

Associação Brasileira de Estatística
21º SINAPE - Simpósio Nacional de Probabilidade e Estatística - 2014

Comunicações Orais – Resumos

Sessão: Estatística Aplicada em Agronomia e Biologia

Título: Análise probabilística do risco de contaminação de batatas por pesticidas

Autores: Rosineide Fernando da Paz; Carlos Alberto Ribeiro Diniz; Lourival Costa Paraíba; Ricardo Antonio Almeida Pazianotto

Resumo: Para os seres humanos, os vegetais são componentes fundamentais em dietas saudáveis. No entanto, eles também podem ser fontes de substâncias não desejáveis, por exemplo, se contiverem concentrações significativas de pesticidas. Neocotinoídeos são pesticidas utilizados largamente em plantações de batatas e são bons exemplos de substâncias que poderiam ser encontradas em tubérculos de plantas de batatas. Um modelo matemático descrito por um sistema de equações diferenciais ordinárias e parciais que simula a absorção de pesticida do solo por batatas foi utilizado para simular concentrações de pesticidas em batatas e um modelo estatístico que descreve a distribuição de batatas no perfil do solo foram combinados para estimar a probabilidade do risco de contaminação de batatas por pesticidas. Os modelos combinados descrevem a probabilidade do risco de batatas colhidas após aplicações de pesticidas apresentarem concentrações de pesticidas acima de valores limites aceitáveis. O método de análise de risco foi ilustrado supondo-se que batatas foram cultivadas com o neocotinoídeo imidacloprid, um inseticida largamente utilizado em todo o mundo em cultivos de batatas.

Palavras-Chave: *análise de risco; inseticidas; contaminação; segurança alimentar; tubérculos.*

Sessão: Estatística Aplicada em Ciências Médicas, Saúde e Meio Ambiente

Título: Modelando dados de contagem sobredisperso

Autores: Emerson Barili; Juliana V.C.M. Perles; Omar C. N. Pereira; Isolde Previdelli; Rosângela G. Santana

Resumo: O objetivo do presente trabalho é investigar os modelos de regressão Poisson, quasi-Poisson e binomial negativo da classe da família exponencial, utilizando dados de contagem sobredispersos da doença Diabetes Mellitus. O artigo apresentado por Hoef e Boveng (2007) sobre modelos de regressão de quasi-verossimilhança norteia a metodologia aqui adotada para responder algumas hipóteses de um experimento conduzido pela Professora Juliana do Departamento de Farmacologia da UEM. Cada vez mais pesquisas na área da saúde têm sido realizadas, no intuito de encontrar mecanismos de prevenção e diagnósticos de doenças crônicas como é o caso da Diabetes Mellitus (DM).

Palavras-Chave: *Poisson; Dispersão; Neurônios.*

Título: Modelo Espaço-Temporal para a Análise da Abundância dos Vetores da Dengue

Autores: Ana Carolina Carioca da Costa; Aline Araújo Nobre; Cláudia Torres Codeço; Nildimar Alves Honório

Resumo: Atualmente, o controle da dengue está centrado na redução da densidade vetorial, uma vez que o mosquito é o único elo vulnerável da cadeia de transmissão. O uso de novas abordagens na vigilância entomológica da dengue é de extrema relevância, tendo em vista a ineficiência dos atuais métodos de vigilância. Sob essa perspectiva, o presente estudo visa analisar, do ponto de vista bayesiano, a distribuição espaço-temporal dos índices de infestação obtidos a partir de armadilhas de oviposição e avaliar a relação entre a abundância de ovos e fatores climáticos e ambientais. A área de estudo abrangeu o campus Manguinhos da Fundação Oswaldo Cruz, localizado no Rio de Janeiro. Na região de estudo foram definidas 8 áreas sentinelas para o monitoramento contínuo de formas imaturas de *Aedes aegypti* e *Aedes albopictus*. Em cada área sentinela foram implantadas 30 armadilhas de oviposição. As coletas foram realizadas semanalmente no período de novembro de 2010 a agosto de 2012, totalizando 89 semanas. A relação entre a abundância de ovos e fatores climáticos e ambientais foi avaliada através de um modelo espaço-temporal inflacionado de zeros. A inferência paramétrica foi realizada de forma aproximada através do algoritmo INLA (Integrated Nested Laplace Approximations). Os índices de infestação indicaram a presença de oviposições ao longo de todo o período de estudo. A temperatura mínima se mostrou significativa para explicar a abundância de ovos em apenas uma área. A distância entre as armadilhas e o limite mais próximo do campus evidenciou que regiões densamente povoadas ou localizadas na proximidade de vias com tráfego intenso apresentam maior abundância de ovos. Os dados foram caracterizados por uma fraca dependência temporal e uma correlação espacial moderada. Foram detectados padrões espaciais consistentes com a densidade populacional humana em todas as áreas sentinelas. Os resultados encontrados sugerem que seria necessário conduzir a vigilância na área de estudo a partir de vistorias semanais em armadilhas

alocadas, no máximo, a cada 20 metros. Embora o monitoramento realizado através de armadilhas de oviposição seja insuficiente para eliminar o principal vetor, *Aedes aegypti*, pode reduzir os índices de infestação. A utilização do método INLA permite que a inferência seja realizada eficientemente do ponto de vista computacional e possibilita a incorporação do modelo proposto em sistemas de vigilância e monitoramento.

Palavras-Chave: *Modelagem espaço-temporal; INLA; Dengue; Vigilância entomológica; Armadilhas de oviposição.*

Título: Modelo unimodal simétrico-asimétrico con aplicación al estudio del VIH

Autores: Guillermo Martínez-Florez; Germán Moreno-Arenas; Solange Lugo-Buitrago

Resumo: Se definen dos nuevas distribuciones de probabilidad: modelo unimodal simétrico con función de riesgo proporcional a la distribución normal y modelo unimodal asimétrico con función de riesgo proporcional a la distribución normal asimétrica. Estos modelos permiten ajustar datos censurados con comportamiento bimodal y altos (o bajos) niveles de curtosis comparado con la curtosis de la distribución normal y altos (o bajos) niveles de asimetría. Además, se estiman los parámetros de los modelos por máxima verosimilitud y se determina la matriz de información observada. La flexibilidad de la nueva distribución se ilustra ajustando un conjunto de datos reales: el número de moléculas de ARN VIH-1 por mililitros de sangre medida en personas con pruebas confirmadas de presencia del VIH.

Palavras-Chave: *Bimodalidad; Límite de detección; Función de riesgo proporcional; ARN VIH-1; HAART.*

Título: O uso da metodologia de dados faltantes em séries temporais com aplicação a dados de concentração de (pm10) observados na região da grande Vitória

Autores: Wanderson de Paula Pinto; Valdério Anselmo Reisen; Taciana Toledo de Almeida Albuquerque

Resumo: Dados da poluição atmosférica apresentam, em geral, observações faltantes. Esta pesquisa apresenta um estudo de metodologias para estimação da função de autocorrelação na presença de dados faltantes, baseados no trabalho de Yajima e Nishino (1999). Contempla também algumas técnicas para imputação de dados faltantes baseadas no uso do algoritmo EM, proposto por Dempster (1977), e nos modelos de séries temporais ARIMA de Box e Jenkins. Ensaio de simulações com quadro proporções de dados faltantes foram realizadas para comparar os erros quadráticos médios dos estimadores propostos. O estudo empírico evidenciou que o método de estimação sugerido apresenta bom desempenho em termos de medidas de erro quadrático médio. Como ilustração da metodologia proposta, duas séries temporais de concentrações de Material Particulado Inalável (PM10) emitida na Região da Grande Vitória, E.S., Brasil, são analisadas.

Palavras-Chave: *Material Particulado inalável; Função de autocorrelação; Dados faltantes*

Título: Obtaining adjusted prevalence ratios from logistic regression model in cross-sectional studies

Autores: Leonardo Soares Bastos, Raquel de Vasconcellos Carvalhaes de Oliveira, Luciane Velasque

Resumo: In the last decades, it has been discussing the use of epidemiological prevalence ratio (PR) rather than odds ratio as a measure of association to be estimated in cross-sectional studies. It discusses the main difficulties in use of statistical models for the calculation of PR: convergence problems, availability of adequate tools and assumptions not appropriate. The goal of this study is to develop a function to calculate the true PR and its confidence interval in statistical software R. We used 3 examples and compared the crude and adjusted estimative of PR, with the estimates obtained by use of log- binomial, Poisson regression and the prevalence odds ratio (POR). The PR obtained from the function resulted in values close enough to those obtained by log- binomial and Poisson, while the POR overestimated the PR. The model here implemented showed the following advantages: no numerical instability; assumes probability distribution appropriate and, it is availability in a free R statistical package.

