BOOK OF ABSTRACTS

Universidade Aberta
Lisbon - PORTUGAL
June 3–5, 2019

Committees

Executive Committee

Christos Kitsos, University of West Attica, Athens, Greece
M. Ivette Gomes, Faculdade de Ciências, Universidade de Lisboa, Portugal
Stanislaw Mejza, Poznan University of Life Sciences, Poland
Teresa A. Oliveira, Universidade Aberta and CEAUL, Portugal

Scientific Committee

Amílcar Oliveira, Universidade Aberta and CEAUL, Portugal
António Araújo, Universidade Aberta, Portugal
Arminda Manuela Gonçalves, University of Minho, Portugal
Carla Santos, Politechnic Institute of Beja, Portugal
Carlos Agra Coelho, Nova University of Lisbon, Portugal
Carlos Braumann, Universidade de Évora, Portugal
Célia Nunes, University of Beira Interior, Portugal
Christos Kitsos, University of West Attika, Greece
Christos Skiadas, Technological University of Crete, Greece
Cristina Dias, Politechnic Institute of Portalegre, Portugal
Dário Ferreira, University of Beira Interior, Portugal
Dora Prata Gomes, Nova University of Lisbon, Portugal
Elisa Henning, University of Estado de Santa Catarina, UDESC, Brazil
Feliz Minhós, University of Évora, Portugal
Fernanda Figueiredo, University of Porto, Portugal
Fernando Carapau, University of Évora, Portugal
Filomena Teodoro, Naval School, Portugal
Francisco Carvalho, Instituto Politécnico de Tomar, Portugal
Frederico Caeiro, Nova University of Lisbon, Portugal
Iwona Mejza, Poznan University of Life Sciences, Poland
James R. Bozeman, American University of Malta, Malta
João Tiago Mexia, Nova University of Lisboa, Portugal
José António Macias-Seijas, University of Corunha, Spain
José António Pereira, University of Porto, Portugal
Karl-Moder, University of Natural Resources and Life Sciences, Austria
Lidia Filus, Northeastern Illinois University, USA
Luís Cavique, Universidade Aberta, Portugal
Luís Grilo, Politechnic Institute of Tomar, Portugal
Manuela Neves, University of Lisbon , Portugal
Manuel L. Esquível, Nova University of Lisbon, Portugal

Maria do Rosário Ramos, Universidade Aberta, Portugal
Marina A. P. Andrade, ISCTE-IUL/ISTAR, Portugal
M. Ivette Gomes, University of Lisbon, Portugal
Nuno Brites, Instituto Superior de Economia e Gestão, Universidade de Lisboa, Portugal
Sandra Ferreira, University of Beira Interior, Portugal
Sandra Nunes, Politechnic Institute of Setúbal, Portugal
Stanislaw Mejza, Poznan University of Life Sciences, Poland
Teresa A. Oliveira, Universidade Aberta and CEAUL, Portugal
Victor Leiva, Pontificia Universidad Católica de Valparaíso, Chile

Organizing Committee

Amílcar Oliveira, Universidade Aberta and CEAUL, Portugal
Dário Ferreira, University of Beira Interior, Portugal
Fernando Carapau, University of Évora, Portugal
Maria do Rosário Ramos, Universidade Aberta, Portugal
Sandra Ferreira, University of Beira Interior, Portugal
Teresa A. Oliveira, Universidade Aberta and CEAUL, Portugal
Virgínia Zaidam, Universidade Aberta, Portugal

Sponsored by

Title
13th Conference on Statistics, Mathematics and Computation – Book of Abstracts

Editor
Universidade Aberta
Rua da Escola Politécnica, 147
1269-001 Lisboa

Authors
Amílcar Oliveira, Universidade Aberta and CEAUL, Portugal
Dário Ferreira, University of Beira Interior, Portugal
Sandra Ferreira, University of Beira Interior, Portugal
Teresa A. Oliveira, Universidade Aberta and CEAUL, Portugal

ISBN: 978-972-674-850-2

Preface

Dear participants and friends,

WELCOME to the 13th International Conference on Statistics, Mathematics and Computation!

We celebrate this meeting in Honour of Professor João Tiago Mexia, for his brilliant career and outstanding contribution to the field of Statistics in Portugal and abroad, and to whom we are deeply grateful for all the support, kindness and precious collaboration in our meetings!

We are delighted to have this celebration, fostering strong interaction between national and international researchers, leading to a successful commitment and enthusiasm on promoting research in and between the broad areas of Statistics, Mathematics and Computation.

It is a great pleasure to receive all our guests and contributors from 3 to 5th June in the Aberta University, and we emphasize our huge gratitude to the Magnificent Rector, for kindly accepting to embrace this challenge of hosting once more this International Conference!

The WSMC was successfully organized in several places along the 13 editions and we believe that our meetings have been creating very nice opportunities for showcasing the growth and development of the focused main areas, in a time underlined by so many new technologies and huge emerging challenges. The scientific exchanges of ideas between participants is always helpful for generating positive impact on propelling the advancement of science and technology and some of these results have already been successfully published in high standard Journals, special issues and Springer Series Books.

Also now, in this 13th edition, the participants have the opportunity to submit their work to high referenced Journals and Book volumes. Selected papers, after peer review, will appear in the Chilean Journal of Statistics (ChJS), Journal of Applied Statistics (JAS), Biometrical Letters, Biometrics & Biostatistics International Journal (BBIJ) and in Springer volumes of the Series Contributions to Statistics. Details will be soon announced in our webpage.

We are highly grateful to all the participants, Invited Speakers, members of Scientific Committee, Session Organizers and Authors who submitted ab-

stracts, for their valuable contribution and for the enthusiastic way how their participation is assumed!

We greatly acknowledge all the sponsors and contributors who made this meeting a reality and we reinforce the importance of our social programme. This year it includes the nice opportunity to visit the National Museum of Natural History and Science and the Botanical Garden, allowed all the week, from 3 to 9 of June - during and after WSMC13. We are really grateful to the Director of the Department of Museums Management Support and IICT, Dr Vitor Lucas and to his team, for kindly offering this outstanding opportunity to all the participants of WSMC13.

A deeply thanks is addressed to all the members of the Organizing Committee, emphasizing the invaluable collaboration of Dr. Virginia Zaidam, the crucial help of the Gabinete de Comunicação e de Relações Internacionais-UAB team, and the very hard work of Amílcar, Sandra and Dário on the organization and supervision of many of the programme and schedule details, preparing the website and the abstracts book in an incredible record time!

Finally, as I guess most of you already noticed, Lisbon is very much enjoyable for offering a rich and varied history, a buzzing nightlife and for being blessed with a glorious year-round climate, where the sky joins the ocean and the city is kissed by the river Tagus. Please take some of your time to enjoy and take profit of it! Don't forget to taste the famous "Pastéis de Belém" and to listen to Fado music!

We wish you all a very productive and successfull meeting, as well as a pleasant and memorable stay in Lisbon, the stunning capital of Portugal!

By the Executive Committee,

Teresa A. Oliveira

Lisbon, 3-5th June 2019

Programme

	June 3^{rd}, 2019 (Monday)
16:00 - 18:30	Registration (**Hall**)
17:00 - 18:00	Formação PORDATA (in Portuguese) (Room: **Auditório**)
18:00 - 19:00	Welcome Reception

	June 4^{th}, 2019 (Tuesday)
09:00 - 15:00	Registration (**Hall**)
09:30 - 10:00	Opening Session (Room: **Auditório**)
10:00 - 10:30	Keynote Session: João Tiago Mexia (Room: **Auditório**) Title: **Estimation for Additive Models: LSE and GLSE**
10:30 - 11:00	Coffee Break
11:00 - 12:00	Organized Session

Organized Session S2; Room: **Auditório**
Chair: Filomena Teodoro
Title: Computational Mathematics, Statistics and Information Management - Part I

Analysis of SMEs Financial Distress Probability: the case of Tâmega and Sousa Region, Ana Borges
Portuguese SME's internationalization and financial performance, Aldina Correia
Conditions for entrepreneurship of municipalities, Eliana Costa e Silva
Modeling KAP about Zika Virus, M. Filomena Teodoro

12:00 - 13:00	Invited Session, (Room: **Auditório**) **High Performance Computing + High Performance Data Analytics - Advancing science with the oblivion supercomputer@U._Evora - ENGAGE SKA**, Miguel de Avillez
13:00 - 14:30	Lunch
14:30 - 15:30	Invited Sessions (I1 & I2), Room: **Auditório**
14:30 - 15:00	Invited Session I1 **On the Quadratic Logit Model**, Christos P. Kitsos
15:00 - 15:30	Invited Session I2 **Tail dependence coefficient semi-parametric estimation through a generalized mean**, M. Ivette Gomes

15:30 - 16:30	Organized Sessions

Organized Session S6; Room: **Auditório** Chair: Teresa Oliveira Title: Statistical Models and Applications	Organized Session S3; Room: **Sala 8** Chair: M. Filomena Teodoro Title: Computational Mathematics, Statistics and Information Management - Part II
Individual growth mixed models with stochastic differential equations, Carlos Braumann	**Impact of using excellence management models in the customer Satisfaction of Brazilian electricity, Distributors - 10 years of studies**, Marina A.P. Andrade
The Extended Skew-Normal Distribution and the Product of two Normal Variables, José António Seijas-Macias	**Daily Load Profile of Portuguese Electricity Demand using Kohonen Self-Organizing Maps**, Ricardo Moura
Structural Equation Modelling - some applications with R tools, Conceição Leal	**Reassessment of Periodontal Inflammation and High Density Lipoproteins Relationship through GAMLSS models**, José António Pereira
Development of a collaborative game to assess statistical literacy: an adventure in medieval times, Elisa Henning	

16:30 - 17:00	Coffee Break and Poster Session I

17:00 - 18:00	Invited Sessions (I3 & I4), Room: **Auditório**
17:00 - 17:30	Invited Session I3 **On single-factor experiments with repeated measures**, Stanislaw Mejza
17:30 - 18:00	Invited Session I4 **Effects of Different Higher Moments on Type One Error Rate in ANOVA**, Karl Moder

	June 5^{th}, 2019 (Wednesday)
09:30 - 12:30	Registration (**Hall**)
09:30 - 10:30	Invited Sessions (I5 & I6), Room: **Auditório**
09:30 - 10:00	Invited Session I5 **Testing the "complete symmetrical equivalence" of two sets of variables**, Carlos Agra Coelho
10:00 - 10:30	Invited Session I6 **On the use of the LINEX loss function in lifetime models when dealing with progressively type-II censored samples**, Inmaculada Barranco-Chamorro

10:30 - 11:00 Coffee Break and Poster Session II

11:00 - 12:00 Invited Sessions (I7 & I8), Room: **Auditório**

11:00 - 11:30 Invited Session I7
Samples with random dimensions in fixed effects ANOVA, Célia Nunes

11:30 - 12:00 Invited Session I8
An application of SEM in Social and Health Sciences using different estimation methods, Luís Grilo

12:00 - 13:00 Organized Sessions

Organized Session S1; Room: **Auditório**
Chair: Ayana Mateus
Title: Statistical Methods and Applications

Inference for the Growth Curve Model with Dual Orthogonal Covariance Structure, Miguel Fonseca

Multivariate collective risk models Inference and special case, Elsa Moreira

Unbalanced Full Factorial and Fractional Factorial Designs Associated with a Commutative Jordan Algebra, Sandra Oliveira

Revisiting the calibration procedure in the extremal index estimation, Dora Prata Gomes

A generalized jacknife estimator of a negative extreme value índex, Frederico Caeiro

Organized Session S4; Room: **Sala 8**
Chair: M. Rosário Ramos
Title: Times Series Modelling and Applications

Modeling monthly temperature data in European Cities, Marco Costa

Change point detection to model recent trends in European cities temperature time series, Magda Monteiro

Trend analysis of climate variables: the example of the Sea Ice Index, M. Rosário Ramos

13:00 - 14:30 Lunch

14:30 - 15:30 Organized Session

Organized Session S4; Room: **Auditório**
Chair: Miguel Fonseca
Title: Advances in Applications of Inference Methodologies

Modelling the dominant height growth of eucalyptus plantations in Portugal, Ayana Mateus
Multiple Additive Models, Sandra Ferreira
Mixed models with independent terms, Dário Ferreira
From a napkin to a most expected book, Ricardo Covas

15:30 - 16:30 Invited Sessions (I9 & I10), Room: **Auditório**

15:30 - 16:00 Invited Session I9
Defining a Stochastic Controller based on a Fuzzy Controller to adjust the mutation probabilities of a genetic algorithm, André Gustavo Pereira

16:00 - 16:30 Invited Session I10
Universal Representation of Bivariate Survival Functions, Lidia Filus

16:30 - 17:00 Coffee Break

17:00 Closing, Room: **Auditório**

Poster Session I

P1 - **Harvesting policies in random environment: logistic growth and profit optimization**, Nuno M. Brites
P2 - **A note on models with commutative orthogonal block structure**, Carla Santos
P3 - **A Time series analysis of road traffic accidents in Luanda**, Patrícia A. Filipe
P4 - **Three-dimensional velocity field for blood flow using the power-law viscosity function**, Fernando Carapau
P5 - **Extinction times for the logistic model with and without Allee effects**, Clara Carlos (jointly with Carlos A. Braumann and Nuno M. brites)
P6 - **Robust Portfolio Selection**, Mickael Borges
P7 - **Data modeling from an experiment with barley malt and some results**, Iwona Mejza (jointly with Katarzyna Ambroży-Deręgowska)

Poster Session II

P8 - **Time Series Analysis for Maximum Wind Speed Forecasting**, Arminda Manuela Gonçalves (jointly with Cláudia Costa and Marco Costa)
P9 - **Statistics Analysis in the Management of a Local Public Company**, Arminda Manuela Gonçalves
P10 - **Degree one models for symmetric Stochastic matrices**, Cristina Dias (jointly with Carla Santos and João Tiago Mexia)
P11 - **Numerical Approach of Mixed Functional Differential Equations: Some Applications**, M. Filomena Teodoro
P12 - **Multinomial Distribution and Discriminant Analysis: An application to discrete variables**, Isaac Akoto (jointly with Dário Ferreira, Gracinda Guerreiro, Sandra Ferreira and João T. Mexia)
P13 - **Cumulant Generating Functions: a simulation study**, Patrícia Antunes (jointly with Sandra Ferreira, Célia Nunes, Dário Ferreira and João T. Mexia)
P14 - **The Potential of Logistic Regression**, Sandra Nunes (jointly with Dina Salvador and Sandra Monteiro)

Contents

Keynote Speaker

Estimation for Additive Models: LSE and GLSE

João T. Mexia[1]

[1] *Center of Mathematics and its Applications, Faculty of Science and Technology, New University of Lisbon, Monte da Caparica, Portugal*

Abstract

The model

$$\boldsymbol{Y} = \mathbf{X}_0\beta_0 + \sum_{l=1}^{w} \mathbf{X}_l\mathbf{U}_l,$$

where β_0 is fixed and the $\mathbf{U}_1, ..., \mathbf{U}_w$ with null mean vectors and variance covariance matrices $\sigma_l^2\mathbf{I}_{c_l}, l = 1, ..., w$, has covariance matrix $\sum_{l=1}^{w} \sigma_l^2\mathbf{M}_l$, with $\mathbf{M}_l = \mathbf{X}_l\mathbf{X}_l^\top, l = 1, ..., w$. Thus this model is additive.
For these models are obtained LSE estimators for the first four order cumulants of the components of the vectors $\mathbf{U}_1, ..., \mathbf{U}_w$ and for the coefficients vectors β_0. Moreover GLSE are obtained for β_0 and the variance components $\sigma_l^2, l = 1, ..., w$, which will be the second cumulants of the components of $\mathbf{U}_1, ..., \mathbf{U}_w$.

Keywords: additive models, cumulants, GLSE, LSE, moments.

Acknowledgements

This work was partially supported by national founds of FCT-Foundation for Science and Technology under UID/MAT/00297/2019.

References

[1]Akhiezer, N. I. (1965), The Classical Moment Problem, N. Kemme, Transl., Hafner, New York.
[2]Balakrishnan, N., Johnson,N. L. and Kotz, S. (1998), A note on relationships between moments, central moments and cumulants from multivariate distributions, Statistics & probability letters, **39**: 49-âĂŤ54, Elsevier.
[3]Dodge Y. and Valentin R. (1999), The Complications of the Fourth Central Moment, The American Statistician, **53** (3): 267–269.
[4]Harvey, J. R., (1972), On Expressing Moments in Terms of Cumulants and Vice Versa. The American Statistician, **26** (4): 38–39.
[5]Kendall, M. and Stuart, A. (1958). The Advanced Theory of Statistics. Charles Griffin & Company.
[6]Landau, H. J. (1987), Moments in Mathematics, Proceedings of Symposia in Applied Mathematics, **37**, Amer. Math. Soc., Providence, RI.
[7]Ruppert, D. (1987) What is Kurtosis? An Influence Function Approach, The American Statistician, **41**: 1–5.
[8]Schultze, J. and Steinebach, J. (1996). On least squares estimates of an exponential tail coefficient. Statistics & Decisions, **14**, 353–372.
[9]Smith, P. J. (1995), A Recursive Formulation of the Old Problem of Obtaining Moments from Cumulants and Vice Versa. The American Statistician, **49** (2): 217–218.

Invited Speakers

Testing the "complete symmetrical equivalence" of two sets of variables

Carlos A. Coelho[1] **and Barry C. Arnold**[2]

[1] *Mathematics Department and Center for Mathematics and Applications (CMA), Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Portugal*

[2] *Statistics Department, University of California Riverside, CA, U.S.A.*

E-mail address: *cmac@fct.unl.pt; barnold@ucr.edu*

Abstract

Let $\underline{X} = [\underline{X}_1', \underline{X}_2']' \sim N_{2p}(\underline{\mu}, \Sigma)$, where both subvectors $\underline{X}_1$ and $\underline{X}_2$ are p-dimensional, and

$$\underline{\mu} = \begin{bmatrix} \underline{\mu}_1 \\ \underline{\mu}_2 \end{bmatrix} \quad \text{with} \quad \underline{\mu}_1 = E(\underline{X}_1) \quad \text{and} \quad \underline{\mu}_2 = E(\underline{X}_2)\,.$$

We will say that $\underline{X}_1$ and $\underline{X}_2$ are "completely symmetrically equivalent" if

$$\Sigma = \begin{bmatrix} \Sigma_1 & \Sigma_2 \\ \Sigma_2 & \Sigma_1 \end{bmatrix} \quad \text{and} \quad \underline{\mu}_1 = \underline{\mu}_2\,, \tag{1}$$

where Σ_1 and Σ_2 are non-specified, but with Σ_1, $\Sigma_1 + \Sigma_2$ and $\Sigma_1 - \Sigma_2$ being positive-definite matrices. We will be interested in testing the null hypothesis in (1), which we will call the "complete symmetrical equivalence" hypothesis.

