício físico, encontrando-se mais praticantes de exercício físico no género masculino e na faixa etária dos 26 aos 35 anos. Verificou-se também que o exercício físico se relaciona com a AFS – os níveis de AFS são maiores para os não praticantes de exercício físico. Contudo, a insatisfação com o corpo (medida através da diferença entre o peso do inquirido e o peso que este gostaria de ter) não se revelou associada com a prática de exercício físico regular. Na prática de exercício físico o género masculino procura tornar-se atrativo, bem como divertir-se; já o género feminino, que tem níveis superiores de AFS e insatisfação corporal, pratica exercício físico com o objetivo de controlar o peso, tonificar o corpo e melhorar a boa disposição. Quanto ao consumo de produtos dietéticos, verificou-se que é o género feminino que mais recorre a este tipo de produtos. Verificamos também que os consumidores de produtos dietéticos têm maiores níveis de AFS e gostariam de ser mais magros.

Conclusões

A prática regular de exercício físico não se revelou significativamente associada com a insatisfação corporal. Contudo, verificou-se uma tendência para menores níveis de AFS de quem pratica exercício físico. Este recurso revelou-se mais prevalente no género masculino e na faixa etária dos 26 aos 35 anos. Relativamente ao recurso a produtos dietéticos, este é mais utilizado pelo género feminino e está associado a maiores níveis de AFS e de insatisfação corporal.

Referências

- HART, E. A., LEARY, M. R. & REJESKI, W. J. (1989) *The measurement of social physique anxiety*. Journal of Sport and Exercise Psychology, 11, 94-104.
- SILBERSTEIN, L.R., STRIEGEL-MOORE, R. H., TIMKO, C., & RODIN, J., (1988) *Behavioral and Psychological Implications of Body Dissatisfaction: Do Men and Women Differ*. Sex Roles, 19, 219-232.

**Sessão Paralela II – Aplicação de Métodos de Regressão – sábado, 11 de abril,
Sala 0.02 (10h25)**

Estudo de biomarcadores de uma doença renal

Ana Cristina Bico Matos¹, Carla Henriques², Bernardo Faria³, Manuel Pestana³

¹*Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Viseu; Centro de Estudos em Educação, Tecnologias e Saúde (CI&DETS), amatos@estv.ipv.pt;*

²*Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Viseu; Centro de Matemática da Universidade de Coimbra (CMUC), carlahenriq@estv.ipv.pt;*

³*Nephrology and Infectious Disease R&D Group, INEB, I3S, Universidade do Porto, faria_bernardo@yahoo.com*

Sumário

Nefropatia da imunoglobulina A é uma doença renal, que apresenta variadas formas e tende a evoluir para estados graves. A investigação sobre os meios de previsão da patogénese e progressão da doença IgAN é de grande interesse clínico.

Recorrendo a modelos multivariados, de regressão logística e de regressão de Cox, podemos verificar que a combinação de dois biomarcadores histológico revela capacidade preditiva sobre a progressão da IgAN, constituindo uma potencial ferramenta para a abordagem clínica da doença.

Palavras-chave: Doença Renal, Regressão de Cox, Regressão Logística, Regressão Múltipla.

Introdução

Nefropatia da imunoglobulina A (IgAN) é uma doença renal que apresenta variadas formas. Na maior parte das situações apresenta-se numa condição benigna ou reversível podendo evoluir para uma doença renal progressiva (aproximadamente 30% dos casos) levando a situações de terapia de substituição da função renal [1,2].

Estudos recentes têm destacado a associação de algumas moléculas com a patogénese e progressão da IgAN. Este estudo envolve quatro desses biomarcadores histológicos, C4d, CD3, p-ERK 1/2, TGase2, a fim de se obter um painel de biomarcadores, e avaliar o seu valor preditivo para o curso da doença.

O estudo envolveu pacientes do Hospital São João do Porto entre 2003 e 2010. Dos 115 pacientes envolvidos registou-se um conjunto de parâmetros clínicos: idade do paciente, sexo, presença ou ausência de hipertensão, nível de creatinina sérica e de proteinúria, presença ou ausência de história de hematuria macroscópica, níveis no sangue de imunoglobulina A (IgA) e a presença ou ausência de comorbidades. Excluíram-se 41 pacientes, conduzindo ao estudo de 74 biópsias de pacientes com nefropatia por IgA, com um follow-up médio após a biópsia renal de 48 meses, dos quais 20 apresentam doença renal progressiva e 54 doença renal não progressiva. Além dos biomarcadores em estudo também se avaliaram outros marcadores já identificados por estarem relacionados com a doença, por exemplo IgG, IgM, C3, C1q,

fibrinogénio, taxa de filtração glomerular (EGFR), proliferação endocapilar (E), esclerose segmentar (S), atrofia tubular/fibrose intestinal (T1_2).

Recorrendo à análise estatística univariada e a modelos de regressão logística e de regressão de Cox, evidenciou-se que a combinação de dois biomarcadores histológicos (C4d e CD3) constitui um bom preditor de progressão IgAN.

Resultados

Os níveis de proteinúria e EGFR são indicadores da gravidade da IgAN. Recorrendo a modelos de regressão múltipla relacionaram-se estes níveis com os parâmetros clínicos. Os parâmetros clínicos que se revelaram associados com proteinúria mais elevadas foram: menores EGFR, idade avançada e presença de hipertensão. Menor EGFR também se encontra associado com idade avançada, hipertensão e proteinúria superior.

A regressão logística e a regressão de Cox, evidenciou que do conjunto dos parâmetros clínicos e dos parâmetros avaliados nas biopsias, revelaram-se associados com a progressão da doença renal a proteinúria basal, EGFR e T1_2.

Relativamente aos quatro marcadores propostos (C4d, CD3, p-ERK ½, TGase2) observou-se associação positiva de C4d com proteinúria superior e doença renal progressiva. Maior pontuação CD3 associada à doença renal progressiva. Modelos de regressão logística e a regressão de Cox revelam que estes dois marcadores permaneceram independentemente associados com a progressão.

A análise conjunta, incluindo além dos marcadores propostos, os parâmetros clínicos e os marcadores já reconhecidos como associados à IgAN, permitiu concluir que os mais determinantes são dois dos propostos, C4d e CD3, em conjunto com EGFR.

Conclusões

Este estudo confirma o papel preditivo de alguns dos fatores de risco estabelecidos para a progressão da doença, incluindo menor EGFR, proteinúria mais elevados, hipertensão e T1-2. Adicionalmente ilustrou-se que a combinação de C4d, CD3 com EGFR constituem um bom preditor de progressão IgAN, revelando-se crucial na abordagem clínica da doença.

Referências

FLOEGE J. EITNER. (2011) Current therapy for IgA nephropathy. *J Am Soc Nephrol*, 22, 1785–94.

WYATT RJ., JULIAN BA. (2013) IgA nephropathy. *N Engl J Med*, 368, 2402–14.

Sessão Paralela II – Métodos de Classificação – sábado, 11 de abril, Anf. 0.01 (9h45)

TweeProfiles@Sapo: relating Spatio-Temporal patterns on Twitter with Sapo content

André Maia¹, Tiago Cunha², Carlos Soares³, Pedro Henriques Abreu⁴.

¹ *Faculdade de Engenharia da Universidade do Porto, ee08163@fe.up.pt*

² *Faculdade de Engenharia da Universidade do Porto, tiagodscunha@fe.up.pt*

³ *Faculdade de Engenharia da Universidade do Porto, csoares@fe.up.pt;*

⁴ *Departamento de Engenharia Informática da Universidade de Coimbra, pha@dei.uc.pt;*

Abstract

TweeProfiles is a clustering tool that analyses tweets over multiple dimensions: spatial, temporal, content and social. TweeProfiles@SAPO is an extension of TweeProfiles that aims to create an interactive visualization tool based on a real-time clustering that relates the profiles in the clusters with content from SAPO, aiming to support journalism activities. A visualization process suitable for data streaming was designed, using multiple widgets to represent all the information. The visualization tool was validated through a user study with journalists.

Keywords: Clustering, Database, Stream, Visualization

Introduction

The massive adhesion to the various social media platforms lead to a growth in the data stored within these services. The initial version of TweeProfiles was focused on identifying profiles with information extracted from Twitter over 4 dimensions (spatial, temporal, social and content). The process involves the manipulation of dissimilarity matrices for each type of data, which are then fed to a clustering algorithm to obtain the desired patterns. The process allows a weighted combination of multiple dimensions and delivers the extracted knowledge through an interactive visualization tool. An extension was created, TweeProfiles2, replacing the original batch-clustering algorithm with a stream method. The contributions of this project are: the implementation of a real-time Data Mining process, using the clustering algorithm from TweeProfiles2 and the social media data extraction module SocialBus and the development of an interactive web application to support journalist in extracting information from social media and from SAPO.

Results

TweeProfiles@SAPO combines two systems: SocialBus and TweeProfiles2 and integrates them with different SAPO tools, such as Sapo News and Máquina do Tempo. We use SocialBus to gather data from Twitter and TweeProfiles2 to perform the clustering. The process is as follows: as the data is retrieved form SocialBus it is forwarded to the clustering algorithms, Den-Stream and DBSCAN. DenStream is responsible for the creation of micro-clusters in the online stage and DBSCAN is responsible to create the macro-clusters that aggregate the previous results and output the final clustering result. Afterwards, the results are displayed in our web-based platform. The system includes 5 core components: 1) a map that

represents tweets and clustering's geographical distributions, 2) a list of the most recent tweets, personalities found and the most recent news for those entities, 3) display/dimension controls that enable query filtering by region selection, time range and text-based queries, 4) a wordcloud that displays the most frequent words and 5) a graphic displaying temporal distribution.

The system was tested with a large-scale dataset over a period of 30 days and focused on tweets labeled in Portuguese. Figure 1.a) shows a clustering with all dimensions except the spatial dimension being ignored. It creates four clusters, containing the majority of tweets. The resulting clusters can be deemed adequate, since the languages being processed are more frequent in Europe and South America. The temporal dimension (Figure 1.b)) shows that the system creates one cluster including the majority of points. We performed a user study with journalists to evaluate the performance of our tool and their feedback points to the easy and effective understanding of the patterns extracted. With a very reduced sample of only 3 journalists, all of them classified the overall system as a 4 out of 5, highlighting the map and the news modules as the most interesting ones.

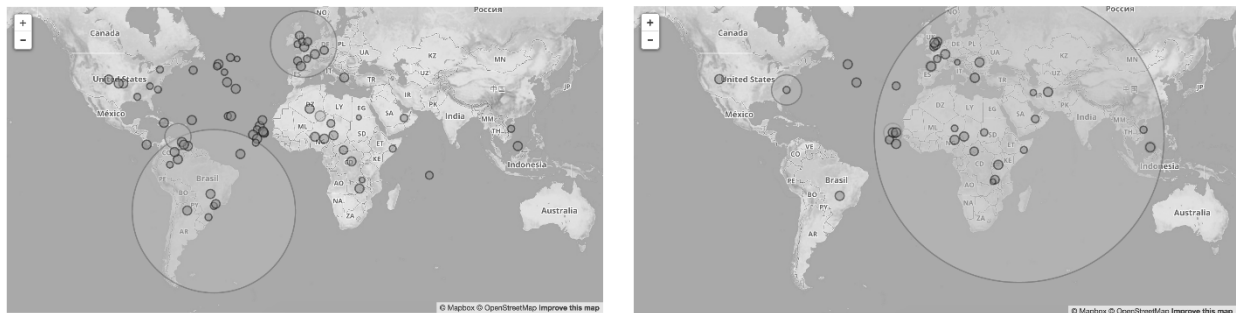


Figure 1.a): Results for Spatial Clustering; 1.b) Results for Temporal Clustering

Conclusions

We developed a framework devoted to the integration of a real-time clustering algorithm and SocialBus, to create a visualization system capable of displaying the profiles found in all different dimensions, as well as different news and entities obtained from SAPO sources. The visualization tool used different visual components for all dimensions. Future work includes improving the stream clustering process and perform an appropriate evaluation strategy.

References

CUNHA, T., SOARES, C. & MENDES RODRIGUES, E. (2014). TweepProfiles: Detection of Spatio-temporal Patterns on Twitter. *Proceedings of the 10th International Conference on Advanced Data Mining and Applications*, 123–136.

Sessão Paralela II – Métodos de Classificação – sábado, 11 de abril, Anf. 0.01 (10h05)

Distribuições de índices de concordância entre agrupamentos

Maria José Amorim¹, Margarida Cardoso²

¹Área Departamental de Matemática, ISEL-Instituto Superior de Engenharia de Lisboa, mjamorim@adm.isel.pt;

²Departamento de Métodos Quantitativos para Gestão e Economia e UNIDE, ISCTE-Instituto Universitário de Lisboa, margarida.cardoso@iscte.pt

Sumário

No presente trabalho estuda-se a distribuição empírica de diversos índices de concordância pareada entre duas partições, sob a hipótese de independência restrita (H_0). A distribuição de cada índice é obtida a partir de um processo de simulação, sendo os seus valores resultantes da geração de tabelas de contingência, sob H_0 . Estas tabelas correspondem à avaliação externa de agrupamentos que se realizam em diferentes cenários, controlando os números de grupos, dimensões relativas e graus de sobreposição dos grupos. Os cenários de agrupamento correspondem a bases de dados simulados (misturas de Normais).

Palavras-chave: Avaliação externa de agrupamentos, Índices corrigidos, Índices de concordância pareada.

Introdução

Os índices de concordância pareada (ICP) são comumente utilizados na avaliação externa de agrupamentos, comparando os resultados de uma Análise de Agrupamento com classes que “a priori” se associam aos dados base. Estes índices podem ser calculados a partir dos valores da tabela de contingência entre as duas partições a comparar. A fim de melhorar a avaliação da concordância entre as partições, deve retirar-se ao valor de um ICP a concordância que se deve ao acaso, obtendo o seu valor ajustado - ICPa. Geralmente, um valor limiar de concordância por acaso corresponde à média do ICP sob hipótese de independência entre as duas partições que se comparam, restrita à igualdade das marginais da respetiva tabela de contingência – hipótese de independência restrita H_0 . O estudo que se propõe tem como objetivo a análise detalhada da distribuição empírica, sob H_0 , de diversos ICP o que permitirá melhorar a seleção dos seus limiares.

Análise experimental

De entre os diversos ICP que se encontram definidos na literatura são utilizados os índices de: Jaccard (Jaccard, 1908); Czekanowski (Czekanowski, 1932); Goodman & Kruskal (Goodman & Kruskal, 1954); SoKal & Sneath ((Sokal & Sneath, 1963), dois índices); Rand (Rand, 1971); Fowlkes & Mallows (Fowlkes & Mallows, 1983) e Gower & Legendre (Gower & Legendre, 1986).

As bases de dados geradas são misturas de Normais, com 2, 3 e 4 grupos e diferenciadas pela dimensão relativa dos grupos e pelos seus níveis de sobreposição (Maitra e Melnykov, 2010). Para cada base de dados gerada (30 em cada cenário) é efetuada uma análise de agrupamento, utilizando-se o algoritmo Expectation-Maximization implementado em (Lebret *et al.*, 2012). Tendo como fixas as marginais da tabela de contingência entre a solução obtida e a

estrutura real da base de dados, são finalmente geradas 17000 tabelas de contingência (17000×30 em cada cenário), viabilizando a obtenção das distribuições empíricas dos ICP sob H_0 (Amorim & Cardoso, 2014).

Conclusões

Independentemente do grau de sobreposição dos grupos, as distribuições empíricas obtidas nos cenários com grupos equilibrados diferem das obtidas nos cenários com os grupos não equilibrados, apresentando as primeiras uma assimetria positiva acentuada e as segundas uma assimetria positiva próxima de zero; por outro lado, os valores dos índices obtidos nos cenários com os grupos não equilibrados apresentam uma maior dispersão (maior amplitude de valores e coeficientes de variação superiores).

Assim, os resultados obtidos sugerem rever a prática de ajustamento dos índices pelo seu valor esperado (sob H_0), propondo-se como alternativa o ajustamento pela mediana.

Referências

- AMORIM, M. J. & CARDOSO, M. G. M. S. (2014) Paired Indices for Clustering Evaluation. Correction for Agreement by Chance. *Proceedings of the 16th International Conference on Enterprise Information Systems*, 164-170.
- CZEKANOWSKI, J. (1932) "Coefficient of racial likeness" and "durchschnittliche Differenz". *Anthropologischer Anzeiger*, 14, 227-249.
- FOWLKES, E. B. & MALLOWS, C. L. (1983) A Method for Comparing Two Hierarchical Clusterings. *Journal of the American Statistical Association*, 78, 553-569.
- GOODMAN, L. A. & KRUSKAL, W. H. (1954) Measures of Association for Cross Classifications. *Journal of the American Statistical Associations*, 49.
- GOWER, J. C. & LEGENDRE, P. (1986) Metric and Euclidean Properties of Dissimilarity Coefficients. *Journal of Classification*, 3, 5-48.
- JACCARD (1908) Nouvelles recherches sur la distribution florale. *Bulletin de la Société Vaudoise de Sciences Naturelles*, 44, 223-370.
- LEBRET, R., LOVLE, S., LANGROGNET, F., BIERNACKI, C., CELEUX, G. & GOVAERT, G. (2012) Rmixmod: The r package of the model-based unsupervised, supervised and semi-supervised classification mixmod library.
- MAITRA, R. & MELNYKOV, V. (2010) Simulating Data to Study Performance of Finite Mixture Modeling and Clustering Algorithms. *Journal of Computational and Graphical Statistics*, 2, 354-376.
- RAND, W. M. (1971) Objective criteria for the evaluation of clustering methods. *Journal of the American statistical association*, 66, 846-850.
- SOKAL, R. R. & SNEATH, P. H. (1963) *Principles of Numerical Taxonomy*, San Francisco CA:Freeman.