Palavras-Chave: *GLM; Binomial; Poisson; Log-Binomial; Prevalence.*

Sessão: Dados Funcionais, Dados em Alta Dimensão e Aprendizado Estatístico de Máquinas

Título: Análise multivariada de dados Poisson composto longitudinais

Autores: Rubem Kaipper Ceratti; Afrânio Márcio Corrêa Vieira

Resumo: On studies about the blend of volatile organic compounds, released due to the damage caused by herbivorous insects, usually produce a number of mass concentration of these compounds (say 30, 50 or higher), correlated each other, what partially describes the metabolic route stimulated by the damage stimulus. These measures are asymmetric with a probability to occur zero values. Furthermore, the stimuli are experiment-based and longitudinal measures are taken, trying to identify significant factors and/or time point that are scientifically relevant. In this work will be presented a proposal of multivariate data analysis, based on generalized linear mixed models where the multivariate response will be assumed as Tweedie's compound Poisson distribution. The modeling strategy adopted was all-pairwise mixed models as well a pseudo-likelihood approach to reconstruct the covariance response matrix, already adjusted for non normality, experimental design structure and longitudinal correlations. It means that traditional multivariate analysis as PCA, factor and cluster analysis can be produced, bringing light to a complex data structure where, usually, univariate analyses are made. The package `pair.mglmm` for R computing environment was produced based on `cplm` and `doParallel` packages. These packages allow fitting compound Poisson mixed models using parallel computing on multicore processors. Simulations were done to evaluate the computational effort and it can be concluded that even for higher number of multivariate responses, the modeling is feasible, where the full multivariate approach isn't possible. These models were applied on a cotton plant study, where 25 chemical compound mass were measured on 4 time points, under a completely randomized experimental design with 5 treatments. The results were promising in the biological control research area as well as methodological approach for biostatisticians.

Palavras-Chave: *Modelos Lineares Generalizados Mistos; Família Tweedie de Distribuições; Análise Multivariada; Pseudo verossimilhança; Controle biológico*

Título: Nonparametric Conditional Density Estimation in a High-Dimensional Regression Setting

Autores: Rafael Izbicki; Ann B. Lee

Resumo: There has been an increased interest in estimating the full density $f(z|x)$ of a random variable Z given a high-dimensional random vector x ; in, e.g., cosmology and economics, the regression $E[Z|x]$ is sometimes not informative enough because of multimodality and asymmetry of $f(z|x)$. Currently, there are no effective ways of estimating $f(z|x)$ when x are complex, high-dimensional objects such as images or spectra. We propose a novel nonparametric estimator of $f(z|x)$ that takes advantage of lower-dimensional structure (e.g., submanifolds and cluster structure) in x . Our estimator is based on an expansion onto the eigenfunctions of a kernel-based operator, which are smooth and adapted to the underlying geometry of the data. These eigenfunctions are also orthogonal, and hence allow a fast implementation and tuning of parameters. Additionally, they provide

a useful tool for data visualization. We derive rates of convergence and demonstrate the effectiveness of the method by numerical examples, including an application to photometric redshift prediction of galaxies.
Palavras-Chave: *High-Dimensional Data; Nonparametric Statistics; Spectral Methods*

Título: Statistical analysis of NIR spectroscopy data

Autores: Ronaldo Dias; Nancy L. Garcia;Guilherme Ludwig; Marley A. Saraiva

Resumo: Calibration and multivariate, simultaneous prediction for NIR spectroscopy data are performed based on a functional interpretation of the Beer-Lambert formula. For each chemical sample, the resulting spectrum is the summation of overlapped absorption spectra from each analyte plus a Gaussian error resulting in a continuous curve. We assume that each individual spectrum can be expanded as a linear combination of B-splines basis. Departures from the standard error models are discussed through simulations. In particular, how correlated errors impact on the calibration step and consequently on the analytes' concentration prediction, and how the estimates can be improved with a proper covariance estimation step. A jackknife method for improved inference on prediction values is developed. Finally, two publicly available datasets are analyzed.

Palavras-Chave: *B-splines; leave-one-out jackknife; square error of prediction.*

Sessão: Análise de Sobrevivência

Título: A COM-Poisson type generalization of the binomial distribution and its properties and applications

Autores: Patrick Borges; Josemar Rodrigues; Narayanaswamy Balakrishnan; Jorge Bazán

Resumo: Shmueli et al. (2005) introduced a new discrete distribution, called COM-Poisson-binomial distribution, by adding a dispersion parameter to the binomial distribution. However, they did not study the mathematical properties of this family of distributions. In this paper, we investigate in detail this probability distribution; we discuss its moments, probability and moment generating functions and how the dispersion parameter affects the asymptotic approximation of the COM-Poisson-binomial distribution by the COM-Poisson distribution. The fitting of this model with three data sets is also considered for illustrative purposes.

Palavras-Chave: *COM Poisson distribution; Weighted Poisson distribution; Exponential family.*

Título: A New Extended Normal Model

Autores: Gauss Moutinho Cordeiro; Maria do Carmo Soares de Lima; Abraão David Costa do Nascimento

Resumo: Various applications in natural science require models more accurate than well-known distributions. In this sense, several generators of new distributions were proposed recently. In this paper, we introduce a new four-parameter extended normal distribution and discuss some of its mathematical properties. The maximum likelihood method is used to estimate the model parameters. Two applications to real data are performed and the results provide evidence that the proposed model outperforms the classical skew normal distribution.

Palavras-Chave: *Exponentiated distribution; Maximum likelihood method; Normal distribution; Quantile function.*

Título: A distribuição Harris extended exponential

Autores: Luis Gustavo Bastos Pinho; Gauss Moutinho Cordeiro; Juvêncio Santos Nobre

Resumo: Uma nova distribuição de tempo de sobrevivência é estudada nesse trabalho. A "Harris extended exponential" é obtida a partir da distribuição Harris, usada como gerador, e da distribuição exponencial, usada como distribuição base. Diversas propriedades da nova distribuição foram obtidas e relatadas. A nova família acomoda falhas crescentes ou decrescentes. Uma aplicação interessante dessa nova família, que é consideravelmente simples, é na área de modelagem de imagens SAR. Comparada a distribuições largamente utilizadas, a nova família obteve um bom desempenho e possui propriedades matemáticas mais atraentes que as distribuições usuais.

Palavras-Chave: *Imagens SAR; Marshall-Olkin; Análise de sobrevivência; Geradores de densidades*

Título: Distribuição Kumaraswamy-Exponencial para Análise de Dados de Sobrevivência

Autores: Amanda Morales Eudes; Vera Lucia Damasceno Tomazella, Cirdêmia Costa Feitosa

Resumo: Na literatura, diversas distribuições conhecidas são utilizadas para acomodar dados de tempos de falha, porém, grande parte destas distribuições não é capaz de acomodar taxas de falha não monótonas. Kumaraswamy (1980) propôs uma nova distribuição de probabilidade capaz de acomodar tais taxas e, baseada nela, mais recentemente Cordeiro e de Castro (2009) propuseram uma nova família de distribuições generalizadas, a Kumaraswamy-Generalizada (Kum-G). Esta distribuição além de ser flexível, contém distribuições com funções de risco unimodal e em forma de banheira, como mostrado por Pascoal et al. (2011). Neste artigo, nós apresentamos a distribuição Kumaraswamy-Exponencial (Kum-Exp) para analisar dados de tempo de vida dos indivíduos em risco, sendo que este modelo é caso particular da família de distribuições Kum-G. Algumas propriedades desta distribuição serão apresentadas, assim como o método adequado de estimação para os parâmetros do modelo, de forma clássica e também bayesiana. A nova distribuição é ilustrada com dois conjuntos de dados encontrados na literatura.

Palavras-Chave: *Análise de sobrevivência; Distribuição Kumaraswamy-Exponencial; abordagem bayesiana.*

Título: Distribuição Weibull Exponencial de Zografos

Autores: Patrícia Silva Nascimento Barros, Kleber Napoleão Nunes de O. Barros, Cícero Carlos R. de Brito, Joseilme Fernandes Gouveia, Wilson Rosa de Oliveira Junior

Resumo: A distribuição Weibull é muito popular para dados em modelagem de vida. Muitas modificações da distribuição Weibull foram propostas nos últimos anos. Zografos apresenta uma introdução a família gama-generalizada de distribuições. Daí tem-se, como proposta, a distribuição Weibull Exponencial de Zografos (WEZ). A WEZ tem a distribuição Weibull, exponencial e Rayleigh como casos particulares. Um tratamento compreensivo das propriedades matemáticas da WEZ é feita, sendo encontradas as expressões para a função de distribuição, função densidade, função de risco, função geradora de momentos, média, variância, coeficiente de assimetria e curtose. Ajustou-se um modelo WEZ para um conjunto de dados reais e comparou-se com os modelos Weibull (W), Fréchet (F), Beta Fréchet (BF), Gama (G) e Beta Exponencial Generalizada (BGE). Para todos os critérios o melhor ajuste foi da WEZ. Dessa forma a distribuição Weibull Exponencial de Zografos pode ser utilizada para ajustar dados de análise de sobrevivência.

Palavras-Chave: *Distribuição Weibull ; Família gama-generalizada; Distribuição Weibull Exponencial de Zografos.*

Título: Exploring ARA and ARI Classes of Imperfect Repair Models

Autores: Enrico A. Colosimo; Maria Luíza G. de Toledo; Marta A. Freitas; Gustavo L. Gilardoni

Resumo: An appropriate maintenance policy is essential to reduce expenses and risks related to equipment failures. A fundamental aspect to be considered when specifying such policies is to understand the behavior of the failure intensity for the systems under study. The usual assumptions of minimal or perfect repair at failures are not adequate for many real world systems, requiring the application of imperfect repair (IR) models. In this paper, the classes ARA and ARI of IR models proposed by Doyen and Gaudoin (2004) are explored. Likelihood functions for such models are derived, assuming Power Law Process and a general memory order. Based on this, punctual and interval parametric estimates were obtained for a real dataset involving failures in trucks used by a mining company for models with different memories. BIC value was used as the criteria for model selection, namely, shape and scale parameters for PLP. They provided evidences that the trucks tend to fail more frequently over time, justifying the necessity for preventive maintenance, and also, that the repairs after failures tend to leave the equipment in a state between as good as new and as bad as old. These results are a valuable information for the mining company, and can be used to support decision making regarding preventive maintenance policy.