The authors derive the likelihood ratio test statistic for the present test and show how it is possible to include the distribution of this statistic under the framework of Theorem 3.2 in [1] and as such have its exact probability density and distribution functions given by Corollary 4.2 in the same reference, in a finite closed form, through those of an EGIG (Exponentiated Generalized Integer Gamma) distribution, with all parameters precisely defined, as simple functions of p and of the sample size. An example of the implementation of the test, using real data, is presented.

Keywords: exact distribution, finite closed form, Meijer G function, likelihood ratio test, product of Beta random variables.

Acknowledgements

This research was partially supported by Portuguese National Funds through Fundação para a Ciência e a Tecnologia, through Project UID/MAT/00297/2019 (CMA/UNL – Centro de Matemática e Aplicações, Universidade Nova de Lisboa).

References

[1]Coelho, C. A., Arnold, B. C. (2019) Finite form representations for Meijer G and Fox H functions – Applied to multivariate likelihood tests using Mathematica®, Maxima and R. Springer Texts in Statistics (accepted for publication).

Defining a Stochastic Controller based on a Fuzzy Controller to adjust the mutation probabilities of a genetic algorithm

Andre G. C. Pereira[1] **and Viviane S.M. Campos**[1]

[1] *Mathematics Department, Universidade Federal do Rio Grande do Norte, Natal/RN, Brazil*

E-mail address: *andre@ccet.ufrn.br; viviane@ccet.ufrn.br*

Abstract

In the theory of fuzzy sets, fuzzy controllers are tools that were developed and have been used since then in a wide amount of practical applications. Controllers are devices that use an output signal of a system to make adjustment in the input signal so as to keep the system stable. Fuzzy controllers use membership functions, a base of rule and an inference method that together determine the way the controller works. The membership functions describe the opinion of the experts in the subject. The domain of these membership functions are divided into sets where the experts agree and into those sets where they do not agree. In this paper we introduce a hybrid inference system, which uses the Mandani system in the set where the experts agree and uses a beta distribution to model the response of the controller in the set where the experts do not agree. We compare the performance of both controllers, the pure fuzzy controller and the hybrid one, called stochastic controller, when they are used to adjust the mutation probability of a genetic algorithm which is run to find the optimum point of some test functions. Many recent papers have found that maintaining high diversity in the population of a Genetic Algorithms improve convergence rate to global optimum. In this paper we consider two different functions in order to keep a population with high diversity while the space of possible solutions is searched through. The values of the functions used to calculate the diversity of the current population are used as the input variable of a fuzzy/stochastic controller which adjusts the mutation probability of an elitist non-homogeneous genetic algorithm. We also present numeric simulations to illustrate that the performance of the stochastic controller is statistically as good as that of the best fuzzy controller in all examples performed.

Keywords: stochastic controller, fuzzy controller, genetic algorithms.

Samples with random dimensions in fixed effects ANOVA

Célia Nunes[1,2]

Department of Mathematics, University of Beira Interior, Covilhã, Portugal[1]
Center of Mathematics and Applications,University of Beira Interior, Covilhã, Portugal[2]

E-mail address: *celian@ubi.pt*

Abstract

Analysis of variance (ANOVA) is one of the most commonly used statistical methods in practical applications. It is often used in several research areas, namely in medical research, social sciences or agriculture, to name just a few. Despite its wide use, it has been applied assuming that sample dimensions are known. The aim of this talk is to present an extension of the theory of fixed effects ANOVA to situations where the samples dimensions are not known in advance. An illustrative example of this, is the collection of observations during a fixed time period in a study comparing, for example, several pathologies of patients arriving at a hospital. In these cases it is more appropriate to consider the sample sizes, $n_1, ..., n_m$, as realizations of independent random variables, $N_1, ..., N_m$. This approach must be based on an adequate choice of the distributions of $N_1, ..., N_m$. We assume the Poisson distribution when the occurrence of observations corresponds to a counting process. The Binomial distribution is the proper choice if we have observations failures and there exist an upper bound for the sample sizes. We conclude this talk by presenting an application on real medical data, considering cancer registries, and we carry out a simulation study to evaluate the proposed method.

Keywords: Fixed effects models, random sample sizes, Poisson distribution, Binomial distribution, cancer registries.

Acknowledgements

This work was partially supported by the Portuguese Foundation for Science and Technology through the project UID/MAT/00212/2019.

References

[1]Mexia, J.T., Nunes, C., Ferreira, D., Ferreira, S.S. and Moreira, E. (2011) Orthogonal fixed effects ANOVA with random sample sizes. Proceedings of the 5th International Conference on Applied Mathematics, Simulation, Modelling (ASM'11), 84-90.

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

[3]Nunes, C., Ferreira, D., Ferreira, S.S. and Mexia, J.T. (2012) F-tests with a rare pathology. *Journal of Applied Statistics* **39**(3): 551-561.

[4]Nunes, C., Capistrano, G., Ferreira, D. and Ferreira, S.S. (2103) ANOVA with Random Sample Sizes: An Application to a Brazilian Database on Cancer Registries. 11th International Conference on Numerical Analysis and Applied Mathematics. AIP Conf. Proc. **1558**: 825-828.

[5]Nunes, C., Ferreira, D., Ferreira, S.S. and Mexia, J.T. (2014) Fixed effects ANOVA: an extension to samples with random size. *Journal of Statistical Computation and Simulation* **84**(11): 2316-2328.

[6]Nunes, C., Capistrano, G., Ferreira, D. and Ferreira, S.S. and Mexia, J.T.(2019) Exact critical values for one-way fixed effects models with random sample sizes. *Journal of Computational and Applied Mathematics* **354**: 112-122.

[7]Scheffé, H. (1959) *The analysis of variance*. Wiley series in Probability and statistics, John Wiley & Sons, New York.

[8]Brazilian National Cancer Institute (INCA) (2010) http://www2.inca.gov.br/wps/wcm/connect/inca/portal/home, (accessed 15.09.05).

On the Quadratic Logit Model

Christos P. Kitsos[1]

[1] *University of West Attica*

E-mail address: *xkitsos@uniwa.gr*

Abstract

The target of this paper is to investigate the Logit model approximated by a second order linear model, i.e

$$LogOdds = p_2(x)$$

with

$$Odds = p/(1-p), \quad p_2(x) = a + bx + cx^2$$

The model is discussed for the point of view of:

1. Design o Experiments, i.e. can we construct an optimal experimental design?

2. The appropriate investigation for the max or min of $p(x)$ i.e can we obtain a 95% Confidence interval for it?

3. Computation difficulties involved on the existence of MLE i.e does MLE always exist?

Universal Representation of Bivariate Survival Functions

Lidia Filus[1,2] **and Jerzy Filus**[3,4]

[1] *Northeastern Illinois University, Chicago, IL 60625, USA*
[2] *Department of Mathematics*
[3] *Oakton Community College, Des Plaines, IL 60016, USA*
[4] *Department of Mathematics and Computer Science*

E-mail address: *L-Filus@neiu.edu; jkfilus98@gmail.com*

Abstract

We present a copula alternative method for construction of any bivariate survival function given marginals. The joint survival functions are given in a product form which is universal so each bivariate model can be expressed in this form. The methods of construction are discussed.

Keywords: bivariate survival function representation, copula, dependence function (joiner).

References

[1]Aalen, O. O. (1989) A Linear Model for the Analysis of the Life Times. Statistics in Medicine 8, 907–925.

[2]Filus, J. K. and Filus, L. Z. (2017) The Cox-Aelen Models as Framework for Construction of Bivariate Probabilty Distributions, Universal Representation. Journal of Statistical Science and Applications 6, 56–63.

[3]Filus, J. K. and Filus, L. Z. (2017) General Method for Construction of Bivariate Stochastic Processes. Presentation at the 7th International Conference on Risk Analysis ICRA7, Chicago, IL,

[4]Gumbel E. J. (1960) Bivariate Exponential Distributions. Journal of the American Statistical Association 55, 698–707

[5]Sklar A. (1959) Funtions de repetition a n dimensions et leurs marges. Publications de l'Institut de Statistique de l"universitete de Paris 8, 229–231.

An application of SEM in Social and Health Sciences using different estimation methods

Luís M. Grilo[1,2]

[1] *Instituto Politécnico de Tomar (IPT), Tomar, Portugal.*
[2] *Centro de Matemática e Aplicações (CMA), FCT, UNL, Portugal.*

E-mail address: *cmac@fct.unl.pt; barnold@ucr.edu*

Abstract

Structural equation modeling (SEM) has become very used in the fields of social and health sciences, namely because it allows to test entire theories and concepts. SEM's have the ability to assess latent variables/constructs at the observation level (measurement or outer model) and test relationships between constructs on the theoretical level (structural or inner model) [2]. In this study, related to the workers' health and wellbeing, the data was collected by a validated Psychosocial Questionnaire, with ordinal variables expressed in a 5 point Likert-type scale (this scale was perceived as symmetric and equidistant and as considered approximated to an interval level measurement). A theoretical SEM was proposed and a model was estimated using two different estimators [1]: the MLE - Maximum Likelihood Estimation (Covariance Based - SEM) and PLS - Partial Least Squares (Variance Based - SEM), that does not impose any distributional assumptions. In SEM the covariance-and-variance based (CB-and-VB) estimators share the same roots, but CB-SEM methods are preferred if the model contains constructs modeled as common factors and VB SEM estimators are favoured if the underlying model consists of constructs modeled as composites [3, 4]. The latest consistent PLS (PLSc) version, which corrects for bias to consistently estimate SEM's with common factors, was applied and the results show strong similarities with MLE as expected, since PLSc was designed to mimic MLE for studies that focus on analyzing and testing the model structure.

Keywords: Maximum Likelihood Estimation, Partial Least Squares, common factors, survey.

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/00297/2019 (Centro de Matemática e Aplicações).

References

[1]Grilo, L. M. and Grilo, H. L. (2019). PLS-SEM to evaluate the impact of stress and work family conflict on the burnout syndrome. 8th International Conference on Risk Analysis and Design of Experiments (ICRA8), Institute of Applied Statistics and Computing, University of Natural Resources and Life Sciences, Vienna, Austria, April 23–26 (p. 38, Book of Abstracts).

[2]Hair, J. F., Hult, G. T. M., Ringle, C. M. and Sarstedt, M. (2017). A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM) (2nd Ed., Thousand Oakes, CA: Sage).

[3]Henseler, J. (2018). Partial least squares path modeling: Quo Vadis. Qual Quant 52: 1–8.

[4]Schuberth, F., Henseler, J. and Dijkstra, T. K. (2018). Partial least squares path modeling using ordinal categorical indicators. Qual Quant 52: 9–35.

Tail dependence coefficient semi-parametric estimation through a generalized mean

M. Ivette Gomes[1,2]

DEIO, Faculdade de Ciências, Universidade de Lisboa, Portugal[1]
Centro de Estatística e Aplicações da Universidade de Lisboa (CEA/UL)[2]

E-mail address: *ivette.gomes@fc.ul.pt*

Abstract

A reliable risk assessment of the amount of tail dependence is important in the most diverse fields, like climatology, environment, finance and insurance, among others. A crucial issue that appears when there is more than one variable is that of dealing with dependence. The study of multivariate extremes can indeed be splited essentially in two parts, the marginal distributions and the dependence structure. The margins are nowadays easy to deal with through univariate *extreme value theory* (EVT). The crucial parameter in univariate EVT is the EV *index* (EVI), which measures the heaviness of the *right-tail function* (RTF) of the model underlying the data. To deal with dependence, also often univariate EV techniques can be used, after an adequate transformation of the available mutivariate data sample. This happens for instance with the estimation of the tail dependence coefficient (TDC). Indeed, and thinking on a bivariate framework, (X,Y), after the standardization of the margins to a simple unit Fréchet distribution, which can be performed for instance by their empirical cumulative distribution function combined with the inverse probability integral transform, the TDC appears as the reciprocal of the regularly varying exponent of a Pareto-type RTF for the differences between the margins. Let us assume that $(X_i, Y_i), 1 \le i \le n$, are independent copies of the random vector (X, Y). In a large variety of cases, setting $Z_i = \min(X_i, Y_i)$, we have

$$\mathbb{P}(Z_i > z) = \mathbb{P}(X_i > z, Y_i > z) = z^{-1/\eta}\mathcal{L}(z),$$

with $\mathcal{L}(\cdot)$ a slowly varying function at infinity, i.e. an ultimately positive function satisfying $\mathcal{L}(tx)/\mathcal{L}(t) \to 1$, for all $x > 0$, as $t \uparrow +\infty$. Thus the TDC, η, can be viewed as the EVI of the minimum of the two components, and hence all the types of EVI-estimators, including *bias-corrected* (BC) EVI-estimators, can be used. BC EVI-estimators were introduced in pioneering

papers by Gomes (1995, 1999), Feuerverger and Hall (1999), Beirlant *et al.* (1999) and Gomes *et al.* (2000), and can be recently found in Caeiro *et al.* (2016), among others. For a recent overview of the topic, see Gomes and Guillou (2015). Such procedures have been considered for the bivariate case in Beirlant and Vandewalle (2002), using an exponential regression model, and in Beirlant *et al.* (2011) under an adequate condition on the slowly varying function $\mathcal{L}$. Goegebeuer and Guillou (2013) reconsidered the bias-issue when estimating the TDC, η, on the basis of a weighted functional estimator for η. These BC TDC-estimators have explicit expressions, making them computationally inexpensive. The nice behaviour of the recently introduced *generalized means* (GMs) EVI-estimators (see Brilhante *et al.*, 2013, among others), also computationally inexpensive, leads us now to suggest the use of the also explicit GMs or even BCGMs, for a reliable estimation of the TDC.

Keywords: semi-parametric estimation, statistics of extremes; tales of the tails.

Acknowledgements

Research partially supported by National Funds through **FCT** — Fundação para a Ciência e a Tecnologia, project UID/MAT/00006/2019 (CEAU/UL)).

References

[1]Beirlant, J., and Vandewalle, B. (2002). Some comments on the estimation of a dependence index in bivariate extreme value in statistics. *Statist. Probab. Lett.* **60**: 265–278.

[2]Beirlant, J., Dierckx, G., Goegebeur, Y., and Matthys, G. (1999). Tail index estimation and an exponential regression model. *Extremes* **2**: 177–200.

[3]Beirlant, J., Dierckx, G., and Guillou, A. (2011). Bias-reduced estimators for bivariate tail modelling. *Insurance Math. Econom.* **49**: 18–26.

[4]Brilhante, M.F., Gomes, M.I., and Pestana, D. (2013). A simple generalization of the Hill estimator. *Computat. Statistics and Data Analysis* **57**(1): 518–535.

[5]Caeiro, F., Gomes, M.I., Beirlant, J., & Wet, T. de (2016). Mean-of-order-p reduced-bias extreme value index estimation under a third-order framework. *Extremes* **19**(4): 561–589.

[6]Feuerverger, A., and Hall, P. (1999). Estimating a tail exponent by modelling departure from a Pareto distribution. *Ann. Statist.* **27**: 760–781.

[7]Goegebeur, Y., and Guillou, A. (2013). Asymptotically unbiased estimation of the coefficient of tail dependence. *Scand. J. of Statistics* **40**(1): 174–189.

[8]Gomes, M.I. (1995). Metodologias jackknife e bootstrap em Estatística de Extremos. *Actas da II Conferência Anual da* S.P.E.: 31–46 (in Portuguese).

[9]Gomes, M.I. (1999). The jackknife and the bootstrap methodologies in the estimation of parameters of rare events. *Revista de Estatística* **99**: 5–23.

[10]Gomes, M.I., & Guillou, A. (2015). Extreme value theory and statistics of univariate extremes: A review. *Intern. Statistical Review* **83**(2): 263–292.

[11]Gomes, M.I., Martins, M.J., and Neves, M.M. (2000). Alternatives to a semi-parametric estimator of parameters of rare events—the Jackknife methodology. *Extremes* **3**(3): 207–229.

Effects of Different Higher Moments on Type One Error Rate in ANOVA.

Karl Moder[1]

University of Natural Resources and Life Sciences[1]
Institut of Statistics
Peter-Jordan-Str. 82, 1190 Vienna, Austria

E-mail address: *karl.moder@boku.ac.at*

Abstract

F-test in analysis of variance is based on three assumptions: normal distributed random variables, independent samples and homogeneity of variances. Deviations from normal distribution are regarded to have almost no effects on type one error rate (Box (1954b), Keppel and Wickens (2004)). Dependencies can be eliminated by means of an appropriate model. Heterogeneity however is regarded to be a serious problem (Keppel et al. (1992); Keppel and Wickens (2004); Moder (2007); Rasch et al. (2009)). To handle this problem various solutions have been proposed. They are either approximative or based on permutations of the samples. Here a method is presented which allows an exact test with inhomogeneous variances (Moder (2010)).

Moments of degree 3 or 4 are usually not really considered in terms of their impact on the performance of the F-test in ANOVA. If these moments are significant, non-parametric tests (such as the Kruskal-Wallis test) are often used.

Various tests (F-test Analysis of Variance, Welch-Test for more than 2 samples, variance weighted ANOVA, Kruskal-Wallis-test, Permutation test using F- statistic or Kruskal-Wallis statistic, and a special kind of Hotelling's T 2 test (Moder, 2007)) were investigated by hand of a simulation study to investigate the influence of higher moments on Type I error rate.

Keywords: higher moments , type I error rate, exact test, approximative test, simulation study

References

[1]Box, G. E. P., (1954b) Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems, II. Effect of Inequality of Variances and of Correlation of Errors in the Two-Way Classification. *Annals of Mathematical Statistics* **25 (3)**: 484–498.

[2]Keppel, G.; Saufley, W. H.; and Tokunaga H.Zedeck S. (1992) Introduction to design and analysis: A students handbook. (2nd ed.), New York: W. H. Freemann.

[3]Keppel, G.; Wickens, T. D. (2004) Design and Analysis: A researcher's handbook. (4th ed.) Upper Saddle River, NJ: Pearson Prentice Hall.

[4]Moder, K. (2007) How to keep the Type I Error Rate in ANOVA if Variances are Heteroscedastic. *Austrian Journal of Statistics* **36 (3)**: 179–188.

[5]Moder, K. (2010) Alternatives to F-Test in One Way ANOVA in Case of Heterogeneity of Variances (a Simulation Study). *Psychological Test and Assessment Modeling* **52**: 343–353.

[6]Rasch, D.; Kubinger K. D.; Moder, K. (2011) The two-sample t-test: pre-testing its assumptions does not pay. *Statistical Papers*, **52 (1)** 219-231.

[7]Rasch, D.; Teuscher, F. and Guiard, V. (2007) How Robust are tests for two independent samples in case of ordered categorical data? *Journal of Statistical Planning and Inference* **133**: 2706-2720.