Sessão Paralela II – Métodos de Classificação – sábado, 11 de abril, Anf. 0.01 (10h25)

Distribuições de Coeficientes de Comparação de Dendrogramas

Fernanda Sousa¹.

¹Faculdade de Engenharia e CITTA, Universidade do Porto, fcsousa@fe.up.pt

Sumário

A necessidade de comparar dendrogramas definidos sobre o mesmo conjunto de elementos surge frequentemente em classificação. Neste trabalho são apresentados coeficientes ordinais adequados para essa comparação e discutidas as suas distribuições.

Palavras-chave: Classificação hierárquica, Coeficientes ordinais de comparação, Dendrograma, Distribuições assintóticas, Distribuições empíricas.

Introdução

É bem conhecido que as escolhas inerentes a um método de Classificação Hierárquica Ascendente (C.H.A.) influenciam o resultado final, cuja expressão mais comum é um dendrograma. Por outro lado não existem escolhas genericamente aceites como as melhores (Sousa, 2000). Tendo por base um quadro de dados e por objetivo classificar os seus indivíduos ou variáveis, usando uma metodologia de C.H.A., é procedimento comum usar várias medidas de comparação e/ou critérios, obtendo assim vários dendrogramas, que carecem de avaliação e comparação. A necessidade de comparar dendrograma pode também advir de supressões, adições de indivíduos ou variáveis ou mesmo perturbações de elementos do quadro de dados.

Este trabalho foca-se na comparação de pares de dendrogramas com o mesmo conjunto de nós terminais. Opta-se por uma abordagem de comparação ordinal, associando a cada um dos dendrogramas uma preordenação total. São considerados vários coeficientes de correlação ordinal para a comparação de pares de preordenações: Spearman, Tau de Kendall e Goodman-Kruskal, discutidas as suas propriedades e apresentadas as distribuições assintóticas sob condição de independência. Contudo este pressuposto de independência é, no contexto deste trabalho, duplamente violado: i) os dendrogramas a comparar são obtidos de um mesmo quadro de dados, ii) os elementos das matrizes ultramétricas associadas, verificam a propriedade ultramétrica, criando ligações de dependência. Na sequência de trabalhos anteriormente desenvolvidos, por Sousa e Tendeiro, avalia-se o impacto da ausência do pressuposto de independência, comparando as distribuições empíricas, deduzidas para os vários coeficientes de comparação por recurso a métodos de geração aleatória de dendrogramas (Sousa, 2000 e Tendeiro, 2005), com as correspondentes distribuições assintóticas conhecidas da literatura.

Resultados

A título ilustrativo apresentam-se apenas alguns dos resultados obtidos para o Coeficiente de Spearman, R_S . Sejam m o número de nós terminais dos dendrogramas a comparar, $P_{k,m}$ o percentil de ordem k da distribuição empírica de R_S , $F(P_{k,m})$ o valor da função de distribuição assintótica de R_S em $P_{k,m}$ e $Dif = \frac{F(P_{k,m}) - k/100}{k/100}$, a diferença relativa entre os dois valores de

distribuição. Nas Tabelas 1 e 2 consideram-se os percentis de ordem 90 e 95, respetivamente, e valores de m entre 4 e 100.

Tabela 1: Comparação das distribuições empírica e assintótica de $R_S - P_{90}$

m	4	5	6	7	8	9	10	11	12	13	14
$F(P_{90,m})$.971	.978	.911	.936	.940	.930	.936	.955	.947	.932	.935
Dif	.079	.086	.012	.040	.045	.034	.041	.061	.052	.036	.039

m	15	20	25	30	35	40	45	50	75	100
$F(P_{90,m})$.937	.973	.971	.976	.983	.978	.988	.989	.996	.997
Dif	.041	.081	.079	.085	.092	.086	.098	.098	.106	.108

Tabela 2: Comparação das distribuições empírica e assintótica de $R_S - P_{95}$

m	4	5	6	7	8	9	10	11	12	13	14
$F(P_{95,m})$.987	.990	.972	.982	.985	.988	.994	.992	.992	.993	.991
Dif	.039	.042	.023	.034	.037	.040	.046	.045	.044	.045	.043

m	15	20	25	30	35	40	45	50	75	100
$F(P_{95,m})$.984	.996	.997	.998	.999	.997	1.000	1.000	1.000	1.000
Dif	.036	.048	.050	.051	.052	.050	.052	.052	.053	.053

Conclusões

Os resultados obtidos para os vários coeficientes analisados são similares e permitem concluir que as distribuições assintóticas conhecidas para os coeficientes em análise não se adequam ao contexto da comparação de dendrogramas, sendo aconselhável o recurso às distribuições empíricas aqui tratadas.

Referências

- SOUSA, F. (2000) *Novas Metodologias e Validação em Classificação Hierárquica Ascendente*, Dissertação de Doutoramento, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa.
- SOUSA, F. & TENDEIRO, J. (2013) Sobre a Comparação de Estruturas de Classificação: Coeficientes e suas Distribuições. IN BACELAR-NICOLAU, H., SOUSA, F. & FERREIRA, C. (Eds.) *Classificação e Análise de Dados: Métodos e Aplicações*, ISSN 0039-0526, pp. 1-14.
- TENDEIRO, J. (2005) *Comparação de Dendrogramas: Obtenção de Distribuições Empíricas de Alguns Coeficientes*, Dissertação de Mestrado, Faculdade de Engenharia da Universidade do Porto.
- TENDEIRO, J. & SOUSA, F. (2006) Metodologia para a obtenção de distribuições empíricas de coeficientes de comparação de dendrogramas, in *Ciência Estatística*, Actas do XIII Congresso Anual da Sociedade Portuguesa de Estatística. Ed. SPE, Lisboa, Portugal, pp. 691-702.

Sessão Paralela II – Análise de Dados em Economia, Gestão e Ensino – sábado, 11 de abril, Sala 0.07 (9h45)

Are callbacks a waste of time in mCATI surveys?

Paula Vicente¹, Catarina Marques², Elizabeth Reis³

¹*Instituto Universitário de Lisboa (ISCTE-IUL), BRU-IUL, paula.vicente@iscte.pt*

²*Instituto Universitário de Lisboa (ISCTE-IUL), BRU-IUL, catarina.marques@iscte.pt*

³*Instituto Universitário de Lisboa (ISCTE-IUL), BRU-IUL, elizabeth.reis@iscte.pt*

Abstract

Survey organizations spend a lot of effort trying to catch respondents for their surveys which usually involves making several call attempts for a given number before giving it up. This procedure is dependent on workforce availability and deadline for survey period completion, and its efficacy depends on careful planning of call scheduling in order to increase successful contacts with less effort. The objective of our analysis is to explore the effect of callbacks on the likelihood of interview and sample composition in a mCATI survey.

Findings reveal that the call history has a marked effect on the efficiency of callbacks on mobile phone numbers. Specifically, callbacks on mobile phone numbers with many call outcomes of busy or voice mail have a very low probability of ending up in a successful interview. The outcomes also reveal that callbacks allow mostly to reaching young people, working people and people living in urbanised areas.

Keywords: Callbacks, mCATI surveys, Non-response bias

Introduction

A key challenge facing mCATI-Mobile Computer Assisted Telephone Interviewing surveys is that of finding ways to minimize the effort required to make contact and to gain cooperation in order to, under constraints of time, budget and resources, achieve high response rates while also minimizing bias.

The odds of making contact and gaining cooperation may be raised by making callbacks on sample units that could not be contacted at first call attempt. Additionally, callbacks may account to control for nonresponse bias. In fixed phone surveys on the general public, respondents contacted on the first call attempt tend to be older, less educated, are more likely to be female, have lower incomes and are more likely to live in rural areas (Gendall & Davis 1993). Callbacks may contribute to reduce this kind of bias by allowing reaching people with a different profile from first call respondents.

In the context of mCATI surveys, the efficacy of a callback strategy both to increase the response rate and/or to reduce nonresponse bias is, to date, a topic insufficiently explored. This paper contributes to this area by analyzing the likelihood of interview at different call attempts and comparing sample composition across call attempts. The data used for this study come from a mCATI survey conducted in Portugal in 2012.

Results

The analysis of calling outcomes reveals that first call attempts are the most efficient; thereafter, both the contact and interview rates decrease steadily. After the 4th attempt, the percentage of interviews gained for the sample by additional call attempts is very low. Mobile phone numbers with many busy or voicemail calls are very difficult to contact and interview; temporarily unavailable and ring with no answer are the most likely noncontact cases to be converted to contacts and yield interviews.

Comparison of the characteristics of respondents interviewed on the first call with those interviewed on subsequent calls over a range of demographic variables reveals that first call respondents are more likely to be older, have a basic educational level, are more likely to be retired or professionally inactive, and more likely to belong to the lowest social class.

Conclusions

Despite the apparent advantage of mobile phones over fixed phones (i.e., people can be easily called with mobile phones due to the mobility of the mobile phone), mCATI surveys do not differ from fixed phone surveys in what concerns difficulty in obtaining response and sample composition (Brick et al. 1996; Hansen 2008). Lifestyles and time occupation are likely to explain the difficulty in reaching the young, the employed and the highly educated, which tend to have a busy life style, with professional and leisure activities that leave them less free time and reduce their availability to take mobile phone calls immediately. The outcomes reveal that these groups should be reached with callbacks, otherwise would be underrepresented in the survey. Although callbacks allow response rate to increase, the biggest share of interviews was obtained at first call attempts (nearly 65%) which is a sign that an effort should be made to improve efficacy of initial calls.

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

References

- BRICK, J., ALLEN, B., CUNNINGHAM, P. & MAKLAN, D. (1996) Outcomes of a calling protocol in a telephone survey. *Proceedings of the American Association for Public Opinion Research Conference*. AAPOR, 142-149. Available at: https://www.amstat.org/sections/srms/proceedings/papers/1996_019.pdf (accessed June 2014).
- GENDALL, P. & DAVIS, P. (1993) Are callbacks a waste of time? *Marketing Bulletin*, 4, 53-57.
- HANSEN, S. (2008) CATI sample management systems. IN LEPKOWSKI, J., TUCKER, C., BRICK, J, DE LEEUW, E., JAPEC, L., LAVRAKAS, P., LINK, M. & SANGSTER, R. (Eds) *Advances in telephone survey methodology*. New Jersey, Wiley.

Sessão Paralela II – Análise de Dados em Economia, Gestão e Ensino – sábado, 11 de abril, Sala 0.07 (10h05)

Uncertainty analysis in conjoint studies

José Gonçalves Dias

Instituto Universitário de Lisboa (ISCTE-IUL), BRU-IUL, jose.dias@iscte.pt

Abstract

One of the main applications of conjoint analysis is market choice simulation based on ‘what-if’ scenarios. This paper extends the methodology of conjoint analysis using the bootstrap technique, which allows a quantification of the uncertainty underlying the typical averaged market shares. An application shows the promising results.

Keywords: Agribusiness, Bbootstrap analysis, Conjoint analysis, Decision support systems, ‘What-if’ scenarios.

Introduction

In many economic analyses, decision making has been framed under the utility theory (FISHBURN, 1968). Consumer’s preferences are viewed as individual tastes about the utility of distinct characteristics that are bundled in a product or service (LANCASTER, 1966, 1971). Thus, products or concepts (e.g., positioning, benefits, and images) are packages of attributes and its utility is derived from its attributes rather than from the product itself. And, then, all the information about attributes is integrated to form a product profile.

One technique that has been used to understand the value of product attributes is conjoint analysis (CA). Developed initially in mathematical psychology (LUCE & TUKEY, 1964; GREEN & RAO, 1971), it operationalizes the Lancaster’s definition of product, i.e., by defining an experimental setting; consumers are asked to assess different hypothetical products with different levels of attributes. Its popularity has increased enormously after an article published in the Harvard Business Review (GREEN & WIND, 1975). Since then, CA has become the favorite technique for understanding how buyers make trade-offs between competitive products (GREEN *et al.*, 2001; RAO, 2014).

While conjoint analysis gives many insights from individual or segment-based results, market simulation is one of its most important applications. Market simulators are decision support systems fed by CA results, which develop what-if competitive scenarios and estimate how consumers will react to them. Given that the structure of preferences of each respondent is known (estimated results), one can simulate individual choices from those results by computing the utility of any product under assessment (all combinations of levels of attributes) for each respondent. Then, the aggregation (over all respondents) of those results for a set of products available on the market provides choice shares for each product being simulated.

We extend conjoint analysis methodology by computing the uncertainty in choice shares using the bootstrap method (EFRON, 1979; EFRON & TIBSHIRANI, 1993). This statistical technique resamples from the original sample with replacement and estimates the distribution of

those choice shares from the CA results. Four choice share rules are discussed: maximum utility rule, Bradley-Terry-Luce rule, Lesourne's rule, and logit rule (TSAFARAKIS et al., 2011).

Application

The application uses conjoint analysis to model the purchasing process of tractors and extracts farmers' structures of preference. Farmers rank combinations of levels of attributes of tractors from the most preferred to the least preferred one. Using this extension of CA, the individual structure of preferences is estimated, the market is simulated, and the uncertainty of choice shares retrieved. Bootstrap distributions and confidence intervals can be obtained. Graphical tools can help depict confidence regions.

Conclusion

This paper introduces uncertainty analysis in conjoint-based market simulations and shows promising results. Our application on the decision of buying a tractor shows the added-value of this extension.

References

- EFRON, B. & TIBSHIRANI, R. (1993) *An Introduction to the Bootstrap*. Boca Raton, FL: Chapman & Hall/CRC.
- EFRON, B. (1979) Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1): 1–26.
- FISHBURN, P.C. (1968) Utility theory. *Management Science*, 14(5), 335–378.
- GREEN, P. E, KRIEGER A. M., & WIND, Y. (2001), Thirty years of conjoint analysis: reflections and Prospects. *Interfaces*, 31(3), S56-S73.
- GREEN, P. E., & RAO, V. R. (1971) Conjoint measurement for quantifying judgmental data. *Journal of Marketing Research*, 8, 355-363.
- GREEN, P. E., & WIND, J. (1975) New way to measure consumer's judgments. *Harvard Business Review*, July-August, 107-117.
- LANCASTER, K. J. (1966) A new approach to consumer theory. *Journal of Political Economy*, 74 (2), 132–157.
- LANCASTER, K.J. (1971) *Consumer Demand: A New Approach*. Columbia University Press, New York.
- LUCE, R. D. & TUKEY, J. W. (1964) Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology*, 1-27.
- RAO, V. R. (2014) *Applied Conjoint Analysis*. Berlin: Springer-Verlag.
- SAFARAKIS, S., GRIGOROUDIS, E., & MATSATSINIS, N. F. (2011) Consumer choice behaviour and new product development: an integrated market simulation approach, *Journal of The Operational Research Society*, 62(7), 1253-1267.

Sessão Paralela II – Análise de Dados em Economia, Gestão e Ensino – sábado, 11 de abril, Sala 0.07 (10h25)

Data Warehouse de suporte à melhoria contínua do ensino na Escola Naval

Martins João¹, Merca Luís²

¹Centro de Investigação Naval, Escola Naval, Marinha Portuguesa, maia.martins@marinha.pt;

²Centro de Investigação Naval, Escola Naval, Marinha Portuguesa, gaspar.merca@marinha.pt;

Sumário

A Escola Naval é um estabelecimento de Ensino Superior Público Universitário Militar e como tal necessitou de desenvolver uma metodologia de autoavaliação, das vertentes ensino-aprendizagem, investigação e estabelecimento de ensino, que sustentasse o processo de melhoria contínua.

No sentido de apresentar essa metodologia de auto-avaliação foi criado um *Data Warehouse* de suporte à melhoria contínua, o qual é apresentado neste resumo.

Palavras-chave: Autoavaliação, *Data Warehouse*, Ensino, Melhoria contínua.

Introdução

Para a criação do *Date Warehouse* de suporte à melhoria contínua do ensino na Escola Naval, desenvolveu-se internamente um sistema de apoio à decisão para auxiliar o Comando, o Director de Ensino, os docentes, o Centro de Investigação Naval (CINAV) e os restantes serviços de apoio da Escola Naval a analisarem e melhorarem o seu desempenho. De modo a obter os indicadores desenhados, criou-se de raiz um *Data Warehouse* (DW) que, alimentado por um Sistema de Gestão Académica, questionários on-line a todos os agentes e bases de dados históricos em Excel e Access permite controlar medidas de melhoria, a evolução da satisfação e das qualidades do corpo discente ao longo da formação na EN, os indicadores de acreditação e a justificação dos planos de estudos e unidades curriculares.