Palavras-Chave: *repairable systems; Poisson process; BIC.*

Título: Modelagem Conjunta de Dados Longitudinais e Dados de Sobrevida de Pacientes com Insuficiência Cardíaca Congestiva

Autores: Diana C. Franco-Soto; Antonio C. Pedroso-de-Lima; Julio M. Singer

Resumo: Consideramos a modelagem conjunta de medidas longitudinais e tempos de sobrevivência, motivados por um estudo realizado no Instituto do Coração (InCor), São Paulo, Brasil, cujo objetivo é avaliar o impacto do Peptídeo Natriurético Tipo-B (BNP) coletado em diferentes instantes sobre a sobrevida de pacientes com Insuficiência Cardíaca Congestiva (ICC). Empregamos um modelo linear de efeitos mistos para explicar a resposta longitudinal e um modelo Birnbaum-Saunders para descrever os tempos de sobrevivência, estendendo algumas das abordagens existentes para acomodar indivíduos sem observações longitudinais. Derivamos os estimadores de máxima verossimilhança dos parâmetros do modelo e testes de hipóteses sobre tais parâmetros são obtidos por meio de resultados de inferência assintótica. O modelo conjunto proposto é aplicado ao conjunto de 1609 pacientes com ICC, dos quais 1080 não têm nenhuma medição de BNP. Os resultados sugerem que a inclusão de medições longitudinais de uma resposta apropriada pode ser utilizada para melhorar a análise dos dados de sobrevivência.

Palavras-Chave: *Dados longitudinais; Dados de sobrevivência; Modelagem conjunta; Modelo Birnbaum-Saunders; Modelo linear de efeitos mistos.*

Título: Modelagem paramétrica na de censura informativa: Um estudo de simulação

Autores: Paulo Cerqueira dos Santos Junior, Enrico Antônio Colosimo, Fábio Nogueira Demarqui

Resumo: O mecanismo de censura é um ponto crucial nas análises de dados de sobrevivência (ver Leung (1997)). A maioria dos métodos em análise de sobrevivência supõe o mecanismo de censura não-informativo, ou seja, os tempos até o evento de interesse e o de censura são independentes (a distribuição dos tempos de censura não carregam informação dos parâmetros da distribuição dos tempos de falha). Entretanto em alguns casos essa suposição é violada, métodos que incorporem estruturas para situações em que o mecanismo de censura é informativo são necessários. Em situações em que pacientes saem do estudo por razões relacionadas ao tipo de terapia e consequentemente tem o fim do seu acompanhamento, pode-se considerar que a censura é informativa, pois a razão da censura está relacionada com o tempo de sobrevivência futuro. Outros exemplos são situações em que a falha é causada por eventos secundários e assim registradas como observações censuradas (Lagakos (1979)). Basicamente, a censura informativa ocorre quando observações censuradas são mais (ou menos) propensas a experimentar um evento específico das observações restantes no futuro (Collett (2003)). Propostas em utilizar modelos de fragilidade sob a suposição de censura informativa também foram consideradas. No trabalho de Huang \emph{et. al} (2002) foi desenvolvido outro modelo de fragilidade para dados em conglomerados, que permite uma maior flexibilidade e grau de dependência entre a falha e censura, e mais tarde em Huang et. al (2004), a proposta de um teste para verificar a presença da censura informativa. Nesse trabalho, será assumido que os tempos de falha T e censura C são condicionalmente independentes dado uma fragilidade Z e as covariáveis. A fragilidade será incluída na modelagem para captar a dependência entre os tempos de falha e de censura associados a cada indivíduo, e dessa forma, incorporar o mecanismo de censura informativo na modelagem. Assumiu-se a distribuição da fragilidade como sendo Gama e realizou-se um estudo de simulação sob uma abordagem completamente paramétrica, com o uso do modelo Weibull. A implementação computacional foi realizada no software OxMetrics 6.

Palavras-Chave: *Fragilidade; Censura informativa; Verossimilhança Marginal.*

Título: Testing inference in accelerated failure time models

Autores: Francisco M. C. Medeiros; Antônio H. M. da Silva-Júnior; Dione M. Valença; Silvia L. P. Ferrari

Resumo: We address the issue of performing hypothesis testing in accelerated failure time models for noncensored and censored samples. The performances of the likelihood ratio test and a recently proposed test, the gradient test, are compared through simulation. The gradient test features the same asymptotic properties as the classical large sample tests, namely, the likelihood ratio, Wald and score tests. Additionally, it is as simple to compute as the likelihood ratio test. Unlike the score and Wald tests, the gradient test does require the computation of the information matrix, neither observed nor expected. Our study suggests that the gradient test is more reliable than the other classical tests when the sample is of small or moderate size.

Palavras-Chave: *Accelerated failure time models; Gradient test; Likelihood ratio test; Random censoring; Simulation*

Título: The Transmuted Log-Logistic Distribution: Modeling, Inference and an Application to a Polled Tabapua Race Time up to First Calving Data

Autores: Daniele Cristina Tita Granzotto; Francisco Louzada

Resumo: In this study, we propose a new lifetime distribution by using a quadratic rank transmutation map in order to add a new parameter to the log-logistic distribution. We provide a comprehensive description of the properties of the proposed distribution along with its reliability study. The usefulness of the transmuted log-logistic distribution for modeling reliability data is illustrated on a polled Tabapua race time up to first calving data.

Palavras-Chave: *Log-logistic distribution; Transmuted map; Reliability analysis.*

Título: The censored bimodal skew-normal model with an application to plasma HIV-1 RNA

Autores: Guillermo Martínez-Flórez; Germán Moreno-Arenas; Solange Lugo-Buitrago

Resumo: We introduce the censored bimodal skew-normal model for adjust censored data with bimodality and high levels of skewness and kurtosis. Various structural properties of the new distribution are derived, including its moments. The model parameters are estimated by maximum likelihood and the observed information matrix is derived. An application of the new model to analysis of the plasma HIV-1 RNA measurement is presented for illustrative purpose.

Palavras-Chave: *Censored; skew-normal; limit of detection; HIV-1 RNA; HAART.*

Sessão: Educação Estatística

Título: Desafios do ensino de Estatística na Licenciatura em Matemática

Autores: Marcos Nascimento Magalhães

Resumo: O ensino de Estatística na Educação Básica faz parte da disciplina de Matemática, conforme está previsto nos Parâmetros Curriculares Nacionais do MEC. Uma boa formação em Estatística de futuros professores é uma ação estratégica importante para a ampliação da presença da Estatística na sociedade. Um desafio a ser enfrentado é a melhoria da qualidade do ensino de Estatística oferecido nas faculdades e universidades que são as formadoras desses professores. Nesse sentido, apresentamos neste artigo algumas atividades desenvolvidas na disciplina Estatística para Licenciatura I, do curso de Licenciatura do Instituto de Matemática e Estatística da Universidade de São Paulo. Discutimos, também, três dificuldades relacionadas com o conhecimento de distribuições: reconhecer a existência de eventos não equiprováveis, compreender a existência da aleatoriedade em dados coletados e distinguir entre distribuições empírica e teórica de variáveis. As atividades desenvolvidas em 2013 parecem ter contribuído para uma melhor aprendizagem dos estudantes.

Palavras-Chave: *licenciatura em matemática; estatística básica; ensino de estatística.*

Título: Ensinar e aprender estatística: uma proposta pedagógica

Autores: Suzi Samá Pinto; Mauren Porciúncula Moreira da Silva

Resumo: Este trabalho tem como objetivo refletir sobre o ensino da estatística através da proposta pedagógica de Projetos de Aprendizagem. A avaliação do desenvolvimento das diversas etapas dos Projetos de Aprendizagem foi fundamental para compreender sua contribuição na construção dos conceitos estatísticos pelos estudantes de graduação. A análise dos dados levantados no instrumento de avaliação envolveu métodos quantitativos e qualitativos. O Ensino da Estatística através dos Projetos de Aprendizagem possibilitou aos estudantes vivenciarem as várias fases de uma pesquisa, partindo de um tema de seu interesse e aplicando os conteúdos que iam sendo trabalhados na sala de aula de Estatística. Também possibilitou conviver e lidar com as inseguranças inerentes ao próprio pesquisar, e contribuir na formação de indivíduos com iniciativa, autonomia, conscientes dos problemas da atualidade e da importância do trabalho em equipe.