On single-factor experiments with repeated measures

Stanislaw Mejza[1] **and Iwona Mejza**[1]

[1] *Poznań University of Life Sciences*
Department of Mathematical and Statistical Methods

E-mail address: *stanislaw.mejza@up.poznan.pl; Iwona.mejza@up.poznan.pl*

Abstract

In many industrial, medical (clinical trials), ecological and agricultural experiments, treatments are applied to each experimental unit sequentially in time or space. There is a possibility that a systematic effect, or trend, influences the observations, in addition to the experimental unit, treatment and carry-over effects. This type of effect should be taken into account both when the experiment is planned and when the results are analyzed. An experiment that takes such issues into account is called a repeated measures experiment. The important feature is that multiple measurements are made on the same experimental unit. A special case of this is the cross-over design, wherein the treatments themselves are switched on the same experimental unit during the course of the experiment. Experiments with repeated measures will be presented here from the point of view of the agricultural sciences. In the agricultural sciences, repeated measures experiments are very often applied when observing, for example, plants subjected to protection treatments and their yields over time (tomatoes, cucumbers, cereals, etc.).

Two fundamental types of repeated measures are common. Repeated measures in time is a situation in which experimental units receive a treatment, and then are simply followed with repeated measures on the response variable at several times. In contrast, experiments may involve administering all treatment levels (in a sequence) to each experimental unit. This type of repeated measures study is a cross-over design. Cross-over designs require the use of a wash-out period between treatment applications to prevent (or minimize) carry-over effects. Carry-over effects occur when the application of one treatment affects the response of the next treatment applied in the cross-over design. The coding for analysis of cross-over designs is very similar to the case of repeated measures in time, with the addition of a 'sequence'

variable added initially to the model to test for the presence of carry-over effects.

There are many approaches to analysis and further inference in connection with repeated measures experiments (cf. e.g. Monlezun et al., 1984, Krzyśko et al., 2014, Pereira et. al., 2012). The approaches are associated with certain assumptions concerning the dispersion structure of a trait with repeated measures. The evolution of the methodology for repeated measures has been driven by the need to consider the nature of potentially correlated residuals. The current state of the art for repeated measures can be found in the rich literature on the subject.

In many cases inherently different experiments share a common analysis, as is the case with some types of split-plot and repeated measures experiments. A split-plot in time approach looks at each subject (experimental unit) as a main plot (receiving a treatment), this then being split into subplots (time periods). The split-plot in time assumes that the correlations among time periods are the same for all treatments and time periods (compound symmetry). It will be assumed that the variance-covariance matrix follows a pattern of compound symmetry. This means that the errors between time points within subjects are correlated-and note that these correlations are presumed to be the same for each set of times, regardless of how distant in time the repeated measures are made. This assumption is not a good one, and often there are other dispersion structures that are more appropriate when using repeated measures. However, the split-plot approach has some advantages. It allows one to plan an experiment in which not all measurements are performed (incomplete experiments due to missing measurements). This is very important in a situation where some measurements cannot be made. Considerations are limited only to planned incompleteness. For such incompleteness we give methods of planning and analyzing the repeated measures experiments.

Keywords: repeated measures, split-plot design, dispersion structure, ANOVA.

References

[1]Krzyśko, M., Śmialowski, T. and Wolyński W. (2014) Analysis of multivariate repeated measures data using a MANOVA model and principal components. Biometrical Letters 51: 103–124.

[2]Monlezun, C. J., Blouin, D. C. and Malone L.C. (1984) Contrasting Split Plot and Repeated Measures Experiments and Analyses. The American Statistician, 38: 21–31.

[3]Pereira, D. G., Rodrigues, P. C., Mejza, I., Mejza, S. and Mexia, J. T. 2012 Analyzing genotype-by-environment interaction using curvilinear regression. Scienta Agricola. 357–363.

High Performance Computing + High Performance Data Analytics - Advancing science with the OBLIVION supercomputer@U.Évora - ENGAGE SKA[1][2]

Miguel A. de Avillez[**,3,4]

[**] *On behalf of the ERASMUS+ HPC+HPDA Consortium*
[3]*Computational Astrophysics Group, Department of Mathematics, University of Évora, Portugal*
[4]*Zentrum für Astronomie und Astrophysik, Technische Universität Berlin, Germany*

E-mail address: *mavillez@astro.physik.tu-berlin.de, mavillez@uevora.pt*

Abstract

In order to address the large data analytics and computational demands of the Square Kilometer Array (SKA; the largest ongoing astronomical project in the world) the University of Évora is acquiring (under the ENGAGE SKA Research Infrastructure[1]) the fastest supercomputer in Portugal with an expected peak performance of 300 TFLOPS in an investment over 1 M€. Although the machine's primary use is for Astrophysics, a large fraction of the computing time will be available for the scientific community to carry out production runs and deal with massive amounts of data (hundreds of terabytes or even petabytes; the "**real**" Big Data).

There are a few major goals defined for this machine: (1) carry out computationally demanding scientific simulations, (2) handle massive volumes of data using high performance computing (HPC), (3) training and transfer of knowledge on HPC and high performance data analytics (HPDA), and (4) establish a HPDA program associated to different data collections and their access by different communities (e.g., astrophysics, atomic and particle physics, engineering, mathematics and statistics, health and medicine, and tourism, just to name a few).

Some of these goals are fostered by the ERASMUS+ "High Performance Computing + High Performance Data Analytics" (HPC+HPDA) Consortium[3] established to advance the use and training on HPC & HPDA among HEIs in Portugal inline with the European Digital Programme.

[1] ENGAGE SKA Research Infrastructure (Ref. POCI-01-0145-FEDER-022217), funded by the FCT and COMPETE2020.

[2] The supercomputer is part of the Iberian Network of Advanced Computing - Portugal (RICA-PT) funded by the FCT/FCCN.

[3] The HPC+HPDA Consortium is presently composed by the Universities of Évora (PI), Coimbra, and Algarve, and the New University of Lisbon.

In this talk I'll present the supercomputer and the expected scalability against other machines in Europe, the goals referred above, the scientific case, as well as the HPC+HPDA consortium.

Keywords: High performance computing – High performance data analytics – Data collections – Supercomputers – Scientific computing

Organized Sessions

Organized Session 1

Statistical Methods and Applications

Organizer: Ayana Mateus (Portugal)

Inference for the Growth Curve Model with Dual Orthogonal Covariance Structure

Miguel Fonseca[1,2] **and Martin Singull**[3]

Department of Mathematics, Faculty of Sciences and Technology, NOVA University of Lisbon, Portugal[1]
Center of Mathematics and Applications, Faculty of Sciences and Technology, NOVA University of Lisbon, Portugal[2]
Mathematics Department, Linköping University, Sweden[3]

E-mail address: *fmig@fct.unl; martin.singull@liu.se*

Abstract

The growth curve model is a well documented multivariate model in literature, with a well established methodology of maximum likelihood estimation. We propose a growth curve model family with an orthogonal covariance structure for lines and columns, proceeding with the derivation of maximum likelihood statistics. Many familiar models fall within this model family, as it will be shown.

Keywords:growth curve model, orthogonal covariance structure, maximum likelihood.

Acknowledgements

The first author would like to state that his work was partially supported by the Portuguese Foundation for Science and Technology through the project UID/MAT/00297/2013 (Center of Mathematics and Applications).

References

[1]Khatri, C. G. (1973). Testing some covariance structures under a growth curve model. *J. Multivariate Anal.*, 3(1):102–116.

[2]Kollo, T. and von Rosen, D.. (2010). *Advanced Multivariate Statistics with Matrices*. Springer.

Multivariate collective risk models Inference and special case

Elsa Moreira[1,2], **Ana Cantarinha**[3,4] **and João T. Mexia**[1,2]

Nova University of Lisbon[1]
Center for Mathematics and its applications - CMA[2]
University of Évora[3]
CIMA - Center for Research on Mathematics and its Applications[4]

E-mail address: *efnm@fct.unl.pt; aicantarinha@gmail.com*

Abstract

In this paper we propose an extension of the theory of univariate collective models to the multivariate case. These models play an important part in Risk Theory thus in Actuarial Mathematics. Up to now mainly the univariate models have been considered but, there are more and more situations requiring a multivariate approach. For instance when considering forest fires, besides their number, the burnt areas are of great interest. Inference based on these models was centered on totals, claims totals being the leading example, we now consider also means, mean claims for instance. The approach presented is based on asymptotic distributions which are very useful, since many modelling problems have large samples. Examples of such situations range from insurances portfolios, to the sets of clients of banks, to Forest Fires, etc. With this purpose, we will consider a special case of the general model where the set of policies is replaced by the set of forests and the claims by forest fires, in order to do inference on two variables of interest for the study of forest fires: the number of fires and the burnt areas. An illustrative application to forest fires in Portugal is presented.

Keywords: Collective models, asymptotic distributions, confidence intervals, risk theory.

Acknowledgements

This work was partially supported by the Portuguese Foundation for Science and Technology through the project UID/MAT/00297/2019.

References

[1]Bowers, N.L.; Gerber, H.U.; Hickman, J.C.; Jones, D.A. and Nesbet, C.J. (1986). *Actuarial Mathematics*. The Society of Actuaries.

[2]BILLINGSLEY, P (1999) *Convergence of Probability Measures*. John Wiley & Sons.

[3]Centeno, M. L. (2003) *Teoria do Risco na Actividade Seguradora*. Celta Editora - Coleçãoo Económicas. Oeiras.

[4]Grandell, J. (1992) *Aspects of Risk Theory*. Springer Series in Statistics.

[5]Khuri A. (2003) *Advanced calculus with applications to statistics*. John Wiley & Sons.

[6]Lukacs, E. and Laha, R. G. (1964) *Applications of characteristic functions*. Charles Griffin & Co. limited. London.

[7]Mexia, J. T. (1990) Best linear unbiased estimates, duality of F tests and ScheffÃľ's multiple comparison method in the presence of controlled heterocedasticity. *Journal Comp. Stat. Data Analysis*. **10(3)**: 271–281.

[8]Mexia, J. T. (1990) *Variance free models*, Trabalhos de Investigação, N^o Departamento de Matemática, Faculdade de Ciências e Tecnologia. Universidade Nova de Lisboa.

[9]Marques, S.; Borges, J. G.; Garcia-Gonzalo, J.; Moreira, F.; Carreiras, J.M.B.; Oliveira, M.M.; Cantarinha, A.; Botequim, B. and Pereira, J. M. C. (2011) Characterization of wildfires in Portugal. *European Journal of Forest Research*, **130(5)**: 775–784.

[10]Pausas, J.P.; Llovet, J.; Rodrigo, A. and Vallejo, R. (2008) Are wildfires a disaster in the Mediterranean basin: a review. *International Journal of Wildland Fire*. **17(6)**: 713–723.

[11]Reis, E. (1993) How long is the surplus below zero? *Insurance: Mathematics and Economics*. **12**: 23–38.

[12]Scheffé, H. (1959). *The Analysis of Variance*. John Wiley & Sons.

Unbalanced Full Factorial and Fractional Factorial Designs Associated with a Commutative Jordan Algebra

Sandra Oliveira[1,2], **Elsa Moreira**[1,3], **Miguel Fonseca**[1,3] **and João Tiago Mexia**[1,3]

Centro de Matemática e Aplicações (CMA) FCT/UNL[1]
Departamento de Economia e Gestão da Escola Superior de Ciências Empresariais - Instituto Politécnico de Setúbal[2]
Departamento de Matemática, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa[3]

E-mail address: *sandra.oliveira@esce.ips.pt; efnm@fct.unl.pt;fmig@fct.unl.pt*

Abstract

The purpose of this work is to show that is possible to extend the study of the full factorial designs and the fractional factorial designs, of fixed effects, to the unbalanced case, that is, when the number of observations taken within each treatment may be different. It is possible combining the linear model associated with a commutative Jordan algebra (CJA) and the L-Model theory. The structure of the factorial design used in this work is based on linear spaces on Galois fields as well as on the relationship between a linear model and a CJA.

Keywords: Design of experiments; Linear models; Unbalanced designs.

References

[1]Fonseca M., Mexia J., Zmyślony R. (2006). Binary operations on Jordan algebras and orthogonal normal models. *Linear Algebra and Its Applications*, 417(1):75-86.

[2]Jesus, V., Mexia, J.T., Fonseca, M., Zmyslony, R. 2009. Binary Operations an Canonical Forms for Factorial and Related Models. Linear Algebra and Its Applications. 430 (10):2781-2797. DOI= 10.1016/j.laa.2009.01.013.

[3]Jesus V., Mexia J.T., Fonseca M., Zmyślony R. (2009) Binary Operations and Canonical Forms for Factorial and Related Models. *Linear Algebra and Its Applications*, 430(10):2781-2797.

[4]Mexia, J.T. 1990. *Best linear unbiased estimates, duality of F tests and the Scheffé multiple comparison method in presence of controled heterocedasticity*, Computational Statistics & Data Analysis, 10(3), pp. 271-281.

[5]Seely J. 1971. Quadratic Subspaces and Completeness. The Annals of Mathematics Statistical, 42(2):710-721.

[6]Moreira, E., Mexia, J. T., Zmyslony, R.. 2009. L-Models and Multiple Regressions Designs. Statistical Papers. 50: 869-885. DOI 10.1007/S00362-009-0255-3.

[7]Moreira, E., Mexia, J. 2011. Multi-treatment regression analysis: the unbalanced case. WSEAS proceedings of the 5th international conference on applied mathematics, simulation, modelling (Corfu, Greece 2011, July), 69-74.

[8]Moreira, E. E., Ribeiro, A. B., Mateus, E. P., Mexia, J. T., Ottosen, L. M.. 2005. Regressional modeling of electrodialytic removal of Cu, Cr and As from CCA treated timber waste: Application to sawdust. Wood Science and Technology 2005; 39(4):291-309. DOI: 10.1007/s00226-004-0267-z.

[9]Moreira E.E., Mexia J. T. 2007. Multiple regression models with cross nested orthogonal base model, Bulletin of the International Statistical Institute 56th Session - Proceedings of ISI.

[10]Oliveira, S., Moreira, E., Fonseca, M., Mexia, J. T. 2018. Multiple Regression Design for a Full Factorial Base Model Associated with a Commutative Jordan Algebra. In Proceedings of the 2018 International Conference on Mathematics and Statistics (pp. 41-45). ACM.

Revisiting the calibration procedure in the extremal index estimation

D. Prata Gomes[1,2] **and M. Manuela Neves**[3,4]

Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa[1],
CMA, Faculdade de Ciências e Tecnologia[2]
Instituto Superior de Agronomia[3], *and CEAUL*[4],
Universidade de Lisboa

E-mail address: *dsrp@fct.unl.pt; manela@isa.ulisboa.pt*

Dedicated to Professor J. Tiago Mexia as a token of friendship and with a thank you

Abstract

In *extreme value theory* we deal essentially with the estimation of parameters of extreme or rare events. Classical methods were first derived when the underlying process is assumed to be a sequence of independent and identically distributed random variables. However, when observations are taken along the time and/or the space the independence is an unrealistic assumption. A relevant parameter that arises in this situation is the extremal index, θ, characterizing the degree of local dependence in the extremes of a stationary series. Most of the semi-parametric estimators of this parameter show a strong dependence on the threshold, u_n, with an increasing bias and a decreasing variance as the threshold decreases. The calibration methodology, considered as a way of controlling the bias of an estimator was revisited. Some models were considered in a simulation study presented in a previous work, see Prata Gomes, Mexia and Neves (2013) and here this technique will be revisited and some new results are discussed such as the construction of confidence intervals for θ.

Keywords: calibration technique, estimation, extreme value theory, extremal index.

Acknowledgements

Research partially supported by National Funds through **FCT** – Fundação para a Ciência e a Tecnologia projects UID/MAT/0297/2013 (CMA/UNL) and UID/MAT/00006/2019 (CEA/UL).

References

[1]Prata Gomes, D., Mexia, J. T. and Neves, M. M. (2013) Simulation Study of the Calibration Technique in the Extremal Index Estimation, *Advances in Regression, Survival Analysis, Extreme Values, Markov Processes and Other Statistical Applications, Studies in Theoretical and Applied Statistics*, Springer-Verlag Berlin Heidelberg: 381-389.

A generalized jacknife estimator of a negative extreme value index

Frederico Caeiro[1,2] **and M. Ivette Gomes**[3,4]

Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Portugal[1]
Centro de Matemática e Aplicações (CMA), Universidade Nova de Lisboa[2]
DEIO, Faculdade de Ciências, Universidade de Lisboa, Portugal[3]
Centro de Estatística e Aplicações da Universidade de Lisboa (CEAUL)[4]

E-mail address: *fac@fct.unl.pt; ivette.gomes@fc.ul.pt*

Abstract

Let $X_1, X_2, \ldots, X_n$ be independent random variables with a common distribution function F in the max-domain of attraction of a non-degenerate distribution function G. Then G is the extreme value distribution with a shape parameter ξ (Gnedenko, 1943). Let us assume that the shape parameter ξ, also know as the extreme value index, is negative ($\xi < 0$). Under a semi parametric framework, inference is usually based on the upper $k+1$ order statistics of the sample of size n. Classical estimators of the extreme value index such as the moment estimator (Dekkers *et al.*, 1989) or Pickand's (1975) estimator have high variance for small values of k and a high bias for large values of k. This leads to the usual bias-variance trade-off problem when we choose the threshold k. In this paper, following the lines of Caeiro and Gomes (2003) and Gomes *et al.* (2013), we use the generalized jackknife methodology to obtain a new asymptotically unbiased estimator of a negative extreme value index. Under a second order condition on the tail function $1-F$ we derive the non degenerate asymptotic behaviour of the new estimator. Using Monte-Carlo simulation techniques, we validate the asymptotic results for finite samples.

Keywords: extreme value index, generalized jackknife methodology, moment estimator, semi-parametric estimation.

Acknowledgements

Research partially supported by National Funds through **FCT** — Fundação para a Ciência e a Tecnologia, projects UID/MAT/00006/2019 (CEAUL) and UID/MAT/00297 /2019 (Centro de Matemática e Aplicações).

References

[1]Caeiro, F. and Gomes, M.I. (2003). Bias reduction for light tail models—the generalized jackknife methodology. In Gomes, M.I., de Haan, L., Pestana, D., Canto e Castro, L. and Fraga Alves, M.I. (eds.). *Extremes, Risk and Resampling Techniques*: 7–12, Edições CEAUL.

[2]Dekkers, A.L.M., Einmahl, J.H.J. and de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution, *The Annals of Statistics* **17**(4): 1833–1855.

[3]Gnedenko, B. V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire, *The Annals of Mathematics* **44**(3): 423–453.

[4]Gomes, M.I., Henriques-Rodrigues, L. and Caeiro, F. (2013). Refined estimation of a light tail: an application to environmental data. In N. Torelli, F. Pesarin, A. Bar-Hen (Eds.), *Advances in Theoretical and Applied Statistics*, Chapter 14:143–152. Springer-Verlag, Berlin and Heidelberg.

[5]Pickands III, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics* **3**(1): 119–131.

