



Conferência Internacional  
da Amazônia em Estatística  
Experimental e Análise de Risco

Manaus, AM – 12 a 15 de agosto de 2014


**Anais**  
*Proceedings*

*Roberval Monteiro Bezerra de Lima*  
*Teresa Paula Costa Azinheira Oliveira*  
Editores Técnicos



# **Conferência Internacional da Amazônia em Estatística Experimental de Risco**

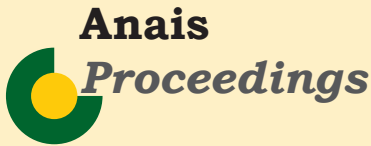
**Anais**  
*Proceedings*





*Empresa Brasileira de Pesquisa Agropecuária  
Embrapa Amazônia Ocidental  
Ministério da Agricultura, Pecuária e Abastecimento*

# **Conferência Internacional da Amazônia em Estatística Experimental de Risco**



*Roberval Monteiro Bezerra de Lima  
Teresa Paula Costa Azinheira Oliveira*  
Editores Técnicos

*Embrapa  
Brasília, DF  
2015*

Exemplares desta publicação podem ser adquiridos na:

**Embrapa Amazônia Ocidental**

Rodovia AM-010, Km 29, Estrada Manaus/Itacoatiara  
CEP 69010-970  
Caixa Postal 319  
Fone: (92) 3303-7800  
Fax: (92) 3303-7820  
www.embrapa.br  
www.embrapa.br/fale-conosco/sac

**Comitê de Publicações da Unidade**

Presidente: *Celso Paulo de Azevedo*  
Secretária: *Gleise Maria Teles de Oliveira*  
Membros: *Maria Augusta Abtibol Brito de Sousa, Maria Perpétua Beleza Pereira e Ricardo Lopes*

Revisor de texto: *Maria Perpétua Beleza Pereira*  
Normalização bibliográfica: *Maria Augusta Abtibol Brito de Sousa*

Diagramação: *Gleise Maria Teles de Oliveira e Lúcio Rogério Bastos Cavalcanti*  
Capa: *Lúcio Rogério Bastos Cavalcanti*

**1ª edição**

On-line (2015)

**Todos os direitos reservados.**

A reprodução não autorizada desta publicação, no todo ou em parte,  
constitui violação dos direitos autorais (Lei nº 9.610).

**Dados Internacionais de Catalogação na Publicação (CIP).**

Embrapa Amazônia Ocidental.

---

Conferência Internacional da Amazônia em Estatística Experimental e Análise de Risco (*I. : 2014: Manaus, AM*).

Anais... / Conferência Internacional da Amazônia em Estatística Experimental e Análise de Risco; editores, Roberval Monteiro Bezerra de Lima, Teresa Paula Costa Azinheira Oliveira. – Brasília, DF: Embrapa, 2015.

110 p.; 14,8 cm x 21 cm.

Modo de acesso: World Wide Web

<<http://ainfo.cnptia.embrapa.br/digital/bitstream/item/116772/1/Anais-Conferencia-Internacional.pdf>>.

Título da página da Web (acesso em 28 jan. 2015).

Só disponível on-line.

ISBN 978-85-7035-413-6

1. Estatística experimental. 2. Análise de risco. 3. Anais. I. Lima, Roberval Monteiro Bezerra de. II. Oliveira, Teresa Paula Costa Azinheira. III. Título. IV. Embrapa Amazônia Ocidental.

CDD 519.5



## **Apresentação** *Presentation*

**Prezados pesquisadores,  
professores, alunos e amigos,**

É com grande satisfação que fazemos a apresentação dessa conferência. Consideramos ser este o melhor lugar para (re)encontrar a comunidade científica da área, compartilhar as nossas experiências, atualizar e multiplicar o conhecimento. Salientamos, em primeiro lugar, que a concretização desse evento só foi possível com o auxílio do Programa de Apoio à Realização de Eventos Científicos e Tecnológicos no Estado do Amazonas (Parev III), da Fundação de Amparo à Pesquisa do Estado do Amazonas (Fapeam), no qual foi aprovado, em dezembro de 2013, o nosso projeto intitulado 1ª Conferência Internacional da Amazônia em Estatística Experimental e Análise de Risco.

A conferência tem como principal objetivo contribuir para a exploração e divulgação de temas relevantes e atuais para a comunidade científica da Amazônia. Como é do conhecimento geral, a estatística tem sido utilizada para a otimização de recursos econômicos, para o aumento

*Dear researchers, teachers,  
students and friends,*

*We are delighted to make the presentation of this conference, since we consider it to be the best place to (re) discover the scientific community, to share our experiences and to update and multiply our knowledge. First we point out that the realization of this event was only possible with the support of the Program for the Implementation of Scientific Events in Amazonas state-Parev III, of Fapeam, which approved our project in December 2013: International Conference of the Amazon in Experimental and Statistical Analysis of Risk.*

*The Conference aims to contribute to the exploitation and dissemination of relevant and current topics for the scientific community in the Amazon. As is well known, statistics has been used for optimization of economic resources, increased quality and productivity and optimization in decision analysis. In this context the achievement of this meeting under the main theme “Experimental Statistics and Risk Analysis” is*

da qualidade e produtividade e em análise de decisões. Neste contexto a realização desse encontro, cujos temas principais são “Estatística Experimental e Análise de Risco”, justifica-se claramente como um complemento do saber: a estatística investiga a tomada de decisões em ambiente de incerteza, procurando minimizar riscos – processo de avaliar a possibilidade de ocorrência e o impacto dos riscos identificados. Por ser fundamental a parte experimental da estatística para a resolução de problemas reais, promovendo a vertente multidisciplinar, serão privilegiadas comunicações que explorem os temas principais na sua vertente computacional e que evidenciem aplicações em áreas como: Ciências Florestais, Ciências Agrônômicas, Ciências da Educação, Ciências do Ambiente, Medicina, Engenharias, Biologia, Economia e Finanças.

A estrutura do encontro contempla sessões convidadas, sessões livres com comunicações orais e em pôster, minicursos e sessão de debate. Nesta, serão discutidas e incentivadas todas as vertentes de investigação científica nas áreas envolvidas. Assim, na concretização do encontro, está patente a perspectiva de estabelecimento de protocolos

*clearly justified as a complement to knowledge: Statistics investigate decision making in an uncertain environment, seeking to minimize the risk - the process of evaluating possibility of occurrence and the impact of identified risks. Being the experimental part fundamental to solving real statistical problems, by promoting multidisciplinary perspective, we will privilege communications about the major current themes and showing their computational applications in areas such as: Forest Sciences, Agricultural Sciences, Science Education, Environmental Sciences, Medicine, Engineering, Biology, Economics and Finance.*

*The structure of the meeting includes invited sessions, free sessions with oral communications and posters, mini-courses and a discussion session. In this, issues related to the event will be discussed, being motivated and encouraged all aspects of scientific research in the areas involved. Thus, in the concretization of this meeting, it is patent the prospect of establishing protocols between Universities and Research Centers Brazilians and other countries, beyond the elaboration of proposals for the organization of international projects.*



entre universidades e centros de pesquisa brasileiros e de outros países, para além da elaboração de propostas com vista à organização de projetos internacionais.

A partilha do conhecimento é, cada vez mais, uma atividade global, ambicionando dar respostas aos problemas da sociedade atual. A divulgação de técnicas usadas em Estatística Experimental e na Análise de Risco conduzirá certamente à motivação e captação de interesses para a análise de problemas reais em diferentes áreas. Apesar da importância e do impacto de sucesso no âmbito internacional, a interação e discussão dos temas propostos, no contexto amazônico, não foram ainda alvos da devida atenção. Estando esses temas sub-representados nesta região, urge, portanto, apostar mais nessas áreas, quer a nível de investigação, quer a nível académico e de sua exploração do ponto de vista das aplicações, sendo evidentes os benefícios.

Expressamos nossa gratidão a todos os oradores, instrutores dos minicursos, aos participantes e autores que submeteram seus trabalhos. Somos também gratos à Fapeam, agência financiadora do evento, e às universidades e

*Sharing knowledge is increasingly a global activity with the ambition to provide answers to the problems of contemporary society. Disclosure of techniques used in Experimental Statistics and Risk Analysis will certainly lead to motivation and attracting interest for the analysis of real problems in different areas. Despite their importance and impact of success on the international level, the interaction and discussion of the themes in the Amazon region, was not yet targeted due attention. Being underrepresented in this region, it is urgent therefore to invest more in these areas, both in terms of research, whether academic level and their operation from the point of view of applications, with the obvious benefits.*

*We would like to express our gratitude to all the speakers, instructors of mini-courses, participants and authors who submitted their work. We are also grateful to Fapeam, funding agency Event and universities and partner institutions in Brazil and abroad Who supported us and contributed to the success of this Conference. We address a special **Thanks** to the members of the Honor Committee and Organizing and Scientific Committees, for making this conference a reality.*

instituições parceiras do Brasil e do exterior, que nos apoiaram e contribuíram para o sucesso dessa conferência.

Agradecemos aos membros do Comitê de Honra e dos Comitês Organizador e Científico, por fazer dessa conferência uma realidade.

Esperamos que esse primeiro encontro seja uma sementeira de novos conhecimentos, que faça frutificar para que possamos nos reunir mais vezes em anos vindouros.

Sejam *muito bem-vindos* à I CIAEEAR e a Manaus, Amazônia, Brasil!

Desejamos a todos uma excelente conferência e uma maravilhosa estadia.

### **A Comissão Organizadora**

Roberval Lima – Embrapa, Brasil  
Amilcar Oliveira – UAb, Portugal

*We hope that this first conference is a seed of new knowledge, which will make it fruitful and we wish to meet you all in our meetings in the near future!*

*You are very welcome to “I CIAEEAR” and Manaus-Amazon-Brazil!*

*Wishing you a excellent conference and a nice stay.*

### ***The Organizing Committee***

Teresa Oliveira – UAb, Portugal  
Cintia de Souza – Embrapa, Brasil

## **Palestras**

Análises e Testes Estatísticos em Trabalhos de Pesquisa Agropecuária: Impactos e Sustentabilidade dos Resultados.....	15
Avaliação e Gestão do Risco de Cheias.....	19
Statistical Risk Methods on Cancer Problems.....	21
Experimental Design and Data Analysis: Methods for the Response Optimization and Applications Using R.....	23
A Family of Autoregressive Conditional Duration Models Applied to Financial Data.....	29
Parametrizações Interpretáveis em Regressão não Linear: Fundamentos e Estudos de Caso.....	31

## **Minicursos**

<b>Curso 1:</b> Análise Exploratória e Modelação com R.....	35
<b>Curso 2:</b> Modelagem Estatística com Aplicações na Análise de Experimentos.....	39
<b>Curso 3:</b> Risk Analysis and Applications.....	41
<b>Curso 4:</b> Modelos para Dados Assimétricos com Aplicações.....	45

## **Trabalhos**

Behavior of <i>Montandoniola confusa</i> Streito & Matocq (Hemiptera: Anthocoridae) Preying Upon Gall-Forming Thrips <i>Gynaikothrips Ficorum Marchal</i> (Thysanoptera: Phlaeothripidae).....	49
Ajustando a Ordem da Cadeia de Markov para Modelagem do Fluxo de Água de Rios.....	51
Exploring Links Between Experimental Design and the Risk of Data Loss on QR Codes.....	53
Statistical Modelling on the Antibiotic Synthesis Process.....	57

Portuguese Purse Seine: A Risk Analysis into Profitability.....	59
Análise Discriminante Aplicada em Alguns Nutrientes dos Tipos de Carne: Bovina, Suína e de Peixe.....	61
Análise de Classificação de Acidentes Automobilísticos nas BRs que Cortam a Paraíba em 2013.....	63
Planejamento de Múltipla Resposta: Análise de Variância Multivariada (MANOVA) Aplicada ao IDHM do Estado da Paraíba.....	65
Fractional Factorial Plans and Taguchi Method: an Application Using R.....	67
The Impact of Outliers on the Power of Randomization Tests: A Study by Simulation with Small Balanced Samples.....	69
Modelos de Regressão Linear: Fatores que Influenciam a Esperança de Vida em Angola.....	73
Determination of the Uncertainty of Experimental Measurements by Statistical Error Propagation.....	75
Análise Conjunta na Produção do Abacateiro <i>Persea americana</i> Mill cv. Hass em Função de Distintas Laminas de Irrigação.....	77
Normalidade de Variáveis Resposta com Excesso de Zeros.....	79
Escolha de Modelos Volumétricos para Clones de <i>Eucalyptus</i> ssp., no Polo Gesseiro do Araripe-PE.....	81
Seleção de Modelos Matemáticos para Estimar Biomassa Florestal em uma Vegetação do Sub-Bosque de um Plantio Comercial de <i>Bertholletia excelsa</i> H&B.....	83
Comparison of Peroxide Index of Olive Oil Suppliers.....	85
Comparação do Teste Ideal das Médias dos Tratamentos no Delineamento Experimental.....	87

A Comparação de Dendrogramas através Métodos de Agrupamentos via Reamostragem "Bootstrap".....	89
Procedimento para Determinação e Validação de Números de Grupos em Análise de Agrupamento.....	91
Processos Autorregressivos com Valores Inteiros Não Negativos – INAR.....	93
Generalized Linear Models and Risk Analysis, an Introductory Approach Using R.....	95
Técnicas de Análise de Sobrevivência Aplicadas em Ensaio Florestal na Amazônia.....	97
Análise Econômica de Sistemas Agroflorestais na Amazônia em Condições de Risco.....	99
Educational Data Mining e Learning Analytics na Melhoria do Ensino Online.....	101
Robust Heritability Estimation in Plant Candidate Genetic Association Studies.....	103





**Palestras**  
*Lectures*





## **Análises e Testes Estatísticos em Trabalhos de Pesquisa Agropecuária: Impactos e Sustentabilidade dos Resultados**

Adroaldo Guimarães Rossetti<sup>1</sup>

*Embrapa Agroindústria Tropical, Pesquisa e Desenvolvimento, Brasil, adroaldo,  
rossetti@embrapa.br*

### **Resumo**

Grande parte da pesquisa agrícola é feita por experimentos normalmente instalados e conduzidos no campo ou em laboratório, onde há efeitos de fatores ambientais que não podem ser controlados individualmente pelo pesquisador, mas que alteram de algum modo os resultados. Além disso, todo experimento é, de fato, uma amostra que se propõe a representar toda a população objeto do estudo, e esta abrange, na maioria dos casos, um universo com grande ou indeterminado número de elementos. A pesquisa experimental é, portanto, uma pesquisa por amostragem que, para produzir resultados sustentáveis, deve ser apoiada em bases científica e metodológica inerentes à temática abordada. É nesse contexto que a estatística experimental, cujo objeto é o estudo dos experimentos (planejamento, execução e análise), associada a uma boa familiaridade com o método científico, torna-se importante. Ela fornece os fundamentos científicos que embasam a interpretação dos resultados e as conclusões extraídas das análises dos experimentos.

O planejamento dos experimentos é a fase mais importante da pesquisa experimental. É nessa fase que o problema de pesquisa é especificado; objetivos, fatores e níveis de tratamentos são estabelecidos; hipóteses e delineamento experimental são formulados; variáveis avaliadas; testes e respectivos níveis de significância são definidos. Dessa forma, o propósito do planejamento é fornecer uma base objetiva para a análise dos dados coletados, os quais estão sujeitos à variação do acaso (MYERS, 1979). Um bom planejamento é essencial para que as hipóteses de interesse possam ser

eficientemente testadas e que os resultados obtidos sejam confiáveis. Nesse contexto, o planejamento do experimento, a análise estatística dos dados e a interpretação de seus resultados estão intimamente associados (SINGH; MASUKU, 2012).

Apesar disso, parece estar havendo maior preocupação com a análise dos dados gerados a partir do experimento do que com o seu planejamento, o que tem repercutido negativamente na qualidade dos resultados da pesquisa. Essa falta de sintonia entre planejamento de experimento e análise estatística provavelmente se deva ao ensino dessa disciplina ou à forma como o conhecimento acerca dela é passado, que não parece suficiente para dar suporte a certos trabalhos de pesquisa. Por outro lado, a existência de softwares “amigáveis”, de fácil utilização, que “dispensam” maiores conhecimentos de estatística, parece também contribuir para que o foco se concentre mais na análise. O resultado é, muitas vezes, a exploração superficial dos dados ou a interpretação equivocada dos resultados obtidos, limitando o alcance das pesquisas (ROSSETTI, 2001). Diante disso, tem sido cada vez mais frequente o emprego inadequado de testes de comparação múltipla, por exemplo, para certos tipos de comparação.