Palavras-Chave: *Ensino de Estatística; Projeto de Aprendizagem; Ensino de Graduação; Análise qualitativa e quantitativa.*

Título: Pessoas com deficiência: o que encontramos atrás da inclusão

Autores: Paulo Tadeu Meira e Silva de Oliveira

Resumo: É considerado, atualmente, como um facto ao longo da história, que pessoas com deficiência sempre existiram (Silva, 1986). Paulatinamente, a sociedade tem percebido que, além de caridade e assistência, tais pessoas devem ser incluídas em programas e políticas públicas que poderiam melhorar o seu potencial produtivo (Garcia, 2010). Desde 1970, em todo o mundo, e desde 1990, no Brasil, tem sido reconhecida a importância do desenvolvimento de raciocínio probabilístico, ou seja, a necessidade de

romper com a cultura determinista, através da inserção de noções de Probabilidade e Estatística no conteúdo de matemática no ensino básico e nos níveis secundários. Além disso, estatísticas é utilizado atualmente em diversas áreas do conhecimento, inclusive em estudos quantitativos sobre educação e deficiência, dando apoio a um melhor acesso à educação, trabalho, saúde e outros direitos humanos, necessário para todas as pessoas, quer tenham ou não deficiência. No século XX, a educação foi assumida como uma responsabilidade social, tornando-se um direito de todos os indivíduos e adicionando novas preocupações, constituindo uma política orientada sob novos paradigmas, a chamada educação inclusiva. Esta tem sido considerada, no Brasil, desde a última década do século XX, e hoje, tornou-se um assunto muitas vezes discutido em vários segmentos da sociedade (Santos, 2010). Em termos estatísticos, o IBGE (Instituto Brasileiro de Geografia e Estatística) em 2010, estima que existiam no Brasil 45.606.048 pessoas com pelo menos uma deficiência, sendo 41.830.812 com idade acima de 15 anos, de acordo com a RAIS (Relação Anual de Informações Sociais) do Ministério do Trabalho, nesse mesmo ano, existiam 44.068.535 trabalhadores registrados, entre os quais 306.048 eram pessoas com deficiência com mais de 15 anos de idade, distribuídos por sexo (masculino e feminino); nível de instrução (fundamental incompleto, fundamental completo entre e secundário incompleto, entre colégio secundário e incompleto e completo grau completo ou mais), e com diferentes tipos de deficiência (física, para ouvir, ver, intelectual, múltipla e reabilitado (somente RAIS)). Com o intuito de corrigir essa distorção, o governo brasileiro, via Congresso Nacional, aprovou em 1993 a Lei de Cotas, uma ação afirmativa exigindo de todas as empresas públicas e privadas contratassem obrigatoriamente pessoas com deficiência em um percentual que varia entre 2-5% do número de funcionários registrados na empresa. Esta lei, se todas as empresas respeitassem em 2010, deveriam haver cerca de 950 mil pessoas com deficiência de acordo com a RAIS, sob a Lei de Cotas. Mesmo assim, este valor representaria cerca de 3% das pessoas com deficiência com carteira assinada, de acordo com os dados do IBGE. A fim de oferecer suporte técnico para que possa melhor incluir essas pessoas, é necessário analisar diversas situações, tais como, possíveis defasagens existentes entre pessoas com deficiência e que não têm deficiência, examinando os resultados obtidos nos dados do Censo do IBGE e da RAIS, a partir de 2010. Então, é necessário ter em conta as diferenças importantes, nomeadamente o grupo de pessoas com deficiência completa, em comparação com aqueles com deficiência severa e, também, outras variáveis tais como fundo de educação e renda, no Brasil (Bussab e Morettin, 2013). É importante destacar que, nos últimos anos, muito tem sido obtido através da mobilização de grupos de pessoas envolvidas neste segmento, direitos para as pessoas com deficiência, tais como educação, emprego, transporte, recreação, saúde e outros. Porém, ainda existe um longo caminho a percorrer antes que as pessoas com deficiência possam ser consideradas, de fato, incluído na sociedade. Neste trabalho, pretendemos mostrar situações, que mesmo depois de várias campanhas, diversos problemas que não deixam claros para que essas pessoas sejam melhores incluídas, como verificado a partir de dados da RAIS e do IBGE.

Palavras-Chave: *Pessoas com deficiência; dados IBGE; dados RAIS; Lei de Cotas; Educação Estatística.*

Sessão: Estatística Aplicada em Ciências Sociais Aplicadas e Demografia e Estatísticas Públicas

Título: Estimando a previsibilidade das decisões de política monetária no Brasil: uma análise através de estudo de eventos

Autores: Caio Gonçalves

Resumo: O objetivo deste artigo é avaliar a previsibilidade das decisões de política monetária no Brasil através das alterações da taxa de juros básica (Selic) e seu impacto sobre a estrutura a termo da taxa de juros. Trabalhou-se com a ideia de através das ações de transparência realizadas pelo Banco Central do Brasil, este envia um sinal aos agentes sobre o seu tipo. Caso os agentes consigam ajustar suas expectativas sem “surpresas”, a sinalização emitida está sendo crível, o problema de assimetria de informação minimizado e a postura do banco central é não oportunista. Utilizando a metodologia de estudo de eventos com dados diários de setembro de 2003 a março de 2014 dividido em dois subperíodos (período I – Meirelles e período II – Tombini), este trabalho encontrou indícios de que sob a presidência de Meirelles a previsibilidade das decisões de política monetária é maior comparado à presidência de Tombini.

Palavras-Chave: *Política Monetária; Sinalização; Estudo de Eventos.*

Título: Explorando alternativas para a calibração dos pesos amostrais da Pesquisa Nacional por Amostra de Domicílios

Autores: Charles Martín Miguel Ruiz; Pedro Luis do Nascimento Silva

Resumo: A Pesquisa Nacional por Amostra de Domicílios (PNAD) é executada pelo Instituto Brasileiro de Geografia e Estatística (IBGE) com o propósito de conhecer as principais características demográficas e socioeconômicas da população brasileira. Os pesos básicos da PNAD são calibrados mediante o método da pós-estratificação, usando como informações auxiliares as projeções da população para cada Unidade da Federação segundo tipo de área (região metropolitana e não metropolitana). Neste artigo explora-se outro método de calibração conhecido como raking que leva em consideração as projeções populacionais utilizadas na pós-estratificação e adicionalmente as projeções por sexo e grupos de idade para o total do país. A calibração por pós-estratificação produz estimativas das frequências relativas por grupos de idade que, ao ser comparadas com as projeções de população, mostram a existência de sub-cobertura das crianças de 0 a 9 anos e de jovens de 20 a 29 anos, e sobre-cobertura dos adultos de 50 anos ou mais de idade. Os problemas de cobertura mencionados foram corrigidos com o método de calibração alternativo baseado no raking. Esse método produz estimativas das frequências relativas por sexo e grupos de idade que coincidem com as respectivas projeções da população. A comparação entre as estimativas dos indicadores obtidas pela calibração por pós-estratificação e raking revelaram que existem diferenças entre as estimativas da taxa de analfabetismo produzidas por ambos os métodos. No caso dos outros indicadores analisados, não há diferenças importantes. Não entanto, a comparação entre os coeficientes de variação das estimativas mostrou que, em geral, eles são menores quando se aplica a calibração por raking, o qual significa um ganho de precisão em relação à pós-estratificação.

Palavras-Chave: *Calibração; Raking; Pós-estratificação; Pesos amostrais.*

Título: Investigação de retornos de escala e análise de eficiência relativa dos terminais de contêineres no Brasil

Autores: Paulo César Coutinho, André Luis Rossi de Oliveira, Alexandre Xavier Ywata de Carvalho, Paulo Augusto Pettenuzzo de Britto, Reinaldo Soares de Camargo

Resumo: A discussão sobre eficiência portuária tem ganhado destaque no Brasil e no mundo, tanto no âmbito acadêmico quanto no de decisões estratégicas dos setores público e privado. Nesse sentido, é primordial promover o uso racional das infraestruturas portuárias e identificar ações específicas para terminais sabidamente ineficientes. Neste artigo estabelecemos um referencial nacional para o desempenho de terminais de contêineres que possa ser utilizado para identificar aqueles com maior potencial de melhoria de desempenho. Com base em dados especificamente coletados para esse estudo junto aos terminais de contêineres brasileiros, com a intermediação da Agência Nacional de Transportes Aquaviários (ANTAQ), realizamos dois estudos sobre a eficiência dos terminais de contêineres: A existência ou não de economias de escala e a análise de eficiência relativa dos terminais nacionais.

Palavras-Chave: *DEA; Retornos de Escala; Eficiência Portuária.*

Título: Imputation and Estimation under Nonignorable Nonresponse for Household Surveys with Missing Covariate Information

Autores: Danny Pfeffermann; Anna Sikov

Resumo: In this paper we develop and apply new methods for handling not missing at random (NMAR) nonresponse. We assume a model for the outcome variable under complete response and a model for the response probability, which is allowed to depend on the outcome and auxiliary variables. The two models define the model holding for the outcomes observed for the responding units, which can be tested. Our methods utilize information on the population totals of some or all of the auxiliary variables in the two models, but we do not require that the auxiliary variables are observed for the nonresponding units. We develop an algorithm for estimating the parameters governing the two models and show how to estimate the distributions of the missing covariates and outcomes, which are then used for imputing the missing values for the nonresponding units and for estimating population means and the variances of the estimators. We also consider several test statistics for testing the model fitted to the observed data and study their performance, thus validating the proposed procedure. The new developments are illustrated using simulated data and a real data set collected as part of the Household Expenditure Survey carried out by the Israel Central Bureau of Statistics in 2005.

Palavras-Chave: *Respondents distribution; Horvitz- Thompson type estimator; Calibration; Bootstrap.*

Título: Modelo de Regressão Beta: abordagens paramétrica e semiparamétrica

Autores: Alisson de Oliveira Silva; Jodavid de Araújo Ferreira; Camila Ribeiro da Silva

Resumo: Com a redução da mortalidade infantil pós-neonatal, principalmente por intervenções nos seus principais fatores a exemplo das doenças diarreicas, a mortalidade infantil neonatal precoce que compreende as mortes no período de 0 a 7 dias de vida,

passou a ser o principal componente responsável por mais da metade dos óbitos no primeiro ano de vida. No Brasil nota-se uma grande disparidade das desigualdades sociais, sendo um fator decisivo para que a mortalidade infantil neonatal precoce ocorra de maneira diferenciada nas regiões do país, concentrando as maiores taxas no eixo Norte-Nordeste e em menores proporções no Centro-Sul. Com base nisso, realizou-se uma comparação dos modelos de regressão beta e beta semiparamétrico com base na estrutura dos modelos aditivos generalizados para posição, escala e forma (GAMLSS) a fim de identificar os principais fatores de risco dessas taxas. Com base nos critérios AIC, BIC, GD e pseudo-R² constatou-se a superioridade do modelo beta semiparamétrico.