Organized Session 2

Computational Mathematics, Statistics and Information Management - Part I

Organizer: M. Filomena Teodoro (Portugal)

Analysis of SMEs Financial Distress Probability: the case of Tâmega and Sousa Region

Ana Borges[1,2], **Fábio Duarte**[1,2] **and Marco Machado**[3,4]

School of Management and Technology of Polytechnic of Porto[1]
CIICESI[2]

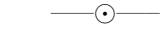
E-mail address: *aib@estg.ipp.pt; fdd@estg.ipp.pt*

Abstract

This study intends to deepen the knowledge of risk factors financial distress of companies, in a region of Portugal that presents different specificities in terms of its business dynamics, characterized by a large number of small and medium enterprises (SMEs): the region of *Tâmega and Sousa*. Since the timely identification of imminent failure of firms is desirable to ensure the stability of financial markets and overall economic prosperity [3]. For that purpose, we analyze historical data on 1190 firms from the manufacturing sector, taken from SABI database, between 2008-18 adopting the Cox proportional Hazard model [1] within survival analysis methodology. Survival analysis allows us to study a distribution of time that elapses between the entrance and the occurrence of an event of interest, which in this study is the financial distress. Based on [4] we consider that a company is in financial distress when its EBITDA (Earnings Before Interest, Taxes, Depreciation and Amortization) is less than financial expenses for two consecutive years, and general liquidity (current assets over current liabilities) is less than one for two consecutive years. We analyze time between firm creation and financial distress. From literature review, and specifically based on the works of [3], [5], [6] and [8] we test for firm *age, size* and *ownership concentration* effects as intrinsic firm characteristics. Additionally, we study if firm survival is influenced by macroeconomic stress conditions [7] by testing the effect of Euribor (6M) and Gross National Product, both measured in the year preceding firm financial stress. We control for spatial effect (*region*), a categorical variable controlling 11 counties, since we expect that different geographic regions idiosyncratic has influence on SMEs' survival. For model robustness check we fitted the flexible parametric Royston and Parmer model [9] to the data to verify if results altered considerably. For model prediction accuracy diagnose,

the Harrell C statistic was calculated. Results show that age, size, ownership concentration have a significant effect on firm probability of financial distress and, as expected, it negatively influenced by macroeconomic stress conditions.

Keywords: Financial Distress, SMEs, Survival Analysis, Cox model



Acknowledgements

The authors thank the support by the Center for Research and Innovation in Business Sciences and Information Systems(CIICESI), ESTG - P.Porto.

References

[1]Cox, D. R. (1972). Regression models and life?tables. Journal of the Royal Statistical Society: Series B (Methodological), 34(2), 187–202.

[2]Fotopoulos, G., & Louri, H. (2000). Determinants of hazard confronting new entry: Does financial structure matter? Review of Industrial Organization, 17(3), 285–300.

[3]Gordini, N. (2014). A genetic algorithm approach for SMEs bankruptcy prediction: Empirical evidence from Italy. Expert Systems with Applications, 41(14), 6433–6445. https://doi.org/10.1016/j.eswa.2014.04.026

[4]Gupta, J., Barzotto, M., & Khorasgani, A. (2018). Does size matter in predicting SMEs failure? International Journal of Finance & Economics, (March), 571–605.

[5]Mata, J., & Portugal, P. (1994). Life Duration of New Firms. The Journal of Industrial Economies, 42(3), 227–245

[6]Mata, J., Portugal, P., & Guimarães, P. (1995). The survival of new plants: Start-up conditions and post-entry evolution. International Journal of Industrial Organization, 13(4), 459–481.

[7]Nunes, P. M., & Serrasqueiro, Z. (2012). Are young SMEs' survival determinants different? Empirical evidence using panel data. Applied Economics Letters, 19(9), 849–855.

[8]Pérez, S. E., Llopis, A. S., & Llopis, J. A. S. (2004). The Determinants of Survival of Spanish Manufacturing Firms. Review of Industrial Organization, 25(3), 251–273.

[9]Royston, P., & Parmar, M. K. (2002). Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. Statistics in medicine, 21(15), 2175–2197.

Portuguese SME's internationalization and financial performance

Sofia Silva[1,2], **Aldina Correia**[1,2] **and Vítor Braga**[1,2]

CIICESI Center for Research and Innovation in Business Sciences and Information Systems[1]
ESTG/P.PORTO School of Management and Technology, Polytechnic of Porto[2]

E-mail address: *8140220@estg.ipp.pt; aic@estg.ipp.pt; vbraga@estg.ipp.pt*

Abstract

The main objective of this work is to study the impact that internationalization can have on the financial performance of a company.

A database collected on the SABI (Iberian Balance Sheet Analysis System) database, which includes Small and Medium Enterprises – SMEs (based on the definition of the European Commission, in recommendation number 2003/361/CE, SMEs are companies with a maximum of 250 employees and a turnover of not more than EUR 50 million), which carry out exports and which had information in relation to the absolute year 2016 in SABI. Thus, the data includes information about 255 SMEs (without missing) and a total of 12 variables (4 qualitative and 8 quantitative).

In order to perform the study, several methodologies were used, namely, an initial descriptive analysis in order to classify the sample, as well as applied a factor analysis (exploratory and confirmatory) and a multiple linear regression.

With this study it was verified that the financial performance of the companies has influence in the internationalization of the company, being the model presented highly significant.

The most significant limitation of this work is that the adjusted coefficient of determination is relatively small. However, this is explained by the literature that the lack of a formal concept of internationalization and financial performance leads to the creation of more subjective models, which can lead to contradictory results.

It should also be noted that by using only exports to represent internationalization, only a small part of the multidimensional phenomenon of internationalization is considered.

We conclude that, for future research, it would be pertinent to conduct questionnaires to the respective companies in order to obtain qualitative information, that together with the quantitative would allow a more correct image of reality. In addition, it would also be interesting to carry out a longitudinal study, since this would allow an effective verification of the relationship and the evolution between financial performance and internationalization.

Keywords: internationalization, financial performance, SABI, multivariate analysis, factor analysis, linear regression.

References

[1]Buckley, P. J., Dunning, J. H., & Pearc, R. D. (1978). The influence of firm size, industry, nationality, and degree of multinationality on the growth and profitability of the World's largest firms, 1962-1972. *Review of World Economics* **114** (2): 243–257.

[2]Calof, J. L., & Beamish, P. W. (1995). Adapting to foreign markets: Explaining internationalization. *International Business Review*, **4** (2): 115–131.

[3]Cavusgil, S. T. (1980). On the internationalization process of firms. *European Research* **8**: 273–81.

[4]Contractor, F. J., Kundu, S. K., & Hsu, C.-C. (2003). A three-stage theory of international expansion: the link between multinationality and performance in the service sector. *Journal of International Business Studies*, **34**(1): 5–18.

[5]Dörrenbächer, C. (2000). Measuring corporate internationalisation. *Intereconomics*, **35**(3): 119-126.

[6]Harris, S., & Wheeler, C. (2005). Entrepreneurs' relationships for internationalization: functions, origins and strategies. *International Business Review*, **14**(2), 187–207.

[7]Hassel, A., Höpner, M., Kurdelbusch, A., Rehder, B., & Zugehör, R. (2003). Two dimensions of the internationalization of firms. *Journal of Management Studies*, **40**(3), 705–723.

[8]Hult, G. T. M., Ketchen, D. J., Griffith, D. A., Chabowski, B. R., Hamman, M. K., Dykes, B. J., ... & Cavusgil, S. T. (2008). An assessment of the measurement of performance in international business research. *Journal of International Business Studies*, **39**(6): 1064–1080.

[9]Ietto-Gillies, G. (1998). Different conceptual frameworks for the assessment of the degree of internationalization: an empirical analysis of various indices for the top 100 transnational corporations. *Transnational corporations*, **7**: 17–40.

[10]Li, L. (2007). Multinationality and performance: A synthetic review and research agenda. *International Journal of Management Reviews*, **9**(2): 117–139.

[11]Ramaswamy, K., Kroeck, K. G., & Renforth, W. (1996). Measuring the degree of internationalization of a firm: A comment. *Journal of International Business Studies*, **27**(1): 167–177.

[12]Reuber, A. R., & Fischer, E. (1997). The influence of the management team's international experience on the internationalization behaviors of SMEs. *Journal of International Business Studies*, **28**(4): 807–825.

[13]Sullivan, D. (1994). Measuring the degree of internationalization of a firm. *Journal of International Business Studies*, **25**(2): 325–342.

[14]Sullivan, D. (1996). Measuring the degree of internationalization of a firm: A reply. *Journal of international Business Studies***27**(1): 179–192.

Conditions for entrepreneurship of municipalities

Aldina Correia[1,2] **and Eliana Costa e Silva**[1,2]

CIICESI Center for Research and Innovation in Business Sciences and Information Systems[1]
ESTG|P.PORTO School of Management and Technology, Polytechnic of Porto[2]

E-mail address: *aic@estg.ipp.pt; eos@estg.ipp.pt*

Abstract

The main objective of this work is to study the conditions of municipalities which favor companies' creation. In particular it is studied if *Tâmega e Sousa* Region has differentiate characteristics for entrepreneurship.

A database concerning several characteristics of the Portuguese municipalities, collected from PORDATA - *Base de Dados de Portugal Contemporâneo*[4], is considered. The database includes municipalities information, such us, population size and age, non-financial companies, unemployment, among others.

In order to perform the study, several methodologies were used. First, an initial descriptive analysis is performed in order to classify the sample. Furthermore a discriminant analysis is applied with the goal of studying the main characteristics of the *Tâmega e Sousa* Region, which distinguishes their municipalities from the others. This is followed by the estimation of a multiple linear regression, aiming at identifying which variables have a larger impact in the companies' creation.

With this study it was to verify that few variables have discriminant power to distinguish the *Tâmega e Sousa* municipalities from the municipalities of other Portuguese NUT III regions. Furthermore, the variables with the highest influence in the companies' creation are also identified, with the presented model highly significant.

Keywords: entrepreneurship, multivariate analysis, discriminant analysis, linear regression.

References

[1]Correia, A., Costa e Silva, E., Lopes, I. C., Braga, A. (2016, December). MANOVA for distinguishing experts' perceptions about entrepreneurship using NES data from GEM. In *AIP Conference Proceedings*, **1790**(1): 140002. AIP Publishing.

[2]Fairlie, R. W. (2013). Entrepreneurship, economic conditions, and the great recession. *Journal of Economics & Management Strategy*, **22**(2): 207–231.

[3]Gnyawali, D. R., & Fogel, D. S. (1994). Environments for entrepreneurship development: key dimensions and research implications. *Entrepreneurship theory and practice*, **18**(4): 43–62.

[4]Fundação Francisco Manuel dos Santos (2019). PORDATA – *Base de Dados de Portugal Contemporâneo*, https://www.pordata.pt/, *Accessed on 08/05/2019*.

[5]Wennekers, S., Uhlaner, L., Thurik, R. (2002). Entrepreneurship and its conditions: a macro perspective. *International Journal of Entrepreneurship Education (IJEE)*, **1**(1): 25–64.

Modeling KAP about Zika Virus

João Faria[1,4], **Rosa Teodósio**[2], **M. Filomena Teodoro**[3,4]

IHMT - Institute of Hygiene and Tropical Medicine, UNL - New University of Lisbon, Lisboa, Portugal[1]
GHTM - Global Health and Tropical Medicine, IHMT - Institute of Hygiene and Tropical Medicine, UNL - Universidade Nova de Lisboa, Rua da Junqueira 100, 1349-008 Lisbon, Portugal[2]
CEMAT - Center for Computational and Stochastic Mathematics, IST-Instituto Superior Técnico, Lisbon University, Avenida Rovisco Pais, n. 1, 1048-001 Lisboa, Portugal[3]
CINAV - Center of Naval Research, Portuguese Naval Academy, Portuguese Navy, Base Naval de Lisboa, Alfeite, 2810-001 Almada, Portugal[4]

E-mail address: *maria.alves.teodoro@marinha.pt, mteodoro64@gmail.com*

Abstract

The objective of this study is to describe the knowledge, attitudes and preventive practices regarding the infection by the Zika virus (ZIKV). We have focused our study on the embarked population. A questionnaire was built and applied to different groups under study: the navigators who will navigate in endemic areas of Zika virus and those that already have traveled to endemic areas of ZIKV. A descriptive analysis of such questionnaire can be found in [1]. In [3,2] was done an exploratory factorial analysis, several factors associated with knowledge were identified. Here, we present the continuation of such work, completing the FA relatively to attitudes and practice. The knowledge, attitudes and Practice (KAP) about ZIKV need to be improved .

Keywords: KAP, questionnaire, statistical approach, factorial analysis, Zika virus.

—⊙—

The Zika cases and the associated disorders with infection have increased significantly in last years. In 2016, the World Health Organization (WHO) declared Zika as a public health emergency of international importance [15,16].

Later, in the same year, declined this status, but promoted precaution and observation measures so can be observed and be taken some actions to control the dissemination of the ZIKA virus due the migration flows between distinct continents, take into account the asymptomatic cases, the persistence of virus.

To evaluate the KAP of emboarded individuals was built a questionnaire based on the validated questionnaire proposed by Rosalino C. et al in [5] and the resource kit of the WHO [7].

The statistical approach is planned to be done in several stages: a detailed descriptive analysis was already done and can be found in [1]. Using more elaborated techniques, we can found in [3,2] an exploratory factorial analysis restricted to Knowledgement issues. To continue the analysis we preview to apply some general linear models to the factors found in [3,2] and extend the analysis to the Practice and Attitude questions. The preliminary results are consistent: there is an important need to improve the literacy bout Zika infection.

Acknowledgements

M. Filomena Teodoro was supported by Portuguese funds through the *Center of Naval Research* (CINAV), Portuguese Naval Academy, Portugal and *The Portuguese Foundation for Science and Technology* (FCT), through the *Center for Computational and Stochastic Mathematics* (CEMAT), University of Lisbon, Portugal, project UID/Multi/04621/2019.

References

[1]Faria, J. (2019) *Conhecimentos, atitudes e práticas de prevenção sobre Zika da população embarcada em navios da Marinha Portuguesa*. Master Thesis. Lisboa, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa. Under evaluation

[2]Faria, J., Teodósio, R. and Teodoro, F. (Submitted) *Zika: literacy and behavior of individuals on board ships. A first approach*. In: Simos, T. and et al. Computational Methods in Science and Engineering (ICCMSE 2019). AIP proceedings, New York, AIP.

[3]Faria, J., Teodósio, R. and Teodoro, F. (In press) *Zika: Knowledge of the Boarded Population About Zika Virus*. In: Misra, S. and et al. Computational Science and Applications (ICCSA 2019). Lectures Notes in Computer Science, Cham, Springer.

[4]Mlakar J., Korva M., Tul N., Popovicć M. and et al. (2016) Zika Virus Associated with Microcephaly. *N Engl J Med.* **374**: 951-958.

[5]Rosalino, C. and et al. (2016) *Conhecimentos, atitudes e práticas sobre Zika de viajantes visitando o Rio de Janeiro, Brasil, durante os Jogos Olímpicos e Paraolímpicos de 2016: Inquérito no Aeroporto Internacional AntÃşnio Carlos Jobim*, Rio de Janeiro.

[6]WHO (2016) WHO statement on the first meeting of the International Health Regulations (IHR 2005), Emergency Committee on Zika virus and observed increase in neurological disorders and neonatal malformations . Available from: `https://www.who.int/en/news-room/detail/`

01-02-2016-who-statement-on-the-first-meeting-of-the-international-health-regulations-(2005)-(ihr-2005)-emergency-committee-on-zika-virus-and-observed-increase-in-neurological-disorders-and-neonatal-malformations. Accessed at 10/4/2019.

[7]WHO. Inquéritos sobre Conhecimentos, Atitudes e Práticas. Doença do Vírus Zika e Potenciais Complicações - Pacote de recursos (2016). Available from: https://apps.who.int/iris/bitstream/handle/10665/204689/WHO_ZIKV_RCCE_16.2_por.pdf?sequence=5. Accessed at 10/4/2019.

Organized Session 3

Computational Mathematics, Statistics and Information Management - Part II

Organizer: M. Filomena Teodoro (Portugal)

Impact of using excellence management models in the customer Satisfaction of Brazilian electricity distributors -10 years of studies

Alexandre Carrasco[1], Marina A. P. Andrade[1,2] and Álvaro Rosa[3,4]

ISCTE - Instituto Universitário de Lisboa[1]
ISTAR - Instituto Universitário de Lisboa[2]

E-mail address: *marina.p.andrade@iscte.pt*

Abstract

The article evaluates the impact of the use of the model of excellence in Brazilian management (MEG) by electricity distribution companies and their impact on customer satisfaction (IASC). It was evaluated 10 years of use of the model in groups of companies with different levels of implantation of the model (users, indifferent, engaged and winning) using statistical methods. As results it is verified the existence of differences between the groups revealing the correct decision by the use of the model seen by this view. The results can be used by similar organizations or other industries

Its importance is reinforced by the coverage of studies that reaches 96% of the sector, for a decade (2007 to 2016), by the use of publicly unavailable sources. The organizations were classified according to the use of the MEG (Excellence in Management Model) and evaluated, through statistical tests, differences between the indicators in these groups, allowing to conclude about the mass adoption of the MEG and impacts on the non-financial indicators more relevant in this sector.

Keywords: Excelence Models, Customer Sattisfaction, Energy distribution, MEG, IASC.

Daily Load Profile of Portuguese Electricity Demand using Kohonen Self-Organizing Maps

Ricardo Moura[1,2,3,4], **João Neves**[3,5], **Francisco Reis**[3,5], **and Anacleto Correia**[1,2]

Portuguese Naval Academy[1]
Naval Research Center (CINAV), Portugal[2]
NOVA University of Lisbon[3]
Center for Mathematics and Aplications (CMA), Portugal[4]
NOVA School of Science and Technology (FCT NOVA), Portugal[5]

E-mail address: *ricardo.pinto.moura@marinha.pt; jpp.neves@campus.fct.unl.pt; fa.reis@fct.unl.pt; cortez.correia@marinha.pt*

Abstract

The massive penetration of distributed renewable energy resources (RER) and an expected more proactive behavior of consumers by taking part in demand response initiatives are, amid others, key drivers that significantly increase the complexity in which the power systems will be operated and planned. To cope with these challenges, system operators need to develop methodologies to assess both current and expected system behaviors. In that regard, load prediction is of paramount importance for the overall assessment of the network. This is affected by several factors from different origins such as weather conditions, seasonal periods, social-economic trends and operational actions by the system operators.

From Portuguese historical data, a Kohonen's Self-Organizing Map (SOM) is used to classify the days of the year according to their load curve profile. The SOM Kohonen's algorithm is an unsupervised neural network clustering technique often used for pattern classification tasks that preserves the topology of data. Through the visualization of the data by projecting it into a bidimensional grid it is possible to identify the 'natural' clusters. The main objective of the classification is to separate the load demand by clusters for forecasting purposes. The classification/clustering via SOM will be the first phase of a so called Artificial Neural Networks (ANN) hybrid model, where two or more ANN models are combined in order to draw more accurate predictions. The historical data used is the Portuguese continental load demand from 2009 to 2017 registered in periods of 15 min, comprising a total of 96

load registries per day. It is important to point out that when carrying with the data preprocessing the authors create cyclical normalized variables, similar to a combination of sine and cosine normalization technique, to define the days and months of a year. It is then presented the profile of the daily load curves per cluster.