Entre os testes mais comuns, está a comparação de médias de tratamentos de natureza quantitativa, em lugar de análise de regressão, que, além da simples relação entre variáveis dependentes e independentes, permite obter estimativas reais de níveis de tais tratamentos e determinar curvas e superfícies de respostas. Não menos raro tem sido o uso inadequado de alguns testes para comparar os demais tratamentos do experimento com um tratamento controle (testemunha), bem como o uso do teste mais apropriado, segundo os objetivos da pesquisa. Nos experimentos fatoriais, a análise e as interpretações das interações têm sido outra questão que não tem sido tratada com a devida atenção. Nos casos de utilização de análise multivariada, às vezes, a complexidade é maior, tanto no tocante à adequação de certos métodos quanto à interpretação dos resultados.

Casos como esses, aparentemente pouco importantes devido à relativa frequência com que são vistos em trabalhos publicados por alguns periódicos da área agrícola, podem ter repercussão econômica significativa se resultados assim determinados forem adotados. Diante disso, a interação de profissionais de áreas biológicas com profissionais de outras áreas, como estatística e economia, por exemplo, parece ser a melhor forma de atender à demanda cada vez mais premente por refinamento da pesquisa agrícola. Ademais, o desenvolvimento de estudos por pesquisadores que atuam na área de estatística em conjunto com os de outras áreas pode contribuir significativamente para o avanço desse refinamento. O uso de amostragem na definição de estimadores para avaliação indireta de certos parâmetros é um exemplo dessa contribuição (ROSSETTI et al., 2014).

O objetivo deste trabalho foi apresentar uma discussão sobre análises, testes e algumas pesquisas em estatística utilizados na pesquisa agrícola e seus impactos na sustentabilidade dos resultados obtidos.

**Palavras-chave:** planejamento de experimentos, testes de comparação múltipla, interpretação de resultados.

## **Referências**

MYERS, J. L. **Fundamentals of experimental design**. 3th. ed. Boston: Allyn & Bacom, 1979. 524 p.

ROSSETTI, A. G. Precisão experimental e tamanho da área de experimentos de campo com fruteiras e outras plantas perenes arbóreas em função da unidade experimental e do número de repetições. **Revista Brasileira de Fruticultura**, Jaboticabal, v. 23, n. 3, p. 704-708, dez. 2001.

ROSSETTI, A. G.; VIDAL NETO, F. das C.; MOISÉS, S. M. M.; COUTINHO, V. T. R. Tamanho de amostra para estimar o peso médio de castanha de caju. In: CONGRESSO BRASILEIRO DE FRUTICULTURA, 23., 2014, Cuiabá. **Anais...** Cuiabá: Sociedade Brasileira de Fruticultura, 2014. 1 Pendrive.

SINGH, A. S.; MASUKU, M. B. An insight in statistical techniques and design in agricultural and applied research. **World Journal of Agricultural Sciences**, Swaziland, v. 8, n. 6, p. 568-584, 2012.

## **Avaliação e Gestão do Risco de Cheias**

Amílcar Oliveira

*Departamento de Ciências e Tecnologia, Universidade Aberta e Centro de Estatística e Aplicações da Universidade de Lisboa, Portugal, amilcar.oliveira@uab.pt*

### **Resumo**

As cheias e inundações ocorrem, na maioria dos casos, após fenômenos de precipitação extrema e de frequência variável. Essas ocorrências, quer sejam originadas por causas naturais, quer pela ação do homem, resultam quase sempre na submersão de terrenos e em elevados prejuízos humanos e materiais, entre outras consequências adversas.

Tendo em vista melhor compreensão desse fenômeno, pretende-se com este trabalho apresentar os principais conceitos associados à análise, avaliação e gestão do risco, no caso particular das cheias, assim como abordar alguns desafios que podem surgir no desenvolvimento e na aplicação do conhecimento existente nessa área.

Questões ligadas à construção e utilização de mapas de inundação e de risco por meio de modelos computacionais, à utilização da modelação estatística e à simulação serão destacadas e exemplificadas como ferramentas de análise e avaliação do risco de cheias.

**Palavras-chave:** cheias e inundações, avaliação do risco, gestão do risco, modelação estatística.

### **Agradecimentos**

Pesquisa parcialmente suportada pela Fundação para a Ciência e a Tecnologia (FCT), Projeto PEst-OE/MAT/UI0006/2014.

## Referências

CAMPOS, R. O. G. **Inundações urbanas**: considerações gerais e modelação matemática com incorporação da obstrução por resíduos sólidos. Rio de Janeiro: Universidade Federal do Rio de Janeiro, 2001.

D'ALTÉRIO, C. F. V. **Metodologia de cenários combinados para controle de cheias urbanas com aplicação à Bacia do Rio Joana**. Rio de Janeiro: Universidade Federal do Rio de Janeiro, 2004.

EM-DAT. The International Disaster Database. 2013. Disponível em: <<http://www.emdat.be/>>. Acesso em: 3 jul. 2014.

IPCC. **Managing the risks of extreme events and disasters to advance climate change adaptation**. Cambridge: Cambridge University, 2012.

QUINTELA, A.; SILVA, M. da. A modelação hidrológica em Portugal nos últimos 25 anos do século XX, nas perspectivas determinística, probabilística e estocástica. **Revista Brasileira de Recursos Hídricos**, Bento Gonçalves, v. 7, n. 4, p. 51-64, out./dez. 2002.

ROCHA, J. **O risco das inundações e a sua gestão**: uma visão nacional e uma visão europeia. [S.l.]: Laboratório Nacional de Engenharia Civil, 1988.

SCHARFFENBERG, W.; FLEMING, E. M. **Hydrologic modeling system HEC-HMS**. User's manual. Davis, CA: U.S. Army Corps of Engineers, 2010. 306 p.

SILVA, D. F. M. e. **Integração de ferramentas de SIG na modelação hidrológica de pequenas bacias hidrográficas**. 2008. 195 f. Dissertação (Mestrado em Engenharia do Ambiente) – Universidade do Porto, Porto.

SISTEMA NACIONAL DE INFORMAÇÃO DE RECURSOS HÍDRICOS. Atlas da Água. 2010. Disponível em: <<http://geo.snirh.pt/AtlasAgua>>. Acesso em: 12 jul. 2014.

## **Statistical Risk Methods on Cancer Problems**

Christos P. Kitsos

*Department of Informatics, Technological Educational Institute of Athens, Greece,  
xkitsos@teiath.gr*

### **Abstract**

The aim of this monograph is to discuss the statistical bioassay of experimental carcinogenesis. The interest in the of the cancer risk assessment has led to various statistical lines of thought, due to different intrinsic biological background and supported by the appropriate statistical analysis.

The biological insight (cell proliferation, mechanism of inhibition in mutagenesis and carcinogenesis) of a cancer risk assessment has not been tackled. Emphasis was given to the statistical models considered in the literature in the area of experimental carcinogenesis as well as on the dose dependent models. Moreover, an optimal experimental design approach is discussed, for evaluating the percentiles of a distribution. The D – optimality is chosen as an optimal design criterion, whereas the very useful method of stochastic approximation is applied for this purpose. We applied this method for the typical One-Hit model.

In this extended paper/presentation the appropriate Statistical background is discussed, were the technicalities are presented, in a compact form. So, we believe, even medical doctors with or without a strong mathematical background can follow the presentation/seminar.

The problem of Mixtures needs a particular statistical treatment, and we discuss a design approach extensively. The mathematical development is simple but important. The effect of covariates (with a typical example being smoking) is discussed also.

There is a different line of thought facing the “Cancer Problem” for different Sciences: There is the Medical approach, Biological, Toxicological and even Mathematical. There is a completely mathematical approach to the Stochastic Models of Carcinogenesis, that despite the excellent mathematical background adopted there, it will be beyond the target of this monograph. We try to bridge the different lines of facing the problem, around the Statistical way of supporting the investigation of the problem.

There are a number of Statistical indexes concerning the various cases of Cancer, evaluated through the Descriptive Statistics. A typical example is usually is that for the women the proportion on Breast Cancer incidences is near to 29%, while a proportion of 13% and 10% concerns colorectal and uterus incident cases. Therefore Breast Cancer incidence is the most common, while for men, the prostate cancer, 20.3% of the total, is the most frequently diagnosed. The second most frequent cases are lung cancer, 17.2% and colorectal, 12.8%. We are devoting particular interest for Breast Cancer problem is this monograph.

We simply want to emphasize that Descriptive Statistics is helpful approach, but a Statistical Analysis is something “more elegant” to face the involved Risk to such an important subject.

## **References**

KITSOS, C. P. **Cancer bioassays: a statistical approach: statistical methods for the cancer problem.** Saarbrücken: Lambert Academic Publishing, 2012. 116 p.

EDLER, L.; KITSOS, C. P. **Recent advances in quantitative methods in cancer and human health risk assessment.** Willey, UK.: Wiley Series in Probability and Statistics, 2006. 502 p.



## **Experimental Design and Data Analysis: Methods for the Response Optimization and Applications Using R**

<sup>1</sup>Teresa A. Oliveira; <sup>2</sup>Conceição Leal; <sup>3</sup>Roberval Lima

<sup>1</sup>*Universidade Aberta, DCeT and CEAUL, Portugal, teresa.oliveira@uab.pt;* <sup>2</sup>*Universidade Aberta, MEMeC, Portugal, conceicao.leal2010@gmail.com;* <sup>3</sup>*Embrapa Amazônia Ocidental, Brasil, roberval.lima@cpaa.embrapa.br*

### **Abstract**

Experimental Design is used by an increasing number of researchers to explore relationships between factors of interest. The factor levels are changed systematically according to the design, in order to evaluate their effects on the process. The aim is to reach a maximum amount of information with minimum cost in a minimal amount of time. This work presents a general review on Experimental Design techniques and a discussion of methods for the response optimization, followed by some applications to real data.

In Response Surface Methodology (RSM), a significant number of studies relates to a single response system or to responses analyzed one-by-one. However, for most of applications it is necessary and useful considering all responses simultaneously, since it is common to find optimization problems with several conflicting responses. The methodologies to multiple responses problems optimization, associated with the RSM, boil down to an informal overlap of contour maps of different responses to select the optimal solution. In most recent research, associated with the increasing computer capabilities, mathematical programming techniques were developed, including heuristic search methods, which allow locating the combinations of variables that lead to the best compromise between the different specifications of multivariate responses. The desirability function is a useful approach to simultaneous optimization of multiple responses.

Considering the experimental designs developed by Lima (2010) in forestry production and by Mondim (2014) in ceramic industry, we will use R to illustrate the role of RSM and of the methodology of genetic algorithms on the optimization of a problem. RSM will be used to optimizing one-response-at-a-time. The desirability function and the genetic algorithm will be used to optimizing one-response-at-a-time and all response simultaneously. Results will be compared and discussed.

**Keywords:** experimental design, response surface methodology, genetic algorithms, desirability, multiple response.

## References

ALVAREZ, M. J.; ILZARBE, L.; VILES, E.; TANCO, M. The use of genetic algorithms in response surface methodology. **Quality Technology and Quantitative Management**, Taiwan, v. 6, n. 3, p. 295-307, 2009.

BILES, W. A response surface method for experimental optimization of multi response processes. **Industrial and Engineering Chemistry, Process Design and Development**, Washington, DC, v. 14, p. 152-158, 1975.

BOX, G. E. P.; DRAPER, N. R. **Empirical model-building and response surfaces**. New York: J. Wiley & Sons, 1987. 688 p. (Wiley Series in Probability and Mathematical Statistics, 157).

BOX, G. E.; DRAPER, N. R. **Response surfaces, mixtures, and ridge analyses**. New York: John Wiley & Sons, 2007. v. 649.

BOX, G. E.; WILSON, K. B. On the experimental attainment of optimum conditions. **Journal of the Royal Statistical Society. Series B: Methodological**, London, v. 13, n. 1, p. 1-45, 1951.

BUSACCA, P. G.; MARSEGUERRA, M.; ZIO, E. Multi-objective optimization by genetic algorithms: application to safety systems. **Reliability Engineering System Safety**, Barking, v. 72, n. 1, p. 59-74, 2001.

CARLYLE, W.; MONTGOMERY, D. C.; RUNGER, G. C. Optimization problems and methods in quality control and improvement. **Journal of Quality Technology**, Milwaukee, v. 32, n. 1, p. 1-17, jan. 2000.

COPELAND, K. A.; NELSON, P. R. Dual response optimization via direct function minimization. **Journal of Quality Technology**, Milwaukee, v. 28, n. 3, p. 331-336, jul. 1996.

DEL CASTILLO, E. Multiresponse process optimization via constrained confidence regions. **Journal of Quality Technology**, Milwaukee, v. 28, n. 1, jan. 1996.

DEL CASTILLO, E.; MONTGOMERY, D. A nonlinear programming solution to the dual response problem. **Journal of Quality Technology**, Milwaukee, v. 25, n. 3, jul. 1993.

DERRINGER, G.; SUICH, R. Simultaneous optimization of several response variables. **Journal of Quality Technology**, Milwaukee, v. 12, n. 4, p. 214-219, oct. 1980.

GOLDBERG, D. E. **Genetic algorithm in search, optimization and machine learning**. Reading, MA: Addison-Wesley, 1989. 432 p.

HOLLAND, J. **Adaptation in natural and artificial systems**. Ann Arbor, MI: The University of Michigan Press, 1975.

HOOKE, R.; JEEVES, T. A. Direct search solution of numerical and statistical problems. **Journal of the ACM**, New York, v. 8, n. 2, p. 212-229, 1961.

KHURI, A. I.; CONLON, M. Simultaneous optimization of multiple responses represented by polynomial regression functions. **Technometrics**, Richmond, v. 23, n. 4, p. 363-375, 1981.

KIM, K. J.; BYUN, J. H.; MIN, D.; JEONG, I. J. **Multiresponse surface optimization**: concept, methods, and future direction (Tutorial). Seoul: Korea Society for Quality Management, 2001.

KIM, D.; RHEE, S.; PARK, H. Modelling and optimization of a GMA welding process by genetic algorithm and response surface methodology. **International Journal of Production Research**, London, v. 40, n. 7, p. 1699-1711, 2002.

LASDON, L. S.; FOX, R. L.; RATNER, M. W. Nonlinear optimization using the generalized reduced gradient method. **RAIRO - Operations Research - Recherche Opérationnelle**, Paris, v. 8, n. 3, p. 73-103, 1974.

LENTH, R. V. Response-surface methods in R, using rsm. **Journal of Statistical Software**, Los Angeles, v. 32, n. 7, p. 1-17, 2009.

LIMA, R. M. B. de. **Técnicas de experimentação estatística**: planos factoriais e aplicações. 2010. 78 f. Dissertação (Mestrado em Estatística, Matemática e Computação) - Universidade Aberta, Lisboa.

MONDIM, A. F. da S. **Metodologias de superfície de resposta**: uma investigação no âmbito da indústria cerâmica e do vidro. 2014. 109 f. Dissertação (Mestrado em Estatística, Matemática e Computação) - Universidade Aberta, Lisboa.

MYERS, R. H.; CARTER, W. H. Response surface techniques for dual response systems. **Technometrics**, Richmond, v. 15, p. 301-317, 1973.

MYERS, R. H.; MONTGOMERY, D. C.; ANDERSON-COOK, C. M. **Response surface methodology**: process and product optimization using designed experiments. New York: John Wiley & Sons, 2009. v. 705.

NELDER, J. A.; MEAD, R. A simplex method for function minimization. **The Computer Journal**, Oxford, v. 7, n. 4, p. 308-313, 1965.

PASANDIDEH, S. H. R.; NIAKI, S. T. A. Multi-response simulation optimization using genetic algorithm within desirability function framework. **Applied Mathematics and Computation**, New York, v. 175, n. 1, p. 366-382, 2006.

PLANTE, R. D. Process capability: a criterion for optimizing multiple response product and process design. **IIE Transactions**, Philadelphia, v. 33, n. 6, p. 497-509, 2001.

R DEVELOPMENT CORE TEAM. **R**: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2012.

SCRUCCA, L. GA: a package for genetic algorithms in R. **Journal of Statistical Software**, Los Angeles, v. 53, p. 1-37, 2012.

ZADBOOD, A.; NOGHONDARIAN, K. Considering preference parameters in multi response surface optimization approaches. **International Journal of Modeling and Optimization**, Jurong West, v. 1, n. 2, p. 158-162, jun. 2011.