Resultados

Para a validação dos dados obtidos através de questionários, utilizam-se atualmente duas variáveis explicativas, a dimensão da turma (as turmas variam entre 1 e 60 alunos) e a avaliação obtida pelo aluno na unidade curricular. Da análise fatorial obtida, consegue-se obter um posicionamento relativo dos docentes, para cada uma das dimensões em análise (eficiência do ensino, nível de transmissão de competências transversais, recursos colocados à disposição dos alunos e justificação do conteúdo programático). A existência de docentes convidados de outros estabelecimentos de ensino pode permitir que o posicionamento dos docentes possa ser generalista. O gráfico 1 representa uma das diversas análises possíveis, onde se observa a evolução da satisfação com o ensino de uma turma ao longo do seu percurso académico.

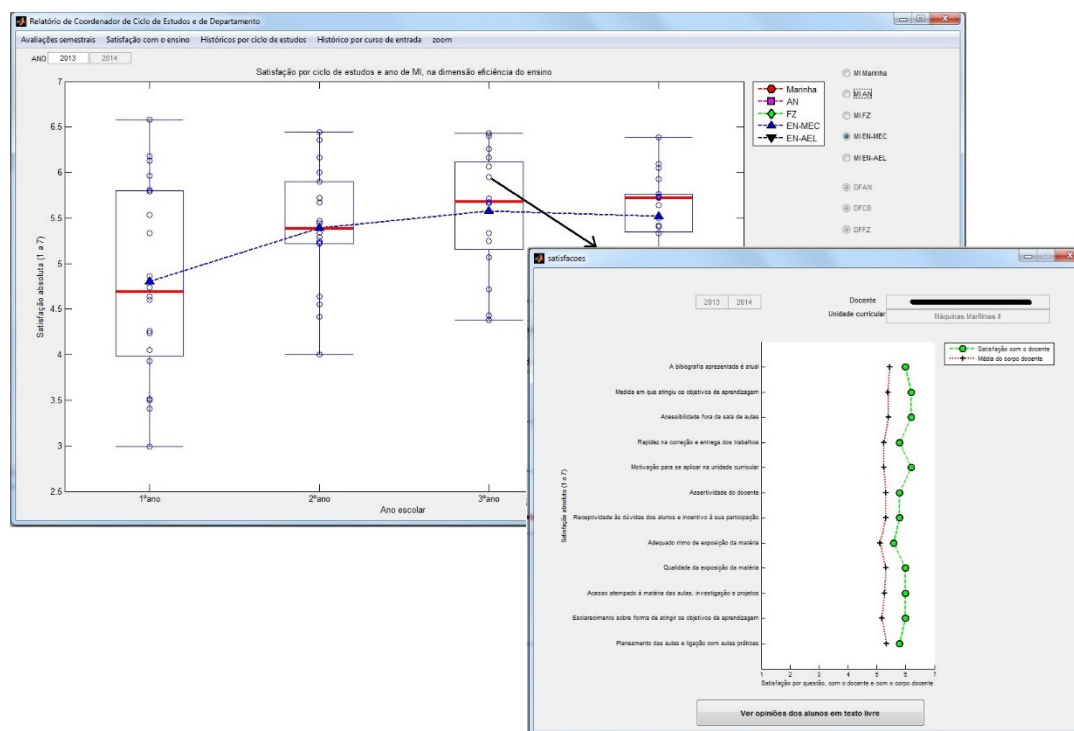


Gráfico 1: Satisfação por ciclo de estudos na dimensão eficiência do ensino

Conclusões

Dispondo de acesso ao mundo empregador, expandiu-se o sistema de melhoria contínua para a envolvente externa da Escola Naval. Para o efeito, identificaram-se competências e objetivos específicos por ciclo de estudos, tendo cada unidade curricular ficado associada, com pesos, a um conjunto de objetivos e competências. A partir da satisfação ou insatisfação do cliente com a proficiência dos oficiais em cada um dos objetivos e competências, é possível obter a cadeia de unidades curriculares a carecer ou não de medidas de melhoria. A eficácia dessas medidas será medida igualmente a nível externo.

Referências

SANTOS, S. (2011) *Análise Comparativa dos Processos Europeus para a Avaliação e Certificação de Sistemas Internos de Garantia da Qualidade*, Lisboa, PORTUGAL, A3ES Readings.

SARRICO, C. (2010). Indicadores de Desempenho para Apoiar os Processos de Avaliação e Acreditação de Cursos, www.a3es.pt, (acedido em 12 de Fevereiro 2015).

AGÊNCIA DE AVALIAÇÃO E ACREDITAÇÃO DO ENSINO SUPERIOR. (2013). Auditoria dos Sistemas Internos de Garantia da Qualidade nas Instituições de Ensino Superior – Manual para o Processo de Auditoria, www.a3es.pt, (acedido em 12 de Fevereiro 2015).

Sessão Paralela III – Modelos Longitudinais – sábado, 11 de abril, Anf. 0.01 (15h00)

Diagnostic tools for bivariate time series of counts

Isabel Silva¹, Maria Eduarda Silva², Cristina Torres³

¹Faculdade de Engenharia da Universidade do Porto, ims@fe.up.pt;

²Faculdade de Economia da Universidade do Porto and CIDMA, mesilva@fep.up.pt;

³ISCAP-IPP and Universidade do Porto, cmptorres@gmail.com

Abstract

An important step in any statistical investigation is the assessment of the adequacy of the model proposed and fitted to the data under analysis. In this work, methods for model diagnostic and validation based on residual analysis, predictive distributions and parametric resampling methods are presented and illustrated with a real dataset modelled by a Bivariate INteger-valued Moving Average (BINMA(1, 1)) model.

Keywords: BINMA models, Model diagnostic, Parametric bootstrap, Predictive distributions, Residual analysis.

Introduction

The evaluation of the quality of a model fit is an important part of any statistical data analysis. For linear models there exist a wide variety of methods for this purpose. However, for discrete-valued time series, and particularly for count time series, this is not the case. As noted by Jung *et al.* (2015), some of the existing model assessment methods for continuous models can be adapted to the discrete-valued framework. In this work, the following tools will be considered: residual analysis through the serial correlation of (standardized) Pearson residuals; parametric bootstrap or resampling methods (Tsay, 1992) and predictive distributions by using a nonrandomized probability integral transform, PIT, histogram (Czado *et al.* 2009).

Results

These procedures will be illustrated on a bivariate dataset consisting of the number of rainy days per week at Bremen and Cuxhaven, which will be denoted by $(X_{1,t})$ and $(X_{2,t})$, $t=1, \dots, n$, respectively ($n=574$ observations, from the year 2000 to 2010), see Scotto *et al.* (2014) for details. A preliminary analysis of the sample mean, variance and autocorrelation function (ACF) indicates that this dataset can be modelled by a BINMA(1, 1) model with Bivariate Poisson distribution for the innovation process, as defined in Silva *et al.* (2014):

$$X_{1,t} = \varepsilon_{1,t} + \beta_{1,1} \circ \varepsilon_{1,t-1}; \quad X_{2,t} = \varepsilon_{2,t} + \beta_{2,1} \circ \varepsilon_{2,t-1},$$

where $(\varepsilon_t) = (\varepsilon_{1,t}, \varepsilon_{2,t})$ follow a Bivariate Poisson distribution, $BP(\lambda_1, \lambda_2, \phi)$, and denotes the binomial thinning operation $(\alpha \circ Y = \sum_{j=0}^Y B_j)$, where Y is a non-negative random variable, $\alpha \in [0,1]$ and $B_j \square \text{Ber}(\alpha)$, independent of Y). Generalized Method of Moment estimates and

their corresponding standard errors (in brackets) are $\hat{\beta}_{1,1} = 0.010(0.054)$, $\hat{\lambda}_1 = 0.970(0.326)$, $\hat{\beta}_{2,1} = 0.161(0.100)$, $\hat{\lambda}_2 = 0.596(0.393)$ and $\hat{\phi} = 2.799(0.155)$.

Although the Pearson residuals exhibit the expected sample mean and variance (zero and one, respectively), there is evidence of serial correlation in some lags. Moreover, the parametric resampling exercise for the ACF also indicates that the fitted BP BINMA(1,1) model does not represent completely the dependence structure in the data. Additionally, the PIT histograms present an U-form, indicating underdispersed distribution. Thus, a BINMA model with a different order for each equation besides a different bivariate distribution for the innovation process (for instance bivariate negative binomial distribution) must be considered in order to improve the model fitting.

Conclusions

The diagnostic and validation of a fitted model must be performed on any time series analysis. In this work, different tools suggest that the BP BINMA(1, 1) do not completely capture the dependence structure of the data. However, as stated by Box: “*All models are wrong, but some are useful.*”

Acknowledgments: For the second author, this work was supported by Portuguese funds through the CIDMA and the Portuguese Foundation for Science and Technology (“FCT–Fundação para a Ciência e a Tecnologia”), within project PEst-OE/MAT/UI4106/2014.

References

- CZADO, C., GNEITING, T. & HELD, L. (2009) Predictive model assessment for count data. *Biometrics*, 65, 1254-1261.
- JUNG, R.C., McCABE, B.P.M & TREMAYNE, A.R. (2015) Model validation and diagnostics. IN *Handbook of Discrete-Valued Time Series*. Chapman & Hall/CRC Press (forthcoming).
- SCOTTO, M., WEIß, C., SILVA, M.E. & PEREIRA, I. (2014) Bivariate binomial autoregressive models. *Journal of Multivariate Analysis*, 125, 233 - 251.
- SILVA, I., TORRES, C. & SILVA, M.E. (2014) Estimating bivariate integer-valued moving average models with the generalized method of moments. IN *Livro de resumos das XXI Jornadas de Classificação e Análise de Dados (JOCLAD 2014)*, 111-114.
- TSAY, R.S. (1992) Model checking via parametric bootstraps in time series analysis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 41, 1-15.

Sessão Paralela III – Modelos Longitudinais – sábado, 11 de abril, Anf. 0.01 (15h20)

O efeito da existência de omissões nas estimativas de um modelo com trajetória latente linear: Um estudo de simulação

Paula C. R. Vicente¹, Maria de Fátima Salgueiro²

¹Universidade Lusófona de Humanidades e Tecnologias – Escola de Ciências Económicas e das Organizações, p951@ulusofona.pt;

²Instituto Universitário de Lisboa (ISCTE-IUL), Business Research Unit, Portugal, fatima.salgueiro@iscte.pt;

Sumário

Este estudo de simulação tem por objetivo analisar o impacto do desvio face à normalidade da distribuição dos dados observados nas estimativas de um modelo com trajetória latente, assim como, analisar o efeito da existência de diferentes padrões de omissão com diferentes percentagens de não resposta. O efeito da dimensão da amostra e do número de momentos temporais constituem ainda questões de interesse.

Palavras-chave: Modelos com trajetória latente, Painel rotativo, *Planned missing design*.

Introdução

Dos modelos com variáveis latentes contínuas mais utilizados na modelação de dados longitudinais destacam-se os modelos com trajetória latente. Estes modelos permitem descrever trajetórias individuais, descrever uma trajetória média para o conjunto dos elementos em estudo, bem como a variabilidade em torno dessa trajetória média, possibilitam ainda, a incorporação de variáveis explicativas dessas trajetórias (BOLLEN & CURRAN, 2006).

O modelo pode ser descrito pela seguinte equação de trajetória, $y_{it} = \alpha_i + \lambda_t \beta_i + \varepsilon_{it}$, em que, y_{it} é o valor da variável observada para o elemento i , no momento t . O intercepto é dado por $\alpha_i = \mu_\alpha + \zeta_{\alpha_i}$, e o declive por $\beta_i = \mu_\beta + \zeta_{\beta_i}$. ε_{it} representa o termo residual da trajetória traçada para o elemento i no momento t , e λ_t é uma constante que assume a forma $(t - 1)$ quando é considerada uma trajetória linear. Os parâmetros que definem o modelo são as médias do intercepto (μ_α) e do declive (μ_β), as variâncias do intercepto ($\psi_{\alpha\alpha}$) e do declive ($\psi_{\beta\beta}$) e a covariância entre intercepto e declive ($\psi_{\alpha\beta}$). Como consequência dos pressupostos do modelo, $\zeta_{\alpha_i} \sim N(0, \psi_{\alpha\alpha})$, $\zeta_{\beta_i} \sim N(0, \psi_{\beta\beta})$, e $\varepsilon_{it} \sim N(0, \theta_{\varepsilon_t})$, em que θ_{ε_t} representa a variância do termo residual, os dados gerados a partir da estrutura imposta seguem distribuição normal.

O Mplus é um pacote estatístico que permite, de forma integrada, gerar dados e estimar os modelos com a estrutura definida pelo utilizador. Neste estudo de simulação foram gerados dados com distribuição normal e dados resultantes de uma mistura de distribuições normais. O objetivo é averiguar o efeito dos desvios face à normalidade da distribuição dos dados, da dimensão da amostra, do número de momentos temporais e da existência de diferentes percentagens de omissões em diferentes momentos temporais, resultantes de um *planned missing design*, nas estimativas de um modelo com trajetória latente linear.

Resultados

O procedimento do Mplus para gerar dados encontra-se no âmbito dos modelos de mistura (MUTHÉN & ASPAROUHOV, 2002a e 2002b). Para gerar dados com distribuição normal foi considerado um modelo de mistura com apenas uma classe, na qual foi definido um modelo com trajetória latente linear, com os seguintes valores para os parâmetros: $\mu_{\alpha}=\mu_{\beta}=0$, $\Psi_{\alpha\alpha}=1$, $\Psi_{\beta\beta}=0.2$ e $\Psi_{\alpha\beta}=0$. Para gerar dados com distribuição resultante de uma mistura de distribuições normais foi utilizado um modelo de mistura com duas classes, com diferentes percentagens de observações (uma classe maioritária e uma classe minoritária), sendo que em cada uma destas classes foi definido um modelo com trajetória latente diferente. A assimetria e o achatamento dos dados gerados resulta da diferente combinação dos valores das médias e variâncias do intercepto e do declive do modelo. Na classe maioritária foram considerados os seguintes valores para os parâmetros: $\mu_{\alpha}=\mu_{\beta}=0$, $\Psi_{\alpha\alpha}=1$, $\Psi_{\beta\beta}=0.2$ e $\Psi_{\alpha\beta}=0$. Na classe minoritária $\mu_{\alpha}=2.5$ e 15 , $\mu_{\beta}=0$ e 4 , $\Psi_{\alpha\alpha}=1$, 5 e 35 , $\Psi_{\beta\beta}=0.2$ e $\Psi_{\alpha\beta}=0$. As amostras de dados foram geradas com dados completos e com omissões que configuram um *planned missing design*, com dimensão 250 e 500 observações, e 3 a 5 momentos temporais.

Os resultados obtidos mostram que, se os dados apresentam assimetria e achatamento fraco ou assimetria suave a moderada e achatamento moderado, o enviesamento na estimativa média da covariância e da média do declive deixa de ser negligenciável, qualquer que seja a dimensão da amostra. Quando existem omissões nos dados, o erro quadrático médio aumenta, independentemente da distribuição dos dados, mas se os dados provêm de uma mistura de distribuições normais o enviesamento e o erro quadrático médio aumentam ainda mais. Os valores de cobertura para a variância do intercepto estão no limite do aceitável, mesmo existindo omissões nos dados, qualquer que seja o n° de momentos temporais. Poucos momentos temporais estão associados a valores elevados do erro quadrático médio, qualquer que seja a dimensão da amostra. Para mais detalhes ver VICENTE (2014).

Referências

- BOLLEN, K.A. & CURRAN, P.J. (2006) *Latent Curve Models- A Structural Equation Perspective*, New Jersey, USA, John Wiley & Sons.
- MUTHÉN, B.O. & ASPAROUHOV, T. (2002a) Using Mplus Monte Carlo simulations in practice: A note on assessing estimation quality and power in latent variable models. *Mplus Web Notes*, n°1, version 2.
- MUTHÉN, B.O. & ASPAROUHOV, T. (2002b) Using Mplus Monte Carlo simulations in practice: A note on non-normal missing data in latent variable models. *Mplus Web Notes*, n°2, version 2.
- VICENTE, P.C.R. (2014) *Modelos com Trajetória Latente em Painéis Rotativos com Dados Omissos*. Tese de Doutoramento. Instituto Universitário de Lisboa (ISCTE-IUL).