Keywords: Self-Organizing Maps, Electric Demand, Short-term Load Forecasting, Artificial Neural Networks, Time Series Clustering.
Acknowledgements

Ricardo Moura thanks FCT (Portuguese Foundation for Science and Technology) project UID/MAT/00297/2013 awarded through CMA/UNL.

References

[1]Fort, J. C., Letrémy, P., & Cottrell, M. (2002). Advantages and drawbacks of the Batch Kohonen algorithm. In ESANN (Vol. 2, pp. 223–230).

[2]Hernandez, L., Baladron, C., Aguiar, J. M., Carro, B., Sanchez-Esguevillas, A. J., Lloret, J., & Massana, J. (2014). A survey on electric power demand forecasting: future trends in smart grids, microgrids and smart buildings. IEEE Communications Surveys & Tutorials, 16(3), 1460–1495.

[3]Hernandez, L., Baladrón, C., Aguiar, J., Carro, B., Sanchez-Esguevillas, A., & Lloret, J. (2013). Short-term load forecasting for microgrids based on artificial neural networks. Energies, 6(3), 1385–1408.

[4]Ismail, S., Shabri, A., & Samsudin, R. (2011). A hybrid model of self-organizing maps (SOM) and least square support vector machine (LSSVM) for time-series forecasting. Expert Systems with Applications, 38(8), 10574–10578.

[5]KOHONEN, T. (2001) Self-organizing Maps, New York, Springer.

[6]Marin, F. J., Garcia-Lagos, F., Joya, G., & Sandoval, F. (2002). Global model for short-term load forecasting using artificial neural networks. IEE Proceedings-Generation, Transmission and Distribution, 149(2), 121–125.

Reassessment of Periodontal Inflammation and High Density Lipoproteins Relationship through GAMLSS models

J.A. Lobo Pereira[1], **Devide Carvalho**[2] **and Teresa Oliveira**[3]

Dentistry School, Universidade do Porto, , Department of Periodontology, Porto, Portugal[1]
Department of Endocrinology Centro Hospitalar Universitário de S. João, Faculty of Medicine, Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal[2]
Universidade Aberta, Lisboa, Portugal[3]

E-mail address: *jpereira@fmd.up.pt*

Abstract

To explain the variability of probing pocket depth(PPD), GAMLSS models were fitted to data of 158 individuals. The results suggest a U-shaped relationship between HDL levels and PPD, together with a significant impact of age, diabetes and HDL on PPD.

Keywords: Periodontitis, HDL, Statistical models, GAMLSS.

1 Introduction

Periodontitis is a chronic inflammatory destructive process in the supporting tissues of the teeth associated with gram negative bacteria present in a biofilm on the surface of the teeth, and is one of the main causes of tooth loss in adulthood. Thus promoting a low level endotoxemia (Andrukhov et al., 2013), that appear to sustain dyslipidemia associated to low-grade inflammation is essential, yet whether periodontitis causes an increase in levels of plasma lipids or whether hyperlipidemia is a risk factor for periodontal infection. The (PPD) is a clinical parameter for periodontal destruction/ inflammation.

Generalized additive models for location, scale and shape (GAMLSS), are a semi-parametric extension of the classical GAM (Hastie and Tibshirani 1990) approach introduced by Rigby and Stasinopoulos (2001, 2005) and Stasinopoulos and Rigby (2007). Besides modelling the conditional mean (location) of the response variable distribution also, model the variance (scale) and the shape parameters (skewness and kurtosis) which may depend on explanatory variables. The aim of this work is to assess non-linear relationships between mean PPD and HDL, age and BMI.

2 Material and methods

Data on lipid profile, anthropometry and sociodemographic from 158 diabetic individual and age and gender matched non-diabetic controls were analysed.

The dependency of mean PPD on HDL levels, age, body mass index (BMI) and diabetic status was modelled using a truncated exGAUS distribution between 0 and 6 addressed by GAMLSS models (Rigby and Stasinopoulos, 2005) in R software Core Team,2018).

3 Background

Under the GAMLSS framework, we assume a random variable Y have probability distribution function of p parameters of Y, being μ,σ,ν and τ parameters for location, scale, skewness, and kurtosis, respectively. A given distribution can be defined by 2, 3 or 4 parameters, which guarantees enough flexibility to fit parameters by a monotonic link function. The exGAUS distribution is a three parameters distribution (μ,σ and ν) that describes the sum of two independent normal and exponential random variables.

4 Results

The best models fitted for the three parameters (μ, σ and ν) of the TexGAUS[0, 6] distribution: location (μ) modelled by smooth functions (P-splines) of HDL, Age and the interaction term of HDL and Diabetes, scale (σ) modelled by a smooth function of Age and shape (ν) by Age.

5 Conclusions and Discussion

Our results suggest a U-shaped relationship between PPD and HDL, similar to that founded by Madsen et al (2017) between HDL and risk of infectious disease. The multivariable analysis suggest significant effect of age, diabetic status and HDL, but BMI, on periodontitis.

The gamlss modelation framework revealed versatile enough to undercover non linear relationships between PPD and the independent variables using TexGAUS[0, 6].

References

[1]Andrukhov O, Haririan H., Bertl K., Rausch W.D., Bantleon H.P., Moritz A., et al (2013). Nitric oxide production, systemic inflammation and lipid metabolism in periodontitis patients: possible gender aspect. J Clin Periodontol;40(10):916–23.

[2]Hastie, T.J. and Tibshirani, R.J. (1990). Generalized additive models 43. CRC Press.

[3]Madsen C.M., Varbo A., Tybjærg-Hansen A., Frikke-Schmidt R. and Nordestgaard B.G. (2017). U-shaped relationship of HDL and risk of infectious disease: two prospective population-based cohort studies. European Heart Journal 0, 1–11.

[4]R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[5]Rigby R.A. and Stasinopoulos D.M. (2001). "The GAMLSS project: a Flexible Approach to Statistical Modelling." In B Klein, L Korsholm (eds.), "New Trends in Statistical Modelling: Proceedings of the 16th International Workshop on Statistical Modelling," pp. 249–256. Odense, Denmark.

[6]Rigby R.A. and Stasinopoulos D.M. (2005). Generalized Additive Models for Location, Scale and Shape. Applied Statistics, 54, 507–554.

[7]Stasinopoulos D. and Rigby R. Generalized Additive Models for Location Scale and Shape (GAMLSS). R. Journal of Statistical Software December 2007, Volume 23, Issue 7.

Organized Session 4

Advances in Applications of Inference Methodologies

Organizer: Miguel Fonseca (Portugal)

Modelling the dominant height growth of eucalyptus plantations in Portugal

Ayana Mateus[1,2] **and Margarida Tomé**[3,4]

Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Portugal[1]
Centro de Matemática e Aplicações (CMA), Universidade Nova de Lisboa[2]
ISA- Instituto Superior de Agronomia, Universidade de Lisboa[3]
Centro de Estudos Florestais (CEF), Universidade de Lisboa[4]

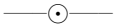
E-mail address: *amf@fct.unl.pt; magatome@isa.ulisboa.pt*

Abstract

Eucalyptus globulus Labill. is one of the most important economic forest species in Portugal, occupying an area of $875,000ha$ of a total forest area of $3,346,000ha$. It is a fast-growing species that is mainly used commercially by the pulp industry. The objective of the research reported here was to model the dominant height growth of eucalyptus plantations in Portugal, where the dominant height is the mean height of 100 thickest trees per hectare. Modeling dominant growth in height involves two problems: estimating the site index, which is defined as the dominant height at some fixed base age, given height at some other age, and estimating height at some desired age, given height at base age, whose objective is obtaining a height prediction equation, having a prior knowledge about the site index. These two problems may be modeled either by individual functions, one for each problem or we may use one only function that estimates height at any desired age, given height at any other known age. When using individual functions, height is assumed to be measured without error when on the right-hand side of the equation (regressor variable), but with error when used on the left-hand side (dependent variable). This latest assumption causes a bias in the estimation of the parameters of the curves. As a result neither the height prediction equation nor the site index prediction equation will have a shape that represents the true functional relationship between height and age across different values of site index. When a unique equation is used, the purpose is to simultaneously optimize the regression of height at base age on height at a given age and of height at a given age on height at base age and avoid parameter estimation

bias. This can be achieved by using all possible sets of differences of observations, built from the available measurements, where we use all possible differences in both directions.

Keywords: Dominant height, Forest planning, Generalized Least Squares, Nonlinear model.



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/00297/2019 (Centro de Matemática e Aplicações).

References

[1]A. Gallant, *Nonlinear Statistical Models*, John Wiley & Sons, 1987.
[2]G.A.F. Seber and C.G. Wild C.G., *Nonlinear Regression*, John Wiley & Sons, 1989.
[3]Mateus A. , Tomé M.. Generalized Least Squares Method in Nonlinear Equations. In T. E. Simos, G. Psihoyios, Ch. Tsitouras and Z. Anastassi (eds.), AIP Conf. Proc. 1479 (ICNAAM 2012), 1118-1120, (2012) doi: 10.1063/1.4756344.

Multiple Additive Models

Sandra S. Ferreira[1,2], Patrícia Antunes[2], Célia Nunes[1,2], Dário Ferreira[1,2] and João T. Mexia[3]

[1] *Department of Mathematics*
University of Beira Interior
[2] *Center of Mathematics and Applications*
University of Beira Interior
[3] *Centre of Mathematics and its Applications*
FCT, New University of Lisbon

Abstract

We consider d individual additive models. These models constitute a multiple model, which will be additive and correspond to d treatments of a base design. Using a classic result on cumulant generation function we show how to obtain least square estimators for cumulants and generalized least squares estimators for estimable vectors in the individual models. We will use the structure of this models to ensure that the vectors of homologue components are approximately homoscedastic which enables us to apply ANOVA.

Keywords: Additive Models, ANOVA, Cumulants, Moments.

Acknowledgements

This work was partially supported by national founds of FCT-Foundation for Science and Technology under UID/MAT/00212/2019 and UID/MAT/00297/2019.

References

[1]Akhiezer, N. I. (1965), The Classical Moment Problem, N. Kemme, Transl., Hafner, New York.

[2]Balakrishnan, N., Johnson,N. L. and Kotz, S. (1998), A note on relationships between moments, central moments and cumulants from multivariate distributions, Statistics & probability letters, **39**: 49–54, Elsevier.

[3]Dodge Y. and Valentin R. (1999), The Complications of the Fourth Central Moment, The American Statistician, **53** (3): 267–269.

[4]Harvey, J. R., (1972), On Expressing Moments in Terms of Cumulants and Vice Versa. The American Statistician, **26** (4): 38–39.

[5]Kendall, M. and Stuart, A. (1958). The Advanced Theory of Statistics. Charles Griffin & Company.

[6]Landau, H. J. (1987), Moments in Mathematics, Proceedings of Symposia in Applied Mathematics, **37**, Amer. Math. Soc., Providence, RI.

[7]Mexia, J. T (1987), Multi-treatment regression designs. Trabalhos de investigação N. 1, Departamento de Matemática, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa.

[8]Ruppert, D. (1987) What is Kurtosis? An Influence Function Approach, The American Statistician, **41**: 1–5.

[9]Schultze, J. and Steinebach, J. (1996). On least squares estimates of an exponential tail coefficient. Statistics & Decisions, **14**, 353–372.

[10]Smith, P. J. (1995), A Recursive Formulation of the Old Problem of Obtaining Moments from Cumulants and Vice Versa. The American Statistician, **49** (2): 217–218.

Mixed models with independent terms

Dário Ferreira[1,2], **Sandra S. Ferreira**[1,2], **Célia Nunes**[1,2] and **João T. Mexia**[3]

Center of Mathematics and Applications, University of Beira Interior, Covilhã, Portugal[1]
Department of Mathematics, University of Beira Interior, Covilhã, Portugal[2]
Center of Mathematics and its Applications, Faculty of Science and Technology, New University of Lisbon, Monte da Caparica, Portugal[3]

E-mail address: *dario@ubi.pt; sandraf@ubi.pt; celian@ubi.pt; jtm@fct.unl.pt*

Abstract

This talk presents an estimation method for the random effect parameters and the variance components in linear mixed models. In particular, least squares estimators and the corresponding confidence regions, based on the estimation of quantiles, are considered. The difference in this method is that it is only assumed that the random effects parameters have null mean vectors and distributions with known dispersion parameters and second order moments. A numerical example is provided to illustrate the proposed method.

Keywords: Dispersion parameters, Inference, Linear mixed models.

Acknowledgements

This work was partially supported by national founds of FCT-Foundation for Science and Technology under UID/MAT/00297/2019 and UID/MAT/00212/2019.

References

[1]Brown, H. and Prescott, R. (1999) Applied Mixed Models in Medicine, Wiley, New York.
[2]Demidenko, E. (2013) Mixed Models: Theory and Applications with R, Wiley, New York.
[3]Searle, S.R. (1995) An overview of variance component estimation. Metrika 42, 215–230.
[4]Schott, J.R. (1997). Matrix Analysis for Statistics. John Wiley & Sons, Inc, New York.

From a napkin to a most expected book

Ricardo Covas[1,2,3]

EDP - Energias de Portugal, S.A., Portugal[1]
Department of Mathematics and Physics, Management School of Tomar, Polytechnic Institute of Tomar, Portugal[2]
Center of Mathematics and Applications, Faculty of Sciences and Technology, NOVA University of Lisbon, Portugal[3]

E-mail address: *ricardo.covas@edp.com*

Abstract

I've met Professor Mexia for the first time in 1996. I still remember the first words he gave to the FCT Math rookies, who had just arrived from high school: "Welcome... I have just been working on the most interesting problem concerning Actuarial Sciences..." And he went on describing and giving hints to solve the problem. That's Professor Mexia! Always sharing and teaching to others what, for him, is the most exciting subject... science.

Happy birthday with some stories and photos!

Keywords:statistical inference, actuarial sciences, mathematical statistics.

Acknowledgements

The author would like to state that his work was partially supported by the Portuguese Foundation for Science and Technology through the project UID/MAT/00297/2013 (Center of Mathematics and Applications).

Organized Session 5

Times Series Modelling and Applications

Organizers: Rosário Ramos (Portugal)

Modeling monthly temperature data in European Cities

Marco Costa[1,2] **and Magda Monteiro**[1,2]

ESTGA - Águeda School of Technology and Management, University of Aveiro[1]
CIDMA - Center for Research & Development in Mathematics and Applications, University of Aveiro[2]

E-mail address: *marco@ua.pt; msvm@ua.pt*

Abstract

According to the United Nations Intergovernmental Panel on Climate Change (IPCC), by 2012, the average global surface temperature was 0.85°C higher than in 1880.

In order to mitigate climate change, we must reduce or prevent the greenhouse gases emissions that result from human activities, since is very likely that warming is due to the observed increase in atmospheric greenhouse gas concentrations. In addition, the analysis of the evolution of climate variables, since we have records, can contribute to the establishment of appropriate goals and their better monitoring.

Since the warming phenomenon must be monitoring in a smaller scale, a better understanding of the evolution of temperature series in Europe can contribute to a more efficient monitoring and identification of climate change patterns.

In this work, we presented time series of monthly average temperatures in several European locations which were statistical analyzed and modeled with appropriated statistical models. Data set comprises sixteen long-term time series of monthly mean temperatures, in °C, from January 1900 until February 2017, available at the Climate Data Online.

Based in a linear regression modeling, all estimates of the linear trends are positive; the smallest slope is equal to 0.66°C per century in Dublin, Ireland, and the largest is equal to 1.90°C in Vilnius, Lithuania. This means that there is a climate change process in all European cities, although with different rates of warming.

Additionally, it is proposed a linear state space model to model monthly temperature times series with a deterministic seasonal component plus a stochastic slope.

In many applications in Engineering, Economics or environmental problems the parameters of a state-space model are estimated by maximum likelihood with Gaussian distribution via the Newton-Raphson method or, more often, by the EM algorithm. Since the normality of errors are not a suitable assumption in several time series, parameters of the state space form are estimated by distribution-free estimators based on the generalized method of moments. This last option was adopted in this work.

Results show that all significant smoother prediction $\widehat{\tau}_{t|n}$ levels have positive values with higher incidence in recent decades. These results are consistent with previous works that identified the period after 1980's with an unprecedented warming in the Central Europe.

Keywords: climate change, monthly data, temperature data, time series analysis, state space modeling.

Acknowledgements

Authors were partially supported through Portuguese funds by the Center for Research and Development in Mathematics and Applications (CIDMA) and the Portuguese Foundation for Science and Technology (Fundação para a Ciência e a Tecnologiaİ [FCT]), within project UID/MAT/04106/2019.

References

[1]Alpuim, T. and El-Shaarawi, A. (2009), Modeling monthly temperature data in Lisbon and Prague, *Environmetrics*, **20**: 835–852.

[2]Costa, M. and Monteiro, M. (2016). Bias-correction of Kalman filter estimators associated to a linear state space model with estimated parameters, *Journal of Statistical Planning and Inference* **176**: 22 – 32.

[3]Costa, M. and Monteiro, M. (2017) Statistical Modeling of an Air Temperature Time Series of European Cities, In Advances in Environmental Research, vol. 59, ed. Justin A. Daniels, 213 - 236. ISBN: 978-1-53612-701-0. New York: Nova Science Publishers.

[4]Costa M. and Monteiro M. (2018). A periodic mixed linear state-space model to monthly long-termtemperature data. *Environmetrics* e2550 https://doi.org/10.1002/env.2550

Change-point detection to model recent trends in European cities temperature time series

Magda Monteiro[1,2] **and Marco Costa**[1,2]

ESTGA - Águeda School of Technology and Management, University of Aveiro[1]
CIDMA - Center for Research & Development in Mathematics and Applications, University of Aveiro[2]

E-mail address: *msvm@ua.pt; marco@ua.pt*

Abstract

The Intergovernmental Panel on Climate Change (IPCC) report of 2018, [2], refers that Human-induced warming reached approximately 1^oC above pre-industrial levels in 2017, increasing at 0.2^oC per decade (high confidence). The effects of global warming are already visible all over the world. The increase of droughts, floods, severe storms, and other weather catastrophes is putting vulnerable human and biological populations at risk. However, global warming is not reflected in the same way around the world and this phenomenon must be monitored in a more small scale, for instance at a Europe cities level. Is in this context that the statistical analysis of local temperature time series, such as cities, has a special interest in order to monitor temperature rise. Suitable statistical time series models must be developed in order to accommodate particular characteristics of this type of correlated data.

The application of state-space models to monthly long-term time series of air temperatures in Lisbon, Porto and Coimbra, Portugal, [3], allowed to conclude that there are different monthly warming rates, beeing the greatest annual mean rise found in Porto with 2.17^oC whereas in Lisbon and Coimbra were, respectively, 0.62^oC and 0.55^oC per century.