Palavras-Chave: *Mortalidade infantil; Mortalidade infantil neonatal precoce; Regressão beta; GAMLSS.*

Título: Métodos de Distribuição Livre para o Ajuste de Modelos de Estrutura de Covariância para Dados Amostrais Longitudinais

Autores: Leandro Vitral Andraos ; Marcel de Toledo Vieira

Resumo: Neste trabalho são estudados métodos estatísticos com o apoio de ferramentas computacionais para o estudo de dados do tipo painel da pesquisa British Household Panel Survey (BHPS). Procedimentos de inferência estatística para parâmetros de modelos de efeitos mistos são avaliados considerando diferentes estruturas longitudinais de covariância. Comparações entre métodos de estimação que consideram o plano amostral complexo e aqueles que o ignoram são realizadas através de um estudo de simulação e o comportamento de estimadores que tem como base funções de ajuste é avaliado. São considerados os seguintes métodos de estimação pontual: máxima verossimilhança, máxima pseudo verossimilhança e mínimos quadrados generalizados de distribuição livre. O desempenho dos estimadores também é avaliado considerando cenários em que os dados são gerados a partir de diferentes distribuições de probabilidade, incluindo tanto distribuições simétricas quanto assimétricas.

Palavras-Chave: *dados do tipo painel; modelos de efeitos mistos; estruturas de covariância; funções de ajuste; inferência para dados amostrais complexos.*

Título: Poder de controle, benefícios privados e normas de proteção ao investidor: uma análise de companhias brasileiras

Autores: Felipe dos Santos Ronco; Wagner Oliveira Monteiro

Resumo: O presente estudo tem por objetivo mensurar o diferencial de preços das ações com e sem direito a voto de emissão de 31 companhias brasileiras, no intervalo de janeiro de 2006 a dezembro de 2011, e identificar o impacto de normas recentes de proteção ao investidor na alteração do referido diferencial de preços, por meio de estimações econométricas, levando em consideração igualmente outras variáveis significativas, de modo a estabelecer, diante dos resultados apresentados, uma relação entre tais normas protetivas e a extração de benefícios privados do poder de controle.

Palavras-Chave: *Poder de Controle; Dados em Painel; Efeitos Fixos.*

Título: Volatilidade no mercado italiano: um estudo comparativo utilizando modelos da família GARCH

Autores: Uriel Moreira Silva; Frank Magalhães de Pinho

Resumo: In this work, a comparison of the performances of several GARCH models for modelling log-return volatility is made for the assets of the 5 most representative companies (as of 30/09/2013) in the benchmark stock market index for the Borsa Italiana, the FTSE-MIB. These companies roughly represent 50.60\% of the index, and are: Ente Nazionale Idrocarburi, UniCredit, Assicurazioni Generali, Ente Nazionale per l'energia ELettrica and Intesa Sanpaolo. The sample used is composed of closing prices for the assets of each company, with data ranging from 1/1/2003 up to 12/30/2011. Model performances are compared in-sample through information criteria, such as AIC, BIC, SIC and HQC, obtained under the hypothesis of gaussian, t-Student, skew-t and GED distributions. The obtained results show that the models with the best in-sample performance belong to the FIAPARCH class, allowing for the presence of long-memory and asymmetry on log-return volatility.

Palavras-Chave: *Volatilidade; Modelos GARCH; FTSE-MIB.*

Sessão: Estatística Aplicada em Engenharia e Ciências Exatas

Título: Aproximações para DCT Baseadas no Algoritmo de Loeffler

Autores: D. F. G. Coelho; R. J. Cintra; F. M. Bayer; S. Kulasekera; A. Madanayake

Resumo: Uma nova classe de aproximações para a transformada discreta do cosseno (DCT) de comprimento 8 é proposta. A nova classe é baseada na parametrização do algoritmo rápido de Loeffler. Por meio de um procedimento de otimização multicritério, baseada em figuras de mérito como medidas de proximidade em relação a DCT e de codificação, novas transformadas eficientes, na fronteira de Pareto, são propostas. Essas transformadas eficientes apresentam boas propriedades em termos de aproximação com a DCT, baixa complexidade aritmética, ortogonalidade e bom desempenho em compressão de imagens. Mostra-se que as transformadas propostas superam aproximações no estado da arte em diversas métricas consideradas.

Palavras-Chave: *DCT; Aproximações; Algoritmo de Loeffler; Otimização Multicritério; Compressão de Imagens.*

Título: Ensaio Computacional sobre o Modelo Gama Generalizado em Processamento de Imagens SAR

Autores: Ramon Lima; Abraão D. C. Nascimento; Renato J. Cintra; Alejandro C. Frery

Resumo: Sistemas de radar de abertura sintética têm sido utilizados em sensoriamento remoto. Tais sistemas podem produzir imagens com alta resolução espacial; entretanto, estas imagens são contaminadas por uma interferência, denominada speckle. Este ruído dificulta o reconhecimento de padrões em imagens SAR por análise visual e/ou pelo uso de métodos clássicos. Assim, a proposta de técnicas especializadas é uma importante etapa de pre-processamento de imagens SAR. Recentemente, uma versão da distribuição gama generalizada foi aplicada com sucesso para descrever dados SAR de intensidade. Este artigo fornece uma discussão detalhada sobre estimação por máxima verossimilhança (MV) para os parâmetros do modelo gama generalizado aplicado a imagens SAR. Em particular, dois estimadores em forma fechada são propostos para os parâmetros do modelo gama generalizado (sendo um deles o número de looks). Um estudo de simulação Monte Carlo é realizado para quantificar a performance das estimativas de MV. Finalmente, uma aplicação a dados reais é realizada. Resultados apresentam evidências em duas direções, (i) o uso do modelo gama generalizado é mais recomendado do que o gama em regiões com textura mais intensa e (ii) as estimativas MV (realização do estimador proposto) para o número de looks são menos viesadas do que as obtidas através do modelo gama.

Palavras-Chave: *Sistemas SAR; Modelo gama generalizada; Estimação MV; Parametrização de Li et al.*

Título: Uma abordagem para monitorar dados de contagem utilizando Regressão poisson por componentes principais

Autores: Danilo Marcondes Filho, Ângelo Márcio Oliveira Sant'Anna

Resumo: Cartas de controle baseadas em modelos de regressão são úteis para monitorar processos em que a variável resposta varia em função de ajustes em variáveis de processo. Sua utilização permite o monitoramento da estrutura de correlação entre as variáveis de processo e a variável resposta através dos resíduos do modelo ajustado a partir de dados históricos do processo. Entretanto, esta estratégia está restrita a dados oriundos de variáveis de processo não correlacionadas significativamente. Caso contrário, variáveis de processo colineares e que guardam informação substancial sobre a variabilidade da variável resposta podem ficar ausentes no ajuste do modelo de regressão. Este artigo propõe uma estratégia para monitorar dados de contagem que combina Regressão de Poisson e Análise por Componentes Principais (ACP). Nesta estratégia, as variáveis colineares são transformadas em variáveis não correlacionadas via ACP, e uma Regressão de Poisson da variável de contagem sobre os escores da ACP é ajustada. Uma carta de controle de resíduos do modelo é utilizada para detectar descontroles presentes no processo que afetam a relação das variáveis de processo com a variável resposta. O procedimento é ilustrado através de um estudo de caso simulado.

Palavras-Chave: *Regressão Poisson; Carta de controle para Regressão; Análise de Componentes Principais; Regressão ACP.*

Sessão: Estatística Computacional

Título: Variações bootstrap do AIC em regressão beta

Autores: Fábio Mariano Bayer; Francisco Cribari-Neto

Resumo: O critério de informação de Akaike (AIC) é um critério de seleção de modelos largamente utilizado em aplicações práticas. O AIC é um estimador do valor esperado da log-verossimilhança, sendo uma medida de discrepância entre o modelo verdadeiro e o modelo candidato estimado. No entanto, em pequenas amostras o AIC é viesado e tende a selecionar modelos com alta dimensionalidade. Para contornar esse problema nós propomos novos critérios de seleção para serem usados em pequenas amostras, denominados *bootstrap likelihood quasi-CV* (BQCV) e sua modificação 632QCV. Comparamos os desempenhos dos critérios propostos, do AIC e de suas diversas variações que utilizam log-verossimilhança bootstrap por meio de um extensivo estudo de simulação. O estudo numérico considera inúmeros cenários para a seleção em pequenas amostras de modelos de regressão beta com dispersão constante e com dispersão variável. Os resultados mostram que as variações bootstrap do AIC em regressão beta são boas alternativas para a seleção de modelos em pequenas amostras. Pode-se verificar que o uso da log-verossimilhança bootstrap diminui o problema da sobre-especificação do AIC na seleção de modelos. Dentre as diversas variações do AIC investigadas os critérios propostos, BQCV e 632QCV, destacaram-se, apresentando os melhores desempenhos na seleção de modelos em muitos cenários considerados. Apresentamos uma aplicação a dados reais, em que a seleção do modelo de regressão beta com dispersão variável é feita utilizando os critérios de seleção propostos.