Sessão Paralela III – Análise de Dados em Controlo de Qualidade – sábado, 11 de abril,
Sala 0.02 (15h00)

A seleção de compostos voláteis como indicadores da qualidade sensorial da carne fresca

C. Saraiva¹, I. Oliveira², C. Martins³, J. Ventanas⁴, C. García⁵

^{1,3} ECAV, DCV- CECAV, Univ. Trás-os-Montes e Alto Douro Apart.1013, P 2, 5001-801 Vila Real crisarai@utad.pt¹; cmartinsutad@gmail.com³;

² ECT, DM- CITAB, Centro de Investigação e de Tecnologias Agro-Ambientais e Biológicas, Univ. Trás-os-Montes e Alto Douro Apart.1013, P 2, 5001-801 Vila Real ioliveir@utad.pt²

^{4,5} Tecnología de Alimentos, Facultad de Veterinaria Universidad de Extremadura, Avenida de la Universidad s/n, 10071 Cáceres, Spain: cgarciag@unex.es⁴; cgarciag@unex.es⁵;

Sumário

Avaliou-se o perfil volátil de carne fresca (m. *Longissimus dorsi*) de bovino, usando HS-SPME-GC-MS, para compreender as alterações bioquímicas que podem ocorrer durante o armazenamento, com implicações na qualidade sensorial da carne, especificamente relacionadas com as medidas sensoriais globais para a perceção do cheiro a deteriorado e avaliação global da frescura. Apresenta-se a questão da seleção de variáveis previamente às análises multivariadas e da escolha do uso das variáveis binárias (presença/ausência) em detrimento das variáveis originais.

Palavras-chave: Análise multivariada, Compostos voláteis, Seleção de variáveis, SPME-GC-MS.

Introdução

Em diversos estudos para a avaliação da qualidade de alimentos, mais especificamente de carne, encontram-se exemplos com a aplicação de técnicas SPME-GC-MS, (*solid-phase microextraction associated to gas-chromatography and mass-spectrometry*) para recolha de múltiplos indicadores voláteis com vista à avaliação da deterioração e determinação da vida útil de um certo produto alimentar. Nestes estudos é usual a aplicação de diversas técnicas multivariadas, como seja a análise em componentes principais (ACP) e no uso de modelos de regressão em componentes principais (RCP) e nos modelos de regressão de mínimos quadrados parciais (PLSR) e da análise discriminante linear (ADL), Berrueta *et al.* (2007).

Neste trabalho avaliaram-se 32 amostras (observações) de bifes de bovino (m. *Longissimus dorsi*) embaladas a vácuo e em MAP (70% O₂: 20% CO₂: 10% N₂) armazenados a 4° C, com um perfil de 54 compostos voláteis (VOCs) (Saraiva, 2008). Tendo em conta que em estudos com mais variáveis que observações é recorrente a existência de multicolinearidade entre variáveis e que na maioria de estudos deste tipo é frequente os resultados dos VOCs, variáveis contínuas, apresentarem distribuição assimétrica com um excesso de zeros, procedeu-se a uma prévia seleção dos VOCs a usar nas análises multivariadas e criaram-se análises paralelas com os indicadores binários de presença/ausência (0/1) destes nas amostras.

Resultados:

Os VOCs foram selecionados considerando-se uma pré-seleção de 31, tendo em conta se os efeitos tempo de armazenamento e tipo de embalagem para cada VOC e avaliando a correlação destes com as medidas sensoriais globais para a perceção do cheiro a deteriorado e avaliação global da frescura das amostras, OFF e OAF respetivamente. Considerou-se a aplicação dos algoritmos existentes no pacote *subselect* do R, Cerdeira *et al.* (2012), utilizando métodos ótimos pelo critério RM, para obtenção de subconjuntos destes 31 VOCs, Duarte Silva (2001, 2002), mas note-se que o critério RM ignora por completo a associação com qualquer variável dependente.

Apresentam-se 6 análises multivariadas aos dados, considerando a seleção inicial de 31 VOCs e ainda a análise dos subconjuntos ótimos de 10 e 20 VOCs nos dados originais e nos dados binários (0/1), tendo em vista a predição de OFF e OAF usando as metodologias PCR e PLSR. Nas análises PCR e PLSR avaliaram-se os coeficientes de correlação e os valores de erro quadrático médio (RMSEP).

Conclusões

Os dados 0/1 podem ser uma opção mais simples de implementar na avaliação da qualidade dos alimentos. Os resultados mostram a simplicidade e utilidade da aplicação de técnicas de seleção de variáveis para a predição do nível de deterioração das amostras em estudos deste tipo. Os resultados nos subconjuntos ótimos apresentaram semelhanças sendo tal um indicador da importância que se deve dar à escolha sobre quais variáveis selecionar à partida.

Agradecimentos: Os autores agradecem à FCT no âmbito dos projetos PEst-OE/AGR/UI0772/2015-2020 e PEst-OE/AGR/UI4033/2015-2020.

Referências

- BERRUETA, L. A., ALONSO-SALCES, R. M., & HÉBERGER, K. (2007). Supervised pattern recognition in food analysis. *Journal of Chromatography A*, 1158, 196-214.
- CERDEIRA, J.O., DUARTE SILVA, P., CADIMA, J. & MINHOTO, M. (2012). *subselect*: Selecting variable subsets. R package version 0.12-2. <http://CRAN.R-project.org/package=subselect>
- DUARTE SILVA, A.P. (2001). Efficient Variable Screening for Multivariate Analysis. *Journal of Multivariate Analysis*, 76, 35-62.
- DUARTE SILVA, A.P. (2002). Discarding Variables in a Principal Component Analysis: Algorithms for All-Subsets Comparisons. *Computational Statistics*, 17, 251-271.
- SARAIVA, C. (2008). Influência do pH final e tipo de embalagem na conservação de carne de bovino da raça Maronesa: parâmetros microbiológicos, físico-químicos, sensoriais e fração volátil. Tese de Doutoramento em Ciências Veterinárias, UTAD, Vila Real.

Sessão Paralela III – Análise de Dados em Controlo de Qualidade – sábado, 11 de abril, Sala 0.02 (15h20)

Controlo estatístico da qualidade em ensaios interlaboratoriais – caso particular de nutrientes em águas salinas

Dora Carinhas¹, Paulo Infante², Pilar Pestana³

¹*Universidade de Évora; Instituto Hidrográfico, dora.carinhas@hidrografico.pt;*

²*CIMA-UE e ECT, Universidade de Évora, pinfante@uevora.pt;*

³*Instituto Hidrográfico, pilar.pestana@hidrografico.pt*

Sumário

Devido à grande concorrência no mercado nacional e internacional, os laboratórios procuram a acreditação dos seus ensaios. A participação em ensaios de comparação interlaboratorial (ECI) permite evidenciar a competência de um laboratório através da avaliação do seu desempenho. Neste trabalho apresentam-se alguns resultados de estudos de reprodutibilidade e da aplicação de outras técnicas de controlo de qualidade aos dados de um ECI, realizado no Instituto Hidrográfico, relativos a nutrientes em águas salinas, no âmbito da acreditação segundo a norma NP EN ISO/IEC 17025.

Palavras-chave: Controlo estatístico de qualidade, Ensaios de comparação interlaboratorial, Reprodutibilidade.

Introdução

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.

A norma NP EN ISO/IEC 17025 descreve os requisitos gerais para o reconhecimento da competência de um laboratório em realizar ensaios e/ou calibrações, incluindo amostragem que é utilizada pelos laboratórios para desenvolver os seus sistemas de gestão para a qualidade e para as atividades administrativas e técnicas. Além da referência aos procedimentos de controlo da qualidade para monitorizar a validade dos ensaios e calibrações realizadas, é referida 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.

Resultados

Na Figura 1, apresenta-se um exemplo de participação do Instituto Hidrográfico nos ECI, organizados pelo laboratório Quasimeme, entre abril de 2004 e abril de 2013, nomeadamente, para a determinação de Nitrito. Para monitorização da estatística Z-score (diferença entre o resultado obtido pelo laboratório participante e o valor de referência dividida pelo desvio padrão do ECI) considerou-se a carta de controlo para valores individuais (Gomes *et al.*, 2010).

Depois de retirados os pontos fora dos limites que correspondiam a causas assinaláveis, obtivemos a carta da Figura 1 que traduz um processo sob controlo com média muito perto do valor alvo para esta estatística (que é zero).



Figura 1 – Carta de controlo para valores individuais sob controlo, após remoção de causas especiais.

Os cálculos da repetibilidade e reprodutibilidade foram realizados com recurso ao método da Análise de Variância (Montgomery, 2012). Concluimos que, para um nível de significância de 5%, o fator amostra teve influência significativa no estudo (*valor p* < 0,001, logo, rejeita-se a hipótese nula de igualdade entre amostras), portanto, o sistema de medição é capaz de detetar a variação natural existente entre amostras. Concluimos também que os laboratórios não tiveram influência significativa nos resultados apresentados, pois não existem diferenças entre os laboratórios (*valor p* = 0,465). De acordo com os componentes de variação calculados, os laboratórios tiveram uma contribuição muito pequena na variação do sistema de medição (reprodutibilidade). A maior parte da variação foi devido às amostras analisadas (99,7%). O índice R&R apresenta um valor igual a 5,35% o que nos indica que estamos na presença de um sistema de medição aceitável.

Conclusões

Neste trabalho foram apresentados alguns resultados de um estudo de implementação de algumas ferramentas de controlo de qualidade na determinação de nitrito em águas salinas, as quais foram aplicadas ao ensaio acreditado. Este exemplo demonstra a competência do laboratório para produzir resultados tecnicamente válidos de acordo com a norma de acreditação, permitindo assim o reconhecimento e a satisfação dos clientes.

Agradecimentos: Ao Instituto Hidrográfico pela cedência dos dados, em particular ao Dr. Carlos Borges e à equipa do laboratório de química orgânica – sector dos nutrientes.

Referências

- GOMES, M. I., FIGUEIREDO, F., BARÃO, M. I. (2010) *Controlo Estatístico da Qualidade*, 2 ed., Edições SPE.
- MONTGOMERY, D. C. (2012) *Introduction to Statistical Quality Control*, 7th. Ed, New York, USA, Wiley.

POSTERS

Sessão de Posters – 6ª feira, 10 de abril (16h15)

A metodologia *Bootstrap* associada ao método de Holt-Winters na previsão de séries temporais

A. Manuela Gonçalves¹, Marco Costa², Joana Silva³

¹CMAT - Centro de Matemática da Universidade do Minho, mneves@math.uminho.pt;

²CIDMA - Centro de Investigação e Desenvolvimento em Matemática e Aplicações, Universidade de Aveiro, marco@ua.pt;

³Universidade do Algarve, joanamarques_gomes@hotmail.com

Sumário

Os métodos de alisamento exponencial são muito utilizados na modelação e previsão de séries temporais, devido à sua versatilidade e à opção de modelos que integram. Na estatística computacional, a metodologia *Bootstrap* é muito aplicada em inferência estatística no âmbito de séries temporais. Este estudo teve como principal objectivo analisar o desempenho do método de Holt-Winters (HW) associado à metodologia *Bootstrap*, como um processo alternativo na modelação e previsão de séries temporais (SILVA, 2014).

Palavras-chave: *Bootstrap*, Holt-Winters, Previsão, Qualidade da água, Séries temporais.

Introdução

A metodologia adoptada neste trabalho resulta da combinação do método de Holt-Winters com a abordagem *Bootstrap*, na aplicação a uma série temporal. Neste trabalho é apresentada a aplicação desta combinação a um caso de estudo de séries temporais relativas a uma variável ambiental de qualidade da água de superfície, o Oxigénio Dissolvido, medida mensalmente desde Março de 2000 a Dezembro de 2011, em oito estações de amostragem de qualidade da bacia hidrográfica do Rio Ave (COSTA & GONÇALVES, 2011). Para isso foi estudado o desempenho do método de alisamento exponencial HW (KALEKAR, 2004) associado à metodologia de reamostragem *Bootstrap* (EFRON & TIBSHIRANI, 1986), que denominamos por metodologia HW-Boot, com o objectivo principal de apresentar um procedimento alternativo para a determinação de estimativas e previsões de valores de uma série temporal, em particular a determinação de intervalos de previsão. Foi também efetuado um estudo comparativo entre o desempenho da aplicação do método Holt-Winters usual e o método HW-Boot. Programou-se em R um algoritmo (com 11 passos), considerando 2000 réplicas no processo de reamostragem dos resíduos.

Resultados

Com o procedimento proposto são obtidas previsões que são comparadas com os valores reais do período observado e deixado para previsão. São utilizadas algumas medidas para avaliar o desempenho da metodologia adoptada quanto à qualidade das previsões dos dois métodos de previsão considerados (o método de Holt-Winters usual e a combinação deste método com a abordagem *Bootstrap*). O estudo comparativo mostrou que a aplicação do *Bootstrap* em associação ao método de Holt-Winters permite obter intervalos de confiança das previsões com menor amplitude do que os obtidos pelos métodos usuais. De modo a aferir se a

redução da amplitude dos intervalos de previsão obtida pelo método HW-Boot implicou um aumento da significância empírica, contabilizou-se o número de observações que no período de previsão entre Janeiro de 2011 e Dezembro de 2011 estavam fora dos intervalos de previsão em ambos os métodos. Verificou-se que para o método HW nenhuma observação, no período respectivo, estava fora do intervalo de previsão. No método HW-Boot, 8% dos intervalos de previsão não contêm as verdadeiras observações. Os resultados obtidos mostram que o método HW usual é conservador na construção dos intervalos de previsão uma vez que apresenta uma significância empírica nula. Por outro lado, o método HW-Boot permitiu obter intervalos de previsão com menores amplitudes obtendo-se, contudo, uma significância empírica próxima da considerada na sua construção.

Conclusões

Com o trabalho desenvolvido espera-se ter contribuído para a obtenção de modelos de previsão mais úteis por forma a obterem-se previsões o mais "correctas" possíveis. Assim, a obtenção de previsões e a construção dos respectivos intervalos de previsão com as propriedades estatísticas estabelecidas num contexto das especificações de um processo de monitorização e do controlo de qualidade das águas de superfície, possibilita a implementação de estratégias a nível de gestão ambiental que permitam a prevenção e a melhoria da qualidade de água de superfície desta bacia.

Agradecimentos: Este trabalho foi parcialmente financiado pelo Centro de Matemática da Universidade do Minho por Fundos Nacionais através da FCT - “Fundação para a Ciência e a Tecnologia”, no âmbito do projecto PEstOE/MAT/UI0013/2014.

Referências

COSTA, M. & GONÇALVES, A.M. (2011) Clustering and forecasting of dissolved oxygen concentration on a river basin. *Stoch. Env. Res. Risk. Ass.*, 25, 151-163.

EFRON, B. & TIBSHIRANI, R. (1986) Bootstrap methods for standard errors, confidence intervals, and another measures of statistical accuracy. *Stat. Sci.*, 1, 54-77.

KALEKAR, P.S. (2004) *Time series forecasting using holt-winters exponential smoothing*. Technical report. Kanwal Rekhi School of Information Technology.

R Development Core Team (2012) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org>.

SILVA, J. (2014) A metodologia *Bootstrap* associada ao método de Holt-Winters na previsão de séries temporais. Tese de Mestrado da Escola de Ciências da Universidade do Minho.

Sessão de Posters – 6ª feira, 10 de abril (16h15)

Alguns benefícios do envelhecimento ativo: um caso de estudo

Paulo Infante¹, Gonçalo Jacinto¹, Anabela Afonso¹, José Conde², Teresa Engana², Jorge Vicente², Pedro Silva², Ana Carla Coelho³, Paula Costa³,

¹*CIMA-UE e ECT, Universidade de Évora, pinfante@uevora.pt*

²*Divisão da Juventude e Desporto, Câmara Municipal de Évora*

³*Agrupamento de Centros de Saúde do Alentejo Central*

Sumário

A Câmara Municipal de Évora, em parceria com várias entidades do Concelho, oferece um programa denominado Seniores Ativos que visa a promoção da atividade física e estilos de vida saudáveis na população sedentária com pelo menos 55 anos de idade. Neste trabalho avaliamos alguns parâmetros físicos e outros relacionados com a saúde dos participantes deste programa. A análise dos resultados, ao fim de 6 meses, com base em 3 grupos que foram constituídos (iniciantes e praticantes com dois níveis de regularidade), permitem comprovar alguns benefícios deste tipo de atividade.

Palavras-chave: Análise de Covariância, Atividade Física, Envelhecimento Ativo, Testes Não Paramétricos, Testes Paramétricos.

Introdução

No decurso do Ano Europeu do Envelhecimento Ativo e da Solidariedade entre Gerações, a Câmara Municipal de Évora (CME), em parceria com a Universidade de Évora (através do CIMA-UE), e com o Agrupamento de Centros de Saúde do Alentejo Central (ACES), realizou um estudo sobre os benefícios da Atividade Física (AF) do programa Seniores Ativos (SA).

Para este estudo construíram-se, por conveniência, 3 grupos de participantes: (1) Iniciados: composto por 15 indivíduos sedentários que iniciaram a AF no programa SA no início do estudo, com uma regularidade de 1 vez por semana; (2) 1x Ativos (grupo padrão): constituído por 37 membros dos SA, mantendo a frequência atual de AF de 1 vez por semana; (3) 3x Ativos: 28 membros dos SA que se voluntariaram para aumentar a frequência da AF para 3 vezes por semana. Os participantes foram submetidos a uma bateria de testes, no início (18-Dez-2012) e no final do estudo (26-Jun e 3-Jul-2013), para avaliar parâmetros físicos e parâmetros médicos.