Costa and Monteiro, [4], applied state space models associated to the Kalman filter and a clustering procedure of smoother trend levels in order to investigate patterns on the temperature rise in Europe, to long temperature time series. The temperature rise rates in Europe seem to have increased in the last decades when compared with longer periods, such as a century, so it is important to find the change-point and use the information from there to get accurate temperature rise rates.

In this work we will analyze monthly average of air temperatures time series in several European locations. Data set comprises sixteen long-term time series of monthly mean temperatures, in ^{o}C, from January 1900 to March 2019 (available at the Climate Data Online CDO).

Keywords: Air temperature, Climate change, State-space models, Change-point detection.

Acknowledgements

Authors were partially supported through Portuguese funds by the Center for Research and Development in Mathematics and Applications (CIDMA) and the Portuguese Foundation for Science and Technology (Fundação para a Ciência e a Tecnologia" [FCT]), within project UID/MAT/04106/2019.

References

[1]A Alpuim, T. and El-Shaarawi, A. (2009), Modeling monthly temperature data in Lisbon and Prague, *Environmetrics*, 20: 835–852.

[2]Allen, M.R., O.P. Dube, W. Solecki, F. Aragón-Durand, W. Cramer, S. Humphreys, M. Kainuma, J. Kala, N. Mahowald, Y. Mulugetta, R. Perez, M.Wairiu, and K. Zickfeld. (2018) Framing and Context. *In: Global Warming of 1.5^{o}C. An IPCC Special Report on the impacts of global warming of 1.5^{o}C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty* [Masson-Delmotte, V., P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J.B.R. Matthews, Y. Chen, X. Zhou, M.I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, and T. Waterfield (eds.)]. In Press.

[3]Costa, M. and Monteiro, M. (2018) A Periodic Mixed Linear State Space Model to Monthly Long-term Temperature Data. *Environmetrics* (0): 1–20. https://doi.org/10.1002/env.2550.

[4]Costa, M. and Monteiro, M. (2017) Statistical Modeling of an Air Temperature Time Series of European Cities. *In Advances in Environmental Research*: 213–236, New York, Nova Science Publishers.

[Jarušková(2018)]Jarušková, D. and Antoch, J. (2018). Changepoint analysis of Klementinum temperature series. Environmetrics. 2019;e2570. https://doi.org/10.1002/env.2570.

[5]Jarušková, D. (1997). Some problems with application of change-point detection methods to environmental data. Environmetrics, 8:469–483.

Trend analysis of climate variables: the example of the Sea Ice Index

M.Rosário Ramos[1,2] **and Clara Cordeiro**[3,4]

Universidade Aberta, Portugal[1]
Centro de Matemática, Aplicações Fundamentais e Investigação Operacional(CMAF-CIO), Universidade de Lisboa, 1749-016 Lisboa, Portugal[2]
Faculdade de Ciências e Tecnologia da Universidade do Algarve[3]
Centro de Estatística e Aplicações, Universidade de Lisboa, 1749-016 Lisboa, Portugal[4]

E-mail address: *MariaR.Ramos@uab.pt; ccordei@ualg.pt*

Abstract

Research on trend analysis is substantial and has originated a variety of methods with the objective of testing for trend in time series. Versions of the tests that use resampling techniques have shown to be versatile and its use is now generalized due to the computational progress

In this study we analyse the time series of a climate variable, namely the Sea Ice Extent Index in the Northern (NH) and Southern (SH) Hemispheres, through its records available by NSIDC: National Snow and Ice Data Center (USA). Sea Ice Index is a source for consistently processed ice extent and concentration images and data values since 1979. The research on this subject is part of the problem of Climate Change. The analysis of NH and SH time series is performed and compared on the base of two trend tests, the t-test and the Mann-Kendall test, applied in its original form and combined with a resampling technique. Data are monthly Sea Ice Index from November 1979 to March 2017. In this mixed approach, the seasonal component is estimated through STL - Seasonal Decomposition of Time Series by Loess, with non-constant seasonality. The comparison between methods is then performed, using resampling to estimate the test statistic and its significance. The order of autocorrelation is estimated by the best fitting model obtained through the AIC information criterion.

Keywords: trend analysis, sieve bootstrapp, STL decomposition, Sea Ice Index

Acknowledgements

Clara Cordeiro's research is partially supported by FCT- Fundação para a Ciência e Tecnologia, through the project FCT Portugal UID/MAT/00006/2019

References

[1]R.B. Cleveland, W.S. Cleveland, J.E. McRae, and I.Terpenning. (1990) Stl: A seasonal-trend decomposition procedure based on loess. *J. Off. Stat* **6** (1): 3–73.

[2]Noguchi, K., Gel, Y.R., Duguay, C.R. (2011) Bootstrap-based tests for trends in hydrological time series, with application to ice phenology data. *J. Hydrology.* **410** 150–161.

[3]Ramos, M.R., Cordeiro C. (2014) Trend Tests: a Tendency to resampling. *Proceedings of ITISE 2014-International Work Conference on Time Series*, 322–328, Granada, Spain, ISBN 978-84-15814-97-9.

[4]Fetterer, F., K. Knowles, W. Meier, and M. Savoie. (2016), updated daily. Sea Ice Index, Version 2. Boulder, Colorado USA. NSIDC: National Snow and Ice Data Center. doi: http://dx.doi.org/10.7265/N5736NV7. [acessed 08/04/17].

[5]McLeod, A.I.(2011) kendall: Kendall rank correlation and Mann-Kendall trend test. R package version 2.2, http://cran.r-project.org/package=kendall

[6]Thorsten Pohlert.(2016) trend: Non-Parametric Trend Tests and Change-Point Detection. R package version 0.2.0. http://cran.r-project.org/package=trend

[7]Hyndman RJ. (2017) forecast: Forecasting functions for time series and linear models. R package version 8.0, http://github.com/robjhyndman/forecast.

Organized Session 6

Statistical Models and Applications

Organizers: Teresa Oliveira (Portugal)

Individual growth mixed models with stochastic differential equations

Carlos A. Braumann[1,2], **Patrícia A. Filipe**[1,2] **and Gonçalo Jacinto**[1,2]

Centro de Investigação em Matemática e Aplicações, Instituto de Investigação e Formação Avançada, Universidade de Évora[1]
Departamento de Matemática, Escola de Ciências e Tecnologia, Universidade de Évora[2]

E-mail address: *braumann@uevora.pt; pasf@uevora.pt; gjcj@uevora.pt*

Abstract

Common growth curves for the weight $X(t)$ of an animal at age t can be described as the solution of a differential equation of the form $dY(t) = \beta(\alpha - Y(t))dt$, where $Y(t) = h(X(t))$ and h is an appropriate strictly increasing C^1 function, $\alpha = h(A)$ (A = maturity weight of the animal), and $\beta > 0$ is a rate of approach to maturity. Adjustment to data was usually done through non-linear regression, an inappropriate methodology that ignores the growth dynamics and the influence of environmental fluctuations on it. We use instead stochastic differential equations (SDEs) models $dY(t) = \beta(\alpha - Y(t))dt + \sigma dW(t)$, where $W(t)$ is a standard Wiener process and σ is an intensity parameter of the fluctuations.

We have previously (see [1], [2], [3]) studied estimation, prediction and optimization issues using cattle weight data from females of Mertolengo cattle breed.

In the present work, we have adjusted and applied the methodologies to the weight data of males of Mertolengo cattle breed and Alentejana cattle breed. Since model parameters may vary from animal to animal and that variability can be partially explained by their genetic differences, we introduce the extension of the study to SDE mixed models. These mixed models incorporate the individual genetic values that are available at the databases of the producer associations.

Keywords: application to cattle data, genetic values, individual growth models, mixed models, stochastic differential equations.

Acknowledgements

The authors belong to the research centre Centro de Investigação em Matemática e Aplicações, Universidade de Évora, supported by FCT (Fundação para a Ciência e a Tecnologia, Portugal, project UID/MAT/04674/2019).

References

[1]Filipe, P. A.; Braumann, C. A., Brites, N. M. and Roquete, C. J. (2010). Modelling animal growth in random environments: An application using nonparametric estimation. *Biometrical Journal* **52(5)**: 653–666.

[2]Filipe, P. A., Braumann, C. A. and Roquete, C. J. (2012). Multiphasic individual growth models in random environments. *Methodology and Computing in Applied Probability* **14(1)**: 49–56.

[3]Filipe, P. A., Braumann, C. A. and Carlos, C. (2015). Profit optimization for cattle growing in a randomly fluctuating environment. *Optimization* **64(6)**: 1393–1407.

The Extended Skew-Normal Distribution and the Product of two Normal Variables.

Antonio Seijas-Macias[1,2,3], **Amílcar Oliveira**[2,4] **and Teresa Oliveira**[2,4]

Universidade da Coruña (Spain)[1]
Centro de Estatística e Aplicações, Faculdade de Ciências, Universidade de Lisboa - Portugal[2]
Universidad Nacional de Educación a Distancia (UNED) - Spain[3]
Universidade Aberta (Portugal)[4]

E-mail address: *antonio.smacias@udc.gal*

Abstract

The product of two normal variables is an open problem from the early years in the XXth century. First approaches of the problem were focused into the product of two standard normal variables (with and without correlation), Bessel type functions were used as functions to approximate the product (Craig, 1936), although presents some problems, especially for the value zero. Other approaches consider the Pearson Type function (Aroian, 1947) or other statistical distributions. Finally, other authors (Ware and Lad, 2003, Glenn et al, 2004, Cui et al. 2016) use the improvement of the computational methods in order to approximate numerical solutions for the integrals of the distribution. But until now, have no successful solution to estimate the product of two normally distributed variables. Nonetheless, the moment-generating function of the product of two normally distributed variables is useful in order to calculate the statistics of the distribution of the product, especially, skewness and kurtosis. We have observed that values of skewness and kurtosis of the product don't coincide with traditional values of the normal distribution: skewness null, and kurtosis three (zero, when we consider the excess of kurtosis).

The Skew-normal distribution is a generalization of the normal distribution in order to consider the existence of skewness, it was developed by Azzalini in 1985 (Azzalini, 2014). The objective is to modeling data with the presence of a certain level of asymmetries. Further developments, in more recent, produce new generalizations like the Extended Skew-normal

distribution (Azzalini, 2014). This new distribution is determined by four parameters: location, scale, shape, and truncation. These distributions are potentially useful for the data presenting high values of skewness and kurtosis. Applications of this type of distributions are very common in the model of economic data, especially when asymmetric models are underlying the data. Although, we have four parameters, only two: shape and truncation determinate the value of skewness and kurtosis of the Extended Skew-Normal distribution.

We focus our work into the study of the relationship between the product of two normally distributed variables and the Extended Skew-Normal (ESN) distribution in order to use a normal approach to the product but considering the existence of levels of skewness and kurtosis, no traditionally commons values for them in Normal Distribution. Our approach has two stages: the first stage, we determinate the values of shape and truncation parameters of the ESN corresponding the values of skewness and kurtosis of the product; at the second stage we determinate the values of the other two parameters: location and scale, using the mean and variance of the product. Although, we have obtained a good approach for the mean and the variance; unfortunately, the approximation of skewness and kurtosis is not so good in every case.

We have analyzed the influence of the parameters of the two normal distributions involve in the product. We have concluded that the presence of one the normal variables with a value for the inverse of the coefficient of variation greater than one corresponds to a better approach than we have two normal variables with values of the inverse of the coefficient of variation less than one. Generally, this situation produces a product with a small skewness and a high kurtosis, and then, ESN Distribution fits no good.

Keywords: Extended Skew-Normal, Inverse of the Coefficient of Variation, Kurtosis, Product Two Normal Variables, Skewness, .

Acknowledgements

Research partially supported by National Funds through FCT - Fundação para a Ciência e a Tecnologia, project UID/MAT/00006/2019 (CEAUL).

References

[1]Aroian, L. A. (1947) The probability function of the product of two normally distributed variables. *The Annals of Mathematical Statistics*, **18**(2), 265-271.

[2]Azzalini, A. (2014) *The Skew-Normal and Related Families*. Cambridge University Press, Cambridge U.K.

[3]Craig, C. C. (1936) On the frequency of the function xy. *The Annals of Mathematical Statistics*, **7**, 1-15.

[4]Cui, G., Yu, X. , Iommelli, S. and Kong, L. (2016) Exact distribution for the product of two correlated gaussian random variables. *IEEE Signal Processing Letters*, **23** (11), 1662-1666.

[5]Glen, A. G., Lemmis, L. M. and Drew, J. H. (2004) Computing the distribution of the product of two continuous random variables. *Computational Statistics & Data Analysis*, **44**, 451-464.

[6]Ware, R. and Lad, F. (2003) *Approximating the distribution for sums of products of normal variables*. Technical report, The University of Queensland.

Structural Equation Modelling - some applications with R tools

Conceição Leal[1,2], **Teresa Oliveira**[1,2] **and Amílcar Oliveira**[1,2]

Universidade Aberta[1]
CEAUL[2]

E-mail address: *conceicao.leal2010@gmail.com; Teresa.Oliveira@uab.pt; Amilcar.Oliveira@uab.pt*

Abstract

Structural Equation Modelling (SEM) is a collection of multivariate statistical techniques, which, in most cases, are used to formulate, adjust, and test a wide variety of models. The main objective of methodology is to evaluate the degree in which a proposed theoretical model is supported by data, making it a potential driving force for knowledge.

The theoretical developments in SEM and in its applications led to the expansion of software capable of dealing with problems of increasing complexity and with large volumes of data. Some of the most popular software to deal with SEM are commercials software like Mplus, AMOS or LISREL. Software R is an alternative to these software because it has a great diversity of tools to deal with SEM, that can compete with most of the commercial software.

A common problem in SEM concerns the frequent existence of missing data, a problem that is transversal to all types of data analysis. Omitting data can reduce the statistical power of a study and produce biased estimates, leading to invalid conclusions. SEM is not immune to these problems either.

In this work, we will present some packages of R that illustrate the potential of this software to deal with various aspects of the implementation of SEM, with special emphasis on resources available to handle with missing data, specifically for SEM analysis. For this purpose, we will present an application to data of PISA 2012, concerning the literacy in Mathematics of Portuguese young people.

Keywords: Structural Equation Modelling, Missing data, SEM package of R, PISA 2012.

Acknowledgements

Research partially supported by National Funds through FCT - Fundação para a Ciência e a Tecnologia, project UID/MAT/00006/2019 (CEAUL).

References

[1]Allison, P. D. (2003). Missing data techniques for structural equation modeling. Journal of abnormal psychology, 112(4), 545.

[2]Beaujean, A. A., & Beaujean, M. A. A. (2018). Package 'BaylorEdPsych'. Version 0.5

[3]Boker, S., Neale, M., Maes, H., Wilde, M., Spiegel, M., Brick, T., ... & Mehta, P. (2011). OpenMx: an open source extended structural equation modeling framework. Psychometrika, 76(2), 306–317.

[4]Buuren, Stef van, Groothuis-Oudshoorn, Karin (2011). mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software, 45(3), 1–67. URL https://www.jstatsoft.org/v45/i03/.

[5]Epskamp, Sacha (2017). semPlot: Path Diagrams and Visual Analysis of Various SEM Packages' Output. An R package version 1.1

[6]Fox, J., Nie, Z., Byrnes, J., Culbertson, M., DebRoy, S., Friendly, M., ... & Fox, M. J. (2017). Package 'sem'.

[7]Gelman, A., Hill, J., Su, Y. S., Yajima, M., Pittau, M., Goodrich, B., ... & Goodrich, M. B. (2015). Package 'mi'.

[8]Graham, J. W., & Coffman, D. L. (2012). Structural equation modeling with missing data. In R. H. Hoyle (Ed.), Handbook of structural equation modeling (pp. 277-295). New York, NY, US: Guilford Press.

[9]Hallquist, M. N., & Wiley, J. F. (2018). MplusAutomation: An R Package for Facilitating Large-Scale Latent Variable Analyses in M plus. Structural Equation Modeling: A Multidisciplinary Journal, 1-18

[10]Honaker, J., King, G., & Blackwell, M. (2011). Amelia II: A program for missing data. Journal of statistical software, 45(7), 1–47.

[11]Jamshidian, M., Jalal, S. J., & Jansen, C. (2014). MissMech: an R package for testing homoscedasticity, multivariate normality, and missing completely at random (MCAR).

[12]Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. (2018).semTools: Useful Tools for Structural Equation Modeling. R package version 0.5–0

[13]Mair, P., Wu, E., & Wu, M. E. (2011). R Package 'REQS'.

[14]Oberski, D. (2014). lavaan. survey: An R package for complex survey analysis of structural equation models. Journal of Statistical Software, 57(1), 1–27. URL http://www.jstatsoft.org/v57/i01/.

[15]Peugh, J. L., & Enders, C. K. (2004). Missing Data in Educational Research: A Review of Reporting Practices and Suggestions for Improvement. Review of Educational Research, 74(4), 525–556.

[16]Rosseel, Y. (2018). The lavaan tutorial. Department of Data Analysis: Ghent University.

Development of a collaborative game to assess statistical literacy: an adventure in medieval times

Henning, E.[1,2], **Santos, L.M.**[1,2], **Siple, I.**[1,2] **and Oliveira, T.**,[3,4]

Santa Catarina State University[1]
College of Technological Sciences[2]
Universidade Aberta[3]
CEAUL[4]

E-mail address: *elisa.henning@udesc.br; luciane.mulazani@udesc.br; ivanete.siple@udesc.br; teresa.oliveira@uab.pt*

Abstract

In the field of mathematics in basic education, the teaching of statistics occupies an important place in knowledge construction that allows the exercise of citizenship through a critical reading of the world. It is understood that children, from the earliest years at school, can develop statistical thinking and thus, be literate in statistics. This thinking is present in the Brazilian curricular guidelines. In this context, it is essential to develop actions that contribute effectively in all aspects of the teaching learning process, from construction to evaluation of results. In this paper, we present some partial results of research that develops resources to assess statistical literacy of students in the early years of elementary school. We proposed a collaborative game in RPG form that helps teachers to identify if the child understood what is expected of the contents of statistics according to the Brazilian curricular guidelines. The game developed involves an adventure in the medieval era in which the children advance together through various stages to find an orchid and save the king's life. Through the game, we intend to verify the children's knowledge about bar graphs.

Keywords: Statistics Education, Elementary School, Collaborative Games.

Acknowledgements

To CNPq (grant 427352/2016-9) for the financial support. We also thank the students Isabela Chagas Luiz, Livia de Souza Monteiro, Pietro Guesser e Ana Maria Henning Codeço who assisted this research.

References

[1]Bidgood P.; Hunt, N.; Jolliffe, F. (2010) Assessment Methods in Statistical Education: An International Perspective, Wiley, New York,

[2]Elizabeth Brondos Fry (2015) Teaching Bits: Statistics Education Articles from 2014 and 2015, Journal of Statistics Education, 23:1.

[3]Boyle, E. A. et al. (2014). A narrative literature review of games, animations and simulations to teach research methods and statistics. Computers & Education 74: 1–14.