# **A Family of Autoregressive Conditional Duration Models Applied to Financial Data**

Victor Leiva

*Universidad de Valparaíso, Chile, victor.leiva@uv.cl*

## **Abstract**

The Birnbaum–Saunders distribution is receiving considerable attention due to its good properties. One of its extensions is the class of scale-mixture Birnbaum–Saunders (SBS) distributions, which. Shares its good properties, but it also has further properties. The autoregressive conditional duration models are the primary family used for analyzing high-frequency financial data. We propose a methodology based on SBS autoregressive conditional duration models, which includes in-sample inference, goodness-of-fit and out-of-sample forecast techniques. We carry out a Monte Carlo study to evaluate its performance and assess its practical usefulness with real-world data of financial transactions from the New York stock exchange.

**Keywords:** Birnbaum-Saunders, forecast techniques





## **Parametrizações Interpretáveis em Regressão Não Linear: Fundamentos e Estudos de Caso**

Walmes Marques Zeviani

*Universidade Federal do Paraná, Brasil, walmes@ufpr.br*

### **Resumo**

Uma das vantagens dos modelos de regressão não linear é ter interpretação para os parâmetros. No entanto, além dos parâmetros do modelo, frequentemente os focos da investigação são parâmetros expressos como função destes, como uma dose letal ou ponto de inflexão. Surge então a preocupação de como fazer inferência sobre eles. Para isso, o método delta, a simulação Monte Carlo e procedimentos *bootstrap* são alternativas frequentes. Uma abordagem menos explorada é a de aplicar uma reparametrização no modelo, de forma a representar os parâmetros de interesse. Além de melhorar a interpretação, a presença do parâmetro alvo estende as possibilidades com relação a especificação de modelos e inferência estatística. O objetivo com este trabalho foi sistematizar o procedimento de aplicar reparametrizações, ilustrar com estudos de caso e destacar as vantagens e cuidados. Ênfase foi dada a modelos não lineares considerados em aplicações dentro das Ciências Agrárias, em que uma lista com 17 modelos foi obtida. No primeiro estudo de caso, o nível de dano econômico da desfolha no algodoeiro foi avaliado com o principal objetivo de propor uma parametrização de modelo que representasse o nível de dano econômico. No segundo estudo de caso, objetivou-se representar os parâmetros I (inflection) e S (slope) da curva de retenção de água do solo no modelo van Genuchten e avaliar o efeito da posição de amostragem e profundidade do solo sobre estes. Por fim, uma comparação entre parametrizações do modelo logístico, extensamente aplicado em curvas de crescimento e progresso de doenças, é feita, enfatizando a qualidade da inferência baseada em argumentos assintóticos contra a inferência baseada na verossimilhança.

**Palavras-chave:** modelo não linear, *Bootstrap*, reparametrização.





**Minicursos**  
*Courses*



## Minicurso 1

### **Análise Exploratória e Modelação com R**

#### **Tópicos**

- 1) Introdução ao R
- 2) Análise exploratória de dados
- 3) Tópicos de inferência estatística
- 4) Introdução à modelação estatística

#### **Docentes**

- Amílcar Oliveira, DCeT  
Universidade Aberta e CEAUL
- Teresa Oliveira, DCeT  
Universidade Aberta e CEAUL

#### **Resumo**

O R é uma aplicação de distribuição gratuita que disponibiliza uma linguagem de programação e um ambiente para análises estatísticas e gráficos, disponível em <http://cran.r-project.org/>.

Nesse minicurso pretende-se proporcionar a todos os participantes uma formação de Introdução ao Sistema R, com

ênfase nos aspectos básicos da linguagem, na estrutura e forma de operar com o programa. Ilustrando o uso das técnicas estatísticas e recorrendo a aplicações práticas e ao software R, o curso permite adquirir competências básicas em análise exploratória, estatística descritiva e visualização gráfica, inferência estatística paramétrica e não paramétrica e modelação estatística. Os participantes devem ter familiaridade com os conceitos e métodos estatísticos básicos, sem necessidade de conhecimentos em R.

Pretende-se oferecer uma formação baseada no rigor científico e na apresentação de exemplos de aplicação de interesse prático para os participantes.

#### **Biografia**

Amílcar Manuel do Rosário Oliveira é doutorando em Matemática, na especialidade



Modelação Estatística, pela Universidade Aberta, Portugal, 2008. É mestre em Estatística e Optimização

pela Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa, 1999, e licenciado em Engenharia Civil pelo Instituto Superior Técnico, Universidade Técnica de Lisboa, 1991. Exerce atividade docente em disciplinas das áreas de matemática e de estatística no ensino superior desde 1992. Desempenha funções docentes na Universidade Aberta em Portugal desde 2001. Atualmente é professor-auxiliar do Departamento de Ciências e Tecnologia da Universidade Aberta. É coordenador do mestrado em Estatística, Matemática e Computação, além de coordenador da seção de Matemática do Departamento de Ciências e Tecnologia da Universidade Aberta. É membro do Conselho Coordenador do mesmo departamento da Universidade Aberta. Foi orientador de alunos de mestrado (5) e de doutoramento (1) e também participou de vários júris de pós-graduação.

É coautor de três capítulos de livros e coeditor de um livro. Tem 25 artigos publicados em revistas internacionais e em atas de conferências. Participa regularmente de conferências nacionais e internacionais com apresentação de comunicações como autor e coautor, assim também como organizador e coorganizador (14).

É membro integrado do Centro de Estatística e Aplicações da Universidade de

Lisboa desde 2009; membro da Sociedade Portuguesa de Estatística, da International Biometric Society (IBS) e do International Statistical Institute (ISI).

## Biografia

Teresa Paula Costa Azinheira Oliveira é doutoranda em Estatística e Investigação Operacional na especialidade Estatística



Experimental e Análise de Dados, pela Universidade de Lisboa, 2000. É mestre em Estatística e Investigação Operacional pela Faculdade

de Ciências da Universidade de Lisboa, 1994, e licenciada em Matemáticas Aplicadas, Universidade Autónoma de Lisboa, 1987. Exerce atividade docente em disciplinas das áreas de Matemática e Estatística no ensino superior desde 1985. Desempenha funções docentes na Universidade Aberta em Portugal desde 1999. Atualmente é professora-auxiliar do Departamento de Ciências e Tecnologia da Universidade Aberta. Coordenadora do mestrado em Bioestatística e Biometria. Foi orientadora de alunos de mestrado (23) e de doutoramento (2). Atualmente orienta alunos de mestrado (14) e doutoramento (3). Participou de vários júris de mestrado e de doutoramento.

É coautora de cinco capítulos de livros e coeditora de dez livros. Tem 35 artigos publicados em revistas internacionais e em atas de conferências. Participa regularmente de conferências nacionais e internacionais com apresentação de comunicações como autora e coautora, e também como organizadora e coorganizadora (34).

É membro integrado do Centro de Estatística e Aplicações da Universidade de Lisboa desde 2000; membro da Sociedade Portuguesa de Estatística, nº 268; membro da International Biometric Society (IBS), nº 106.577; membro do International Statistical Institute (ISI), nº 16.803, onde, desde janeiro de 2014, exerce o cargo de secretária científica do Comitê de Análise de Risco (Secretary of ISI-CRA).





## Minicurso 2

### **Modelagem Estatística com Aplicações na Análise de Experimentos**

#### **Tópicos**

- 1) Modelos para análise de respostas não gaussianas (modelos lineares generalizados)
- 2) Modelos de regressão não linear
- 3) Modelos mistos e para medidas repetidas
- 4) Métodos computacionais para inferência baseada em verossimilhança

#### **Docente**

- Walmes Marques Zeviani, Universidade Federal do Paraná

#### **Resumo**

Na análise de experimentos não raramente ocorrem situações que se afastam dos métodos tradicionais de análise. Por exemplo, respostas métricas discretas (e.g.: contagens, proporções), limitadas (e.g.: ao intervalo unitário) e categóricas

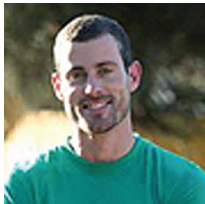
(e.g.: notas em escala) em geral não se aderem aos pressupostos dos modelos tradicionais de análise de variância, algumas vezes nem mesmo mediante transformação. Além disso, delineamentos experimentais complexos, como em vários locais/safras ou com medidas ao longo do tempo ou no perfil do solo, são cada vez mais comuns, bem como experimentos onde houve perda de observações ou com tratamentos adicionais. Até mesmo em casos simples, os resultados podem ser melhor representados se modelos mais interpretáveis forem considerados, como, por exemplo, modelos de regressão não linear. O objetivo do curso é explorar, por meio de experimentos reais, aplicações de modelagem estatística para as situações mencionadas. As aplicações contemplam respostas não gaussianas, modelos não lineares e modelos de efeitos mistos, além da exposição de métodos computacionais para inferência baseada na verossimilhança. Nesse sentido, o foco principal é a especificação, o ajuste e a interpretação dos modelos desenvolvidos para cada caso. Conhecimento em inferência



estatística e planejamento de experimentos são os pré-requisitos para um bom aproveitamento. Os estudos de caso serão feitos com aplicativo R para computação estatística.

participa regularmente de conferências nacionais e internacionais com apresentação de comunicações como autor e coautor.

## Biografia



Walmes Marques Zeviani tem mestrado e doutorado em Estatística e Experimentação Agropecuária pela Universidade Federal de Lavras, Brasil, 2009 e 2013, respectivamente. Bacharel em Agronomia pela Universidade Federal da Grande Dourados, 2007. Atualmente é professor-assistente do Departamento de Estatística da Universidade Federal do Paraná. Tem diversos cursos ministrados, tais como: curso em modelos de regressão não linear (58 RBras, 2013); aplicação de modelos mistos por meio da ferramenta R (Embrapa/CNPAF, 2012); curso em análise de experimentos com afastamento das pressuposições (Embrapa/CPAO, 2012); métodos computacionais para inferência com aplicações em R (Esalq/57 Rbras, 2012); estatística experimental aplicada (Embrapa/CNPAF, 2011). Tem 21 artigos publicados em revistas internacionais e em atas de conferências. Par-

## Minicurso 3

### Risk Analysis and Applications

#### Topics

- 1) Optimal design in cancer problem
- 2) Analysis of mixtures in cancer problem
- 3) The use of covariates in bioassays

#### Professor

- Christos Kitsos – TEI-ATHENS, Greece
- President of ISI – Committee on Risk Analysis

#### Abstract

#### ***Statistical Risk Methods on Cancer Problems***

Christos P. Kitsos

*Department of Informatics, Technological  
Educational Institute of Athens, Greece,  
xkitsos@teiath.gr*

The aim of this Seminar is to discuss the statistical bioassay of experimental carcinogenesis.

The interest in the of the cancer risk assessment has led to various statistical lines of thought, due to different intrinsic biological background and supported by the appropriate statistical analysis.

The biological insight (cell proliferation, mechanism of inhibition in mutagenesis and carcinogenesis) of a cancer risk assessment has not been tackled. Emphasis was given to the statistical models considered in the literature in the area of experimental carcinogenesis as well as on the dose dependent models. Moreover, an optimal experimental design approach is discussed, for evaluating the percentiles of a distribution. The D – optimality is chosen as an optimal design criterion, whereas the very useful method of stochastic approximation is applied for this purpose. We applied this method for the typical One-Hit model.

In this extended paper/presentation the appropriate Statistical background is discussed, were the technicalities are presented, in a compact form. So, we believe, even



medical doctors with or without a strong mathematical background can follow the presentation/seminar.

The problem of Mixtures needs a particular statistical treatment, and we discuss a design approach extensively. The mathematical development is simple but important. The effect of covariates (with a typical example being smoking) is discussed also.

There is a different line of thought facing the “Cancer Problem” for different Sciences: There is the Medical approach, Biological, Toxicological and even Mathematical. There is a completely mathematical approach to the Stochastic Models of Carcinogenesis, that despite the excellent mathematical background adopted there, it will be beyond the target of this monograph. We try to bridge the different lines of facing the problem, around the Statistical way of supporting the investigation of the problem.

There are a number of Statistical indexes concerning the various cases of Cancer, evaluated through

the Descriptive Statistics. A typical example is usually is that for the women the proportion on Breast Cancer incidences is near to 29%, while a proportion of 13% and 10% concerns colorectal and uterus incident cases. Therefore Breast Cancer incidence is the most common, while for men, the prostate cancer, 20.3% of the total, is the most frequently diagnosed. The second most frequent cases are lung cancer, 17.2% and colorectal, 12.8%. We are devoting particular interest for Breast Cancer problem is this monograph.

We simply want to emphasize that Descriptive Statistics is helpful approach, but a Statistical Analysis is something “more elegant” to face the involved Risk to such an important subject.

## References

KITSOS, C. P. **Cancer bioassays: a statistical approach: statistical methods for the cancer problem.** Saarbrücken: Lambert Academic Publishing, 2012. 116 p.

**EDLER, L.; KITSOS, C. P. Recent advances in quantitative methods in cancer and human health risk assessment.** Willey, UK.: Wiley Series in Probability and Statistics, 2006. 502 p.

## Biography

Education: Professor Christos Kitsos got his first degree in Mathematics, from the



University of Athens, Greece (1973), his Master's, in Math & Statistics, from the University of New Brunswick,

Canada(1978), and his PhD in Statistics, from the University of Glasgow, Scotland, UK (1986).

Research interests: Optimal Experimental Design (Linear and Non-Linear), Industrial Statistics, Cancer Risk Assessment, Information Theory, etc. He has participated at various research Programs in Greece, as well as at the European programs ERASMUS and TEMPUS.

Teaching experience: Professor Christos Kitsos has a huge experience on lecturing activities, in a number of Greek Universities

(U. of Patras, U. of Athens, Athens U. of Economics, U. of Thessaly). He is currently lecturing as a Professor at the Technological Educational Institute of Athens.

Publications: Professor Christos Kitsos has published 43 papers at international journals, 75 papers at volumes of Proceedings of International Conferences, and he presented a number of 88 various papers either at international meetings such as COMPSTAT, MODA, ISI, ICCRA, WSMC, etc or Greek meetings. He has published 8 Greek books on Statistics and Mathematics, and he is the editor of 5 international editions (with Physica-Verlag, Wiley, Springer, etc). He also published 3 monographs and a number of Greek volumes of Proceedings, either by himself or as a co-author.

Conferences: International conferences in Greece, such as MODA, a satellite (to ISI) meeting on Industrial Statistics, ISPIM, ICCRA1-3 and a Summer School for the IASC.

He was Co-Chair and member of Committees of WSMC7 and ICRA5, both meetings held in Tomar, Portugal in 2013.

Societies: Christos Kitsos is a member of ISI – International Statistical Institute and currently he is the Chairman of ISI

Committee on Risk Analysis (CRA-ISI). He is a member of IASC - International Association of Statistical Computing (council member 1998-2001).

Projects: Christos was the Institutional Project Leader of “ARHIMIDES” (2005-2008), the Scientific Director of the “Liaison Office” (2004-2008) of TEI Athens, and the Head of the Department of Mathematics (2005-2010).

Professor Chistos Kitsos was consultant of Greek Ministry of Education in INES/OECD, Eurostat (1997-2000) and currently he is a member of the Council of the Greek Statistical Institute.

## Minicurso 4

### Modelos para Dados Assimétricos com Aplicações

#### Tópicos

- 1) Introdução e motivação
- 2) Distribuições assimétricas
- 3) Modelos para dados assimétricos

#### Docente

- Víctor Leiva, Universidade de Valparaíso, Chile

#### Objetivo

Apresentar distribuições assimétricas e sua modelagem com ênfase na distribuição Birnbaum-Saunders.

#### Resumo

Distribuições assimétricas e sua modelagem estatística têm sido amplamente estudadas e aplicadas há décadas. Modelos gamma, Gaussiana inversa, lognormal e Weibull são usados frequentemente para análises de dados e, em particular, o modelo Birnbaum-Saunders está recebendo atenção especial. Embora quase tudo desses modelos tenha a sua origem na física

e engenharia, sua aplicabilidade tem sido considerada também em outras áreas, tais como ciências da terra, médicas e financeiras. Neste minicurso, são apresentadas algumas distribuições assimétricas com ênfase na distribuição Birnbaum-Saunders e discutidas as ferramentas de bondade de ajuste e diagnóstico. Além disso, serão formulados modelos de regressão para dados assimétricos com ênfase na distribuição Birnbaum-Saunders e mencionadas as extensões e generalizações dessa distribuição. Por fim, serão esboçadas algumas reflexões sobre as metodologias e aplicações recentes baseadas em modelos Birnbaum-Saunders e implementação no software R.