Palavras-Chave: *AIC; bootstrap; critérios de seleção; regressão beta; validação cruzada.*

Sessão: Inferência Bayesiana

Título: A mixture model for rare and clustered populations under adaptive cluster sampling

Autores: Kelly Cristina M. Gonçalves; Fernando A. S. Moura

Resumo: Rare populations, such as endangered species, drug users and individuals infected by rare diseases, tend to cluster in regions. Adaptive cluster designs are generally applied to obtain information from clustered and sparse populations. The aim of this work is to propose a unit level mixture model for clustered and sparse population when the data are obtained from an adaptive cluster sample. Our approach allows taking into account heterogeneity among units belonging to different clusters. The proposed model is evaluated using simulated data and a real experiment where adaptive samples were drawn from an enumeration of a waterfowl species in a 5,000 km² area of central Florida.

Palavras-Chave: *Informative sampling; mixture models; RJMCMC.*

Título: Cognitive-Constructivism, Quine, Dogmas of Empiricism, and Muenchhausen's Trilemma

Autores: Julio M. Stern

Resumo: The Bayesian research group at University of Sao Paulo has been exploring a specific version of Cognitive Constructivism - Cog-Con - that has, among its most salient features, a distinctive "objective" character. Cog-Con is supported by a specially designed measure of statistical significance, namely, $ev(H | X)$ - the Bayesian epistemic value of sharp hypotheses H , given the observed data X . This article explores possible parallels or contrasts between Cog-Con and the epistemological framework developed by the philosopher Willard van Orman Quine.

Palavras-Chave: *Bayesian Statistics; Sharp Hypotheses; Significance Measures; Epistemology ; Philosophy of Science.*

Título: Detecção de outliers e análise de resíduos bayesiana em modelos espaciais

Autores: Viviana das Graças Ribeiro Lobo; Thaís Cristina Oliveira da Fonseca

Resumo: Neste trabalho, são propostos métodos de diagnóstico e análise de resíduos para detecção de outliers num contexto espacial. O resíduo bayesiano padronizado espacial é escrito para três modelos (Normal, T-student e GLG). Outros métodos propostos na literatura para detecção de outliers tais como concordância preditiva e teste de Savage-Dickey são abordados e comparados em exemplos simulados.

Palavras-Chave: *detecção de outliers; estatística espacial; análise de resíduos.*

Título: Inferência Bayesiana para uma extensão assimétrica do modelo de Grubbs

Autores: Fábio Rocha da Silva; Lourdes C. Montenegro

Resumo: O modelo de Grubbs é utilizado para comparar vários instrumentos de medição,

e é comum supor que os termos aleatórios seguem uma distribuição normal. Neste trabalho estudamos aspectos de inferência bayesiana para o modelo de Grubbs, onde o valor não observado da covariável (variável latente) segue uma distribuição normal assimétrica (Montenegro et. al. (2010)). As estimativas a posteriori dos parâmetros do modelo em estudo são calculados através do amostrador de Gibbs. Os resultados e métodos desenvolvidos neste trabalho são ilustrados através de um conjunto de dados.

Palavras-Chave: *normal assimétrica; amostrador de Gibbs ; modelo de Grubbs.*

Título: Modelo Hierárquico Bayesiano Multivariado para Mapeamento da Taxa de Mortalidade por Câncer nas Microrregiões do Estado de São Paulo

Autores: Suelen Umbelino da Silva; Aparecida D. P. Souza; Vilma M. Tachibana

Resumo: A modelagem de dados de área tem sido tema de diversas pesquisas em Estatística nas últimas décadas. Os modelos espaço-temporais têm sido utilizados para lidar com esse tipo de dados de um modo natural, uma vez que muitas vezes envolvem processos que têm transições no tempo e no espaço. O avanço da tecnologia e, simultaneamente, de métodos estatísticos têm permitido a elaboração de modelos cada vez mais estruturados para a descrição de fenômenos aleatórios complexos, cuja ideia é descrever, de forma realista, a estrutura de correlação presente nos dados, o que pode ser feito através do uso de modelos hierárquicos dinâmicos. Dada a importância do tema modelagem de fenômenos espaço-temporais, neste trabalho são estudadas propostas recentes apresentadas na literatura, envolvendo modelos autoregressivos condicionais multivariados para capturar a estrutura espacial e modelos dinâmicos para capturar a estrutura temporal. Como aplicação da metodologia em estudo explora-se a distribuição espacial da taxa de mortalidade pelos cânceres de maior importância quantitativa, a saber o de traqueia, brônquios e pulmão, feminino de mama, de lábios, cavidade oral e faringe, de estômago e de cólon, por microrregiões administrativas do estado de São Paulo, considerando o período 1998 até 2010. Os resultados da aplicação de um modelo bayesiano hierárquico multivariado para os dados evidenciaram as regiões centro e norte como sendo as de maior risco de mortalidade pelos cânceres estudados, além de mostrar que existe forte correlação espacial entre quatro deles, o que constitui um resultado importante para os órgãos do sistema de saúde, que tem como função direcionar e alocar recursos para o tratamento e diagnóstico de tais doenças. Como se trata de modelos com estrutura complexa, para o procedimento de inferência, são utilizados métodos de Monte Carlo via Cadeias de Markov.

Palavras-Chave: *Modelo Hierárquico Bayesiano; Modelo Autorregressivo Condicional (CAR) Multivariado; Simulação Estocástica; Taxa de Mortalidade por Câncer.*

Título: Split-merge procedure to estimate mixture models

Autores: Daiane A Zuanetti; Luiz A Milan

Resumo: A mixture model with dependent variables generalizes some others specific models. We propose the split-merge MCMC procedure to estimate, jointly, its number of components, parameters and the component from which each observation comes from. We employ a Binomial mixture model as our main example, but the methodology is not specific to it. The performance of the method is checked on simulated and real datasets.

Palavras-Chave: *mixture models; dependent variables; split-merge MCMC.*

Sessão: Inferência Estatística

Título: An improved Wald test in dispersion models

Autores: Tiago M. Magalhães; Denise A. Botter; Mônica C. Sandoval

Resumo: In this work, we find the second-order covariance matrix of the bias-corrected maximum likelihood estimator of the regression parameter β in dispersion models. Based on this matrix, we modify the Wald test, which improves its performance. We evaluate the results by using a Monte Carlo simulation.

Palavras-Chave: *Bias estimator; Covariance matrix; Improved Wald test.*

Título: Bayesian semi-parametric symmetric models for binary data

Autores: Marcio Augusto Diniz; Carlos Alberto de Braganca Pereira; Adriano Polpo

Resumo: This work proposes a general Bayesian semi-parametric model to binary data. It is considered symmetric prior probability curves as an extension for discussed ideas from Basu and Mukhopadhyay (1998) using the Blocked Gibbs sampler which is more general than the Polya Urn Gibbs sampler. The semi-parametric approach allows to incorporate the uncertainty around the F distribution of the latent data and modeling heavy-tailed or light-tailed distributions than that prior proposed. In particular, the Bayesian semi-parametric Logistic model is introduced which enables one to elicit prior distributions for regression coefficients from information about odds ratios what is quite interesting in applied research. Then, this framework opens several possibilities to deal with binary data in the Bayesian perspective.

Palavras-Chave: *Bayesian Semi-parametric; Logistic model; Dirichlet process.*

Título: Distribuição Birnbaum-Saunders Tri-paramétrica

Autores: Terezinha K. de A. Ribeiro; Michelli Barros; Areli M. da Silva; Manoel Santos-Neto

Resumo: Neste trabalho, estudamos uma generalização da distribuição Birnbaum-Saunders, proposta por Owen (2006). Inferência estatística para os parâmetros do modelo são apresentadas, bem como, análise de diagnóstico baseado no método de influência local. O comportamento dos estimadores foram avaliados por meio de estudos de simulação de Monte Carlo. Além disso, realizamos uma comparação entre as estimativas intervalares assintóticas e as estimativas intervalares obtidas via bootstrap. Por fim, aplicamos a teoria estudada a um conjunto de dados reais.

Palavras-Chave: *distribuição Birnbaum-Saunders generalizada; método de máxima verossimilhança; bootstrap.*

Título: Likelihood Based Inference for Quantile Regression Using the Asymmetric Laplace Distribution

Autores: Luis Benites Sánchez; Victor Hugo Lachos; Filidor Vilca Labra

Resumo: To make inferences about the shape of a population distribution, the widely popular mean regression model, for example, is inadequate if the distribution is not approximately Gaussian (or symmetric). Compared to conventional mean regression (MR), quantile regression (QR) can characterize the entire conditional distribution of the outcome variable, and is more robust to outliers and misspecification of the error distribution. We present a likelihood-based approach to the estimation of the regression quantiles based on the asymmetric Laplace distribution (ALD), a choice that turns out to be natural in this context. The ALD has a nice hierarchical representation which facilitates the implementation of the EM algorithm for maximum likelihood estimation of the parameters at the p th level with the observed information matrix as a byproduct. Inspired by the EM algorithm, we develop case-deletion diagnostics analysis for QR models, following the approach of Zhu et al. (2001). This is because the observed data log-likelihood function associated with the proposed model is somewhat complex (e.g., not differentiable at zero) and by using Cook's well-known approach it can be very difficult to obtain case-deletion measures. The techniques are illustrated with both simulated and real data. In particular, in an empirical comparison, our approach out-performed other common classic estimators under a wide array of simulated data models and is flexible enough to easily accommodate changes in their assumed distribution. The proposed algorithm and methods are implemented in the R package ALDqr()

Palavras-Chave: *Quantile regression model; EM algorithm; Case-deletion model; asymmetric Laplace distribution.*

Título: Zero-Modified Power Series Distributions

Autores: Katiane S. Conceição; Marinho G. Andrade; Francisco Louzada

Resumo: In this paper we introduce the Zero-Modified Power Series distributions family, a new class of discrete distributions for modeling datasets exhibiting overdispersion or underdispersion beyond zero-inflated or zero-deflated proportion. The new class of distributions is very flexible and has the Power Series and Zero-Truncated Power Series distributions as particular cases. The probability mass function and characterizations of this new class are presented. The distribution parameters are estimated via maximum likelihood method. The use of this new family of distributions is illustrated in three real datasets.