Organized Session 6 (Poster Session)

Applications of Statistics Mathematics and Computation

Organizers: Nuno Brites (Portugal)

Harvesting policies in random environment: logistic growth and profit optimization

Nuno M. Brites[1] **and Carlos A. Braumann**[1,2]

Centro de Investigação em Matemática e Aplicações, Instituto de Investigação e Formação Avançada, Universidade de Évora[1]
Departamento de Matemática, Escola de Ciências e Tecnologia, Universidade de Évora[2]

E-mail address: *brites@uevora.pt; braumann@uevora.pt*

Abstract

We model fish population growth subjected to harvesting taking into account the effect of the environmental random fluctuations on the growth dynamics. For that we use stochastic differential equations (SDE). In particular, we consider a logistic model plus random environmental perturbations for the natural growth and subtract a harvesting yield term based either on a variable or on a constant fishing effort.

There is previous work on the optimal design of the harvesting policy with the purpose of maximizing the expected accumulated profit (discounted by a depreciation rate) over a finite time horizon. We consider a profit structure which includes revenues to be proportional to the yield and costs to be quadratic on the effort per unit time. The harvesting efforts of the optimal policies vary with the randomly varying population size and such policies can, under certain conditions, even be of bang-bang type. These policies, borrowed from the financial world where data is abundant, are not applicable to harvesting since they require constant evaluation of the population size and they have frequent random changes in harvesting effort incompatible with the logistics of fishing.

Our approach, based on sustainable and applicable fishing policies with constant effort, leads to sustainability of the population and to a stationary distribution (see [1]) of the population size and do not require evaluation of population size. We determine the constant harvesting effort policy that optimizes the expected sustainable profit per unit time and check what we lose profitwise when using this policy instead of the optimal inapplicable policy with variable effort. Applying Monte Carlo simulations and using population parameters based on real data, we show that our approach is almost as profitable as the first (see, for instance, [2], [3], [4], [5] and [6]).

Keywords: stochastic differential equations, fisheries management, profit optimization, logistic growth.

Acknowledgements

The authors belong to the Centro de Investigação em Matemática e Aplicações, Universidade de Évora, a research centre supported by FCT (Fundação para a Ciência e a Tecnologia, Portugal, ref. UID/MAT/04674/2019). The second author had a PhD grant from FCT (ref. SFRH/BD/85096/2012).

References

[1]Braumann, C. A., Variable effort fishing models in random environments, *Mathematical Biosciences*, 156, 1–19, 1999.

[2]Brites, N. M. e Braumann, C. A., Fisheries management in randomly varying environments: comparison of constant, variable and penalized efforts policies for the Gompertz model, *Fish. Res.*, 216, 196–203, 2019.

[3]Brites, N. M. e Braumann, C. A., Harvesting in a random varying environment: optimal, stepwise and sustainable policies for the Gompertz model, *Stat., Optim. Inf. Comput.*, 2019 (in press).

[4]Brites, N. M. e Braumann, C. A., Harvesting policies with stepwise effort and logistic growth in a random environment. In: âĂIJDynamical Systems in Biology and Natural SciencesâĂİ (Ventorino, E., Aguiar, M. A. F., Stollenwek, N., Braumann, C. A., Kooi, B. Pugliese, A., Editores), Springer-SIMAI series, 2019 (in press).

[5]Brites, N. M., Stochastic differential equation harvesting models: Sustainable policies and profit optimization, Ph. D. thesis, Universidade de Évora, Portugal, 2017.

[6]Brites, N. M. e Braumann, C.A., Fisheries management in random environments: Comparison of harvesting policies for the logistic model, *Fish. Res.*, 195, 238–246, 2017.

A note on models with commutative orthogonal block structure

Carla Santos[1,2], Cristina Dias[2,3], Célia Nunes[4] and João Tiago Mexia[2,5]

Polytechnic Institute of Beja, Portugal[1]
CMA - Center of Mathematics and its Applications, FCT- New University of Lisbon, Portugal[2]
Polytechnic Institute of Portalegre, Portugal[3]
University of Beira Interior and Center of Mathematics and Applications[4]
Department of Mathematics -FCT- New University of Lisbon, Portugal[5]

E-mail address: *carla.santos@ipbeja.pt; cpsd@ipportalegre.pt; celian@ubi.pt*

Abstract

In this work we focus on a special class of linear mixed models, named models with orthogonal block structure, in which the variance-covariance matrix is a linear combination of known pairwise orthogonal orthogonal projection matrices that sum to the identity matrix [5][6]. Models with orthogonal block structure allow optimal estimation for variance components of blocks and contrasts of treatments, however, aiming to obtain least square estimators that are the best linear unbiased estimators whatever the variance components, we must impose that the orthogonal projection matrix on the space spanned by the mean vector commutes with the matrices of the principal basis [2]. Resorting to B-matrices [9], we address an alternative to the commutativity condition introduced in [2].

Keywords: B-matrices, Commutativity, Jordan Algebra, Linear mixed models.

Acknowledgements

This work was partially supported by national founds of FCT - Foundation for Science andTechnology under UID/MAT/00297/2019 and UID/MAT/00212/2019.

References

[1]Carvalho, F. Mexia, J.T Santos,C. and Nunes, C. (2014) Inference for types and structured families of commutative orthogonal block structures. Metrik, 78: 337–372, doi: 10.1007/s00184-014-0506-8.

[2]Fonseca, M., Mexia, J. T. and Zmy?lony, R. (2008), Inference in normal models with commutative orthogonal block structure, Acta et Commentationes Universitatis Tartuensis de Mathematica 12: 3–16.

[3]Houtman, A.M. and Speed, T.P. (1983). Balance in designed experiments with orthogonal block structure. Ann. Statist 11(4): 1069–1983.

[4]Jordan, P., von Neumann, J. and Wigner, E. (1934). On the algebraic generalization of the quantum mechanical formalism. Annals of Mathematics 36: 26–64.

[5]Nelder, J.A. (1965a) The analysis of randomized experiments with orthogonal block structure I. Block structure and the null analysis of variance. Proceedings of the Royal Society, Series A, 283: 147–162.

[6]Nelder, J.A. (1965b) The analysis of randomized experiments with orthogonal block structure II. Treatment structure and the general analysis of variance. Proceedings of the Royal Society, Series A, 283: 163–178.

[7]Nunes, C., Santos, C. and 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, ISSN 0972-0502, doi: 10.1080/09720502.2008.10700581.

[8]Santos, C., Nunes, C. and Mexia, J. T. (2007) OBS, COBS and Mixed Models associated to commutative Jordan Algebra. In proceedings of 56th session of the International Statistical Institute. Lisboa, Portugal, August 22–29.

[9]Santos, C. (2012) Error orthogonal models. Structure, Operations and Inference. PhD thesis, Universidade da Beira Interior.

[10]Seely, J. (1971) Quadratic subspaces and completeness. Ann. Math. Stat.,42: 710–721.

[11]Seely, J. and Zyskind, G. (1971). Linear spaces and minimum variance estimators. The Annals of Mathematical Statistics, 42(2).

[12]Zmyslony, R. (1978) A characterization of Best Linear Unbiased Estimators in the general linear model. Lecture Notes in Statistics 2: 365-373.

A Time series analysis of road traffic accidents in Luanda

Manuel Alberto[1], Patrícia A. Filipe[2,3] and Dulce Gomes[2,3]

Instituto Superior Politécnico Internacional de Angola[1]
Centro de Investigação em Matemática e Aplicações, Instituto de Investigação e Formação Avançada, Universidade de Évora[2]
Departamento de Matemática, Escola de Ciências e Tecnologia, Universidade de Évora[3]

E-mail address: *d11837@alunos.uevora.pt; dmog@uevora.pt; pasf@uevora.pt*

Abstract

In this work, time series models are applied to explain and forecast the rate of traffic accidents, deaths and injureds in Luanda, Angola. Monthly Luanda data from 2002 to 2015 are used to fit models and to make predictions. The data was collected from Direcção Nacional de Viação e Trânsito (DNVT) and Gabinete de Estudos, Informação e Análise (GEIA) databases. Road accidents in Angola are currently one of the major causes of death in the country. Particularly Luanda, the capital, is the province that shows the highest rate in terms of accidents, deaths and injureds [4]. However, in recent years there has been a decrease in the accidents rate, with average growth rates of -6.73%, 0.19% and -2.54% for accidents, deaths and injureds respectively.

In early work [1] we have presented some results on modelling and predicting the rates of road accidents, deaths and injureds in Angola through classic seasonal ARIMA models (SARIMA), exploring and comparing two different approaches. One in which all observations are treated the same way, and the other that identifies outliers, taking into account its magnitude, and adjust SARIMA models excluding the significant outliers [2]. A Seasonal-Trend decomposition based on a locally-weighted regression smoothing (Loess) approach [3] was also applied. Since, Luanda is the capital and the most problematic area, here we intend to develop a more detailed analysis for this province adapting and applying similar methodology as used before. The best models were identified and used to fit and predict road accidents, deaths and

injureds in Luanda. The SARIMA models that take into account the extreme values revealed to fit and predict better than the pure SARIMA models time series of traffic accident data.

Keywords: Road accidents, Outliers, Time series, Seasonal ARIMA models, Decomposition.

Acknowledgements

The second and third authors belong to the research centre Centro de Investigação em Matemática e Aplicações, Universidade de Évora, supported by FCT (Fundação para a Ciênncia e a Tecnologia, Portugal, project UID/MAT/04674/2019).

References

[1]Alberto, M., Gomes, D. and Filipe, P. A. (2018). A time series approach to road accidents in Angola from 2002 to 2015. *International Journal of Applied Mathematics and Statistics*, **57**, 63 – 73.

[2]Chen, C. and Liu, L. (1993). Joint Estimation of Model Parameters and Outlier Effects in Time Series. *Journal of American Statistical Association*, **88**, 284 – 297.

[3]Cleveland, R. and Cleveland, W. and Mcrae, J. and Terpenning, I. (1990). STL: A Seasonal-Trend Decomposition Procedure Based on Loess. *Journal of Official Statistics*, **6**, 3 – 73.

[4]DNVT (2018). *Relatório sobre a Sinistralidade Rodoviária em Angola*. Direcção Nacional de Viação de Trânsito, Angola.

Three-dimensional velocity field for blood flow using the power-law viscosity function

Fernando Carapau[1,2]

Departamento de Matemática, Escola de Ciências e Tecnologia, Universidade de Évora[1]
Centro de Investigação em Matemática e Aplicações, Instituto de Investigação e Formação Avançada, Universidade de Évora[2]

E-mail address: *flc@uevora.pt*

Abstract

The three-dimensional model associated with blood flow where viscosity depends on shear-rate, such power-law type dependence, is a complex model to study in terms of computational optimization, which in many relevant situations becomes infeasible. In order to simplify the three-dimensional model and as an alternative to classic 1D models, we will use the Cosserat theory related with fluid dynamics to approximate the velocity field and thus obtain a one-dimensional system consisting of an ordinary differential equation depending only on time and on a single spatial variable, the flow axis. From this reduce system, we obtain the unsteady equation for the mean pressure gradient depending on the volume flow rate, Womersley number and the flow index over a finite section of the tube geometry. Attention is focused on some numerical simulations for constant and non-constant mean pressure gradient using a Runge-Kutta method and on the analysis of perturbed flows. In particular, given a specific data we can get information about the volume flow rate and consequently we can illustrate the three-dimensional velocity field on the constant circular cross-section of the tube. Moreover, we compare the three-dimensional exact solution for steady volume flow rate with the corresponding one-dimensional solution obtained by the Cosserat theory.

Keywords: shear-thinning fluid; Cosserat theory; one-dimensional model; volume flow rate; mean pressure gradient.

Extinction times for the logistic model with and without Allee effects

Clara Carlos[1,2], **Carlos A. Braumann**[2,3] **and Nuno M. Brites**[3]

Escola Superior de Tecnologia do Barreiro, Instituto Politécnico de Setúbal[1]
Departamento de Matemática da Universidade de Évora[2]
Centro de Investigação em Matemática e Aplicações, Instituto de Investigação e Formação Avançada, Universidade de Évora[3]

E-mail address: *clara.carlos@estbarreiro.ips.pt; braumann@uevora.pt; brites@uevora.pt*

Abstract

We propose a general model of population growth that incorporates the effect of random environmental fluctuations on the *per capita* growth rate, in the form of a stochastic differential equation. We study this stochastic model with and without Allee effects (see [1]). In the case where there are Allee effects, we consider two types, weak Allee effects (see [4]) and strong Allee effects (see [2]).

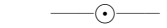
For these general models, in [5] we have proved the existence and uniqueness of solution in all these cases. In the cases of no Allee effects or weak Allee effects, we have also proved the existence of a stationary density and a zero probability of "mathematical extinction". In the case of strong Allee effects, we have proved that the probability of "mathematical extinction" is equal to one. However, we have shown that "realistic extinction" occurs with probability one for all cases.

Based on the results obtained in [3] for the first passage times and, in particular, for the extinction times, we obtain explicit expressions for the mean and standard deviation of the extinction times of this general model.

We consider next a particular case of the general model, which is the stochastic version of a very popular logistic-like deterministic model with Allee effects. However, we use a parametrization slightly different from the usual one, in order to facilitate comparisons with the logistic model by having a parameter A that "measures" the strength of the Allee effects and converging to the logistic model (model without Allee effects) when $A \to -\infty$.

For these particular logistic-like stochastic differential equation models, we show figures illustrating the behaviour of the mean and the standard deviation of extinction times for various combinations of parameters. This allows us to see the influence of Allee effects on the population extinction times by comparing extinction times under the model without Allee effects (the stochastic logistic model) and under the models with weak Allee effects and with strong Allee effects.

Keywords: Population growth, random environments, extinction times, stochastic differential equations, Allee effects.



Acknowledgements

The authors belong to research centre Centro de Investigação em Matemática e Aplicações UID/MAT/04674/2019, Universidade de Évora, supported by FCT-Fundação para a Ciência e a Tecnologia, Portugal.

References

[1]Allee, W. C.,Emerson, A. E., Park, O., Park, T. and Schmidt, K. P. (1949) *Principles of Animal Ecology*. Saunders, Philadelphia.

[2]Braumann, C. A. and Carlos, C. (2013) Allee effects in randomly varying environments. *Proceedings 2013 International Conference on Computational and Mathematical Methods in Science and Engineering* 304–307.

[3]Carlos, C., Braumann, C. A. and Filipe, P. A. (2013) Models of individual growth in a random environment: study and application of first passage times. *Advances in Regression, Survival Analysis, Extreme Values, Markov Processes and Other Statistical Applications*, Springer, Berlin 103–111.

[4]Braumann, C. A. and Carlos, C. (2015) Weak Allee effects population growth models in a random environment. *Proceedings 60th ISI World Statistics Congress* 2455–2460.

[5]Carlos, C. and Braumann, C. A. (2017) General population growth models with Allee effects in a random environment. *Ecological Complexity* **30**, 26–33.

Robust Portfolio Selection

Mickael Borges[1], **Loveleen Kaur**[2], **M. Teresa Godinho**[3] **and Fernando Teixeira**[3]

Université Montesquieu - BORDEAUX IV - IUT Périgueux[1]
Thomas More University, Mechelen[2]
Instituto Politécnico de Beja[3]

Abstract

The Robust Portfolio Selection Problem has been wildly studied, with contributions from various fields of mathematics and finance. In this paper, we revisit the classic Markowitz model [2], where we aim to find the portfolio with the lowest risk for a given level of expected return.

This problem can be modeled as a constrained optimization problem where a risk objective function has to be minimize considering the existence of a minimum expected return constraint. In the Markowitz model, the risk is defined as the variance of the portfolio return and the expected return as the mean of the probability distributions of the returns of the assets.

We implement the mean-variance framework of Markowitz [2], choosing the optimal portfolio weights in a two-step procedure. First, the mean and covariance matrix of the asset returns are estimated based on the observed datafor the first T periods. Under the assumption that observed return is i.i.d. normal, we calculate the standard estimates for the mean vector and the covariance matrix. Second, these sample estimates are then treated as if they were the true parameters, and then used to compute the optimal portfolio weights for period T+1.

This procedure was applied to data originated from the stock market value of the shares of 23 Portuguese companies, gathered in 517 consecutive days. By varying the expected return, we were able to obtain the Markovitz efficient frontier.

Keywords: Robust Portfolio, Constrained Optimization , Markovitz Model.

References

[1]Fabozzi, F.J., Huang, D. & Zhou, G. (2010) Robust portfolios: contributions from operations research and finance. Ann Oper Res 176: 191.
[2]Markowitz, H. M. (1952). Portfolio selection. Journal of Finance, 7, 77–91.
[3]Roman, S. (2004). Introduction to the Mathematics of Finance, Springer-Verlag.

Data modeling from an experiment with barley malt and some results

Iwona Mejza[1], **Katarzyna Ambroży-Deręgowska**[1]

Poznań University of Life Sciences, Poland[1]
Department of Mathematical and Statistical Methods

E-mail address: *iwona.mejza@up.poznan.pl; katarzyna.ambrozy@up.poznan.pl*

Abstract

The aim of this study is to present different ways of modeling and analyzing data related to malting. The malting industry is located between agriculture and the brewing industry, and is of great importance in the production of beer. The literature on the topic is very rich, and concerns primarily the practical, agricultural and brewing aspects. An important element of the preparation of malt-in pilsner, for example-is the grain of, among others, malting barley, coming from a field experiment lasting many years. The method of obtaining grain samples as experimental material is of fundamental importance in statistical analysis. The next stage is the malting of the obtained grains, carried out under laboratory conditions, for example in a brewery, using various methods of steeping, germination and drying (a three-step process). Germination time is counted from the end of the steeping cycle, and lasted 3-6 days in the described example. This paper is concerned chiefly with the theoretical aspects. Various models of observations are considered, in which 1) the split-block structure of the two-factor field experiment (cultivars and nitrogen fertilization) is taken into account, and 2) there is no knowledge about this structure. The models presented relate to observations from both field and brewery experiments, and include: 1) fixed effects of the factors (cultivars, nitrogen fertilization, germination time); 2) fixed or random effects of years ([2], [3]). The considerations are illustrated by a three-year experiment from the Agricultural Research Station in Paw?owice in Poland. The presented results relate to selected traits, such as extractivity and productivity, which are among the most important parameters determining the malt value, and thus the quality of the beer that is the final product [1].

Keywords: fixed effects model, mixed model, brewing barley, extractivity, malting productivity. .

References

[1]Blażewicz J., Liszewski M., Zembold-Gula A. (2007): Usability of Bishop formula in evaluation of malting quality of barley grain. Pol. J. Food Nutr. Sci. 57 4(A) : 37–40.

[2]Mejza I., Ambroży-Deręgowska K., Bocianowski J., Blażewicz J., Liszewski M., Nowosad K., Zalewski D. (2018): A method for analysis and interpretation of data from field and brewery trials. Colloquium Biometricum 48 : 53–63.

[3]Mejza I., Ambroży-Deręgowska K., Bocianowski J., Blażewicz J., Liszewski M., Nowosad K., Zalewski D. (2019): On modeling and analyzing barley malt data in different years. Biometrical Letters (DOI: 10.2478/bile-2019-0004).