#### Biografia



Professor titular do Instituto de Estatística da Universidade de Valparaíso, Chile, Víctor Leiva é graduado em Estatística pela Universidade Católica de Valparaíso, doutorado em Estatística e

Pesquisa Operacional pela Universidade de Granada, Espanha, e pós-doutorado pela Universidade McMaster, Canadá. Membro sênior do IEEE, membro eleito do Instituto Internacional de Estatística (ISI), membro da comissão de finanças da Sociedade Internacional de Biometria e editor de Modelos Estocásticos Aplicados em Negócios e na Indústria, Revista Brasileira de Probabilidade e Estatística, Estatística Computacional e Análises de Dados, Comunicação e Estatística (em duas versões: teoria e métodos e simulação e computação), Revista de Estatística Aplicada, Revista Paquistanesa de Estatística, Revstat Revista de Estatística, e Pesquisa de Estocástica Ambiental e Avaliação de Riscos. Recebeu vários reconhecimentos internacionais, como os prêmios: “Juarez-Lincoln-Martí”, concedido pelos Estados Unidos, destacando-se como “O educador do ano de 2012” da América Latina, Brasil, Espanha e Portugal; e “Excelência em Pesquisa e Produtividade Científica”, concedido pela Universidade de Valparaíso (UV-Chile), recebido em sete ocasiões consecutivas. Fundador e diretor do programa internacional credenciado de doutorado em estatística da UV-Chile e membro de várias comissões científicas e acadêmicas do governo do Chile. O interesse pela pesquisa inclui distribuições de dano cumulativo e teoria de distribuições, métodos de bondade de ajuste e diagnóstico de influência, meto-

dologia de censura, confiabilidade e análise de sobrevivência, modelagem estatística e métodos estatísticos para a qualidade e pesquisa operacional.





**Trabalhos**  
*Papers*



## **Behavior of *Montandoniola Confusa* Streito & Matocq (Hemiptera: Anthocoridae) Preying Upon Gall-Forming Thrips *Gynaikothrips Ficorum* Marchal (Thysanoptera: Phlaeothripidae)**

<sup>1</sup>Adauto M. Tavares; <sup>2</sup>Jorge B. Torres; <sup>2</sup>Christian S.A. Silva-Torres; <sup>3</sup>Alessandra M. Vacari

<sup>1</sup>*Embrapa Amazônia Ocidental, Brazil, [adauto.tavares@embrapa.br](mailto:adauto.tavares@embrapa.br); <sup>2</sup>Universidade Federal Rural de Pernambuco, Brazil, [jtorres@depa.ufpe.br](mailto:jtorres@depa.ufpe.br); <sup>2</sup>Universidade Federal Rural de Pernambuco, Brazil, [christian.torres@depa.ufpe.br](mailto:christian.torres@depa.ufpe.br); <sup>3</sup>Faculdade de Ciências Agrárias e Veterinárias / Unesp Brazil, [amvacari@gmail.com](mailto:amvacari@gmail.com)*

### **Abstract**

*Gynaikothrips ficorum* is associated its host plant, *Ficus microcarpa*. In this present study of three aspects of *M. confusa* predation upon all three stages of thrips: (i) functional response of the predator as a function of different densities of thrips eggs, immature and adults separately. Based on a logistic regression of the number of prey consumed/density, was determined through the CATMOD of SAS. These parameters were estimated with nonlinear least square regression was determined through the PROC NLIN of SAS. The functional response curve, which is based on a logistic regression of the number of prey consumed/density, was determined through the CATMOD of SAS. These were estimated with nonlinear least square regression based on the proportion of prey eaten/initial prey densities (PROC NLIN of SAS). (ii) The adult thrips interference on predation of eggs by *M. confusa* was analyzed by using PROC ANOVA of SAS to compare the treatments as function of egg densities and the interaction of egg densities/treatments. The number of preyed upon eggs/initial egg density was calculated using the regression analysis performed by PROC REG of SAS. The models of egg predation/egg densities were compared between two treatments using PROC MIXED of SAS to test the equality of linear slopes; and (iii) adult *M. confusa* prey preferences when all thrips stages occurred. The data were tested under the

null hypothesis that no prey preference implies the expectation of an equal percentage of predation regardless of prey proportion. These analyses were performed using the PROC FREQ of SAS, and chi-square; (iii) *M. confusa* prey preferences when all thrips stages occurred simultaneously. For all three thrips life stages tested, the predator exhibited a type II functional response. The data were tested under the null hypothesis that no prey preference implies the expectation of an equal percentage of predation regardless of prey proportion performed using the PROC FREQ of SAS, and chi-square

**Keywords:** functional response, prey defense, prey preference.

**Puppport:** Embrapa, UFRPE, Capes

## References

CHESSON, J. Estimation and analysis of preference and its relationship to foraging models. **Ecology**, Tempe, v. 64, p. 1297-1304, 1983.

DI STEFANO, J. Effect size estimates and confidence intervals, an alternative focus for the presentation and interpretation of ecological data. In: BURK, A. R. (Ed.). **New trends in ecology research**. New York: Nova Science, 2005. p. 71-102.

HOLLING, C. S. Some characteristics of simple types of predation and parasitism. **Canadian Entomologist**, Ottawa, v. 91, p. 385-398, 1959.

JULIANO, S. A. Nonlinear curve fitting, predation and functional response curves. In: SCHEINER, S. M.; GUREVITCH, J. (Ed.). **Design and analysis of ecological experiments**. New York: Chapman and Hall, 1993. p. 159-182.

SAS INSTITUTE. **SAS/STAT user's guide**: version 8.02, TS level 2 MO. Cary, NC: SAS Institute Inc., 1999.

## **Ajustando a Ordem da Cadeia de Markov para Modelagem do Fluxo de Água de Rios**

<sup>1</sup>André G.C. Pereira; <sup>2</sup>Bernardo B. de Andrade; <sup>3</sup>Francisco de A. S. de Souza<sup>3</sup>

<sup>1</sup>*UFRN, Departamento de Matemática, Brasil, andre@ccet.ufrn.br*; <sup>2</sup>*UnB, Departamento de Estatística, Brasil, bbandrade@unb.br*; <sup>3</sup>*UFCEG, Unidade Acadêmica de Ciências Atmosféricas, Brasil, souza2011@gmail.com*

### **Resumo**

As cadeias de Markov são utilizadas como ferramenta principal em vários trabalhos de modelagem de precipitação de chuvas e de fluxo de água de rios. Entretanto, nenhum teste sobre a ordem da cadeia que melhor se ajusta aos dados é feito, gerando a seguinte questão: as cadeias de ordem são, de fato, a melhor ordem a ser utilizada? Neste trabalho mostra-se como deve ser feito o teste para detectar a melhor ordem da cadeia a ser usada (testes AIC e BIC) e como deve ser o procedimento depois de identificada a melhor ordem (geração da matriz de transição da cadeia). Finaliza-se aplicando o procedimento mencionado a dados de alguns rios brasileiros, obtidos da Agência Nacional de Águas (ANA), e comparando os resultados com os procedimentos da literatura que não usam o ajuste da ordem da cadeia.

**Palavras-chave:** cadeias de Markov, curvas de ascensão e recessão, fluxo de água de rios, modelagem.

### **Referências**

AKSOY, H. Markov chain-based modeling techniques for stochastic generation of daily intermittent streamflows. **Advances in Water Resources**, Southampton, v. 26, p. 663-671, 2003.

AKSOY, H. Use of gamma distribution in hydrological analysis. **Turkish Journal Engineering and Environmental Sciences**, Istanbul, v. 24, p. 419-428, 2000.

EVORA, N. D.; ROUSSELLE, J. Hybrid stochastic model for daily flows simulation in arid climates. **ASCE Journal of Hydrologic Engineering**, New York, v. 5, n. 1, p. 33-42, 2000.

KOTTEGODA, N. T. **Stochastic water resources technology**. New York: John Wiley & Sons, 1980.

# Exploring Links Between Experimental Design and the Risk of Data Loss on QR Codes

<sup>1</sup>Carla Francisco; <sup>1,2</sup>Teresa A. Oliveira

*<sup>1</sup>Departamento de Ciências e Tecnologia, Universidade Aberta, Portugal; <sup>2</sup>Centro de Estatística e Aplicações, Universidade de Lisboa*

## Abstract

Experimental design utilizing blocks was proposed by Ronald A. Fisher in the beginning of the 20th century with applications in agriculture. It became one of the areas of mass development for research in various fields, from agriculture to medicine, playing a central role not only on the research process but also on the new technological advances. As a branch of block designs, research on Balanced Incomplete Block Designs (BIBD) arose several interesting and defying problems within combinatorial mathematics. BIBD designs can be constructed and analysed more easily with the aid of computer tools, such as the The R Project for Statistical Computing, which, as will be shown, contains several different tools just for this purpose. Hadamard matrices give rise to a class of block designs named Hadamard configurations. It is easy and current to find their different applications based on new technologies and codes of figures such as Quick Response Codes (QR Codes) which are present in our daily life. These are bi-dimensional bar codes that can be easily read by common devices which have image capture function, such as mobile phones.

The use of such codes is very popular nowadays for example to send a simple text message, a picture or for a batch of information regarding a product or whatever one desires. Risk is the potential of losing something of value, weighed against the potential to gain something of value. There

are several types of risk, however in this work we will devote particular attention to the risk of data loss on QR Codes. Connections between the various methodologies and QR Codes will be discussed.

**Keywords:** experimental design, R Statistics, Hadamard Matrices, Quick Response Codes (QR Codes), risk.

### **Acknowledgements**

This research was partially sponsored by national funds through the Fundação Nacional para a Ciência e Tecnologia, Portugal - FCT under the project PEst-OE/MAT/UI0006/2014.

### **References**

BAUMERT, L. D.; GOLOMB, S. W.; HALL JR., M. Discovery of an Hadamard matrix of order 92. **Bulletin of the American Mathematical Society**, Providence, RI, v. 68, n. 3, p. 237-238, 1962.

BOSE, R. C. On the construction of balanced incomplete block designs. **Annals of Eugenics**, London, v. 9, 358-399, 1939.

CALINSKI, T.; KAGEYAMA, S. **Block designs: a randomization approach**. New York: Springer, 2000. v. 1: Analysis. (Lecture Notes in Statistics, 150).

CALINSKI, T.; KAGEYAMA, S. **Block designs: a randomization approach**. New York: Springer, 2000. v. 2: Design. (Lecture Notes in Statistics, 170).

FOODY, W.; HEDAYAT, A. On theory and applications of BIB designs with repeated blocks. **Annals of Statistics**, San Francisco, v. 5, n. 5, p. 932-945, 1977.



GEORGIU, S.; KOUKOUVINOS, C.; SEBERRY, J. Hadamard matrices, orthogonal designs and construction algorithms. In: WALLIS, W. D. (Ed.). **Designs 2002: further combinatorial and constructive design theory**. Norwell, Massachusetts: Kluwer Academic Publishers, 2002. p. 133-205.

HALL JR., M. **Combinatorial theory**. 2nd ed. New York: Wiley, 1986.

HEDAYAT, A.; WALLIS, W. D. Hadamard matrices and their applications. **Annals of Statistics**, San Francisco, v. 6, n. 6, p. 1184-1238, 1978.

HEDAYAT, A.; HWANG, H. **Construction of BIB designs with various support sizes**. Chicago: University of Illinois, 1980.

HEDAYAT, A. S.; HWANG, H. L. BIB (8,56,21,3,6) and BIB(10,30,9,3,2) designs with repeated blocks. **Journal of Combinatorial Theory, Series A**, New York, v. 36, p. 73-91, 1984.

MATHON, R.; ROSA, A. **CRC Handbook of combinatorial designs**. Boca Raton: CRC Press, 2010.

OLIVEIRA, T. A.; CERANKA, B.; GRACZYK, M. The variance of the difference of block effects in the balanced incomplete block designs with repeated blocks. **Colloquium Biometryczne**, [s.l.], v. 36, p. 115-124, 2006.

OLIVEIRA, T. A. BIB designs with repeated blocks: review and perspectives. In: ISLAMIC COUNTRIES CONFERENCE ON STATISTICAL SCIENCES, 10., 2010, Cairo. **Proceedings...** Cairo: The Islamic Countries Society of Statistical Sciences, 2010. p. 82-96.

OSTROM, L. T.; WILHELMSSEN, C. A. **Risk assessment: tools, techniques, and their applications**. New York: Wiley, 2012.

PLACKETT, R. L.; BURMAN, J. P. The design of optimum multifactorial experiments. **Biometrika**, Oxford, v. 33, n. 4, p. 305-325, 1946.

R DEVELOPMENT CORE TEAM. **R**: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2012.

RAGHVARAO, D. **Constructions and combinatorial problems in design of experiments**. New York: Wiley, 1971.

REJDA, G. E.; McNAMARA, M. **Principles of risk management and insurance**. 12th. ed. Englewood Cliffs: Prentice-Hall, 2013. (Series in Finance).

## **Statistical Modelling on the Antibiotic Synthesis Process**

<sup>1</sup>Carla Noronha; <sup>2</sup>Teresa Oliveira; <sup>3</sup>Ruy Costa

<sup>1</sup>*Universidade Aberta, Portugal, carla.noronha6@gmail.com;* <sup>2</sup>*Universidade Aberta e CEUL- Center of Statistics and Applications, Portugal, teresa.oliveira@uab.pt;*

<sup>3</sup>*Mathematics Dept., Faculdade de Ciências e Tecnologia – Universidade Nova de Lisboa, Portugal, ryac@fct.unl.pt*

### **Abstract**

The investigation of a problem associated with the complex process of industrial synthesis of an antibiotic - Minocycline – was carried out using statistical analysis and statistical modelling. Minocycline is indicated for the treatment of infections caused by microorganisms sensitive to tetracyclines, as well as infections caused by certain strains of staphylococci resistant to tetracyclines. It is particularly indicated for the treatment of acne and rosacea. The Portuguese pharmaceutical company who kindly provided us the data set is the world's largest producer of this antibiotic.

The synthesis of minocycline is a process that involves multiple steps which are carried out sequentially in this company. The 2nd step of the synthesis process involves a key piece - a metallic catalyzer in activated carbon that since the year when the process was initially developed until now has revealed some inconsistent behavior: the process of synthesis varies greatly in its levels of income. For environmental reasons, in 2009 a change occurred in the solvent used in the process which may have contributed to this variability.

The investigation focused on trying to understand the process, reaching to a possible model in order to obtain higher yields with lower costs. Our work was concentrated on the study of the lives and of the utilizations of rhodium metal piece that, after being prepared is divided into two parts to

optimize the metal. The observed sample consists of 78 yields catalyzers, each having two parts, A and B and a comparison of the number of lives and of the utilizations of different catalyzers was carried out. Kolmogorov-Smirnov and Chi-Square tests were used to compare the performance of the catalyzer before and after the change of the solvent and also to compare the two halves of the metal piece. Probability distribution fitting and goodness of fit tests were also conducted to the samples of the yields, from the several lives and of the several utilizations performed by the catalyzers. The purpose was generating pseudorandom numbers in order to develop a simulation model capable to allow the comparison of different policies for the use of the rhodium catalyzer.

**Keywords:** mynociline, rhodium, catalyzer, non-parametric tests, statistical analysis, statistical modelling.

## References

COSTA, R. A. **Investigação operacional**. Lisboa: Edições Universidade Aberta, 2002.

DAVID, H. A three sample Kolmogorov-Smirnov Test. **The Annals of Mathematical Statistics**, Ann Arbor, v. 29, n. 3, p. 842-851, 1958.

KRISHNAMOORTHY, K. **Handbook of statistical distributions with applications**. London: Chapman & Hall, 2006.

OLIVEIRA, T. P. C. **Estatística aplicada**. Lisboa: Edições Universidade Aberta, 2004.