Quase todos os grupos são compostos maioritariamente por mulheres (Iniciados: 67%, 1x Ativos: 86%, 3x Ativos: 68%). A idade dos participantes varia entre os 55 e os 90 anos de idade. Um em cada 3 participantes dos grupos Iniciados e 3x Ativos indicaram ter praticado AF, nos 6 meses que antecederam o estudo, fora do programa SA. Mais de 90% dos participantes de todos os grupos tomam medicamentos. Uma caracterização mais completa dos grupos, bem como parâmetros analisados e respetivos resultados descritivos, pode ser vista em Infante *et al.* (2014).

Resultados

Para identificar, em cada um dos grupos, os parâmetros com melhorias significativas ao fim de 6 meses, aplicou-se o teste T para amostras emparelhadas ou, perante a violação da normalidade, o teste de Wilcoxon ou o teste dos sinais (aquando da violação da simetria da distribuição). No grupo Iniciados verificaram-se melhorias significativas nas variáveis médicas HDL ($P=0,02$) e Não HDL ($P=0,04$). Nos grupos iniciados e 3x Ativos verificou-se uma melhoria significativa da tensão arterial sistólica ($P<0,01$). Relativamente às variáveis físicas, no grupo Iniciados houve uma melhoria significativa nos resultados de todos os testes, ao nível de significância de 5%, enquanto no grupo 3x Ativos apenas não se detetaram melhorias significativas no teste de agilidade.

Para a análise entre grupos, e uma vez que temos dois momentos de medição com grupos que não se podem considerar completamente aleatorizados, seguimos as recomendações de Dimitrov e Rumrill (2003) e usámos a ANCOVA, considerando a primeira medição como covariável, seguida da aplicação de contrastes nos casos em que o factor grupo se revelou significativo. De notar que estes modelos foram apenas considerados para as variáveis médicas ou físicas que apresentaram fraca correlação entre elas. Para as variáveis médicas, ocorreu uma redução significativa nos valores do mau colesterol (LDL) nos iniciantes de AF em comparação com os 1x Ativos ($P=0,05$). Nas variáveis físicas o grupo que aumentou a frequência de AF mostrou melhorias significativas relativamente ao grupo que manteve a intensidade da AF nos exercícios de resistência ($P=0,04$) e força dos membros superiores ($P<0,01$). Interessante foi notar o acentuado efeito da AF em apenas 6 meses no grupo Iniciados, pois este apresentou melhorias significativas na resistência ($P=0,04$) relativamente ao grupo que manteve a AF. O grupo 3x Ativos apresentou uma maior força dos membros superiores ($P<0,01$) e resistência ($P<0,01$) do que o grupo 1x Ativos.

Conclusões

Tendo presente que este estudo tem limitações, como sejam um delineamento desequilibrado e unidades amostrais seleccionadas por conveniência, foi possível concluir que a prática regular de AF permite melhorar os níveis de colesterol, bem como a resistência e a força dos membros superiores. Note-se que devido à idade avançada dos indivíduos, muitas vezes a não existência de diferenças significa uma desaceleração da perda de funções (Envelhecimento Ativo).

Referências

INFANTE, P.; AFONSO, A.; JACINTO, G., CONDE, J., ENGANA, T., VICENTE, J., SILVA, P., COELHO, A., C., MANO, P. (2014) *Alguns Benefícios do Programa Seniores Ativos*, Portugal, CME/CIMA-UE/ACES.

Disponível em: http://www2.cm-evora.pt/evorajuventudedesporto/Estudos/livro_final.pdf.

DIMITROV D., RUMRILL P. (2003) Pretest-posttest designs and measurement of change. *Work: A Journal of Prevention, Assessment and Rehabilitation*, 20 (2), 159-165.

Sessão de Posters – 6ª feira, 10 de abril (16h15)

Alguns motivos para indivíduos entre os 40 e os 49 anos quererem ter filhos*

Anabela Afonso¹, Maria Filomena Mendes², Paulo Infante¹

¹ CIMA-UE e ECT, Universidade de Évora, aafonso@uevora.pt / pinfante@uevora.pt;

² CIDHEUS e ECS, Universidade de Évora, mmendes@uevora.pt;

Sumário

A fecundidade portuguesa tem-se tornado tema de debate alargado não só entre a comunidade académica, como também na sociedade civil, na comunicação social e mesmo no seio dos governo e oposição, pois nunca se registaram tão poucos nascimentos em Portugal como atualmente. Com base nos dados do Inquérito à Fecundidade realizado pelo INE em 2013, neste trabalho caracterizamos a população do grupo etário dos 40 aos 49 anos que ainda pensa vir a ter filhos e apresentamos alguns motivos que estão na base dessa decisão.

Palavras-chave: Análise de correspondências múltiplas, Demografia, Fecundidade.

Introdução

A realização, em 2013, do Inquérito à Fecundidade (IFEC), no âmbito de um protocolo celebrado entre a Fundação Francisco Manuel dos Santos (FFMS) e o Instituto Nacional de Estatística (INE), veio compensar a escassez de informação que existia acerca desta temática (INE, 2014). Um processo de amostragem complexa permitiu obter uma amostra representativa ao nível regional (NUTS II) para as mulheres, dos 18 aos 49 anos, e ao nível nacional para os homens, dos 18 aos 54 anos.

Neste trabalho vamos focar-nos no grupo etário dos 40 aos 49 anos em que as mulheres revelam uma menor intenção de vir a ter filhos relativamente aos homens ($4,6\% \pm 0,5\%$ das mulheres e $15,7\% \pm 1,2\%$ dos homens), sendo também entre elas que se verifica, uma maior proporção dos que transitaram para a parentalidade ($87,4\% \pm 0,9\%$ das mulheres e $82,2\% \pm 1,2\%$ dos homens). Para averiguar associações entre os motivos mencionados para ainda vir a ter filhos realizou-se uma análise de correspondências múltiplas (Greenacre & Blasius, 2006) e identificaram-se alguns perfis. Posteriormente foi usada uma análise de *clusters* para identificar grupos que correspondessem a estes perfis.

Resultados

De uma forma geral, quanto maior o nível de escolaridade da mulher, maior é a proporção das que ainda pretendem fazer a transição para a maternidade, bem como aumentar a sua descendência. Entre as mulheres que pretendem aumentar a sua família, independentemente de já terem ou não filhos, a maioria encontra-se numa situação profissional estável (59% possui contrato sem termo) ao passo que entre os homens que pretendem vir a ter mais filhos o mais

* Este resumo e as respetivas conclusões não podem ser citados.

comum é serem desempregados (35%) ou trabalhadores sem contrato (31%). Entre quem não tem filhos, mais de metade das mulheres que ainda pensam vir a ter filhos vive atualmente numa união legal ou em coabitação (63%), contrapondo-se ao que se regista entre as que já não pensam vir a ter filhos, onde a grande maioria (66%) não vive em conjugalidade. Entre os homens é surpreendente a proporção daqueles que não vivendo em conjugalidade ainda esperam vir a ser pais (67%).

Neste grupo etário, os motivos considerados mais importantes indicados entre quem pretende alterar a sua fecundidade realizada são essencialmente familiares ou associados à realização familiar. Os motivos considerados importantes são similares entre quem tem e não tem filhos, mas observam-se algumas diferenças entre os sexos. Se por um lado, há uma maior percentagem de homens do que de mulheres que considera importante que ter um filho único não é bom, bem como o diminuir as hipóteses de solidão na velhice, por outro lado, mais de metade das mulheres com filhos indicaram ser um motivo importante quererem ter uma filha.

Numa análise multivariada sobre a importância dos motivos para pretender vivenciar ou repetir a experiência de parentalidade, foi possível resumir a informação relacional entre as variáveis em três fatores que explicam 58% da variância. As três dimensões obtidas, tendo em conta as variáveis que mais contribuem para a sua definição, podem ser nomeadas como: 1ª Relacionamentos (pessoais e conjugais); 2ª Consolidação da família (através dos filhos); 3ª Negociação entre o casal (na constituição e crescimento da família). Foram identificados quatro perfis de pessoas cujas principais características são: 1) desvalorizam os relacionamentos pessoais; 2) valorizam a consolidação da família mas desvalorizam os relacionamentos; 3) atribuem importância à consolidação da família e à negociação entre o casal e desvalorizam os relacionamentos; 4) valorizam a negociação entre o casal.

Conclusões

Apesar do grupo etário dos 40 aos 49 anos ter um impacto residual em termos de fecundidade intencional, a sua caracterização e a identificação de motivos que levam pessoas nesta classe etária a quererem ainda ter filhos ou vir a ter mais filhos parece-nos muito importante para a compreensão dos comportamentos de fecundidade dos portugueses. Este trabalho constitui um contributo nesse sentido.

Agradecimentos: Este trabalho insere-se no projeto “Determinantes da Fecundidade em Portugal”, financiado pela Fundação Francisco Manuel dos Santos e apresenta um extrato limitado dos resultados preliminares com vista à sua discussão inter-pares.

Referências

GREENACRE, M.J, BLASIUS, J. (2006) *Multiple Correspondence Analysis and Related Methods*, Boca Raton, Chapman & Hall/CRC.

INE (2014) *Inquérito à fecundidade 2013*, Lisboa, Instituto Nacional de Estatística e Fundação Francisco Manuel dos Santos.

Sessão de Posters – 6ª feira, 10 de abril (16h15)

Análise das perceções sobre a liderança num organismo público dos Açores com recurso a métodos quantitativos

Áurea Sousa¹, Maria da Graça Batista², Ana Carolina Arruda³

¹*Departamento de Matemática; CEEApla; Universidade dos Açores, aurea@uac.pt;*

²*Departamento de Economia e Gestão; CEEApla, Universidade dos Açores, mbatista@uac.pt;*

³*Departamento de Economia e Gestão, Universidade dos Açores, carolina.arruda19@gmail.com*

Sumário

O objetivo geral deste trabalho é o de identificar e caracterizar as qualidades de liderança das chefias de um organismo público da Região Autónoma dos Açores (RAA). Apresentam-se as principais conclusões obtidas com base na análise dos dados recolhidos a partir de um questionário previamente testado e validado. Conclui-se, em termos gerais, que os dirigentes não apresentam comportamentos de liderança totalmente positiva e eficaz.

Palavras-chave: Liderança, Organizações positivas, Organismos públicos, Métodos quantitativos.

Introdução

Os estudos recentes sobre liderança têm enfatizado a ideia de uma liderança positiva (Cameron, 2008), que requer que os líderes conheçam e sejam capazes de mobilizar três tipos de capital (o humano ou intelectual, o psicológico positivo e o social). Algumas atitudes dos líderes tornam as pessoas mais produtivas e satisfeitas com as ações de liderança, nomeadamente, quando estes indicam o caminho a seguir, inspiram uma visão conjunta, procuram novas oportunidades, testam novas soluções, correm riscos, estimulam a colaboração e transmitem-lhes força e motivação (Cockerell, 2008; Kouzes e Posner, 2002). No entanto, alguns pensamentos e emoções negativas, resultantes da interpretação da realidade, podem destruir o impacto de uma liderança positiva. Assim, procura-se aferir as perceções dos colaboradores sobre as características dos líderes, de modo a que possam ser desenvolvidas medidas que contribuam para uma liderança positiva e eficaz. Sempre que possível as conclusões obtidas são comparadas com as da literatura existente.

Metodologia

Os dados, relativos a 153 funcionários de um organismo público da RAA, foram recolhidos através de um questionário, devidamente testado e validado. Este questionário, além de 6 variáveis de caracterização geral da amostra, contém, numa das suas subsecções, uma escala, constituída por 17 itens, que foi utilizada para avaliar as características específicas do líder (seu líder) e as características do líder eficaz. Cada funcionário selecionou uma e uma só de 6 modalidades possíveis de resposta (1 = *Não Concordo Totalmente*, 2 = *Não Concordo*, 3 = *Não Concordo nem Discordo*, 4 = *Concordo Pouco*, 5 = *Concordo* e 6 = *Concordo Totalmente*)

em relação a cada um dos itens que avaliam as duas variáveis latentes em estudo (“*Características do Líder*” e “*Características do Líder Eficaz*”).

Podem-se destacar, de entre os diversos métodos estatísticos aplicados, alguns testes de hipóteses não paramétricos (Testes de Independência do Qui-quadrado, de Wilcoxon, Mann-Whitney, Kruskal-Wallis e de comparações múltiplas de Dunn), a Análise Classificatória Hierárquica Ascendente (*ACHA*), a Análise em Componentes Principais Categórica (*CatPCA*) e a Regressão Ordinal.

A *ACHA* foi efetuada com o software *CLASSIF* (Nicolau *et al.*, 2009) e permitiu identificar duas tipologias de variáveis ligadas, respetivamente, à caracterização do líder e à caracterização do líder eficaz. A *CatPCA* foi utilizada de forma a possibilitar a representação gráfica das proximidades entre as variáveis num espaço bidimensional. De modo a relacionar as variáveis dependentes (alguns dos 17 itens que caracterizam o líder e o líder eficaz) com as variáveis independentes, “*Habilitações literárias*”, “*Vínculo ao organismo*” e “*Categoria*” foi utilizada a Regressão Ordinal.

Resultados e Conclusões

As principais conclusões relativas à caracterização dos líderes e dos líderes eficazes pelos colaboradores de um organismo público da RAA são sintetizadas em tabelas/figuras. Conclui-se que, na sua maioria, os colaboradores não consideram que os seus líderes se comportam como líderes totalmente eficazes, embora estejam satisfeitos com o facto de os dirigentes não serem extremamente controladores e de procurarem fazer o que a maioria deseja.

A dimensão da amostra utilizada constitui uma limitação à generalização dos resultados. A expansão deste estudo, nomeadamente no que concerne às eventuais diferenças de perceções quando o líder é do género feminino poderá ser útil para expandir o conhecimento relativo a diferenças de género na prática da liderança.

Referências

- CAMERON, K. (2008) *Positive Leadership: Strategies for Extraordinary Performance*, San Francisco, Berrett-Koehler Publishers.
- COCKERELL, L. (2008) *Creating Magic: 10 Common Sense Leadership Strategies from a Life at Disney*, New York, Doubleday Publishing Group.
- KOUZES, J. M. & POSNER, B. Z. (2002) *The Leadership Challenge* (3rd ed.), San Francisco, CA., Jossey-Bass/Wiley.
- NICOLAU, F. C., BACELAR-NICOLAU, H., SOUSA, F., SOUSA, Á. & SILVA, O. (2009) *CLASSIF: Software de Análise Classificatória - Abordagens Clássica e Probabilística VL*. Estudos e Aplicações do LEAD, FP-UL.

Sessão de Posters – 6ª feira, 10 de abril (16h15)

Análise de um Teste de Motivação: Comparação em Países Lusófonos

Ana Lorga da Silva¹, Artur Parreira²

¹DEG-ECEO & CPES, ULHT, ana.lorga@ulusofona.pt;

²FUNDAÇÃO CESGRANRIO & CPES, ULHT, arturmparreira@gmail.com;

Sumário

O artigo apresenta um teste de motivação, baseado em escalas de Likert, como instrumento de avaliação dos padrões de motivação, comparando os resultados obtidos em países lusófonos. É um trabalho que vem na sequência de outros realizados por Parreira e Silva (2013, 2014), nos quais se tem vindo a construir e validar esse teste. Neste estudo, mostra-se o interesse dos resultados em termos de diagnóstico de competências que se podem considerar significativas na condução de pessoas e equipas. Os padrões motivacionais são determinados com base no cálculo dos valores atribuídos.

Palavras-chave: Competência de Motivação; Maturidade; Motivação para o Conhecimento e a Inovação, Testes Paramétricos e Não paramétricos.

Introdução

No presente estudo, foi aplicado um teste de motivação a indivíduos de diversos Países Lusófonos com formação académica não inferior à licenciatura. Os resultados obtidos foram comparados ao nível do país, tendo-se aplicado, para o efeito, testes paramétricos e não paramétricos (Sheskin, 2011).

O referido teste de motivação apresenta diversas questões com resposta numa escala de Likert, efectuando a comparação entre a intensidade de dois motivadores, na maioria das questões.

Para a elaboração do questionário realizaram-se entrevistas abertas, no sentido de identificar temas e expressões a utilizar na sua construção.

Foi também realizado um pré-teste do questionário elaborado, sujeitando as respostas obtidas ao tratamento estatístico apropriado para verificar a sua consistência interna e a sua validade discriminante.

Com base nos dados obtidos no pré-teste foi realizada a revisão do formato do questionário e adaptados alguns dos seus itens. Neste trabalho apresentam-se os resultados da aplicação do questionário no formato e formulação atual.

Foram utilizados modelos de regressão adequados a dados ordinais para compreender alguns dos motivadores presentes no questionário.