Palavras-Chave: *Power Series Distribution; Zero-Truncated Distributions; Zero-Inflated Dataset; Zero-Deflated Dataset.*

Sessão: Modelos de Regressão

Título: A Semiparametric Approach for Joint Modeling of Median and Skewness

Autores: Luis Hernando Vanegas; Gilberto A. Paula

Resumo: In this paper we show through Monte Carlo simulation that ignoring the skewness of the response variable distribution in nonlinear regression models may introduce biases on the parameters estimation and/or on the estimation of the associated variability measures. Therefore, this paper proposes a semiparametric regression model suitable for data set analysis in which the distribution of the response is strictly positive and asymmetric. In this setup, both median and skewness of response variable distribution are explicitly modeled, the median using a nonlinear function and the skewness using a semiparametric function. The proposed model allows for the description of the response using the log-symmetric distribution, which is a generalization of the log-normal distribution and is flexible enough to consider bimodal distributions in particular cases as well as distributions having heavier or lighter tails than those of the log-normal. An iterative estimation process as well as some diagnostic methods (overall and individual goodness-of-fit measures and local influence) are derived. Two data sets previously analyzed under parametric models are reanalyzed using the proposed methodology.

Palavras-Chave: *skewness; asymmetric responses; maximum penalized likelihood estimates; semiparametric models; robust estimates.*

Título: A general class of zero-or-one inflated logit-skew-normal regression models

Autores: Guillermo Martínez-Flórez; Germán Moreno-Arenas; Solange Lugo-Buitrago

Resumo: This paper proposes a general class of regression models for continuous proportions when data is inflated with zeros or/and ones. The proposed models assumes that the response variable has a mixture continuous--discrete distribution with covariates in both the discrete and continuous parts of the model. As revealed by real data applications, models investigated seem to be a valid alternative to the modeling proportions and rates with zero or one inflated.

Palavras-Chave: *Proportions; zero-or-one inflated; alpha-power-normal distribution; regression models.*

Título: Bayesian Inference and Diagnostics in Zero-Inflated Generalized Power Series Regression Model

Autores: Gladys D.C. Barriga, Vicente G. Cancho

Resumo: The paper provides a Bayesian analysis for the zero-inflated regression models based on the generalized power series distribution. The approach is based on Markov chain Monte Carlo methods. The residual analysis is discussed and case-deletion influence diagnostics are developed for the joint posterior distribution, based on the ψ -divergence, which includes several divergence measures such as the Kullback-Leibler, J-distance, L1 norm and χ^2 -square in zero-inflated general power series models. The methodology is reflected in a dataset collected by wildlife biologists in a state park in California.

Palavras-Chave: *Bayesian analysis; count data; regression model; zero-inflated model.*

Título: Cook's distance for generalized linear mixed models

Autores: Luis Gustavo Bastos Pinho; Juvêncio Santos Nobre; Julio da Motta Singer

Resumo: Consideramos uma extensão da distância de Cook para modelos lineares mistos generalizados com o objetivo de identificar observações com alta influência no valor condicional predito para a variável resposta. A distância proposta pode ser decomposta em três fatores que ajudam a distinguir entre influência na estimação de efeitos fixos e efeitos aleatórios. Influência conjunta e condicional também foram avaliadas. Uma aproximação de primeira ordem para agilizar os cálculos é apresentada e uma simulação de Monte Carlo ilustra a capacidade da técnica proposta. Uma aplicação em um conjunto de dados usado na literatura é apresentada para exemplificar como a técnica proposta é utilizada na prática.

Palavras-Chave: *GLMM; Diagnóstico; Alavancagem; Distância de Cook.*

Título: Diagnostics for censored mixed-effects models using the multivariate t-distribution

Autores: Dipankar Bandyopadhyay; Larissa A. Matos; Luis M. Castro; Victor H. Lachos

Resumo: In biomedical studies on HIV RNA dynamics, the viral loads generate repeated measures that are often subjected to (upper and lower) detection limits, and hence these responses are either left- or right-censored. Linear and non-linear mixed-effects censored (LMEC/NLMEC) models are routinely used to analyze these longitudinal data, with normality assumptions for the random effects and residual errors. However, the derived inference may not be robust when these underlying normality assumptions are questionable, specially presence of outlier- s and thick-tails. Motivated by this, Matos et al. (2013b) recently proposed an exact EM- type algorithm for LMEC/NLMEC models using a multivariate Student's-t distribution, with closed-form expressions at the E-step. In this paper, we develop influence diagnostics for LMEC/NLMEC models using multivariate Student's-t density, based on the conditional ex- pectation of the complete data log-likelihood which eliminates the complexity associated with the approach of Cook (1977, 1986) for censored mixed-effects models. The new method- ology is illustrated through an application to a longitudinal HIV dataset using the NLMEC framework. In addition, a simulation study is presented, which explores the accuracy of the proposed measures in detecting influential observations in heavy-tailed censored data under different perturbation schemes.

Palavras-Chave: *Censored data; Case-deletion diagnostics; ECM algorithm; Linear mixed-effects model; Multivariate t.*

Título: Fitting non-linear mixed models to messy longitudinal data

Autores: Julio M. Singer; Carmen D.S. André; Francisco M.M. Rocha; Talita Zerbini

Resumo: Non-linear mixed models are useful in many practical longitudinal data problems, especially when they are derived from subject matter theoretical considerations, usually in the form of differential equations, or when the response approaches an asymptote. Although many alternatives are available in the literature, examples are generally based on regular and well behaved data. Fitting such models to messy data may be associated with problems such as failure of convergence of the computational algorithms or lack of diagnostic tools. We consider data from an observational study

conducted at the School of Medicine of the University of São Paulo in which a new method to estimate the time since death (postmortem interval) is proposed. In particular, the attenuation of the density of intra-cardiac hypostasis (concentration of red cells in the vascular system by gravity) obtained from a series of tomographic images was observed in the thoraces of 21 bodies of hospitalized patients with known time of death. The images were obtained at different instants and not always at the same conditions for each body, generating a set of messy data. We use segmented regression, Mitscherlich and exponential linearizable mixed models to represent the evolution of the response along time since death, commenting on the advantages and caveats of each approach. We show that the results lead to very similar conclusions but that the linearizable exponential mixed model is more flexible and may profit from the well established theory, fitting algorithms and diagnostic tools.

Palavras-Chave: *Autopsy; Computer tomography; Diagnostics; Hypostasis; Postmortem interval.*

Título: Modelling performance of students with bivariate generalized linear mixed models

Autores: Gabriel Franco ; Hildete P. Pinheiro; Mariana R. Motta

Resumo: We propose a bivariate generalized linear mixed models (GLMM) to evaluate the performance of undergraduate students from the State University of Campinas (Unicamp). For each student we have the final GPA score as well as the number of courses he/she failed during his/her Bachelor's degree. The courses are separated in three categories: Required (R), Elective (E) and Extracurricular courses (Ex). Therefore, for each variable each student may have at most three measures. In this model we need to take into account the within student correlation between required, elective and extracurricular courses as well as the correlation between the GPA score and the number of courses failed. The main purpose of this study is the sector of High School education from which college students come - Private or Public. Because of affirmative action programs being implemented by the Brazilian government to include more students from Public Schools in the Universities, there is a great interest in studies of performance of undergraduate students according to the sector of High School of which they come from. The data comes from the State University of Campinas (Unicamp), a public institution, in the State of S~ao Paulo, Brazil and one of the top universities in Brazil. The socioeconomic status and academic data of more than 10,000 students admitted to Unicamp from 2000 through 2005 forms the study database.

Palavras-Chave: *multivariate generalized mixed models; multivariate analysis; zero inflated models.*

Título: Modelo de Regressão Beta com Erro de Medida Multiplicativo Log-Normal

Autores: Eveliny Barroso da Silva; Carlos Alberto Ribeiro Diniz; Jalmar Manuel Farfan Carrasco

Resumo: Neste trabalho, propomos o modelo de regressão beta com erro de medida multiplicativo log-normal. Alguns métodos de estimação são estudados. Tais métodos têm como princípio a estimação por máxima pseudo-verossimilhança. Para controlar problemas de identificabilidade do modelo, que é usual é modelos com erro de medida nas variáveis, estimamos a variância do erro de medida via dados replicados. Um estudo de simulação foi feito para ilustrar os resultados das estimações para cada método.

Palavras-Chave: *Modelos de regressão beta; modelos com erros nas covariáveis; pseudo-verossimilhança.*

Título: Modelo de regressão autoregressivo Binomial bivariado inflacionado de zeros

Autores: Natália Manduca Ferreira; Carlos Alberto Ribeiro Diniz

Resumo: Problemas envolvendo dados de contagem podem resultar em conjunto de dados com uma grande quantidade de zeros. Quando utilizamos distribuições usuais, como Poisson, Binomial ou Binomial Negativa, em conjuntos com excesso de zeros, análises estatísticas podem apresentar-se errôneas. As distribuições mais indicadas para este caso são as compostas por uma mistura de distribuições, sendo uma degenerada no ponto zero e outra que se adequaria aos dados caso não houvesse a inflação de zeros. Neste artigo utilizaremos a distribuição Binomial bivariada inflacionada de zeros para construir o modelo de regressão autoregressivo Binomial bivariado inflacionado de zeros. A estrutura de regressão é modelada por meio de uma função de ligação. Uma estrutura de correlação autoregressiva é assumida nos componentes aleatórios que compõem o modelo a ser utilizado. A estimação dos parâmetros é feita via algoritmo EM, sendo a parte computacional realizada em linguagem Ox.