## **Portuguese Purse Seine: A Risk Analysis into Profitability**

<sup>1</sup>Carlos Moura; <sup>2</sup>Teresa Oliveira; <sup>3</sup>Amílcar Oliveira

<sup>1</sup>MEMeC, Universidade Aberta, Portugal, [cjmoura@gmail.com](mailto:cjmoura@gmail.com); <sup>2</sup>Universidade Aberta e CEAUL-Center of Statistics and Applications, Portugal, [teresa.oliveira@uab.pt](mailto:teresa.oliveira@uab.pt);

<sup>3</sup>Universidade Aberta e CEAUL-Center of Statistics and Applications, Portugal, [amilcar.oliveira@uab.pt](mailto:amilcar.oliveira@uab.pt)

### **Abstract**

European Pilchard (Sardine) historically is one of the most representative species for the Portuguese fisheries (alongside with Codfish, caught in the distant waters of the Northern seas). In 2008 the Sardine represented around 33% of total catches. Purse seiners are traditionally dependent on pilchard and are responsible for over 98% of this catches in Portugal. In 2010 the Portuguese Sardine purse seine obtained certification from MSC (Marine Stewardship Council). Alongside other measures, this certification implies the sustainability of the fishery. The goal of the administration and the fishery representatives to maintain the certification (suspended in 2012 and reinstated in 2013), alongside degrading stocks, due to poor recruitment, result in the implementation of a management plan which strongly restricted Sardine catches. As a consequence, the 71 thousand tons observed in 2008 were reduced to less than 28 thousand in 2013, with obvious impacts on the activity of the purse seiners. The study focused on assessing the impact of these restrictions in the profitability of the fleet, according to different scenarios for the status of Sardine stocks in 2015. The data used in the study was available at DGRM – Directorate-General for Natural Resources, Security and Maritime Services. Catch data is collected by administrative procedure from the auctions (Docapesca) and economic data is collected under DCF (Data Collection Framework). In the first part of the study the different risks are identified (variables of cost and income), analyzed and correlated with sardine catches, and statistical models are created to use

in the risk analysis. On the second part of the study, 3 possible scenarios were built and the models were applied in a Monte Carlo simulation of the profitability.

**Keywords:** statistical modelling, fisheries, european pilchard, risk analysis, Monte Carlo, DCF.

## References

COMMISSION decision of 18 December 2009 adopting a multiannual Community programme for the collection, management and use of data in the fisheries sector for the period 2011-2013. **Official Journal of the European Union**, Luxembourg, v. 53, n. 41, p. 8-71, 16 Feb. 2010. Disponível em: <[http://eur-lex.europa.eu/legalcontent/EN/TXT/?uri=uriserv:OJ.L\\_.2010.041.01.0008.01.ENG](http://eur-lex.europa.eu/legalcontent/EN/TXT/?uri=uriserv:OJ.L_.2010.041.01.0008.01.ENG)>. Acesso em: 15 ago. 2014.

DGRM. Planos de recuperação/gestão/ação e ajustamento de esforço de pesca. Disponível em: <[http://www.dgrm.min-agricultura.pt/xportal/xmain?xpid=dgrm&selectedmenu=1469969&xpgid=genericPageV2&conteudoDetalhe\\_v2=209429](http://www.dgrm.min-agricultura.pt/xportal/xmain?xpid=dgrm&selectedmenu=1469969&xpgid=genericPageV2&conteudoDetalhe_v2=209429)>. Acesso em: 23 jul. 2014.

ESTATÍSTICAS da pesca 2008. Lisbon: Instituto Nacional de Estatística, 2009.

FRONTLINE Solver. Risk analysis. Disponível em: <[www.solver.com/riskanalysis](http://www.solver.com/riskanalysis)>. Acesso em: 25 maio 2014.

MARINE STEWARDSHIP COUNCIL. Portugal sardine purse seine. Disponível em: <<http://www.msc.org/track-a-fishery/fisheries-in-the-program/certified/northeast-atlantic/portugal-sardine-purse-seine>>. Acesso em: 23 jul. 2014.

MORGAN, M. G.; HENRION, M.; SMALL, M. **Uncertainty**: a guide to dealing with uncertainty in quantitative risk and policy analysis. Melbourne: Cambridge University Press, 1990.

ROBERT, C. P.; CASELLA, G. **Introducing Monte Carlo Methods with R.**  
New York: Springer, 2010.

VOSE, D. **Risk analysis** - a quantitative guide. 3th ed. Chippenham, Wiltshire:  
John Wiley & Sons, 2008.





## **Análise Discriminante Aplicada em Alguns Nutrientes das Carnes Bovina, Suína e de Peixe**

<sup>1</sup>Edwirde Luiz Silva; <sup>2</sup>Mácio Augusto Albuquerque; <sup>3</sup>Jorge Alves de Sousa;

<sup>4</sup>Dalila Camêlo Aguiar

*<sup>1</sup>Universidade Estadual da Paraíba, Departamento de Estatística, Brasil, edwirde@uepb.edu.br; <sup>2</sup>Universidade Estadual da Paraíba, Departamento de Estatística, Brasil, macioaa@uol.com.br; <sup>3</sup>Universidade Federal de Campina Grande, Unidade Acadêmica de Educação, Brasil, jorgeas@ufcg.edu.br; <sup>4</sup>Fundação Universitária de Apoio, Pesquisa e Extensão, Brasil, dalila.camelo@hotmail.com*

### **Resumo**

Os alimentos, pela importância de seus nutrientes, suprem as necessidades do organismo, mantendo seu bom funcionamento. Todos os seres vivos dependem da correta assimilação dos nutrientes que ingerem. Segundo Slywitch (2012), o surgimento de diversos problemas de saúde, como as doenças cardiovasculares, está relacionado ao consumo de carnes. Neste estudo, realizou-se descrição dos dados e Análise Discriminante (AD) com saída gráfica avançada, usando o pacote Candisc (FRIENDLY; FOX, 2012), aplicada a teores de colesterol, proteínas, calorias, ferro e cálcio em carnes bovina, suína e de peixe, em preparos sob diversas formas (frita, cozida, assada, grelhada e crua), com objetivo de entender as diferenças existentes entre dois ou mais grupos de objetos, estimando os pesos dessas variáveis que permitirão atribuir os tipos de carne a seus respectivos grupos, minimizando o erro de classificação. Segundo Aguiar et al. (2014), houve associação de alguns tipos de carne, peixe, crustáceos e moluscos por quantidade de colesterol e calorias. Em se tratando de colesterol e calorias, observou-se a predominância de gordura e calorias na carne suína; na carne bovina sobressaíram os nutrientes ferro, proteínas e colesterol; e cálcio associado ao peixe; e fósforo relativamente próximo da carne bovina e do peixe. O modelo explicou 71,9% da classificação. As variáveis calorias, colesterol

e ferro, com relação ao tipo carne de peixe, ficaram bem discriminadas, mostrando que tiveram ótima classificação em relação às demais regressões. Entre colesterol e gordura, houve grande interseção entre os três tipos de carnes, apresentando grande variação causada principalmente pela carne de porco. Entre calorias e gordura, houve correlação positiva entre os três tipos. Quanto ao fósforo e à caloria, o tipo carne de porco está relacionado negativamente com as demais. Foram diagnosticadas, portanto, similaridades e diferenças entre os tipos de carne quanto a seus nutrientes.

**Palavras-chave:** discriminante, carnes, nutrientes.

## Referências

AGUIAR, D. C.; NÓBREGA, L. L. de S.; SILVA, E. L. Classificação de alguns tipos de carnes, peixes, crustáceos e moluscos por quantidade de colesterol e calorias: uma análise discriminante. In: CONGRESSO NACIONAL DE CIÊNCIAS DA SAÚDE, 1., 2014, Cajazeiras. Avanços, interfaces e práticas integrativas. **Anais CONACIS**, João Pessoa, v. 1, 2014.

ELMADFA, I.; AIGN, W.; FRITZSCHE, D. **Tabla de valores nutricionales de los alimentos**. Barcelona: Editorial Hispano Europea, 2012. 128 p.

FRIENDLY, M.; FOX, J. Package "candisc". In: R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2012.

SLYWITCH, E. **Alimentação sem carne** - guia prático. São Paulo: Alaúde, 2012.

## **Análise de Classificação de Acidentes Automobilísticos nas BRs que Cortam a Paraíba em 2013**

<sup>1</sup>Edwirde L. Silva; <sup>2</sup>Mácio A. Albuquerque; <sup>3</sup>Jorge A. Sousa; <sup>4</sup>Dalila Camêlo Aguiar

<sup>1</sup>*Universidade Estadual da Paraíba, Departamento de Estatística, Brasil, edwirde@uepb.edu.br;* <sup>2</sup>*Universidade Estadual da Paraíba, Departamento de Estatística, Brasil, macioaa@uol.com.br;* <sup>3</sup>*Universidade Federal de Campina Grande, Unidade Acadêmica de Educação, Brasil, jorgeas@ufcg.edu.br;* <sup>4</sup>*Fundação Universitária de Apoio, Pesquisa e Extensão, Brasil, dalila.camelo@hotmail.com*

### **Resumo**

Os testes de hipótese contrastados são frequentemente relativos aos valores paramétricos de populações com distribuição conhecida. A problemática que se aborda concerne à própria distribuição dos dados. Concretamente, a hipótese nula se anuncia em termos gerais como “a distribuição da população se ajusta a determinado modelo de probabilidade” e refere-se ao contraste consequente como contraste de bondade de ajuste. Para aplicar um teste não paramétrico, não é necessário admitir hipóteses sobre distribuições de probabilidade da população da qual tenham sido extraídas amostras para análise. Em teste de homogeneidade, testa-se a afirmativa de que populações diferentes têm a mesma proporção de alguma característica (TRIOLA, 2008). O estudo do número e do tipo de classificação de acidentes automobilísticos (com ou sem vítimas, ferido e ignorado) em quatro BRs que cortam a Paraíba é de importância para conscientizar a sociedade a respeito das causas dos acidentes de trânsito e sua prevenção. Os dados constituem uma tabela cruzada de tamanho  $7 \times 5$ , dos tipos de acidente versus os sete dias da semana. Segundo Araújo Neto e Silva (2014), houve associação no número de acidentes em relação aos diferentes dias da semana, visto que o valor da estatística  $\chi^2 = 568,04$  ( $p$  – valor  $< 0,05$ ) foi menor que o crítico, permitindo assim rejeitar a hipótese nula de que a classificação dos tipos de acidente se distribui igualmente nos sete dias da semana. Confirmou-se,

portanto, a existência de associação nos tipos de acidente em diferentes dias da semana, principalmente nos fins de semana, com o tipo de acidente com vítimas fatais.

**Palavras-chave:** Qui-quadrado, acidentes, BRs.

## **Referências**

ARAÚJO NETO, J. F. de; SILVA, E. L. Estatística descritiva e teste qui-quadrado aplicados a acidentes de trânsito ocorridos em rodovias federais na Paraíba em 2012. **Revista da Estatística UFOP**, Campina Grande, PB, v. 3, n. 3, p. 430-435, 2014.

TRIOLA, M. F. **Introdução à estatística**. 10. ed. Rio de Janeiro: LTC, 2008.

## **Planejamento de Múltipla Resposta: Análise de Variância Multivariada (MANOVA) Aplicada ao IDHM do Estado da Paraíba**

<sup>1</sup>Edwirde Luiz Silva; <sup>2</sup>Mácio Augusto Albuquerque; <sup>3</sup>Jorge Alves de Sousa;

<sup>4</sup>Dalila Camêlo Aguiar

*<sup>1</sup> Universidade Estadual da Paraíba, Departamento de Estatística, Brasil, edwirde@uepb.edu.br; <sup>2</sup> Universidade Estadual da Paraíba, Departamento de Estatística, Brasil, macioaa@uol.com.br; <sup>3</sup> Universidade Federal de Campina Grande, Unidade Acadêmica de Educação, Brasil, jorgeas@ufcg.edu.br; <sup>4</sup> Fundação Universitária de Apoio, Pesquisa e Extensão, Brasil, dalila.camelo@hotmail.com*

### **Resumo**

Os modelos de análise de variância e covariância múltipla têm como característica fundamental a presença de diversas variáveis dependentes métricas e independentes não métricas. A MANOVA permite avaliar a significância estatística de diferença entre grupos (HAIR et al., 1998). Com interesse de enfrentar as desigualdades sociais, o Programa das Nações Unidas para o Desenvolvimento (PNUD) fornece dados da realidade socioeconômica internacional, como dos municípios brasileiros, com mais de 200 indicadores. Para este trabalho considerou-se apenas os três pilares (longevidade, educação e renda) do Índice de Desenvolvimento Humano Municipal (PNUD, 2013) dos 223 municípios paraibanos agrupados em suas respectivas mesorregiões (Sertão, Agreste, Borborema e Zona da Mata). Pelas estatísticas Lambda de Wilks, Traço de Hotelling e Roy's Largest Root sobre os fatores, existem diferenças significativas nos índices de longevidade, renda e educação entre as mesorregiões, ao nível de 5% de significância. O coeficiente ETA quadrado parcial e a potência observada apresentaram valores baixos, eles certificaram que os fatores explicam muito pouco a variabilidade total. As estimações dos parâmetros do modelo multivariado da variância apresentaram em geral p- valores não aceitáveis. As comparações

por par entre as médias dos três índices (longevidade, educação e renda) delimitados nas variáveis dependentes pelo fator mesorregiões mostram que, pelo teste de Scheffé e Bonferroni, a educação para Borborema foi significativa, diferindo estatisticamente da Zona da Mata e Agreste. É importante que a gestão pública municipal defina políticas e programas específicos capazes de enfrentar e superar essas diferenças educacionais e buscar continuamente melhorias na qualidade de vida dos paraibanos.

**Palavras-chave:** índice, mesorregiões, comparações.

## Referências

HAIR, J. F.; ANDERSON, R. E.; TATHAM, R. L.; BLACK, W. **Multivariate data analysis**. 5th. ed. New Jersey: Prentice Hall, 1998. 333 p.

PNUD. Índice de Desenvolvimento Humano Municipal. 2013. Disponível em <<http://www.pnud.org.br/IDH/DH.aspx>>. Acesso em: 28 abr. 2014.

# Fractional Factorial Plans and Taguchi Method: an Application Using R

<sup>1</sup>Elisa Henning; <sup>2</sup>Amilcar Oliveira; <sup>3</sup>Teresa Oliveira

*<sup>1</sup>Santa Catarina State University, College of Technological Science, Brazil, elisa.henning@udesc.br; <sup>2</sup>DCeT-Universidade Aberta e CEAUL, Portugal, Teresa.Oliveira@uab.pt; <sup>3</sup>DCeT-Universidade Aberta e CEAUL, Portugal, Amilcar.Oliveira@uab.pt*

## Abstract

This paper discusses statistical techniques of design of experiments with emphasis on fractional factorial designs with mixed levels. Taguchi designs and fractionated factorial designs were obtained with functions of the R were applied to a real case. The objective is to analyze the effects on the micro hardness of a composite resin, testing two types of resin, and four photopolymerizers in four different depths. The analyses of fractionated designs are compared with the results of a full factorial design. The results indicate that these fractional techniques were effective for the analysis of the experiment and present themselves as viable alternatives. The three factors (resin, photopolymerizers and depth) influence on the micro hardness. From the Taguchi design, it was possible to establish a combination of factors and levels that maximizes the micro hardness. Packages and functions for the proposed applications were also investigated and some routines in the R language were developed for implementing the Taguchi method.

**Keywords:** design of experiments, fractionated factorial designs, Taguchi method.

## References

KOBILINSKY, A.; BOUVIER, A.; MONOD, H. **PLANOR**: an R package for the automatic generation of regular fractional factorial designs. R package version 0.1-9. Disponível em: <<http://w3.jouy.inra.fr/unites/miaj/public/logiciels/planor/>>. Acesso em: 08 jan. 2014.

PINTO, J. C.; SCHWAAB, M. **Análise de dados experimentais**. Rio de Janeiro: E-papers, 2011, v. 2.

ROTH, T. **QualityTools**: statistical methods for quality science. R package version 1.53. Disponível em: <<http://CRAN.R-project.org/package=qualityTools>>. Acesso em: 15 fev. 2013.



# The Impact of Outliers on the Power of Randomization Tests: A Study by Simulation with Small Balanced Samples

<sup>1</sup>Fernando Branco; <sup>2</sup>Teresa Azinheira Oliveira

<sup>1</sup>ULHT, EPCV, Portugal, branco.fernando@netcabo.pt; <sup>2</sup>U. Aberta, CEAUL, Portugal, toliveir@univ-ab.pt

## Abstract

Our aim is to present some results of an ongoing research on the impact of outliers on the power of randomization tests to compare two independent groups, in the case of small balanced samples. We have evaluated the power of these tests in various conditions, manipulating: sample size (4, 5 or 6 elements per group), the model for the generation of outliers (contaminated normal distributions, 1-wider distributions), the number of groups with outliers (1 or 2), the scale of the contaminating distribution (3, 5 or 10) and effect size (from -4 to 4, by 0.25 steps). As a comparison standard, we also have, for each sample size, a condition in which the two groups of data come from standard normal distributions. For each condition we have simulated 10 000 samples. We have established a significance level ( $\alpha$ ) of 0.05, and for each sample we have evaluated the significance of the two-tailed randomization test using all the data permutations. The power of these tests was estimated calculating the proportion of the 10 000 samples with significance values less than or equal to  $\alpha$ . The results show that the presence of outliers decreases the power of the tests, with an inverse relationship between the power and the size of the outliers. It is also observed that under the conditions in which only one of the groups has outliers the power is lower than under those in which the two groups contain them.