Contributed Poster

Time Series Analysis for Maximum Wind Speed Forecasting

A. Manuela Gonçalves[1,2], **Cláudia Costa**[1] **and Marco Costa**[3]

Department of Mathematics, University of Minho, Guimarães, Portugal[1]
Center of Mathematics, University of Minho, Guimarães, Portugal[2]
Águeda School of Technology and Management and Center for Research and Development in Mathematics and Applications, University of Aveiro, Aveiro, Portugal[3]

E-mail address: *mneves@math.uminho.pt; claudiacosta_mf@hotmail.com; marco@ua.pt*

Abstract

Time series analyses of meteorological data have an increasing interest in many fields. Series are especially interesting for a better understanding of atmospheric phenomena in order to model them, to determine the climate of a geographical area, or to forecast possibilities of extreme weather occurrences. Predicting and forecasting weather has always been a difficult field of research analysis with a very slow progress rate over the years. Weather data consider the noises and outliers; therefore, investigation in this area may not be accurate. Wind is one of the weather variables which are very difficult to predict.

This work presents an environmental study of maximum wind speed at a particular site (a farm located in Vila Real County, in northern Portugal), in the field of agriculture irrigation and proposes models to forecast daily maximum wind speed data. This research investigates the forecasting performances of two forecasting methods: Box-Jenkins SARIMA models, and TBATS (Trigonometric Seasonal, Box-Cox Transformation, ARMA errors, Trend and Seasonal Components) models. These methods are chosen because of their ability to model trend and seasonal fluctuations present in environmental data, in particular when dealing bearing time series with complex seasonal patterns (multiple seasonal patterns).

Our data source are the records registered in the period from January 23rd, 2015 to August 11th, 2018, on a daily basis. The main goal is to forecast this environmental variable at a location (in this case, at the farm) where

there are historical observations but current measurements are not available (including various steps for forecasting (i.e., 10 days)).

Keywords: maximum wind speed, time series, SARIMA, TBATS, forecasting.

Acknowledgements

This research was partially funded by Portuguese funds through the Center for Research and Development in Mathematics and Applications (CIDMA) and the Portuguese Foundation for Science and Technology ("Fundação para a Ciência e a Tecnologia" - FCT), within the framework of project UID/MAT/04106/2019. This research was partially funded by Portuguese funds through the Portuguese Foundation for Science and Technology ("Fundação para a Ciência e a Tecnologia" - FCT), within the framework of project UID/MAT/00013/2013.

References

[1]Benth, J. S. and Benth, F. E. (2010) Analysis and modelling of wind speed in New York. *Journal of Applied Statistics* **37** (6): 893–909.

[2]Bradley, T. E., Kruse, J. B. and Schroeder, J. S. (2006) Time series analysis of wind speed with time-varying turbulance. *Environmetrics* **17**: 119–127.

[3]Hyndman, R. J., Koehler, A. B., Ord, J. K. and Snyder, R. D. (2008) *Forecasting with Exponential Smoothing: the State Space Approach*, Berlin, Springer.

[4]R Core Team. (2018) R: a language and environment for statistical computing. Version 3.5.0. Vienna, Austria: R Foundation for Statistical Computing. *URL https://www.R-project.org/*.

Statistics Analysis in the Management of a Local Public Company

A. Manuela Gonçalves[1,2], **Vítor Hugo Silva**[1,3] **and João Pedro Castro**[3]

Department of Mathematics, University of Minho, Guimarães, Portugal[1]
Center of Mathematics, University of Minho, Guimarães, Portugal[2]
VITRUS AMBIENTE, EM, S.A., Guimarães, Portugal[3]

E-mail address: *mneves@math.uminho.pt; vitsilva25@hotmail.com; joaopedro.castro@vitrusambiente.pt*

Abstract

Cars occupy space when they circulate, but most of the time they are parked, occupying private and public space, frequently in an abusive and disorderly way. Most of the times, the demands for parking space occur on locations where it is in shortage, as is the case of the central area of the municipality of Guimarães (located in the northwest of Portugal). Public space is indeed a finite valuable asset, and it should be shared by all citizens in a balanced and orderly manner, considering its different urban functions. Parking lots and their respective management are, for a number of reasons, the most complex component of the policies relating to urban mobility, for it involves different types of users with very distinct needs and interests, often in conflict. It is crucial to try to find ideal solutions for the needs identified by the users and by those who choose other means of mobility.

VITRUS is a local public company that operates at different levels. This study focused on parking lot services with the goal of evaluating the quality of these services. A first study was developed and statistical methodologies were applied in order to evaluate the level of satisfaction of the parking lots users, since the demand for covered parking infrastructures which allow a prolonged usage without time limits presents a significant evolution. A poll was conducted, centered on an universe comprised of monthly users and rotative users of the five parking lots under study. A sampling process was established, and the sample was stratified by parking lot and by user type (monthly or rotative users). The satisfaction level and the quality level were studied by means of a questionnaire created for that specific purpose, consisting of twenty one questions. The data was subjected to statistical

treatment, particularly by applying statistical methodologies from the field of Statistical Inference (hypothesis tests). The results show positive results for the company, although some points can be improved for the full success of the service.

In a second study, the evolution of the revenue obtained from the Limited Duration Parking Zones (LDPZs) was analyzed by means of an exploratory analysis and by time series analysis of the data and of the revenue collected weekly from the parking meters in the period from January 2015 to February 2017. These methodologies served as support for the company's management and decision-making regarding LDPZs, a service that VITRUS manages with the aim of improving urban mobility.

Keywords: management, sampling, statistical inference, time series analysis.

Acknowledgements

This research was partially funded by Portuguese funds through the Portuguese Foundation for Science and Technology ("Fundação para a Ciência e a Tecnologia" - FCT), within the framework of project UID/MAT/00013/2013.

References

[1]Kuyter, K. and Bloemer, J. (1998) On the relationship between perceived service quality, service loyalty and switching costs. *International Journal of Service Industry Management* **19**: 436–453.

[2]Lorh, S. L. (2010) *Sampling Design and Analysis*, Second Edition. Boston, Michelle Julet.

[3]R Core Team. (2018) R: a language and environment for statistical computing. Version 3.5.0. Vienna, Austria: R Foundation for Statistical Computing. *URL https://www.R-project.org/*.

[4]Shumway, R. H. and Stoffer, D. S. (2006) *Time series analysis and its applications: with R Examples*, Second Edition. New York, Springer.

Degree one models for symmetric stochastic matrices

Cristina Dias[1,2] **Carla Santos**[3,4] **and João Tiago Mexia**[5,6]

Polytechnic Institute of Portalegre, Portugal[1]
CMA/FCT/UNL, Portugal[2]
Polytechnic Institute of Beja, Portugal[3]
CMA/FCT/UNL, Portugal[4]
Departamento de Matemática da Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa, Portugal[5]
CMA/FCT/UNL, Portugal[6]

E-mail address: *cpsd@ipportalegre.pt; carla.Santos@ipbeja.pt; jtm@fct.unl.pt*

Abstract

The rank of the fixed effects part of a model is its degree. Hence, degree one models are those that are described by $M = \lambda\alpha\alpha^t + \bar{E}$, form where $\bar{E}$ is the model's random component, λ and α are the eigenvalue and corresponding eigenvector, respectively. The matrices of this family models are based on a specific base design model. These models were successfully used in various applications, namely in the first phase of the STATIS methodology, the inter-structure. As the Hilbert-Schmidt matrices are cross product matrices they can be used in our applications and if needed the model degree can be increased providing the degree one models do not adjust. Simulations were carried out to validate the procedure adopted to estimate α. To carry out inference we will assume that vec(E) as normal distribution with null mean and variance-covariance matrix given by $\sigma^2 I_{n^2}$.

Keywords: STATIS methodology, Hilbert-Schmidt matrices, Simulations, inference.

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).

References

[1]Areia, A. (2009) Séries Emparelhadas de Estudos. Ph.D. Thesis, University of Évora.
[2]Dias, C. (2013) Models and Families of Models for Symmetric Stochastic Matrices. Ph.D. Thesis, University of Évora, Évora.
[3]Dias, C. (2000) Análise Conjunta Pesada de Regressões. Master Thesis, University of Évora, Évora.
[4]Dias, C., Oliveira, M. M. and Mexia, J. T. (2012) Models of Symmetric Stochastic Matrices. Proceedings of the International Conference of Numerical Analysis and Applied Mathematics 1479(1) : 1666-1669.
[5]Mexia, J. T. (1990) Best Linear Unbiased Estimates, Duality of F Tests and the Scheffé Multiple Comparison Method in Presence of Controlled Heterocedasticity. Comp. Stat and Data Analysis, 10(3) : 271-281.
[6]Mexia, J. T. (1995) Introdução à Inferência Estatística Linear. Centro de Estudos de Matemática Aplicada. Edições Universitárias Lusófonas, Lisboa.

Numerical Approach of Mixed Functional Differential Equations: Some Applications

M. Filomena Teodoro[1,2]

CEMAT - Center for Computational and Stochastic Mathematics, IST-Instituto Superior Técnico, Lisbon University, Avenida Rovisco Pais, n. 1, 1048-001 Lisboa, Portugal[1]
CINAV - Center of Naval Research, Portuguese Naval Academy, Portuguese Navy, Base Naval de Lisboa, Alfeite, 2810-001 Almada, Portugal[2]

E-mail address: *maria.alves.teodoro@marinha.pt, mteodoro64@gmail.com*

Abstract

In a high number of knowledge areas, many mathematical models show up functional differential equations with delayed and advanced arguments, the mixed type functional differential equations (MTFDEs). MTFDEs appear in biology, quantum physics, economy, control, acoustics, aerospace engineering, etc Some recent numerical methods to approximate the solution of a linear MTFDE were introduced in [1] and improved in [2] [3]. More recently, these methods were adapted and used to solve numerically a nonlinear MTFDE [4]. This paper provides a technique to solve a functional nonlinear mixed differential equation that allows to solve numerically some equations from physiology. Here we consider examples from nervous conduction and from vocal phonation. The performance of the numerical method is verified and the accuracy of the results is adequate.

Keywords: Mixed-type functional differential equations, numerical approximation, physiology.

Acknowledgements

M. Filomena Teodoro was supported by Portuguese funds through the *Center of Naval Research* (CINAV), Portuguese Naval Academy, Portugal and *The Portuguese Foundation for Science and Technology* (FCT), through the *Center for Computational and Stochastic Mathematics* (CEMAT), University of Lisbon, Portugal, project UID/Multi/04621/2019.

References

[1]Iakovleva V., Vanegas C. (2005) On the Solution of differential equations withe delayed and advanced arguments. *Journal of Differential Equations*, Conference **13**: 57–63.

[2]Lima P. M., Teodoro M. F., Ford N. J., Lumb P. M. (2010) Analytical and Numerical Investigation of Mixed Type Functional Differential Equations. *Journal of Computational and Applied Mathematics* **234(9)**: 2826–2837.

[3]Lima P. M., Teodoro M. F., Ford N. J., Lumb P. M. (2010) Finite Element Solution of a Linear Mixed-Type Functional Differential Equation. *Numerical Algorithms* **55**: 301–320.

[4]Teodoro M. F. (2017) Numerical Solution of a Delay-Advanced Differential Equation from Physiology. *Applied Mathematics & Information Sciences* **11(5)**: 1287-1297.

Multinomial Distribution and Discriminant Analysis: An application to discrete variables

Isaac Akoto[1,5], **Dário Ferreira**[2,3], **Gracinda Guerreiro**[1,4], **Sandra S. Ferreira**[2,3] **and João T. Mexia**[4]

Faculty of Science and Technology, Nova University of Lisbon, Monte da Caparica, Portugal[1]
Center of Mathematics and Applications, University of Beira Interior, Covilhã, Portugal[2]
Department of Mathematics, University of Beira Interior, Covilhã, Portugal[3]
Center of Mathematics and its Applications, Faculty of Science and Technology, New University of Lisbon, Monte da Caparica, Portugal[4]
Department of Mathematics and Statistics, University of Energy and Natural Resources, Sunyani, Ghana[5]

E-mail address: *i.akoto@campus.fct.unl.pt; dario@ubi.pt; sandraf@ubi.pt; jtm@fct.unl.pt*

Abstract

When dealing with continuous predictor variables, feature selection is a very common theme in the literature. However, this is no longer the case when we are dealing with methods proposed for Discriminant Analysis with discrete predictor variables.

In this work, we use results in Multinomial Distribution in connection with Discriminant Analysis applying them to a case with four categorical variables. The aim is to identify any tendency, considering two priori defined classes. An application is presented.

Keywords: Classification, Decision theory, Discriminant analysis.

Acknowledgements

This work was partially supported by national founds of FCT-Foundation for Science and Technology under UID/MAT/00297/2019 and UID/MAT/00212/2019.

References

[1]Brown, P.J., Fearn, T. and Haque, H.S. (1999), Discrimination with many variables, Journal of the American Statistical Association **94**, 448,1320–1329.

[2]Cooley, P.L. (1975), Bayesian and Cost Considerations for Optimal Classification with Discriminant Analysis, The Journal of Risk and Insurance, 277–287.

[3]Ferreira, S.S., Ferreira, D., Nunes, C. and Mexia, J.T. (2011), Discriminant Analysis and Decision Theory, Far East Journal of Mathematical Sciences **51**, 1, 69–79.

[4]Flury, L., Bourkai, B. and Flury, B.D. (1997), The discrimination subspace model. Journal of the American Statistical Association **92**, 438, 758–766.

[5]Graham, R.E. and Johnson, L. W. (1977), A unified Bayesian Approach to Optimal Classification with Discriminant Analysis, The Journal of Risk and Insurance **44**, 2, 316–319.

[6]Mexia, J.T. (1992), Assymptotic Chi-Squared Tests, Designs, and Log-Linear Models, Trabalhos de Investigação 1, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa.

Cumulant Generating Functions: a simulation study

Patrícia Antunes[1], **Sandra S. Ferreira**[1,2], **Célia Nunes**[1,2], **Dário Ferreira**[1,2] **and João T. Mexia**[3]

Center of Mathematics and Applications, University of Beira Interior, Covilhã, Portugal[1]
Department of Mathematics, University of Beira Interior, Covilhã, Portugal[2]
Center of Mathematics and its Applications, Faculty of Science and Technology, New University of Lisbon, Monte da Caparica, Portugal[3]

E-mail address: *antunes.patricia03@gmail.com; sandraf@ubi.pt; celian@ubi.pt;dario@ubi.pt;jtm@fct.unl.pt*

Abstract

Moments and cumulants have been intensively studied over the years. In this talk we will give an overview of some results related to higher-order moments and cumulants. Then, using cumulant generating functions we will derive estimators for cumulants in mixed models. A simulation study is also provided to illustrate our developments.

Keywords: Additive Models, Cumulants, Mixed Models, Moments.

Acknowledgements

This work was partially supported by national founds of FCT-Foundation for Science and Technology under UID/MAT/00297/2019 and UID/MAT/00212/2019.

References

[1]Akhiezer, N. I. (1965), The Classical Moment Problem, N. Kemme, Transl., Hafner, New York.

[2]Balakrishnan, N., Johnson,N. L. and Kotz, S. (1998), A note on relationships between moments, central moments and cumulants from multivariate distributions, Statistics & probability letters, **39**: 49–54, Elsevier.

[3]Dodge Y. and Valentin R. (1999), The Complications of the Fourth Central Moment, The American Statistician, **53** (3): 267–269.

[4]Craig, C. C. (1931). On A Property of the Semi-Invariants of Thiele. Ann Math Statist, 2 (2): 154–164.
[5]Harvey, J. R., (1972), On Expressing Moments in Terms of Cumulants and Vice Versa. The American Statistician, **26** (4): 38–39.
[6]Kendall, M. and Stuart, A. (1958). The Advanced Theory of Statistics. Charles Griffin & Company.
[7]Smith, P. J. (1995), A Recursive Formulation of the Old Problem of Obtaining Moments from Cumulants and Vice Versa. The American Statistician, **49** (2): 217–218.

The Potential of Logistic Regression

Dina Salvador[1,2], Sandra Monteiro[1,2] and Sandra Nunes[1,2]

Polytechnic Institute of Setúbal[1]
Centre for Mathematics and Applications of Faculty of Sciences and Technology of NOVA University of Lisbon[2]

E-mail address: *dina.salvador@estsetubal.ips.pt; sandra.ines.monteiro@esce.ips.pt;sandra.nunes@esce.ips.pt*

Abstract

Logistic Regression, as well as all regressions, are used to predict a data value based on prior observations of a data set. Logistic Regression allows estimating the relation between several variables when the dependent variable is categorical. At the base of this technique is the logistic function that began to be used to describe the growth of a population, originally due to Pierre-François Verhulst in 1838, see [1].

The Logistic Regression is currently widely used in the most diverse areas of knowledge, see [7], [8] and [3]. This regression can be binomial, ordinal or multinomial, depending on the number of categories that the dependent variable assumes. In this study only the Binary Logistic Regression is considered.

In this work we intent to show the importance of this technique in rapidly expanding areas such as Big Data.

According Wang et all (2016) "Big Data are data on a massive scale in terms of volume, intensity, and complexity that exceed the capacity of standard analytic tools" (p.399).

The role of statisticians in Big Data studies has been under-recognized. In our opinion it is important to change it. With this work we intent to contribute to this change, showing the important role that Logistic Regression already has in Big Data analysis. We believe that the research in Big Data present opportunities as well as challenges to statisticians.

Keywords: Logistic Regression, Logistic Regression Model, Binary Data, Big Data.

References

[1]Bacaër N. (2011) Verhulst and the logistic equation (1838). In: *A Short History of Mathematical Population Dynamics* (35–39), Springer, London.

[2]Baesens, B. (2012) *Analytics in a Big Data World: The Essential Guide to Data Science and Its Applications*, Jonh Wiley and Sons, New York.

[3]El aboudi, N. and Benhlima, L. (2018) Big Data Management for Healthcare Systems: Architecture, Requirements, and Implementation. *Advances in Bioinformatics.* Article ID 4059018, 10 pages.

[4]Hilbe, J. M. (2009) *Logistic Regression Model*, Chapman and Hall, New York.

[5]Hosmer, D. W. and Lemeshow, S. (2000) *Applied Logistic Regression*, 2nd ed.,Jonh Wiley and Sons, New York.

[6]Hosmer, D. W., Lemeshow, S. and Sturdivant, R. X. (2013) *Applied Logistic Regression*, 3rd ed., Wily, Chichester.

[7]Jiang, Y.(2018) Using Logistic Regression Model to Predict the Sucess of Bank Telemarketing. *International Journal on Data Science and Technology* **4** (1): 35–41.

[8]Nie, G., Rowe, W., Zang, L., Tian, Y. and Shi, Y. (2011) Credit card churn forecasting by logistic regression and decision tree. *Expert Systems with Applications* **38**: 15273–15285.

[9]Wang, C., Chen, M., Schifano, E., Wu, J. and Yan, J. (2016) Statistical methods and computing for big data.*Stat Interface* **9(4)**: 399–414.