Neste trabalho o questionário foi aplicado apenas a indivíduos com nível de ensino mínimo licenciatura.

Resultados / Conclusões

Os resultados apresentados analisam a influência da faixa etária, mas também do género e do país de origem na intensidade motivacional demonstrada nos vários padrões.

Como limitação, associada ao objectivo deste estudo em particular, considera-se o facto de todos os inquiridos possuírem nível de escolaridade superior.

Referências

PARREIRA, A. & SILVA, A.L. (2013). The use of numerical value of adverbs of quantity and frequency in the measurement of attitudes: transforming ordinal scales into interval scales. In *Communication to ECDA Conference. Univ. Luxembourg.*

SHEKIN, D. J.(2011) *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall

SILVA, A.L. & PARREIRA, A. (2014) A Study of an Interval Scale for a Motivation Test, Incluído nos *Anais do Congresso SMTDA* 14 June, Lisbon , Portugal.

Sessão de Posters – 6ª feira, 10 de abril (16h15)

Building up complex COBS

Carla Santos¹, Célia Nunes², Cristina Dias³, João Tiago Mexia⁴.

¹*Departamento de Matemática e Ciências Físicas do Instituto Politécnico de Beja e CMA – Centro de Matemática e Aplicações – Universidade Nova de Lisboa, carla.santos@ipbeja.pt;*

²*Departamento de Matemática e Centro de Matemática e Aplicações – Universidade da Beira Interior, celian@ubi.pt*

³*Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Portalegre e CMA– Universidade Nova de Lisboa, cpsilvadias@gmail.com;*

⁴*Departamento de Matemática da Faculdade de Ciências e Tecnologia e CMA– Universidade Nova de Lisboa, jtm@fct.unl.pt*

Abstract

A model with Orthogonal Block Structure (OBS) is a mixed model whose variance-covariance matrix is a linear combination of known Pairwise Orthogonal Orthogonal Projection Matrices (POOPM) that add up to the identity matrix. When the orthogonal projection matrix, on the space spanned by the mean vector, commutes with the variance-covariance matrices, we have a special type of OBS, named COBS (models with Commutative Orthogonal Block Structure). Using, as background, Commutative Jordan Algebras (CJA) we build up complex COBS from simpler ones, through joining.

Keywords: COBS, Commutative Jordan Algebra, Mixed models, Operations with COBS.

Introduction

Mixed models provide a flexible approach in situations of correlated data, for example, due to repeated measurements in experiments, see Khuri *et al.*, (1998), in biological and medical research, animal and human genetics, agriculture or industry.

Let us consider the mixed model $\mathbf{y} = \sum_{i=0}^w \mathbf{X}_i \boldsymbol{\beta}_i$, where $\boldsymbol{\beta}_0$ is fixed and $\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_w$ are independent random vectors with null mean vectors and variance-covariance matrices $\sigma_1^2 \mathbf{I}_{c_1} \dots \sigma_w^2 \mathbf{I}_{c_w}$, where $c_i = \text{rank}(\mathbf{X}_i)$, $i = 1, \dots, w$. \mathbf{y} has mean vector $\boldsymbol{\mu} = \mathbf{X}_0 \boldsymbol{\beta}_0$ and variance-covariance matrix $\mathbf{V} = \sum_{i=1}^w \sigma_i^2 \mathbf{M}_i$, where $\mathbf{M}_i = \mathbf{X}_i \mathbf{X}_i^T$, $i = 1, \dots, w$. We are interested in a special class of mixed models, named COBS (Nunes *et al.*, 2008, Fonseca *et al.*, 2008), whose variance-covariance matrices are all the positive semi-definite linear combinations of known POOPM, $\mathbf{Q}_1 \dots \mathbf{Q}_m$, such that $\sum_{j=1}^m \mathbf{Q}_j = \mathbf{I}_n$, and moreover, $\mathbf{Q}_1 \dots \mathbf{Q}_m$ commute with the orthogonal projection matrix on the space spanned by the mean vector, \mathbf{T} . This commutativity condition is a necessary and sufficient condition for the least square estimators (LSE) to be best linear unbiased estimators (BLUE) whatever the variance components (Zmyślony, 1978). When the matrices of $\{\mathbf{M}_1 \dots \mathbf{M}_w\}$ commute, they will generate a CJA, A^0 , with principal basis $\{\mathbf{Q}_1^0, \dots, \mathbf{Q}_m^0\}$. Using CJA, as background, our treatment of the models will be based on their algebraic structure.

Results

Model crossing and model nesting enable us to build up complex models from simple ones (Mexia *et al.*, 2010). Model joining is a possible alternative to these operations, in which the observations vector for the final model is the overlap of initial models' observations vectors. The technic used to join models rests on the algebraic structure of COBS and the Cartesian product of CJA, (Fernandes *et al.*, 2010). We prove that joining the COBS

$$\mathbf{y}(l) = \sum_{i=1}^{w(l)} \mathbf{X}_i(l) \boldsymbol{\beta}_i(l), l = 1, 2,$$

with the corresponding CJA, $A(l)$, with principal basis $Q(l) = \{\mathbf{Q}_1(l), \dots, \mathbf{Q}_{m(l)}(l)\}$, we obtain a new COBS, whose corresponding CJA is the cartesian product $A(1) \times A(2)$ with principal basis $\{D(Q_1(1), 0_{n(2) \times n(2)}) \dots D(Q_{m(1)}(1), 0_{n(2) \times n(2)})\} \cup \{D(0_{n(1) \times n(1)}, Q_1(2)) \dots D(0_{n(1) \times n(1)}, Q_{m(2)}(2))\}$.

Acknowledgements

This work was partially supported by the Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) through the project UID/MAT/00297/2013 (Centro de Matemática e Aplicações) and by the Center of Mathematics, University of Beira Interior, through the project PEst-OE/MAT/UI0212/2014.

References

- FERNANDES, C., RAMOS, P., & MEXIA, J.T. (2010) Algebraic structure of step nesting designs. *Discussiones Mathematicae. Probability and Statistics*, 30(2), 221-235.
- FONSECA, M., MEXIA, J.T. & ZMYŚLONY, R. (2008) Inference in normal models with commutative orthogonal block structure. *Acta et Commentationes Universitatis Tartuensis de Mathematica*, 12, 3–16.
- KHURI, A.I., MATHEW, T., & SINHA, B.K. (1998) *Statistical Tests for Mixed Linear Models*. New York :Wiley.
- MEXIA, J.T., VAQUINHAS, R., FONSECA, M. & ZMYŚLONY, R. (2010) COBS: segregation, matching, crossing and nesting. *Latest Trends and Applied Mathematics, Simulation, Modelling*, 4-th International Conference on Applied Mathematics, Simulation, Modelling (ASM'10), 249– 255.
- NUNES, C., SANTOS, C. & MEXIA, J.T. (2008) Relevant statistics for models with commutative orthogonal block structure and unbiased estimator for variance components. *Journal of Interdisciplinary Mathematics*, 11(4), 553-564.
- ZMYŚLONY, R. (1978) A characterization of best linear unbiased estimators in the general linear model. *Mathematical Statistics and Probability Theory*, 2, 365-373.

Sessão de Posters – 6ª feira, 10 de abril (16h15)

Caracterização da praxe académica com recurso à análise de dados

Silva, O.¹, Caldeira, S. N.², Mendes, M.³, Botelho, S. P.⁴, Martins, M. J.⁵

¹Universidade dos Açores, CICS.NOVA, osilva@uac.pt;

²Universidade dos Açores, CICS.NOVA, snc@uac.pt;

³Universidade dos Açores, macmendes1@hotmail.com;

⁴Universidade dos Açores, susanapinhobotelho@hotmail.com;

⁵Instituto Politécnico de Portalegre, mariajmartins@esep.pt.

Sumário

Este estudo teve como intuito conhecer o modo como os estudantes da Universidade dos Açores percecionam a praxe, pretendendo-se aferir se adotam a perspetiva de a verem como um forma de acolhimento e integração na nova instituição ou como uma dificuldade ativadora de sentimentos de apreensão, ansiedade e humilhação. Apresentam-se as principais conclusões obtidas a partir da análise de dados, com recurso a um questionário.

Palavras-chave: Análise de dados, Ensino superior, Praxe académica.

Introdução

A praxe académica é um ritual, que pode desempenhar um papel facilitador da integração do recém-chegado no meio académico, contribuindo para a sua adaptação (Vieira, 2013; Dias & Sá, 2013), oferecendo-lhe suporte afetivo, possibilidade de estabelecer novas amizades e promovendo o sentimento de pertença à nova comunidade (Dias & Sá, 2013; Vieira, 2013). Mas pode, também, constituir um constrangimento sério à transição e adaptação ao ensino superior, caso ocorram atividades ofensivas, intimidatórias, degradantes e violentas sobre o caloiro que ponham em risco a sua saúde e segurança, ameaçando, até, a sua vida (Klerk, 2013; Nirh, 2014).

Neste estudo, participaram 247 estudantes do 1º ciclo de estudos da Universidade dos Açores, de ambos os sexos, com idades entre os 18 e os 43 anos, de diferentes anos, cursos, departamentos e contextos sócio familiares e culturais. A amostra foi selecionada com recurso à amostragem por quotas, atendendo ao género e ao departamento. Foi utilizada a “Escala de Avaliação das Situações de *Bullying* nas Praxes do Ensino Superior” (EASBES) de Matos *et al.* (2010) a qual é constituída por 15 itens de autorresposta, numa escala de Likert onde cinco dos itens são de leitura invertida. Este instrumento possui uma estrutura fatorial em 3 fatores, os quais são designados, respetivamente, por “*Relação positiva com as praxes*” (seis itens), “*Relação negativa com as praxes*” (seis itens) e “*Dimensão Social*” (três itens).

Resultados

Para além de algumas técnicas de estatística descritiva, foram utilizados alguns testes de estatística não paramétrica, a análise de correspondências (simples e múltipla) e algumas técnicas de visualização para a apresentação dos resultados.

Foram os alunos do sexo masculino, com vinte anos ou menos, do 1.º ano e de cursos da área da saúde (Medicina e Enfermagem), aqueles que expressaram opiniões mais favoráveis em relação à praxe académica, reconhecendo o seu papel integrador no meio universitário, e os que mais pontuaram nos itens do fator “Relação *positiva com a praxe*”. Por outro lado, foram os alunos do sexo masculino, de vinte anos ou menos, do 1.º ano e dos departamentos de Biologia e de Economia e Gestão, os que manifestaram uma mais acentuada relação negativa com as praxes. Ao nível da “*Dimensão Social*” foram os alunos do 1.º ano, do sexo masculino, com vinte anos ou menos e do departamento de Biologia (curso de Medicina) e da Escola de Enfermagem foram mais pontuarem nos itens desse fator.

Verificámos que os estudantes da Universidade dos Açores percecionam a praxe como um acontecimento que não configura elevada atratividade, manifestando fraca adesão às atividades propostas, embora prevaleça a opinião de que as ações inerentes à praxe não são violentas.

Referências

DIAS, D. & SÁ, M. (2013) Rituais de transição no ensino superior português: A praxe enquanto processo de reconfiguração identitária. *Revista Galego-Portuguesa de Psicoloxía e Educación*. Vol. 21, (n.º1) Ano 18º, ISSN 1138-1663, http://ruc.udc.es/bitstream/2183/12618/1/RGP_21_2013_art_2.pdf (acedido em 12 de fevereiro 2015).

KLERK, V. (2013) Initiation, Hazing or Orientation? A case study at a south african university. *International Research in Education*, 1 (1), 86-100, <http://dx.doi.org/10.5296/ire.v1i1.4026> (acedido em 12 de fevereiro 2015).

MATOS, F., JESUS, S., SIMÕES, H., & NAVE, F. (2010) Escala para avaliação das situações de *bullying* nas praxes do ensino superior. *Psyc@w@re*, 3 (1), <http://www1.ci.uc.pt/ipc/2007-2010/revista/c6944bceb08cb00930b00b6645171101.pdf>, (acedido em 12 de fevereiro 2015).

NIRH, J. (2014) *Explanations of college students for engaging in hazing*. Dissertação para o grau de Doutor em Filosofia com especialização em Ensino Superior, Universidade do Arizona, Arizona, USA.

VIEIRA, P. (2013) *Vivências da praxe académica: percepção de integração e ansiedade na transição para o ensino superior*. Dissertação de Mestrado em Psicologia da Educação, Desenvolvimento e Aconselhamento, Universidade de Coimbra, Coimbra.

Sessão de Posters – 6ª feira, 10 de abril (16h15)

Consumo Distrital de Antidiabéticos em Portugal – Exemplo Prático da Análise Classificatória

José Guerreiro¹, Carla Torre², Zilda Mendes³

¹ *Centro de Estudos e Avaliação em Saúde (CEFAR), Associação Nacional das Farmácias (ANF), jose.guerreiro@anf.pt;*

² *Centro de Estudos e Avaliação em Saúde (CEFAR), Associação Nacional das Farmácias (ANF), carla.torre@anf.pt;*

³ *Centro de Estudos e Avaliação em Saúde (CEFAR), Associação Nacional das Farmácias (ANF), Zilda.mendes@anf.pt*

Sumário

Portugal é dos países Europeus com maior prevalência da diabetes. A doença constitui uma das áreas prioritárias do plano nacional de saúde. A utilização da Análise Classificatória em dados sobre o consumo distrital de medicamentos na terapêutica da diabetes sugere a existência grupos distintos de distritos que se diferenciam pela utilização desta terapêutica. Planos de intervenção localizados podem assim levar a melhor controlo da doença na população.

Palavras-chave: Análise Classificatória, Diabetes

Introdução

A diabetes é uma doença crónica, metabólica caracterizada pelo aumento dos níveis de glicose no sangue e por complicações que aumentam a morbilidade e mortalidade precoce dos doentes. Estima-se que, em 2013, a prevalência da Diabetes na população portuguesa seja de 13,0%, embora em 44% das pessoas com diabetes a doença permaneça ainda por diagnosticar.

A Análise Classificatória tem como objectivo construir grupos ou classes, partindo de uma recolha de características dos sujeitos ou objectos, sendo os objectos do mesmo grupo mais semelhantes entre si do que objectos situados em grupos diferentes. Este trabalho tem como objectivo analisar a sua aplicação no contexto da investigação associada à utilização de medicamentos na terapêutica farmacológica da diabetes, com exemplos reais recolhidos a partir de uma base de dados nacional sobre o consumo de medicamentos em ambulatório.

Metodologia

A informação sobre o consumo de medicamentos foi obtida a partir da base de dados hmR / SICMED, um sistema de informação representativo do consumo de medicamentos em ambulatório em Portugal, entre 2010 e 2014. Os medicamentos seleccionados fazem parte do grupo terapêutico das insulinas e do grupo de outros antidiabéticos.

Os resultados foram analisados em função do número de Doses Diárias Definidas (DDD; Organização Mundial de Saúde) por 1000 habitantes e por dia, ao nível do país e por distrito.

Na Análise Classificatória, foram utilizados 5 métodos hierárquicos aglomerativos na formação de grupos de distritos, tendo sido seleccionado o método de ligação média. A análise foi efectuada utilizando o software SAS Guide v4.1.

Resultados

Entre 2010 e 2014, o consumo de Insulinas cresceu 19,4% e o consumo de outros antidiabéticos cresceu 5,1%, em Portugal. Foi possível agrupar de forma diferente o consumo de insulinas e o consumo de outros antidiabéticos, por distrito, com reflexo na proporção de consumo de insulinas sobre o consumo total de medicamentos para a diabetes.

Conclusões

A Análise Classificatória permitiu observar grupos de distritos que se diferenciam quer pelo consumo de insulinas e de outros antidiabéticos, quer na proporção de utilização de insulinas no consumo total dos antidiabéticos, sugerindo a importância da eventual criação de programas de intervenção locais para o aumento da taxa de insulinização na diabetes tipo 2, optimizando o controlo da glicémia.

Referências

SOCIEDADE PORTUGUESA DE DIABETOLOGIA (2014), *Relatório Anual do Observatório Nacional da Diabetes - Diabetes: Factos e Números 2014, Portugal*.

BRANCO, J. (2004) *Uma introdução à Análise de Clusters, Portugal, Sociedade Portuguesa de Estatística*.

Sessão de Posters – 6ª feira, 10 de abril (16h15)

Determinants of firm performance: a structural equation approach

Ricardo Jorge Correia¹, José Gonçalves Dias², Mário Sérgio Teixeira³

¹*Instituto Politécnico de Bragança, ricardojorge@ipb.pt;*

²*Instituto Universitário de Lisboa (ISCTE-IUL), BRU-IUL, jose.dias@iscte.pt*

³*Universidade de Trás-os-Montes e Alto Douro, mariosergio@utad.pt*

Abstract

This research analyzes specific determinants of firm Performance taking the Resource Based View (RBV) theory to explaining how firms achieve sustained competitive advantages based on their VRIN (valuable, rare, inimitable and not replaceable) resources and capabilities (BARNEY, 1991). TEECE et al. (1997) introduced the concept of Dynamic Capabilities (DC) as an RBV extension. Our conceptual model defines these capabilities as the result of firm's Learning, Market and Entrepreneurial Orientations, which mediate the development of Competitive Advantages (CA) and improving firm Performance.