**Keywords:** randomization tests, outliers, normal contaminated distributions, 1-wild distributions.

## References

COHEN, J. The statistical power of abnormal-social psychological research: a review. **Journal of Abnormal and Social Psychology**, Albany, N.Y., v. 65, p. 145-153, 1962.

COHEN, J. **Statistical power analysis for the behavioral sciences**. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum, 1988.

DAVID, H. A.; KENNEDY, W. J.; KNIGHT, R. D. Means, variances and covariances of the normal order statistics in the presence of an outlier. In: SOBEL, M. **Selected tables in Mathematical Statistics**. Providence, RI: American Mathematical Society, 1977. v. 5. p. 75-204.

EGDINGTON, E. S.; ONGHENA, P. **Randomization tests**. 4th ed. Boca Raton, FL: Chapman & Hall : CRC, 2007.

GASTWIRTH, J. L.; COHEN, M. L. The small sample behavior of some robust linear estimators of location. **Journal of the American Statistical Association**, Washington, DC, v. 65, p. 946-973, 1970.

HAWKINS, D. M. **Identification of outliers**. London: Chapman and Hall, 1980.

NEYMAN, J.; PEARSON, E. On the problem of the most efficient tests of statistical hypothesis. **Philosophical Transactions of the Royal Society of London**, Series A, London, v. 231, p. 289-337, 1933.

R CORE TEAM. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2014. Disponível em: <<http://www.R-project.org/>>. Acesso em: 15 ago. 2014.

ZIMMERMAN, D. W. A note on the influence of outliers on parametric and nonparametric tests. **Journal of General Psychology**, Provincetown, v. 121, n. 4, p. 391-401, 1994.

ZIMMERMAN, D. W. Increasing the power of nonparametric tests by detecting and down weighting outliers. **Journal of Experimental Education**, Madison, v. 64, n. 1, p. 71-78, 1995.

ZIMMERMAN, D. W. Invalidation of parametric and nonparametric statistical tests by concurrent violation of two assumptions. **Journal of Experimental Education**, Madison, v. 67, n. 1, p. 55-68, 1998.



## **Modelos de Regressão Linear: Fatores que Influenciam a Esperança de Vida em Angola**

<sup>1</sup>Francisco Chimuco; <sup>2</sup>Célia Nunes; <sup>3</sup>Amílcar Oliveira

*<sup>1</sup>Mestrando em Estatística, Matemática e Computação, Universidade Aberta, Portugal;*

*<sup>2</sup>Departamento de Matemática, Universidade da Beira Interior, Portugal, celian@ubi.pt;*

*<sup>3</sup>Universidade Aberta e CEAUL, Portugal, Amilcar.Oliveira@uab.pt*

### **Resumo**

A esperança de vida ao nascer ou esperança de vida à nascença é o número médio de anos que um grupo de indivíduos nascidos no mesmo ano pode esperar viver, se mantidas, desde o seu nascimento, as taxas de mortalidade verificadas no ano de observação. A esperança de vida no nascimento é também um indicador de qualidade de vida de um país, região ou localidade. Pode também ser utilizada para aferir o retorno de investimentos feitos na melhoria das condições de vida e para compor vários índices, tais como o IDH. Os objetivos deste trabalho foram identificar as principais causas que contribuem para a diminuição dos índices da esperança de vida em Angola, construir e propor um modelo de regressão linear para prevenir fenômenos futuros que permitam tirar conclusões relativas aos fatores que contribuem para o aumento da esperança de vida em Angola.

Para tal recorreu-se aos dados estatísticos publicados em relatórios pelos órgãos oficiais angolanos, como o Instituto Nacional de Estatística, e internacionais, como a Organização das Nações Unidas, entre outros, a fim de coletar dados que permitissem caracterizar o fenômeno em estudo em Angola. Por outro lado, fez-se o estudo da regressão linear simples e múltipla para criar um modelo estatístico que permitisse fazer o tratamento dos dados para verificar a correlação entre as diversas variáveis visando identificar aquelas que possuem forte correlação positiva e propor um modelo de regressão linear simples ou múltipla que permita prever e

prevenir o fenômeno em estudo. Com o emprego do software IBM SPSS 20 for Windows serão calculados: as estatísticas descritivas, os coeficientes de correlação de Pearson das variáveis independentes com a esperança de vida e as análises de regressão linear múltipla que permitirão gerar modelos preditivos para cada uma das regiões angolanas.

**Palavras-chave:** esperança de vida, regressão linear múltipla, saúde coletiva.

### **Agradecimentos**

Investigação parcialmente suportada pela Fundação para a Ciência e a Tecnologia (FCT), Projeto PEst-OE/MAT/UI0006/2014 (CEAUL).

## **Determination of the Uncertainty of Experimental Measurements by Statistical Error Propagation**

<sup>1</sup>João C. F. Diogo; <sup>1</sup>Helena M. N. T. Avelino; <sup>2\*</sup>Fernando J. P. Caetano; <sup>2</sup>João M. N. A. Fareleira

Centro de Química Estrutural, Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais, 1049-001 Lisbon, Portugal.

<sup>1</sup> *Instituto Superior de Engenharia de Lisboa, Departamento de Engenharia Química, R. Conselheiro Emídio Navarro, 1, 1959-007 Lisbon, Portugal;* <sup>2</sup> *Universidade Aberta, R. da Escola Politécnica, 147, 1269-001 Lisbon, Portugal;* \* *To whom correspondence should be addressed. Email: fcaetano@uab.pt*

### **Abstract**

Thousands of experimental measurements are performed worldwide every day. Beyond the information obtained in each measurement it is also very important to know the uncertainty of the property obtained from that measurement. Very often the published experimental measurements have an underestimated uncertainty which is determined only by taking into account its repeatability or reproducibility. It is very important to include in the uncertainty determination the errors of all parameters relevant for the measurement.

Thermophysical properties are determinant for chemists to understand chemical phenomenon and chemical engineers to optimize industrial chemical processes. We report how to apply the statistical error propagation to determine the uncertainty of thermophysical properties measured in our laboratory particularly the density of several liquids with a large range of applications in the industry and their derived properties, the isothermal compressibility and the isobaric thermal expansion.

Mathematical correlations of thermophysical properties are very common, from which it is possible to obtain information by interpolation or other properties by derivation or integration. Since the parameters of the correlations are commonly obtained by fitting they have an associated uncertainty and the uncertainty of the parameters can be obtained from the information of the covariance, variance and Jacobian matrixes. Combining these informations and the uncertainties of the experimental measurement it is possible to obtain the overall uncertainty of the determined property by statistical error propagation.



## **Análise Conjunta na Produção do Abacateiro (*Persea americana* Mill), cv. Hass em Função de Distintas Lâminas de Irrigação**

<sup>1</sup>Jorge A. de Sousa; <sup>2</sup>Eduardo A. H. Hoces; <sup>3</sup>Edwirde L. Silva; <sup>4</sup>Mácio A. de Albuquerque

<sup>1</sup> *Universidade Federal de Campina Grande, Unidade Acadêmica de Educação, Brasil, jorgeas@ufcg.edu.br;* <sup>2</sup> *Universidade de Concepción, Departamento de Engenharia Agrícola, Chile, e-mail ehholzapf@udec.cl;* <sup>3</sup> *Universidade Estadual da Paraíba, Departamento de Estatística, Brasil, edwirde1@hotmail.com.;* <sup>4</sup> *Universidade Estadual da Paraíba, Departamento de Estatística, Brasil, macioaa@uol.com.br*

### **Resumo**

Experimentos agronômicos são muitas vezes replicados ao longo do tempo e do espaço, para avaliar como se comportam tratamentos conduzidos em distintas situações ambientais. A análise dos experimentos realizados por mais de uma temporada (anos) e/ou lugares (locais) é comumente denominada de análise conjunta de experimentos. Segundo Barbin (2003), busca-se, nesse tipo de estudo, tomar sempre que possível decisões abrangentes a uma grande área, caso se trabalhe com um mesmo experimento em diversas áreas, ou a um grande período de tempo, caso se trabalhe com um mesmo experimento ao longo do tempo, apontando o mesmo ou similar comportamento por parte dos tratamentos em questão, sempre que a interação se apresentar não significativa. Pimentel-Gomes (2009) afirma que quando a relação entre quadrados médios residuais das análises individuais for além do estipulado, ou seja, sete, convém considerar separadamente subgrupos de experimentos com quadrados médios residuais não muito heterogêneos para construir as análises conjuntas. Barbin (2003) reforça tal ideia, afirmando que nos casos em que essa homogeneidade não é satisfeita se pode eliminar os experimentos cujo QMRes sejam discrepantes, ou ainda, agrupar os experimentos que possuam QMRes similares, levando em consideração critérios enumerados por Campos (1984), citados nesse trabalho. Diante do

exposto realizou-se análise conjunta para avaliar a significância na produção de abacates no Vale de Peumo, Província de Cachapoal, Sexta Região, Chile, em um delineamento em blocos ao acaso com quatro tratamentos (lâminas de irrigação), quatro repetições e com seis plantas por parcela, no período de 2004 a 2008, excluindo-os anos de 2004 e 2005 por não satisfazerem o critério citado anteriormente. Os tratamentos foram significativamente ( $p < 0,05$ ) crescentes para os anos 2006, 2007 e 2008, ou seja, à medida que se aumentava a lâmina de irrigação, aumentava-se a produção de abacates.

**Palavras-chave:** análise conjunta, quadrado médio, experimentos.

## **Referências**

BARBIN, D. **Planejamento e análise estatística de experimentos agrônômicos**. Araçongas: Midas, 2003. 208 p.

PIMENTEL-GOMES, F. **Curso de estatística experimental**. 15. ed. Piracicaba: FEALQ, 2009. 451 p. (Biblioteca de Ciências Agrárias Luiz de Queiroz, 15).

## Normalidade de Variáveis Resposta com Excesso de Zeros

<sup>1</sup>Jorge A. de Sousa; <sup>2</sup>Eduardo A. H. Hoces; <sup>3</sup>Edwirde L. Silva; <sup>4</sup>Mácio A. de Albuquerque

<sup>1</sup> *Universidade Federal de Campina Grande, Unidade Acadêmica de Educação, Brasil, jorgeas@ufcg.edu.br;* <sup>2</sup> *Universidade de Concepción, Departamento de Engenharia Agrícola, Chile, e-mail ehholzapf@udec.cl;* <sup>3</sup> *Universidade Estadual da Paraíba, Departamento de Estatística, Brasil, edwirde1@hotmail.com;* <sup>4</sup> *Universidade Estadual da Paraíba, Departamento de Estatística, Brasil, macioaa@uol.com.br*

### Resumo

Muitos resultados importantes em análise estatística seguem a partir da suposição de que a população a ser amostrada ou investigada é normalmente distribuída. Quando essa suposição para um método selecionado é aproximadamente satisfeita, usuais procedimentos podem ser aplicados sem restrições para fazer inferências sobre parâmetros de interesse desconhecidos. Entretanto, quando a normalidade não é constatada, Graybill (1976) sugere várias opções, dentre as quais, transformação dos dados ou filtragem de alguns pontos suspeitos, denominados outlier. Muitos pesquisadores têm utilizado essa opção para conseguir a normalidade dos dados (HOYLE, 1973; THOENI, 1967). Neste contexto, o potencial da transformação paramétrica proposta por Box e Cox (1964) é utilizada para encontrar uma métrica em que as suposições teóricas sejam satisfeitas em variáveis com excesso de zeros. A normalidade dos erros é uma exigência constante, entretanto em dados de produção essa exigência pode, muitas vezes, ser negligenciada, principalmente em regiões expostas a grandes variações climáticas, como no caso do Vale de Peumo, Província de Cachapoal, Sexta Região, Chile. Aplicou-se a transformação de Box e Cox (1964) para um experimento em blocos ao acaso, com uma taxa de frequência de zeros de 20%. Por meio do teste de Shapiro e Wilk (1965), verificou-se a potência da transformação, constatando-se, que para dados com baixa frequência de zeros a transformação de Box e Cox foi eficiente.

**Palavras-chave:** transformação, suposições, análise estatística.

## Referências

BOX, G. E. P.; COX, D. R. An analysis of transformations. **Journal of the Royal Statistical Society**, Series B, London, v. 26, p. 211-252, 1964.

HOYLE, M. H. Transformations: an introduction and a bibliography. **The International Statistical Review**, The Hague, v. 41, p. 203-223, 1973.

GRAYBILL, F. A. **Theory and application of the linear model**. Massachusetts: Ouxburg Press, 1976. 704 p.

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). **Biometrika**, London, v. 52, p. 591-611, 1965.

THOENI, H. **Transformation of variables used in the analysis of experimental and observational data**: a review. Ames: Iowa State University, 1967. (Technical Report, 7).

## **Escolha de Modelos Volumétricos para Clones de *Eucalyptus ssp.*, no Polo Gesseiro do Araripe, PE**

<sup>1</sup>Jorge A. Sousa; <sup>2</sup>Mácio A. Albuquerque; <sup>3</sup>Edwirde L. Silva; <sup>4</sup>Joseilme F. Gouveia

<sup>1</sup> *Universidade Federal de Campina Grande, Unidade Acadêmica de Educação, Brasil, jorgeas@ufcg.edu.br;* <sup>2</sup> *Universidade Estadual da Paraíba, Departamento de Estatística, Brasil, macioaa@uol.com.br;* <sup>3</sup> *Universidade Estadual da Paraíba, Departamento de Estatística, Brasil, edwirde1@hotmail.com;* <sup>4</sup> *Universidade Federal da Paraíba, Centro de Ciências Aplicadas e Educação, Brasil, joseilme@dce.ufpb.br*

### **Resumo**

Há certo tempo já se tem percebido crescente desenvolvimento dos setores sociais e industriais. Como consequência de tal desenvolvimento, aumentou-se a necessidade por recursos energéticos, cuja carência que tem sido suprida por meio de recursos próprios das regiões de matas nativas. Tal fato pode ser justificado pela procura da biomassa de menor valor e fácil aquisição. Desse modo, como resultado preocupante, temos o aumento de áreas desertificadas que crescem em ritmo acelerado, motivado pela exploração indiscriminada e irracional dos recursos florestais (OLIVEIRA et al., 2006). Segundo Barros et al. (2010), o *Eucalyptus* representa uma das melhores formas de conhecer o potencial produtivo em florestas, e o conhecimento do seu estoque é essencial para todas as atividades de planejamento, sendo o volume individual a base para se conhecer todo um estoque que a floresta pode guardar. Deste modo, o objetivo deste estudo foi estimar o volume dos clones de *Eucalyptus* utilizando os principais modelos volumétricos não lineares. A base de dados é proveniente do povoamento de clones de *Eucalyptus*, localizada no Município de Araripina, no Semiárido de Pernambuco, implantado no ano de 2002. O modelo volumétrico de Schumacher e Hall foi o que melhor se ajustou aos dados de volume de *Eucalyptus*, pois apresentou menores valores de AIC, BIC e EPAM, seus parâmetros foram significativos e na análise de resíduos nenhuma pressuposição do modelo foi violada. Assim, espera-

se que com o uso de modelos volumétricos para o inventário florestal seja possível contribuir para a redução dos custos do inventário, ou até mesmo melhorar a precisão dos inventários, e assim evitar prejuízos, tanto para o produtor quanto para o trabalhador rural. Além disso, os resultados obtidos poderão propiciar o desenvolvimento florestal da região trazendo benefícios sociais, econômicos e ambientais.

**Palavras-chave:** *Eucalyptus*, modelos volumétricos, polo gesseiro do Araripe-PE.

## Referências

BARROS, B. C. de; SILVA, J. A. A. da; FERREIRA, R. L. C.; REBOUÇAS, A. C. M. N. Volumetria e sobrevivência de espécies nativas e exóticas no polo gesseiro do Araripe, PE. **Ciência Florestal**, Santa Maria, v. 20, n. 4, p. 641-647, out.-dez. 2010.

OLIVEIRA, E.; VITAL, B. R.; PIMENTA, A. S.; DELLA LUCIA, R. M.; LADEIRA, A. M. M.; CARNEIRO, A. de C. O. Estrutura anatômica da madeira e qualidade do carvão de *Mimosa tenuiflora* (willd.) Poir. **Revista Árvore**, Viçosa, v. 30, n. 2, p. 311-318, mar./abr. 2006.

## **Seleção de Modelos Matemáticos para Estimar Biomassa Florestal em Vegetação do Sub-Bosque de Um Plantio Comercial de *Bertholletia excelsa* H&B**

<sup>1</sup>José Cordeiro Neto; <sup>2</sup>Celso P. Azevedo; <sup>3</sup>Roberval M. B. Lima

<sup>1</sup>Universidade Federal do Amazonas, mestre em Ciências Florestais e Ambientais, Brasil, josecordeironeto@hotmail.com; <sup>2</sup>Embrapa Amazônia Ocidental, pesquisador, Brasil, celso.azevedo@embrapa.br; <sup>3</sup>Embrapa Amazônia Ocidental, Pesquisador, Brasil, roberval.lima@embrapa.br

### **Resumo**

As estimativas de biomassa dos diversos tipos florestais da Amazônia hoje disponíveis na literatura vêm de estudos que utilizam métodos destrutivos e não destrutivos. Assim, pela dificuldade em determinar a biomassa de cada árvore pelo método destrutivo, é necessário realizar análise de regressão para gerar modelos estatísticos que possam ser aplicados para estimar biomassa em regiões maiores. Este estudo teve como objetivo ajustar modelos matemáticos para estimar a biomassa florestal do sub-bosque de um plantio comercial de *Bertholletia excelsa* H&B, onde não foram aplicados tratamentos culturais, conseqüentemente nascendo a vegetação de estudo. Foram ajustados oito modelos de simples e dupla entrada para um grupo de nove espécies florestais que foram selecionadas pelo maior valor do Índice de Valor de Importância Ampliado (IVIA), no qual foram pesadas dez árvores de cada espécie. Os modelos logaritmos tiveram o erro recalculado pelo Índice de Furnival para que pudessem ser comparados com os demais. Todas as equações apresentaram o erro padrão da estimativa (Syx %) abaixo de 10%, já o coeficiente de determinação ajustado ( $R^2_{aj}$ ) variou entre 78,54% e 95,54%. A equação de simples entrada  $PS = 0,1263 d2,4245$  apresentou os melhores resultados,  $R^2_{aj}$  de 94,01% e erro padrão da estimativa de 3,98%, além de boa distribuição dos resíduos. Já Nelson et al.

(1999) ajustaram modelos para oito espécies de uma floresta secundária, e o modelo que melhor se ajustou para todas as espécies foi  $\text{LnP} = -1,9968 + 2,4128 \cdot \text{Ln DAP}$  com o  $R^2$  98,4% e  $S_{yx} = 0,0610$ .

**Palavras-chave:** regressão, biomassa florestal, floresta secundária.

### **Referência**

NELSON, B. W.; MESQUITA, R.; PEREIRA, J. L. G.; SOUZA, S. G. S.; BATISTA, G. T.; COUTO, L. Allometric regressions for improved estimate of secondary forest biomass in the Central Amazon. **Forest Ecology and Management**, Amsterdam, v. 117, n. 1-3, p. 149-167, May 1999.



## Comparison of Peroxide Index of Olive Oil Suppliers

<sup>1</sup>Luis M. Grilo; <sup>2</sup>Helena L. Grilo

<sup>1</sup>*Unidade Departamental de Matemática e Física, Instituto Politécnico de Tomar e Centro de Matemática e Aplicações (CMA), FCT, UNL, Portugal;* <sup>2</sup>*Centro de Sondagens e Estudos Estatísticos, Instituto Politécnico de Tomar, Portugal.*

### Abstract

The peroxide index is a variable that measures the initial oxidation of olive oil as well as its deterioration. It is analyzed in laboratory in order to assess its quality and to characterize the type of oil (eg, refined, virgin, extra virgin). To compare the variable means of three different suppliers of virgin olive oil we used one-way ANOVA with an adequate Box-Cox transformation, since its assumptions (normality and homoscedasticity) are not clearly verified. Afterwards, we also applied a non-parametric test, which in this case lead us to the same conclusions.

**Keywords:** industrial process, one-way ANOVA, Box-Cox transformation, non-parametric test, robustness.



## **Comparação do Teste Ideal das Médias dos Tratamentos no Delineamento Experimental**

<sup>1</sup>Mácio A. Albuquerque; <sup>2</sup>Jorge A. Sousa; <sup>3</sup>Edwirde L. Silva; <sup>4</sup>Joseilme F. Gouveia

<sup>1</sup> *Universidade Estadual da Paraíba, Departamento de Estatística, Brasil, macioaa@uol.com.br;* <sup>2</sup> *Universidade Federal de Campina Grande, Unidade Acadêmica de Educação, Brasil, jorgeas@ufcg.edu.br;* <sup>3</sup> *Universidade Estadual da Paraíba, Departamento de Estatística, Brasil, edwirde1@hotmail.com;* <sup>4</sup> *Universidade Federal da Paraíba, Centro de Ciências Aplicadas e Educação, Brasil, joseilme@dce.ufpb.br*

### **Resumo**

O melhoramento genético e as extensões de terras no país para a atividade florestal ajudam a garantir o abastecimento de indústrias que utilizam a lenha de maneira direta ou indireta como fonte energética. Segundo Bertoldo et al. (2007), na prática para escolha do teste ideal para realização da comparação das médias dos tratamentos no delineamento experimental, não existe uma maneira padrão ou manual ou solução que defina o teste ótimo. Em geral, os autores de cada teste estabelecem apenas algumas restrições teóricas ou pequenos detalhes, mas nenhum deles determina a que experimento o seu teste é indicado, dificultando assim para o experimentador conhecer o teste que deveria adotar. Para Silva (2009), como existe a necessidade de estabelecer os testes mais indicados e ilustrar uma aplicação, utilizou-se, neste trabalho, um delineamento localizado na região do polo gesseiro do Araripe, com 15 clones de eucalyptus que foram mensurados a cada seis meses durante sete anos, no período de 2002 a 2009. Os testes utilizados foram os mais difundidos na literatura, sendo aplicadas as suas metodologias originais com a finalidade de analisar as semelhanças entre cada teste nos níveis de significância, 1% e 5%, bem como estabelecer nesse delineamento qual(is) clone(s) possui(em) a(s) média(s) de produção mais significativa(s).

De acordo com os resultados desta pesquisa, concluiu-se que o teste de Tukey e Conagin são os mais recomendados e indicados para produção de biomassa na região do polo gesso do Araripe, Pernambuco.

**Palavras-chave:** experimentais, produtividade, comparação de médias.

## **Referências**

BERTOLDO, J. G.; ROCHA, F. da; COIMBRA, J. L. M.; ZITTERELL, D.; GRAH, V. de F. Teste de comparação de médias: dificuldades e acertos em artigos científicos. **Revista Brasileira de Agrociência**, Pelotas, v. 13, n. 4, p. 441-447, out.-dez. 2007.

SILVA, J. A. A. da. Potencialidades de florestas energéticas de Eucalyptus no Pólo Gesso do Araripe - Pernambuco. **Anais da Academia Pernambucana de Ciência Agrônômica**, Recife, v. 5-6, p. 301-319, 2008-2009.

## **Comparação de Dendrogramas por Métodos de Agrupamentos via Reamostragem "Bootstrap"**

<sup>1</sup>Mácio A. Albuquerque; <sup>2</sup>Jorge A. Sousa; <sup>3</sup>Edwirde L. Silva; <sup>4</sup>Joseilme F. Gouveia;  
<sup>5</sup>Carla C. C. Silva

<sup>1</sup> *Universidade Estadual da Paraíba, Departamento de Estatística, Brasil, macioaa@uol.com.br;* <sup>2</sup> *Universidade Federal de Campina Grande, Unidade Acadêmica de Educação, Brasil, jorgeas@ufcg.edu.br;* <sup>3</sup> *Universidade Estadual da Paraíba, Departamento de Estatística, Brasil, edwirde1@hotmail.com;* <sup>4</sup> *Universidade Federal da Paraíba, Centro de Ciências Aplicadas e Educação, Brasil, joseilme@dce.ufpb.br;* <sup>5</sup> *Universidade Estadual da Paraíba, Departamento de Estatística, Brasil, carlacarolinecarvalhosilva@gmail.com.*

### **Resumo**

O processo de agrupamento envolve basicamente duas etapas. A primeira se refere à estimação de uma medida de dissimilaridade entre os indivíduos; a segunda, à adoção de uma técnica de formação de grupos. Segundo Barroso e Artes (2003), para a realização dessa tarefa existe grande número de métodos disponíveis, dentre os quais o pesquisador deverá escolher o mais adequado ao seu propósito. Objetivou-se propor uma sistemática para o estudo e a interpretação da estabilidade dos métodos em análise de agrupamento, por meio de vários algoritmos de agrupamento em dados de vegetação. Albuquerque et al. (2006) utilizaram dados provenientes de levantamento na Mata da Silvicultura, da Universidade Federal de Viçosa, MG. Para análise de agrupamento foram estimadas as matrizes de distância de Mahalanobis com base nos dados originais e via reamostragem "bootstrap" e aplicados os métodos da ligação simples, ligação completa, médias das distâncias, e do Ward. Para a detecção de associação entre os métodos foi aplicado o teste qui-quadrado (2) a 1% e 5% de probabilidade. Para os diversos métodos de agrupamento foi obtida a correlação cofenética. Os resultados de associação dos métodos foram semelhantes, indicando em

princípio que qualquer algoritmo de agrupamento estudado está estabilizado e existem, de fato, grupos entre os indivíduos observados. No entanto, observou-se que os métodos são coincidentes, com base nas matrizes dos dados originais e "bootstrap". A sistemática proposta é promissora para o estudo e a interpretação da estabilidade dos métodos em dados de vegetação.

**Palavras-chave:** multivariada, bootstrap, agrupamentos.

## **Referências**

ALBUQUERQUE, M. A. de; FERREIRA, R. L. C.; SILVA, J. A. A. da; SANTOS, E. de S.; STOSIC, B.; SOUZA, A. L. de. Estabilidade em análise de agrupamento: estudo de caso em ciência florestal. **Revista Árvore**, Viçosa, v. 30, n. 2, p. 257-265, mar./abr. 2006.

BARROSO, L. P.; ARTES, R. Análise multivariada. In: REUNIÃO ANUAL DA RBES, 48.; REUNIÃO ANUAL DA SEAGRO, 10., 2003, Lavras. **Anais...** Lavras: Ufla, 2003. p. 155.

## **Procedimento para Determinação e Validação de Números de Grupos em Análise de Agrupamento**

<sup>1</sup>Mácio A. Albuquerque; <sup>2</sup>Jorge A. Sousa; <sup>3</sup>Edwirde L. Silva; Carla C. C. Silva<sup>5</sup>

<sup>1</sup> *Universidade Estadual da Paraíba, Departamento de Estatística, Brasil, macioaa@uol.com.br;* <sup>2</sup> *Universidade Federal de Campina Grande, Unidade Acadêmica de Educação, Brasil, jorgeas@ufcg.edu.br;* <sup>3</sup> *Universidade Estadual da Paraíba, Departamento de Estatística, Brasil, edwirde1@hotmail.com;* <sup>4</sup> *Universidade Estadual da Paraíba, Departamento de Estatística, Brasil, carlacarolinecarvalhosilva@gmail.com.*

### **Resumo**

Dentre a família de técnicas multivariadas, a análise de agrupamento é uma das mais utilizadas, dada sua utilidade pragmática. Neste trabalho, objetivou-se desenvolver nova abordagem para a análise de agrupamento, a partir da combinação de características das técnicas hierárquicas. Conforme Meyer et al. (2004), procurou-se fornecer uma análise exploratória mais completa dos dados, visando facilitar o trabalho dos pesquisadores quanto a outliers, número de grupos, técnicas de agrupamento e de validação dos grupos, e aumentar o conhecimento que pode ser obtido com a aplicação de um conjunto de sentenças lógicas em análise de agrupamento. Este estudo foi realizado a partir de um banco de dados obtidos por Costa Júnior (2006), em um remanescente de Mata Atlântica denominado Mata das Caldeiras, localizado no Município de Catende, PE, que está situado na mesorregião da Mata Pernambucana. Foram aplicadas, com base na distância euclidiana, as técnicas hierárquicas de ligação simples, de ligação completa, de ligação média e de Ward. Para validação dos métodos foram utilizados os coeficientes cofenético e Rand ajustado. A metodologia apresentada foi aplicada a conjuntos de dados artificiais, comparando os resultados com os obtidos por outros métodos de agrupamento. Observou-se que o índice de Rand ajustado

tem o potencial de melhorar significativamente a tomada de decisões sobre o número de grupos, tornando-se, assim, uma técnica recomendável para buscar o número de grupos ideal de forma criteriosa.

**Palavras-chave:** análise multivariada, agrupamento, validação.

## **Referências**

COSTA JÚNIOR, R. F. **Caracterização estrutural de um remanescente de Mata Atlântica do município de Catende-PE**. 2006. 52 f. Dissertação (Mestrado em Ciências Florestais) - Universidade Federal Rural de Pernambuco, Recife.

MEYER, A. da S.; GARCIA, A. A. F.; SOUZA, A. P. de; SOUZA JUNIOR, C. L. de. Comparison of similarity coefficients used for cluster analysis with dominant markers in maize (*Zea mays* L.). **Genetics and Molecular Biology**, São Paulo, v. 27, n. 1, p. 83-91, 2004.



## **Processos Autorregressivos com Valores Inteiros Não Negativos – INAR**

<sup>1</sup>Nataniel Barros; <sup>2</sup>Maria do Rosário Ramos; <sup>3</sup>Amílcar Oliveira

<sup>1</sup>*Mestrando em Estatística, Matemática e Computação, Universidade Aberta, Portugal;*

<sup>2</sup>*Universidade Aberta, Portugal, marosram@uab.pt;* <sup>3</sup>*Universidade Aberta e CEAUL, Portugal, Amilcar.Oliveira@uab.pt*

### **Resumo**

Neste trabalho, pretende-se analisar os desenvolvimentos recentes no estudo dos Processos autorregressivos com valores inteiros não negativos, uma classe de modelos particularmente aplicável à análise de séries de contagens. Começou-se por definir as séries temporais, as suas componentes de tendência e sazonalidade e a estacionaridade. Estudaram-se igualmente os processos lineares (ARMA), a estimação dos parâmetros e a predição dos processos autorregressivos. Para o estudo do modelo autorregressivo com valores inteiros não negativos de primeira ordem – INAR (1), definiu-se primeiramente as suas propriedades analisando-se a estimação dos parâmetros do modelo pelos métodos de Yule-Walker, dos Mínimos Quadrados e de Máxima Verosimilhança. A generalização dos processos INAR (1) foi tratada a uma ordem superior  $p$  – INAR ( $p$ ) e para tal, estudou-se o critério de seleção de ordem e a estimação dos parâmetros do modelo pelos métodos considerados anteriormente e foram estabelecidas as suas propriedades assintóticas, analisando a consistência dos estimadores via simulação. Em seguida, foram apresentados os métodos de inferência e previsão em processos INAR de ordem 1 e superior. Finalmente, aplicou-se um caso prático a uma série econômica de Cabo Verde.

## Agradecimentos

Investigação parcialmente suportada pela Fundação para a Ciência e a Tecnologia (FCT), projeto PEst-OE/MAT/UI0006/2014 (CEAUL).

## Referências

ALZAID, A. A.; AL-OSH, M. An integer-valued  $p$ th-order autoregressive structure (INAR (p)) process. **Journal of Applied Probability**, Sheffield, v. 27, n. 2, p. 314-324, 1990.

BU, R.; McCABE, B.; HADRI, K. Maximum Likelihood Estimation of Higher-Order Integer-Valued Autoregressive Processes. **Journal of Time Series Analysis**, Avon, v. 29, n. 6, p. 973-994, nov. 2008.

OLIVEIRA, V. L.; SILVA, M. E. Processos para séries temporais de contagem. In: CONGRESSO ANUAL DA SOCIEDADE PORTUGUESA DE ESTATÍSTICA, 7., 2001, Coimbra. **Actas: um olhar sobre a estatística**. Coimbra: Edições SPE, 2001. p. 470-482.

SILVA, I.; SILVA, M. E. Selecção de ordem em modelos INAR. Actas do X Congresso Anual da SPE. Disponível em: <[http://paginas.fe.up.pt/~ims/sel\\_ordem\\_spe.pdf](http://paginas.fe.up.pt/~ims/sel_ordem_spe.pdf)>. Acesso em: 08 ago. 2014.

SILVA, I. M. M. da. **Contributions to the analysis of discrete-valued time series**. 209 f. Tese (Doutorado em Matemática Aplicada) – Faculdade de Ciências, Universidade do Porto, Porto.