Keywords: Business performance, Latent variables, Structural equation models.

Introduction

LIN et al. (2008) argue that Market Orientation (MO) is heavily determined by Learning Orientation (LO), and this is our first hypothesis (H1). Firms that learn quickly improve their ability to respond to the market and identify potential fields of entrepreneurial activities. Consequently, we test how LO affects Entrepreneurial Orientation (EO) (H2). A dynamic learning mechanism is a distinctive process that promotes the firm DC (LEE et al., 2011), i.e., we assume the relation between LO and DC (H3). One of the key firm resources and capabilities is MO, and its relation with DC defines a new hypothesis to be tested (H4). TEECE (2007) suggests that one of the DC antecedents is the entrepreneurial component, thus, it is suggested the relationship between EO and DC (H5). DC are one of the key antecedents to CA. This suggests testing the relation between DC and Differentiation (H6) and Cost Leadership (H7). Finally, the relationship between the development and/or the maintenance of a CA and the firm Performance has already been proven in literature. Thus, we assume that Differentiation and Cost Leadership positively influence the firm Performance (H8 and H9, respectively). All hypotheses are defined for cause-effect relationships. The operationalization of the constructs in our model is based on battery of items available in the literature. All these items use 7-point Likert-type scales (1 = totally disagree to 7 = totally agree).

Results

The empirical testing of the conceptual model is conducted on the population of 1,168,964 Portuguese firms in 2010 (INE, 2012). The sample of 1,190 Portuguese firms was obtained from a Portuguese database of approximately 91,000 firms. CEOs or other top manager from all those companies were contacted by e-mail and the scope and objectives of the study were clarified.

These scales were validated through an exploratory factor analysis (EFA) using software SPSS v20. Based on the EFA results, items with low factor loadings were eliminated as they were unreliable indicators to measure the construct. Then, the computation of Cronbach's alpha, composite reliability (CR), and average variance extracted (AVE) for all the scales show satisfactory levels of reliability, except for Cost Leadership, with a variance extracted below 0.5.

The structural equation model is estimated by the maximum likelihood method using software MPlus 6.0. The estimation of the structural relations shows that: LO positively affects MO, EO and DC (H1, H2 and H3); MO and EO positively influence DC (H4 and H5); the same conclusion can be drawn from the relationship between DC and CA differentiation (H6) and cost leadership (H7). Finally, both CAs (Differentiation and Cost Leadership) have a positive impact on Performance (H8 and H9).

Conclusion

For the first time, the mediator role of DC between major corporate strategic orientations – MO, EO and LO – and the development of CA and firm Performance is analyzed. The study supports the view that DC enhances CA (Differentiation and Cost Leadership) as well the chance to improve firm Performance.

References

- BARNEY, J. (1991) Firm resources and sustained Competitive Advantage. *Journal of Management*, 17, 99-120.
- INE - Instituto Nacional de Estatística (2012). Portal de Estatísticas Oficiais – Indicadores de 2010. <http://www.ine.pt>
- LEE, P., LIN, H., CHEN, H., & SHYR, Y. (2011) Dynamic capabilities exploitation of market and hierarchy governance structures: An empirical comparison of Taiwan and South Korea. *Journal of World Business*, 46, 359–370.
- LIN, C., Peng, C., & Kao, D. (2008) The innovativeness effect of Market Orientation and Learning Orientation on business performance. *International Journal of Manpower*, 29, 752 - 772.
- TEECE, D. (2007) Explicating dynamic capabilities: the nature and microfoundations of (sustainable) enterprise performance. *Strategic Management Journal*, 28, 1319–1350.
- TEECE, D., PISANO, G., & SHUEN, A. (1997) Dynamic capabilities and strategic management. *Strategic Management Journal*, 18(7), 509–533.

Sessão de Posters – 6ª feira, 10 de abril (16h15)

Estudo do Papel Mediador de Variáveis Psicológicas na Relação entre a Intenção e a Prática de Exercício Físico

A. Manuela Gonçalves¹, Rui Gomes²

¹CMAT - Centro de Matemática da Universidade do Minho, mneves@math.uminho.pt;

²Escola de Psicologia da Universidade do Minho, rgomes@psi.uminho.pt

Sumário

Aplicando Modelos de Equações Estruturais pretendeu-se avaliar os efeitos da teoria do comportamento planeado (TPB) na explicação do comportamento de exercício físico, mediada por um conjunto de variáveis psicológicas relacionadas com o modelo transteórico (TTM), o modelo de ação na saúde (HAPA) e as experiências subjetivas de exercício (SEE). O modelo teórico proposto foi avaliado numa amostra de 454 participantes que praticavam exercício num centro de *fitness*.

Palavras-chave: Exercício Físico, Mediação, Modelos de Equações Estruturais, Teoria do Comportamento Planeado.

Introdução

Os benefícios do exercício físico são bem conhecidos (HARDMAN & STENSEL, 2009). Contudo, mais de 60% da população mundial é fisicamente inativa ou ativa de forma insuficiente para atingir tais vantagens e 50% dos adultos que iniciam um programa de exercício físico regular desistem nos primeiros 6 meses (MARCUS *et al.*, 2000). Este estudo pretendeu avaliar os efeitos da teoria do comportamento planeado (TPB) (MADDUX, 1993) na explicação da intenção e do comportamento de exercício físico mediados por um conjunto de variáveis psicológicas (TTM, HAPA e SEE).

O estudo incluiu 454 participantes (n = 248 do sexo feminino, 54,6%) que praticavam exercício físico num centro de *fitness*. Foram recolhidas, no início da prática desportiva, medidas de TPB-DB (atitudes, normas subjetivas, percepção de controle comportamental, intenção), TTM (balança decisional), HAPA (planeamento), SEE (bem-estar, mal-estar e fadiga) e avaliado no período seguinte de três meses o comportamento de exercício físico.

Foram analisados modelos de mediação múltipla e simples. A estimação dos modelos foi desenvolvida em modelos de equações estruturais (BYRNE, 2001; ANDERSON e GERBING, 1988) recorrendo ao AMOS 21, por aplicação do método de máxima verosimilhança juntamente com procedimento *bootstrap* com correcção de enviesamento (1000 amostras *bootstrap*) para obter intervalos de estimação a 95%.

Resultados

Os modelos de equações estruturais apoiam parcialmente o modelo da TPB (explicando 10% da variância do comportamento de exercício físico) e a dimensão bem-estar (SSE) medeia a relação entre a intenção e o comportamento de exercício físico (explicando 11% da variância do comportamento de exercício físico). Na sua forma actual, a TPB não permite predizer de

forma adequada o comportamento de exercício físico, e mesmo quando considerada em conjunto com as outras construções conceptuais (TTM, HAPA e SSE) também foi encontrada evidência limitada na sua utilidade na explicação do comportamento de exercício físico.

Conclusões

Analisando os modelos obtidos, a TPB representa uma proposta interessante para explicar a intenção de fazer exercício físico, mas não para explicar o comportamento de exercício físico, mesmo quando considerada em conjunto com outras construções conceptuais. Isto sugere que pesquisas futuras devem tentar acomodar novas variáveis entre a intenção e comportamento de exercício, assumindo como referencial teórico a proposta da TPC.

Agradecimentos: Este trabalho foi parcialmente financiado pelo Centro de Matemática da Universidade do Minho por Fundos Nacionais através da FCT - “Fundação para a Ciência e a Tecnologia”, no âmbito do projecto PEstOE/MAT/UI0013/2014.

Referências

ANDERSON, J.C. & GERBING, D.W. (1988) Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, 103(3), 411-423.

BYRNE, B.M. (2001) *Structural equation modeling with AMOS: Basic concepts, Applications, and programming*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

HARDMAN, A.E. & STENSEL, D.J. (2009) *Physical activity and health: The evidence explained*, (2nd ed.). New York: Routledge.

MADDUX, J.E. (1993) Social cognitive models of health and exercise behavior: An introduction and review of conceptual issues. *Journal of Applied Sport Psychology*, 5(2), 116-140.

MARCUS, B.H., DUBBERT, P.M., FORSYTH, L.H., MCKENZIE, T.L., STONE, E., DUNN, A.L. & BLAIR, S.N. (2000) Physical activity behavior change: Issues in adoption and maintenance. *Health Psychology*, 19(1), 32–41.

Sessão de Posters – 6ª feira, 10 de abril (16h15)

Models of stochastic symmetric matrices with truncation

Cristina Dias¹, João Tiago Mexia², Carla Santos³, Célia Nunes⁴.

¹Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Portalegre e CMA–Centro de Matemática e Aplicações da Universidade Nova de Lisboa, cpsilvadias@gmail.com;

²Departamento de Matemática da Faculdade de Ciências e Tecnologia e CMA–Centro de Matemática e Aplicações da Universidade Nova de Lisboa, jtm@fct.unl.pt;

³Departamento de Matemática e Ciências Físicas do Instituto Politécnico de Beja e CMA–Centro de Matemática e Aplicações da Universidade Nova de Lisboa, carla.santos@ipbeja.pt;

⁴Departamento de Matemática e Centro de Matemática –Universidade da Beira Interior, celian@ubi.pt

Abstract

The models we developed are k -degree models of the form $\mathbf{M} = \boldsymbol{\mu} + \bar{\mathbf{E}}$ for symmetric stochastic matrices \mathbf{M} , with mean matrix $\boldsymbol{\mu}$, and $\bar{\mathbf{E}}$ is a symmetric stochastic matrix with null mean. For the models with a degree $k > 1$, we still consider the possibility of truncating the model, when there are eigenvalues, $\theta_1, \dots, \theta_k$ much greater than the other myth.

Keywords: Power tests, Symmetric matrices, Truncation of the model.

Introduction

Given a matrix $\mathbf{M} = [m_{ij}]$ symmetric random $k \times k$, the respective average matrix $\boldsymbol{\mu}$ when set, will be symmetrical with the spectral decomposition $\boldsymbol{\mu} = \sum_{i=1}^k \lambda_i \boldsymbol{\alpha}_i \boldsymbol{\alpha}_i^t$, with $\mathbf{M} = \sum_{i=1}^k \lambda_i \boldsymbol{\alpha}_i \boldsymbol{\alpha}_i^t + \bar{\mathbf{E}}$, where $\bar{\mathbf{E}}$ is symmetric stochastic matrix with zero mean matrix. The degree of \mathbf{M} will be $k = \text{car}(\boldsymbol{\mu})$, which is the characteristic of the average matrix (supposedly set). In models that are considered below, it is assumed that $\mathbf{Z} = \text{vec}(\mathbf{M}) \sim N(\boldsymbol{\eta}, \sigma^2 \mathbf{L})$, is (approximately) normal, with mean vector $\boldsymbol{\eta}$ and covariance matrix $\sigma^2 \mathbf{L}$ (Areia et al., 2008). In the general theory we begin by considering the structure of vectors $\boldsymbol{\beta}_i = \lambda_i \boldsymbol{\alpha}_i$, $i = 1, \dots, k$ and residue vectors. We will see then how to determine the degree of the model. From the potential estimators we will obtain confidence regions for these parameters.

Results

Consider the spectral analysis of the matrix

$$\mathbf{M} = \sum_{j=1}^n \theta_j \gamma_j \gamma_j^t.$$

If the Ω_h e Ω_h^\perp are the spaces generated by $\gamma_1, \dots, \gamma_h$, and for the $\gamma_{h+1}, \dots, \gamma_n$, $h=1, \dots, n-1$ respectively, then $\theta_h = \min_{x \in \Omega_h} \frac{x^t \mathbf{M} x}{x^t x}$, $h=1, \dots, n-1$, and $\theta_h = \max_{x \in \Omega_h^\perp} \frac{x^t \mathbf{M} x}{x^t x}$, $h=1, \dots, n-1$. To

choose the degree of our model we are admitting, that $\tilde{\mathbf{M}} = \sum_{j=1}^{k'} \theta_j \gamma_j \gamma_j^t$, is the adjusted model.

So, the adjustment of the model can be measure by

$$v = \frac{\|\tilde{\mathbf{M}}\|^2}{\|\mathbf{M}\|^2} = 1 - \frac{\|\mathbf{R}\|^2}{\|\mathbf{M}\|^2} = \frac{\sum_{j=1}^{k'} \theta_j^2}{\sum_{j=1}^n \theta_j^2}, \quad 0 \leq v \leq 1, \text{ where } \mathbf{R} = \mathbf{M} - \tilde{\mathbf{M}} = \sum_{j=k'+1}^n \theta_j \gamma_j \gamma_j^t \text{ is the matrix of}$$

errors.

Final remarks

A direct consequence of our methodology is that degree-one models can be applied to cross products matrices and Hilbert-Schmidt scalar products matrices.

Acknowledgements

This work was partially supported by the Portuguese Foundation for Science and through the project UID/MAT/00297/2013 (Centro de Matemática e Aplicações).

References

- AREIA, A., OLIVEIRA, M. M & MEXIA, J. T. (2008) Models for a Series of Studies Based on Geometrical Representation. *Statistical Methodology*, 5, 277-288.
- AZZALINI, A. (1996) *Statistical Inference*, Monographs on Statistics and Applied Probability 68, Chapman & Hall, New York.
- SEELY, J. (1977) Minimal sufficient statistics and completeness for multivariate normal families. *Sankhya* 39, 170-185.
- SALVADOR, D.; MONTEIRO, S. & MEXIA, J.T. (2012) Structured families of stochastic symmetric matrices. *Proceedings of the Numerical International Conference of Numerical Analysis and Applied Mathematics, Volume 1479*. Rhodes, Greece.

Sessão de Posters – 6ª feira, 10 de abril (16h15)

O Papel Mediador da Avaliação Cognitiva na Relação entre Stress e Saúde Psicológica

Susana Faria¹, A. Rui Gomes²

¹ CMAT – Centro de Matemática, Departamento de Matemática e Aplicações, Universidade do Minho, sfaria@math.uminho.pt

² Departamento de Psicologia Aplicada. CIPsi, Centro de Investigação em Psicologia. Universidade do Minho, Portugal, rgomes@psi.uminho.pt

Sumário

Aplicando Modelos de Equações Estruturais pretende-se avaliar o papel mediador da avaliação cognitiva na relação entre o stress ocupacional e a saúde psicológica em enfermeiros.

Os resultados evidenciam o efeito mediador da avaliação cognitiva nessa relação, no entanto, os efeitos diretos do stress sobre a saúde psicológica não podem ser ignorados. A avaliação cognitiva é assim um importante mecanismo para explicar a adaptação no trabalho.

Palavras-chave: Avaliação cognitiva, Mediação, Modelo de equações estruturais, Stress ocupacional.

Introdução

A atividade dos profissionais de saúde está associada a alto stress ocupacional uma vez que estes profissionais enfrentam diferentes fontes de pressão, tais como longos dias de trabalho, carga de trabalho, relações interpessoais, atendimento ao paciente, privação de sono, baixa tolerância para o erro e recursos limitados (Lee & Wang, 2002).

Estes problemas têm demonstrado efeitos significativos sobre os enfermeiros (depressão, esgotamento, insatisfação no trabalho e fadiga crónica grave) justificando a importância de analisar os mecanismos que podem explicar o aparecimento e desenvolvimento do stress ocupacional. Um destes mecanismos, refere-se aos processos de avaliação cognitiva (Lazarus & Folkman, 1984), que nos indicam até que ponto uma dada situação de stress pode ser avaliada pelo indivíduo como mais ameaçadora ou mais desafiadora.

Este estudo pretende averiguar se a avaliação cognitiva poderá interferir na relação entre a perceção das condições de trabalho enfrentadas pelos profissionais de saúde (por exemplo, fontes de stress) e a forma como eles se sentem mentalmente (por exemplo, saúde psicológica). Mais especificamente, pretende-se analisar o papel mediador da avaliação cognitiva na relação entre o stress ocupacional e a saúde psicológica.

Metodologia

Neste estudo participaram 2302 enfermeiros portugueses, sendo 1895 do sexo feminino (82.3%) e 407 do sexo masculino (17.7%) com idades entre os 21 e os 66 anos (M = 33.75; DP = 9.42). A maioria dos enfermeiros trabalha por turnos rotativos (64.3%) e 69.5% tem um contrato permanente de trabalho.

O modelo de mediação da avaliação cognitiva foi estimado com o *software* AMOS 21 (Byrne, 2010). No ajustamento do modelo, utilizou-se uma estratégia em duas etapas: na primeira etapa ajustou-se um modelo de medida e na segunda etapa ajustou-se um modelo estrutural. Para avaliar a qualidade de ajustamento do modelo utilizaram-se a estatística χ^2 , o índice TLI, NFI e CFI e o RMSEA. O método de reamostragem *Bootstrap* foi aplicado para se obterem os intervalos de confiança das estimativas dos parâmetros.

A amostra foi dividida aleatoriamente em duas subamostras: a amostra 1 (n = 1149) para testar o ajustamento do modelo e a amostra 2 (n = 1150) para validar o modelo.