## **Generalized Linear Models and Risk Analysis, an Introductory Approach Using R**

Pedro Rebelo<sup>1</sup>, Teresa Oliveira<sup>2</sup>

<sup>1</sup>*Universidade Aberta MEMeC, Portugal, trashmanp1@gmail.com;* <sup>2</sup>*Universidade Aberta and Center of Statistics and Applications (CEAUL), Portugal, teresa.oliveira@uab.pt*

### **Abstract**

In this work the authors address a brief survey on the theme General Linear Models (GLM) and its practical usefulness, particularly in the used logistic regression by using R software. The poster structure is composed by two parts: (part 1)- a brief theoretical introduction to this type of models is presented, namely the concept of link function, relative risk and odds ratio; (part 2) here practical aspects of R software use is presented to obtain models, which are illustrated with simulated data (a disease that leads to blindness), and real data is presented (exploratory data analysis of a file of hypertensive patients), particularly in situations where the predictors are not continuous variables, or when they do not show a normal distribution. The created GLM can be used in exploratory data analysis, and easily, from these, we obtain useful elements in the risk analysis, including the odds ratio, which is a mental tool of undeniable interest in the medical field, and intuitively approaches the concept of relative risk in some specific situations, additionally to the usefulness of GLM prediction, they also can indirectly help to clarify and substantiate the differences between risk options such as therapeutic strategies against a set of predictors variables.

**Keywords:** General Linear Models (GLM), risk analysis, logistic regression, Odds Ratio, quantal response.

## References

DEEKS, J. When can odds ratios mislead?: odds ratios should be used only in case-control studies and logistic regression analyses. **British Medical Journal**, London, v. 317, n. 7166, p. 1155, Oct. 1998.

NELDER, J. A.; BAKER, R. J. **Generalized linear models**. London: John Wiley & Sons, 1972.

SULLIVAN, L. M.; MASSARO, J. M.; D'AGOSTINO, R. B. Presentation of multivariate data for clinical use: the Framingham Study risk score functions. **Statistics in Medicine**, Chichester, v. 23, n. 10, p. 1631-1660, 2004.

ZHENG, B.; AGRETI, A. Summarizing the predictive power of a generalized linear model. **Statistics in Medicine**, Chichester, v. 19, n. 13, p. 1771-1781, 2000.

## **Técnicas de Análise de Sobrevivência Aplicadas em Ensaio Florestal na Amazônia**

<sup>1</sup>Roberval M. B. de Lima; <sup>2</sup>Cintia R. de Souza; <sup>3</sup>Edluz Gabriel da C. N. de Lima

<sup>1</sup>Embrapa Amazônia Ocidental, Brasil, [roberval.lima@embrapa.br](mailto:roberval.lima@embrapa.br); <sup>2</sup>Embrapa Amazônia Ocidental, Brasil, [cintia.souza@embrapa.br](mailto:cintia.souza@embrapa.br); <sup>3</sup>Universidade Paulista (Unip), Brasil, [edluz.gabriel@globocom.com](mailto:edluz.gabriel@globocom.com)

### **Resumo**

Estudos de sobrevivência na ciência florestal são importantes e frequentemente realizados, pois essa variável está intimamente relacionada com a adaptação e produtividade das espécies florestais. Os dados para realização deste estudo foram coletados na área experimental da Embrapa Amazônia Ocidental, com as espécies *Cordia goeldiana* (freijó), *Sclerolobium paniculatum* (taxi-branco) e *Schefflera morototoni* (morototó), iniciando-se em 1998 (tempo zero) até 60 meses após o plantio. Os objetivos deste trabalho consistiram na aplicação das Técnicas de Análise de Sobrevivência, para estudar o comportamento das três espécies florestais quanto ao tempo de sobrevida das árvores, em dois locais (Km 30 e Caldeirão) no Estado do Amazonas, e determinar a influência das covariáveis “espécie” e “local”. A metodologia consistiu em descrever a função de sobrevivência das árvores utilizando método não paramétrico de Kaplan – Meier (GIOLO; SHIMAKURA, 2004; KLEIN; MOESCHBERGER, 1997) com posterior ajuste dos modelos de Weibull, logístico e exponencial. Os principais resultados obtidos indicam a técnica de análise de sobrevivência como uma ferramenta adequada para estimar as curvas de sobrevida das árvores. A espécie *C. goeldiana* (freijó) apresentou melhor adaptação ao local Km 30, considerando o tempo de sobrevida até 60 meses de idade, e comparando-se a sobrevivência de três espécies florestais para o local Km 30, obteve-se a seguinte ordem de importância: taxi-branco > freijó > morototó; e para o local Caldeirão: taxi-branco > morototó > freijó. Entre os modelos testados, o de Weibull

e logístico foram os que apresentaram os melhores ajustes para descrever a função de sobrevivência das espécies florestais, e recomenda-se o seu emprego em ensaios dessa natureza.

**Palavras-chave:** Kaplan-Meier, modelagem, logístico, Weibull, exponencial, *Cordia goeldiana*, *Sclerolobium paniculatum*, *Schefflera morototoni*.

## Referências

GIOLO, S. R.; SHIMAKURA, S. E. **Nonparametric method for analyzing interval-censored survival data**. Curitiba: UFPR, Department of Statistics, 2004. Mimeografado.

KLEIN, J. P.; MOESCHBERGER, M. L. **Survival analysis: techniques for censored and truncated data**. 3. ed. New York: Springer-Verlag, 1997.

## **Análise Econômica de Sistemas Agroflorestais na Amazônia em Condições de Risco**

<sup>1</sup>Simão C. Silva; <sup>2</sup>Antonio D. Oliveira; <sup>3</sup>Elisa V. Wandelli

*<sup>1</sup>Professor Doutor da Universidade Federal de Lavras, Ufla, Departamento de Ciências Florestais, Lavras, MG, Brasil, simao@dcf.ufla.br; <sup>2</sup>Professor Doutor da Universidade Federal de Lavras, Departamento de Ciências Florestais, Lavras, MG, Brasil, donizete@dcf.ufla.br; <sup>3</sup>Pesquisadora Doutora da Embrapa Amazônia Ocidental, Manaus, AM, Brasil, elisa.wandelli@embrapa.br*

### **Resumo**

Os sistemas agroflorestais (SAFs) apresentam-se como alternativas de cultivos mais adequadas para a região Amazônica, principalmente quando se observam as questões socioambientais dos sistemas. Outro aspecto importante que deve ser analisado é a questão econômica. Neste estudo, analisou-se a viabilidade econômica e compararam-se quatro SAFs implantados na região Amazônica levando-se em consideração os riscos de investimento nessa atividade. Foram avaliados dois sistemas agrossilviculturais (AS1 e AS2) e dois sistemas agrossilvipastoris (ASP1 e ASP2). Os SAFs estudados estão situados no campo experimental da Embrapa Amazônia Ocidental, no Município de Manaus, Estado do Amazonas. A base de dados para modelagem da análise de risco foi adaptada a partir do trabalho de Santos (2000). Para a análise econômica em condições de risco, elaboraram-se fluxos de caixa para os sistemas e definiram-se como variáveis de risco a taxa de juros, o custo de implantação do sistema, a produção e o preço dos diversos produtos do sistema. A distribuição de probabilidade triangular foi adotada como base para modelagem das variáveis de risco. A variável de saída do modelo foi o Valor Anual Equivalente. O risco foi analisado por meio de simulações utilizando-se o método de Monte Carlo. A utilização de um modelo probabilístico permitiu considerar o fator incerteza na análise econômica, o que melhorou a compreensão e aumentou a segurança para a

tomada de decisão em relação à viabilidade econômica dos SAFs estudados. Os quatro SAFs mostraram-se viáveis economicamente, e a decisão de investir neles pode ser tomada de forma segura, do ponto de vista econômico, já que não há chances ou riscos de se ter prejuízo no investimento. O sistema AS2 foi o que apresentou melhor desempenho econômico, seguido pelos sistemas AS1, ASP1 e ASP2.

**Palavras-chave:** análise de risco, método de Monte Carlo, simulação.

## **Apoio**

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes).

## **Referência**

SANTOS, M. J. C. **Avaliação econômica de quatro modelos agroflorestais em áreas degradadas por pastagens na Amazônia Ocidental**. 2000. 75 f. Dissertação (Mestrado em Ciências) – Escola Superior de Agricultura “Luiz de Queiroz”, Piracicaba.



## **Educational Data Mining e Learning Analytics na Melhoria do Ensino Online**

<sup>1</sup>Susana Faria; <sup>2</sup>Angel Juan; <sup>3</sup>Amílcar Oliveira

<sup>1</sup>MEMeC – Universidade Aberta, Portugal, [susanafaria76@gmail.com](mailto:susanafaria76@gmail.com); <sup>2</sup>IN3 - Universidad Oberta de Catalonia, Espanha, [ajuanp@uoc.edu](mailto:ajuanp@uoc.edu); <sup>3</sup>Universidade Aberta, Portugal, [Amilcar.Oliveira@uab.pt](mailto:Amilcar.Oliveira@uab.pt)

### **Resumo**

A educação é um processo de aquisição de conhecimento e/ou aptidões que tem sofrido várias transformações ao longo do tempo. A aprendizagem online, metodologia cada vez mais escolhida, cria trilhos de dados dos estudantes que podem fornecer informações valiosas no decorrer do processo de aprendizagem, nomeadamente informações sobre estudantes que estão em risco de abandono escolar ou que precisam de atividades para aumentar o sucesso e a confiança. Educational data mining e Learning analytics são duas áreas em desenvolvimento que trabalham especificamente com grandes volumes de dados em educação.

Este trabalho baseia-se na aplicação de três técnicas de Educational data mining (análise de clusters, cartas de controle e regressão logística binária) a fim de desenvolver um modelo que permita aos professores de cursos de ensino online supervisionar de forma eficiente a atividade e o desempenho dos estudantes. Foi aplicado um questionário a 30 estudantes do ensino online e fez-se a análise estatística dos dados utilizando as referidas técnicas. Com o modelo desenvolvido foi possível obter uma ferramenta de trabalho mais eficaz para o professor, que a qualquer momento do processo de ensino/aprendizagem pode analisar as perspectivas de sucesso e insucesso de um estudante e procurar minimizar o número de casos de abandono escolar.

**Palavras-chave:** educational data mining, Learning analytics, ensino online.

## **Agradecimentos**

Investigação parcialmente suportada pela Fundação para a Ciência e a Tecnologia (FCT), projeto PEst-OE/MAT/UI0006/2014 (CEAUL).

## **Referências**

BIENKOWSKI, M.; FENG, M.; MEANS, B. **Enhancing teaching and learning through educational data mining and learning analytics: an issue brief**. Washington, D.C.: Office of Educational Technology, 2012.

DYCKHOFF, A. L.; ZIELKE, D.; BÜLTMANN, M.; CHATTI, M. A.; SCHROEDER, U. Design and implementation of a learning analytics toolkit for teachers. **Journal of Educational Technology & Society**, Palmerston North, v. 15, n. 3, p. 58-76, 2012.

JUAN, A.; DARADOUMIS, T.; FAULIN, J.; XHAFI, F. SAMOS: a model for monitoring students' and groups' activities in collaborative e-learning. **International Journal of Learning Technology**, Geneva, v. 4, p. 53-72, 2009.

ROMERO, C.; VENTURA, S. Data mining in education. **Data Mining Knowledge Discovery**, Boston, v. 3, p. 12-27, jan./feb. 2013.

SIEMENS, G., BAKER, R. S. J. D. Learning analytics and educational data mining: towards communication and collaboration. In: INTERNACIONAL CONFERENCE ON LEARNING ANALYTICS AND KNOWLEDGE, 2., 2012, Vancouver. **Proceedings...** New York: Association for Computing Machinery, 2012. p. 252-254.

# **Robust Heritability Estimation In Plant Candidate Genetic Association Studies**

<sup>1</sup>Vanda M. Lourenço; <sup>2</sup>Paulo C. Rodrigues; <sup>3</sup>Ana M. Pires; <sup>2</sup>Miguel Fonseca

<sup>1</sup>*FCT/UNL and CMA-FCT/UNL, vmml@fct.unl.pt; <sup>2</sup>CMA-FCT/UNL; <sup>3</sup>CMA IST/UL and CEMAT-IST/UL*

## **Abstract**

Heritability is key in plant studies to help achieve better yield and other agronomic traits of interest. In candidate gene studies regression models are used to test for associations between phenotype and candidate SNPs. SNP imputation guarantees that marker information is complete so both the coefficient of determination,  $R^2$ , and broad-sense heritability are equivalent. However, when the normality assumption is violated, the classical  $R^2$  may be seriously affected. Recently two  $R^2$  alternatives with good properties were proposed for the linear mixed model. We evaluate their performance under contamination and step forward a robust version of these coefficients assessing their adequacy for heritability estimation via simulation. An example of application with a real plant data set is also presented.

**Keywords:** Robust coefficient of determination, single nucleotide polymorphism, association studies, robust regression.

## **Acknowledgements**

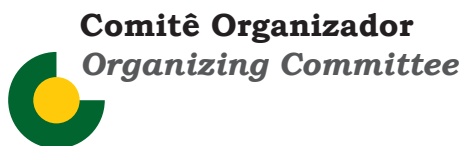
This work received partial financial support from Portuguese National Funds by the Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) through projects PTDC/MAT-STA/0568/2012 and PEst-OE/MAT/UI0297/2014 (Centro de Matemática e Aplicações).





**Comitê Organizador**  
*Organizing Committee*





## **Comitê Organizador**

- **Amílcar Oliveira**  
UAb – Portugal
- **Cintia Rodrigues de Souza**  
Embrapa – Brasil
- **Roberval Lima**  
Presidente da Comissão Organizadora  
Embrapa – Brasil
- **Teresa Oliveira**  
UAb – Portugal

## **Comitê de Honra**

- Chefe da Embrapa Amazônia Ocidental
- Chefe do Departamento de Estatística – Ufam
- Coordenadora do Centro de Estatística e Aplicações da Universidade de Lisboa
- Diretor da Faculdade de Ciências Agrárias – Ufam
- Diretora da Universidade Paulista – Campus Manaus
- Presidente da Fapeam
- Presidente da Sociedade Polaca de Biometria

- Presidente do ISI – CRA (International Statistical Institute-Committee on Risk Analysis)
- Reitor da UAb – Portugal
- Reitor da Ufam

## **Comitê Científico**

- **Amílcar Oliveira**  
UAb – Portugal
- **André Gustavo Pereira**  
Universidade Federal do Rio Grande do Norte
- **Antónia Turkman**  
Faculdade de Ciências da  
Universidade de Lisboa, Portugal
- **Celso Paulo de Azevedo**  
Embrapa – Brasil
- **Christos Kitsos**  
Technological Educational Institution of Athens (TEI) – Greece
- **Iwona Mejza**  
Poznan University of Life Sciences – Poland

- **José Raimundo Gomes Pereira**  
Departamento de Estatística da  
Universidade Federal do Amazonas –  
Brasil
- **José Silva**  
Ulbra – Ji-Paraná, Brasil
- **M. Ivette Gomes**  
Faculdade de Ciências da  
Universidade de Lisboa – Portugal
- **Mauro Oliveira**  
Ulbra – Ji-Paraná, Brasil
- **Max Sousa de Lima**  
Departamento de Estatística da Ufam  
– Brasil
- **Michel Silva**  
Ulbra – Ji-Paraná, Brasil
- **Roberval M. B. de Lima**  
Embrapa, Brasil
- **Stanislaw Mejza**  
Poznan University of Life Sciences –  
Poland
- **Teresa Oliveira**  
UAb – Portugal
- **Thyago Borges**  
Ulbra – Ji-Paraná, Brasil

- **Victor Leiva**  
Universidad de Valparaíso – Chile
- Comissão Organizadora Local**

- **Adriana Ribeiro**  
Embrapa – coordenadora
- **Jucélia Vidal**  
Embrapa – membro
- **Lúcio R. B. Cavalcanti**  
Embrapa – membro
- **Victor Souza**  
Embrapa – membro

**Participantes convidados**

- **Adroaldo Rosseti**  
Embrapa – Brasil
- **Amílcar Oliveira**  
UAb – Portugal
- **Christos Kitsos**  
TEI-Athens – Greece
- **Teresa Oliveira**  
UAb – Portugal
- **Victor Leiva**  
Universidad de Valparaíso – Chile
- **Walmes Zeviani**  
Universidade Federal do Paraná –  
Brasil





Patrocínio  
Sponsored by



Apoio  
Supported by



Realização  
Organized by