Resultados

O modelo de mediação parcial da avaliação cognitiva na relação entre o stress ocupacional e a saúde psicológica apresentou um bom ajustamento ($\chi^2(799)=2380.34$; RMSEA = 0.042; CFI = 0.93; TLI = 0.92). Todos os itens apresentam pesos fatoriais elevados e a percentagem de variância da saúde psicológica explicada pelo modelo é 32%.

Aplicando validação cruzada, o modelo ajustado na amostra 1 continua a apresentar um bom ajustamento na amostra 2 (RMSEA = 0.043; CFI = 0.92; TLI = 0.91).

Conclusões

Este estudo confirmou a importância da avaliação cognitiva na relação entre stress ocupacional e saúde psicológica dos enfermeiros portugueses.

O modelo de mediação da avaliação cognitiva em que o caminho direto entre o stress e a saúde psicológica foi mantido, confirma a influência da avaliação cognitiva na relação entre o stress ocupacional e a saúde psicológica dos enfermeiros; no entanto, também reforçou a influência do stress ocupacional na saúde psicológica.

Agradecimentos: S. Faria foi financiada pelo Centro de Matemática da Universidade do Minho com Fundos Nacionais através da FCT – Fundação para a Ciência e a Tecnologia no âmbito do projeto PEst-OE/MAT/UI0013/2014.

Referências

BYRNE, B. M. (2010) *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. New York: Routledge.

LAZARUS, R. S., & FOLKMAN, S. (1984) *Stress, appraisal, and coping*. New York: Springer.

LEE, I., & WANG, H. H. (2002) Perceived occupational stress and related factors in public health nurses. *Journal of Nursing Research*, 10(4), 253-260.

Sessão de Posters – 6ª feira, 10 de abril (16h15)

Participação de crianças e adolescentes no planeamento das suas localidades

Isabel Preto¹, Mário Freitas², Luís Dourado³, Michael McCall⁴.

¹*Instituto da Educação da Universidade do Minho, id2415@alunos.uminho.pt;*

²*Universidade Federal de Santa Catarina, pmariofreitas@gmail.com;*

³*Instituto da Educação da Universidade do Minho, ldourado@ie.uminho.pt;*

⁴*UNAM, mccall@ciga.unam.mx*

Sumário

A participação ativa de crianças e adolescentes no planeamento das localidades é, atualmente, considerada de grande relevância. No âmbito de um estudo participativo com crianças e adolescentes procurou perceber-se, entre outras coisas, quais as transformações que estes sugeririam para as suas localidades caso fossem convidados a pronunciar-se. A caracterização e análise das sugestões referidas, através de análise de conteúdo e ACM, permitiu distinguir, entre os participantes, três perfis distintos.

Palavras-chave: Participação infantojuvenil, localidade, perfis, ACM.

Introdução

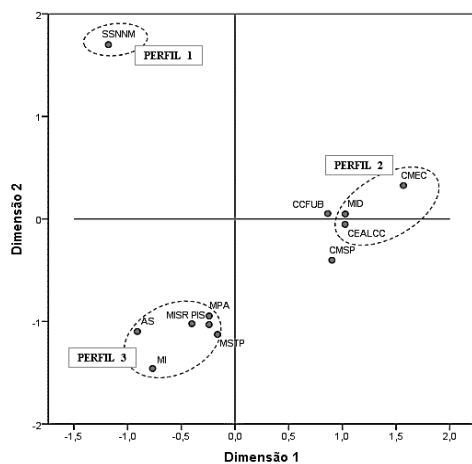
Para construir sociedades mais sustentáveis é importante que aos mais novos seja dada a oportunidade de influenciar assuntos que lhes concernem. Jovens, e sobretudo as crianças mais pequenas, são muitas vezes condicionados na sua ação sobre o meio envolvente, vivendo sob uma redoma, sobre-controladas, sem terem real noção do meio onde se inserem e dos problemas deste (Béneker *et al.*, 2010; Frank, 2006; Hart, 1979). Por isso, tem sido atribuído cada vez maior ênfase à importância de promover a participação infantojuvenil em processos de planeamento dos seus entornos habituais.

No âmbito de um estudo mais alargado relativo à participação infantojuvenil em processos de planeamento urbano procurou perceber-se quais as transformações que crianças e adolescentes sugeririam para as suas localidades, caso fossem chamados a pronunciar-se pelas autoridades locais. Participaram neste estudo 825 alunos, com idade maioritariamente entre os 10 e os 14 anos, de escolas da Póvoa de Varzim, Porto, Póvoa de Lanhoso e também de ilhas dos Açores. Estas localidades foram escolhidas mediante um critério geográfico e os alunos foram selecionados de acordo com a disponibilidade das escolas convidadas a participar.

Para questionar os alunos relativamente às suas sugestões de transformação para as suas localidades utilizou-se uma pergunta de resposta aberta, na qual os alunos podiam indicar até um máximo de três sugestões. O tratamento inicial das respostas dadas pelos alunos efetuou-se através de uma categorização com base num critério semântico (Bardin, 2009), que redundou na criação de um sistema de categorias. Para estudar a relação entre as diferentes categorias foi efetuada uma Análise de Correspondências Múltiplas (ACM), sobre o conjunto de todos os resultados (Carvalho, 2008; Greenacre & Blasius, 2006).

Resultados

Na ACM realizada, verificou-se que as duas primeiras dimensões são muito mais relevantes que qualquer uma das restantes. A estrutura obtida, no plano de cruzamento das duas primeiras dimensões deixa em evidência três configurações bem definidas, correspondentes a três perfis distintos, como se pode ver na Figura 1.



Legenda:

AS – Aumentar a segurança; **CCFUB** – Construção de ciclovias e facilidade de utilização de bicicletas; **CEALCC** – Criar ou melhorar espaços e/ou atividades de lazer/convívio e/ou cultura; **CMEC** – Criar mais estabelecimentos comerciais; **CMSP** – Criar ou melhorar serviços públicos; **CCFUB** – Construção de ciclovias e facilidade de utilização de bicicletas; **MI** – Melhorar infraestruturas físicas e equipamentos; **MID** – Criar ou melhorar infraestruturas desportivas; **MISR** – Melhorar infraestruturas relacionadas com o sistema rodoviário; **MPA** – Maiores preocupações ambientais; **MSTP** – Melhorar o serviço de transportes públicos; **SSNNM** – Sem sugestões ou não necessita melhorias.

Figura 1: Configuração do espaço SUG

Conclusões

As sugestões referidas pelos alunos participantes caracterizaram-se como concretas, contextualizadas no quotidiano e, em muitos casos, suscetíveis de discussão e implementação. Por outro lado, a ACM efetuada sobre as categorias definidas permitiu distinguir três perfis distintos, quanto às sugestões de transformação: alunos que não indicam qualquer sugestão ou consideram que não há necessidade de melhorias na localidade (Perfil 1), alunos que indicam sugestões maioritariamente relacionadas com lazer e cultura (Perfil 2) e alunos que indicam sugestões maioritariamente relacionadas com melhorias a nível de infraestruturas e melhorias ambientais (Perfil 3). Concluiu-se, também, via ACM, da ligação dos alunos da Póvoa de Varzim e Açores ao Perfil 2 e dos alunos do Porto e Póvoa de Lanhoso ao Perfil 3.

Referências

- BARDIN, L. (2009) *Análise de Conteúdo* (4ª ed.), Lisboa, Portugal, Edições 70.
- CARVALHO, H. (2008) *Análise multivariada de dados qualitativos: utilização da Análise de Correspondências Múltiplas com o SPSS* (1ª ed.), Lisboa, Portugal, Edições Sílabo.
- GREENACRE, M., & BLASIUS, J. (Eds.). (2006) *Multiple Correspondence Analysis and Related Methods*, Boca Raton, USA, Chapman & Hall/CRC.
- HART, R. (1979) *Children's Experience of Place*, New York, USA, Irvington.
- BÉNEKER, T., SANDERS, R., TANI, S., & TAYLOR, L. (2010) Picturing the city: young people's representations of urban environments. *Children's Geographies*, 8(2), 123-140.
- FRANK, K. I. (2006) The Potential of Youth Participation in Planning. *Journal of Planning Literature*, 20(4), 351-371.

Sessão de Posters – 6ª feira, 10 de abril (16h15)

Perceção de barreiras na carreira e diversidade de género: Uma abordagem multivariada numa amostra de trabalhadores do sector energético

Ana Sousa Ferreira¹, Rosário Lima²

¹Faculdade de Psicologia da Universidade de Lisboa e BRU-IUL, asferreira@psicologia.ulisboa.pt;

²Faculdade de Psicologia da Universidade de Lisboa, rosariolima@netcabo.pt

Sumário

A presente investigação tem como objetivo estudar a perceção de barreiras na carreira e a sua relação com o género. Contou-se com a participação de 325 colaboradores de uma empresa do sector energético em Portugal e procedeu-se à aplicação do *Inventário de Perceção de Barreiras na Carreira- Revisto (IPBC-R)*. A análise dos resultados iniciou-se com a análise fatorial das respostas ao inventário tendo sido possível concluir que as mulheres percecionam mais barreiras na carreira do que os homens, nomeadamente ao nível da mentalidade individual.

Palavras-chave: Análise Multivariada, Desenvolvimento de Carreira, Diversidade de Género, Perceção de Barreiras na Carreira.

Introdução

Compreender as barreiras que interferem com o processo de desenvolvimento de carreira, é importante para determinar os fatores que podem influenciar o sucesso dos indivíduos e das organizações das quais fazem parte (Swanson & Woitke, 1997). Este objetivo, associado ao aumento do número de mulheres no mercado de trabalho, torna cada vez mais pertinente o estudo da perceção das barreiras na carreira desta população, de forma a compreender melhor não só o papel da mulher como também a influência da sua participação nas organizações (Subramaniam, Arunugam & Akeel, 2013). Vários estudos têm demonstrado que as mulheres percecionam mais barreiras na carreira do que os homens, sendo que a identificação dos fatores que constituem as principais barreiras na carreira das mulheres, pode contribuir significativamente para a aprendizagem das organizações.

Resultados

A presente investigação foi desenvolvida numa empresa portuguesa do sector energético, onde a diversidade da força de trabalho constitui uma prioridade estratégica para a gestão do capital humano. O *IPBC – R* foi aplicado a 325 trabalhadores, tendo-se estudado a sua estrutura fatorial permitindo concluir que se trata de um instrumento multidimensional que avalia um vasto universo de barreiras da carreira. No *IPBC – R* as respostas são dadas numa escala de *Likert* em 7 pontos. A análise da estrutura fatorial deste tipo de escalas são habitualmente analisadas em Psicologia por recurso à Análise Fatorial Exploratória. Neste trabalho, procura-se comparar esses resultados com a aplicação de outros métodos de análise fatorial adequados à natureza dos dados.

Os resultados revelaram diferenças significativas na percepção das barreiras na carreira nos dois géneros, sendo as mulheres que percecionam mais barreiras do que os homens, nomeadamente ao nível da mentalidade individual, discriminação sexual e racial, desaprovação por pessoas significativas, dificuldades de socialização/*networking*, e conflito trabalho-família. As barreiras identificadas poderão traduzir a importância de intervir ao nível da gestão dos recursos humanos em aspetos como a preparação inadequada das mulheres, os seus níveis de satisfação com a carreira, falta de confiança e dificuldades na tomada de decisão. Os resultados representam ainda um alerta face aos obstáculos estruturais e de conciliação trabalho-família identificados na organização, que apelam à revisão de algumas das suas políticas e práticas de recursos humanos.

Conclusões

Com base nos resultados obtidos, torna-se evidente a necessidade das empresas apostarem em medidas que atenuem as barreiras percecionadas pelas mulheres, providenciando a este grupo minoritário as ferramentas necessárias ao seu desenvolvimento de carreira. Uma vez que a diversidade de género, no âmbito do planeamento e desenvolvimento estratégico dos recursos humanos (Cardoso & Marques, 2008) pode ser considerado um fator crítico para a implementação eficaz de estratégias que contribuem para o sucesso das organizações, pode-se afirmar que da investigação desenvolvida resultou a definição de linhas estratégicas a assumir pela empresa alvo do estudo, no âmbito da diversidade de género. Nesse sentido, o foco deve incidir na identificação dos problemas que sistematicamente afetam negativamente a diversidade de género, e prejudicam o desempenho organizacional (Bukhari & Sharma, 2014).

Referências

- BUKHARI, S. S., & SHARMA, B. C. (2014). Workplace Gender Diversity & Inclusive Growth in Public & Private Organizations. *Indian Journal of Industrial Relations*, 49(3), 551.
- CARDOSO, P., & MARQUES, J. F. (2008). Perception of career barriers: The importance of gender and ethnic variables. *International Journal for Educational and Vocational Guidance*, 8(1), 49-61.
- SUBRAMANIAM, I. D., ARUMUGAM, T., & AKEEL, A. B. A. A. B. (2013). Demographic and Family Related Barriers on Women Managers' Career Development. *Asian Social Science*, 10(1), 86.
- SWANSON, J. L., & WOITKE, M. B. (1997). Theory into practice in career assessment for women: Assessment and interventions regarding perceived career barriers. *Journal of career assessment*, 5(4), 443-462.

Índice de Autores

- Abreu, António, 51
Abreu, Pedro Henriques, 73
Afonso, Anabela, 67, 97, 99
Amado, Conceição, 49, 61
Amorim, Maria José, 75
Antunes, Joaquim, 69
Antunes, Patrícia, 31
Arruda, Ana Carolina, 101
Batista, Maria da Graça, 101
Batista, Rodrigo, 33
Botelho, S. P., 107
Caldeira, S. N., 107
Cardoso, Margarida, 75
Carinhas, Dora, 91
Carrasquinha, Eunice, 49
Castro, Conceição, 53
Celeux, Gilles, 9, 19
Coelho, Ana Carla, 97
Colaço, Ricardo, 33
Conde, José, 97
Correia, Ricardo Jorge, 111
Costa, Marques, 95
Costa, Paula, 97
Cunha, Tiago, 73
Dias, Cristina, 105, 115
Dias, João, 43
Dias, José Gonçalves, 81, 111
Dias, Otília, 59
Dourado, Luís, 119
Duarte, Sónia, 69
Engana, Teresa, 97
Faria, Bernardo, 71
Faria, Susana, 57, 117
Ferreira, Ana Sousa, 121
Ferreira, Fernanda A., 53
Figueiredo, Adelaide, 45
Figueiredo, Esperança, 27
Figueiredo, Fernanda, 17, 43, 45
Flores, Isabel, 39
Freitas, Mário, 119
Freitas, Rita, 43, 67
Garcia, C., 89
Góis, Eduarda, 27, 29
Gomes, A. Rui, 117
Gomes, Rui, 113
Gonçalves, A. Manuela, 95, 113
Grilo, Helena L., 41
Grilo, Luís M., 41
Guerreiro, José, 109
Henriques, Carla, 69, 71
Infante, Paulo, 43, 67, 91, 97, 99
Jacinto, Gonçalo, 43, 97
Lima, Rosário, 121
Loura, Luísa Canto e Castro, 35
Maciel, Andreia, 67
Maia, André, 73
Maia, Pedro, 51
Marques, Catarina, 79
Martins, C., 89
Martins, João, 83

Martins, M. J. , 107	Salgueiro, Maria de Fátima , 87
Matos, Ana Cristina Bico Matos , 71	Santos, Carla , 105, 115
McCall, Michael , 119	Saraiva, S. , 89
Medeiros, Luís , 61	Sardinha, Boguslawa , 59
Mendes, M. , 107	Silva, Ana Lorga da , 103
Mendes, Maria Filomena , 63, 65, 67, 99	Silva, Isabel , 85
Mendes, Zilda , 109	Silva, Joana , 95
Merca, Luís , 83	Silva, Maria Eduarda , 85
Mexia, João Tiago , 105, 115	Silva, O , 107
Neves, Cristina , 23	Silva, Pedro , 97
Nogueira, Isabel M. , 41	Soares, Carlos , 73
Nunes, Célia , 105, 115	Sousa, Áurea , 101
Oliveira, I. , 89	Sousa, Beatriz , 61
Oliveira, M. Rosário , 55	Sousa, Fernanda , 77
Parreira, Artur , 103	Sousa, Olívia , 37
Pereira, Raquel , 59	Teixeira, Mário Sérgio , 111
Pestana, Manuel , 71	Tomé, Lídia Patrícia , 63
Pestana, Pilar , 91	Torre, Carla , 109
Pires, Ana M. , 49	Torres, Cristina , 85
Pisco, Rita , 31	Torres, Sónia , 25
Portela, Conceição , 57	Ventanas, J. , 89
Preto, Isabel , 119	Vicente, Jorge , 97
Reis, Elizabeth , 79	Vicente, Paula , 79
Ribeiro, Filipe , 65	Vicente, Paula C. R. , 87
Ribeiro, Rui M. , 15	Vieira, Jéssica , 61
Rocha, Cristina , 11	Vilela, Margarida , 55, 61