Palavras-Chave: *Modelos zero inflacionados; Binomial bivariada; efeito aleatório.*

Título: The Sinh-Normal/Independent Nonlinear Regression Model

Autores: Filidor Vilca Labra; Camila Borelli Zeller; Gauss M. Cordeiro

Resumo: The Normal/Independent family of distributions is an attractive class of symmetric heavy-tailed density functions. They have a nice hierarchical representation to make inferences easily. We propose the Sinh-Normal/Independent distribution which extends the Sinh-normal distribution (Rieck, 1989). We discuss some of its structural properties that generalize those of the Birnbaum-Saunders distribution. Further, we introduce the Sinh-Normal/Independent nonlinear regression model based on a similar set-up of Lemonte and Cordeiro (2009). We develop an EM-algorithm for maximum likelihood estimation of the model parameters. In order to examine the robustness of this flexible class against outlying observations, we perform a simulation study and analyze a real data set to illustrate the usefulness of the new model.

Palavras-Chave: *EM-algorithm; Nonlinear regression; Robust estimation; Scale mixture of normal distributions; Sinh-normal distribution.*

Título: Um novo resíduo em modelos de regressão inflacionados no zero

Autores: Gustavo H. A. Pereira; Denise A. Botter; Mônica C. Sandoval

Resumo: Modelos de regressão inflacionados no zero são desenvolvidos para ajustar variáveis resposta que podem assumir o valor zero com probabilidade maior do que a permitida por conhecidas distribuições de probabilidade. Neste trabalho, o interesse recai em variáveis resposta que podem assumir o valor zero com probabilidade positiva e apresentam distribuição contínua em algum intervalo do conjunto dos números reais positivos. Diversos trabalhos teóricos nessa classe de modelos têm sido desenvolvidos e essa classe tem aplicações em várias diferentes áreas. Porém, ainda não há na literatura um resíduo com boas propriedades para realizar análise de diagnóstico do componente contínuo dos modelos de regressão inflacionados no zero. Este trabalho introduz um resíduo para realizar análise de diagnóstico do componente contínuo de modelos de

regressão inflacionados no zero. Estudos de simulação de Monte Carlos são utilizados para comparar o resíduo proposto com o existente na literatura para o caso do modelo de regressão beta inflacionado no zero. Os resultados sugerem que o resíduo proposto tem melhores propriedades que o resíduo existente.

Palavras-Chave: *Análise de diagnóstico; Modelos de regressão inflacionados; Regressão beta.*

Sessão: Planejamento de Experimento e Controle Estatístico de Qualidade

Título: A combined npx Xbar control chart to monitor the process mean in two-stage sampling

Autores: Elvis dos S. Sampaio; Linda Lee Ho; Pledson G. de Medeiros

Resumo: This paper proposes a new control chart to monitor a process mean employing a combined npx-X control chart. Basically the procedure consists of splitting the sample of size n into two sub-samples n_1 and $n_2 = (n - n_1)$ determined by an optimization search. Units of the first sub-sample are evaluated by attributes and plotted in npx control chart. If this chart signs then units of second sub-sample are measured and the monitored statistic plotted in X control chart. If both control charts sign then the process is stopped for adjustment. The possibility of non-inspection in all n items may promote a reduction not only in the cost but also the time spent to examine the sampled items. Performances of the current proposal and single control charts of X and npx are compared. In this study the proposed procedure presents many competitive options for the X control chart for a sample size n and a shift δ from the target mean. The average time to sign (ATS) of the current proposal lower than the values calculated from a single X control chart (or npx control charts) points out that the combined control chart is an efficient tool in monitoring process mean.

Palavras-Chave: *Process monitoring; control chart; combined control chart.*

Título: Análise do desempenho do gráfico de controle por grupos (GG) em processos com fluxos correlacionados

Autores: Max Brandão de Oliveira; Hérika Brenda Santana da Silva

Resumo: Os gráficos de controle (GC) foram propostos por Walter A. Shewhart em 1931, tendo como objetivo monitorar um processo de fabricação em série, onde os produtos são originados de um único fluxo de produção. Com o surgimento dos processos paralelos (PP), aqueles cujo produto é produzido simultaneamente em vários fluxos de produção, o monitoramento por meio do GC tradicional de Shewhart (GCS) requer a construção de duas cartas para cada fluxo, uma para o nível da qualidade e outra para variabilidade, significando uma análise individual para cada um deles. No entanto, em 1950, Boyd apresenta os gráficos de controle por grupos (GG), uma nova alternativa com o objetivo de facilitar a inspeção de processos paralelos. O GG consiste na elaboração de um único gráfico para o monitoramento do nível da qualidade e outro para a variabilidade do processo, permitindo efetuar o monitoramento de várias linhas de produção com apenas duas cartas. Porém, ambos os GG e o GCS desconsideram a existência de uma estrutura de correlação entre as linhas produtivas. Este artigo, portanto, tem o objetivo de analisar o desempenho do GG na presença e na ausência de correlação entre os fluxos através de uma algoritmo de simulação. O indicador de desempenho é o número médio de amostras (NMA) até a ocorrência do primeiro alarme falso e o algoritmo foi elaborado com base no método de monte carlo (MMC). Além disso, o estudo considera somente o GG para a média. Resultados apontam que o desempenho do GG com correlação é inferior ao sem correlação, chegando a ser 94; 3% inferior para $r = 0; 1$ quando comparado com fluxos independentes.

Palavras-Chave: *Gráficos por Grupos; Processos paralelos; Fluxos correlacionados; Desempenho.*

Título: Gráfico EWMA das diferenças em relação ao nível-base do processo para o controle da média em Processos Multicanal

Autores: Simões, Bruno Francisco Teixeira ; Epprecht, Eugenio Kahn

Resumo: Os Gráficos de Controle de Grupos (GCG), desenvolvidos por Boyd (1950), são as técnicas mais utilizadas para o Controle Estatístico de Processos Multicanal (Pyzdek, 1992 e Montgomery, 2001). A proposta deste artigo é propor e analisar o desempenho da versão EWMA do GCG proposto em Epprecht et al. (2011) para o controle da componente individual de variação de Processos Multicanal. Em termos de desempenho, o esquema proposto será comparado aos esquemas EWMA de R_t de Mortell e Runger (1995) e MEWMA de S^2 de Runger et al. (1996) no caso de alterações na média de um canal individual. São apresentadas e comparadas as medidas de desempenho considerando o controle por observações individuais e por amostras de tamanho igual a quatro. Na maioria dos casos, o GCG de EWMA proposto é mostrado como um dos mais eficientes para alterações de magnitude pequena a moderada na média ($0.5 \leq \delta \leq 1.5$).

Palavras-Chave: *Processos Multicanal; Controle estatístico da média; Componente individual de variação; Gráfico de Controle de Grupos; EWMA.*

Título: Improving Shewhart-type Generalized Variance Control Charts for Multivariate Process Variability Monitoring using Cornish-Fisher Quantile Correction, Meijer-G Function and Other Tools

Autores: Emanuel P. Barbosa ; Mario A. Gneri ; Ariane Meneguetti

Resumo: This paper presents an improved version of the Shewhart-type generalized variance $|S|$ control chart for multivariate process dispersion monitoring, based on the Cornish-Fisher quantile formula for non-normality correction of the traditional normal based 3-sigma chart limits. Also, the exact sample distribution of $|S|$ and its quantiles (chart exact limits) are obtained through the Meijer-G function, and an auxiliary control chart is introduced in order to avoid non detection of certain changes in the process covariance matrix. The performance of the proposed new chart is compared in terms of false alarm risk, with the traditional normal chart and with the exact distributed chart for dimensions 2 and 3. The proposed new chart is illustrated with two numerical examples.

Palavras-Chave: *Control Chart; Variability Monitoring; Multivariate Process; Cornish-Fisher.*

Título: Métodos estatísticos aplicados a uma indústria automobilística

Autores: Raphael Henrique Teixeira da Silva; Daniela Carine Ramires de Oliveira

Resumo: Visitas técnicas a empresas da região de São João del Rei-MG têm mostrado que o monitoramento de características do processo vem sendo realizado sob uma perspectiva univariada, isto é, as empresas monitoram o processo construindo gráficos de controle de qualidade univariados para cada característica individualmente. No entanto, os processos analisam várias variáveis e as mesmas podem estar correlacionadas. Embora a aplicação de gráficos de controle univariados seja uma solução possível, a literatura mostra que isso não é eficaz e pode levar a conclusões errôneas. Além disso, verifica-se na rotina desses processos o uso de tamanho de amostra variável. A teoria e implementação computacional das versões multivariadas dos gráficos de controle são

apresentadas em sua maioria quando o tamanho amostral é fixo. Devido a dificuldade de se obter essas duas informações simultaneamente na literatura, as empresas apresentam resistência ao uso dessas versões multivariadas para o controle de qualidade de seus produtos. Nesse contexto, esse trabalho apresenta a teoria e implementação computacional no software livre R do gráfico T^2 de Hotelling (versão multivariada do gráfico Xbarra de Shewhart), quando o tamanho da amostra é variável. Adicionalmente, o gráfico T^2 de Hotelling é empregado de forma satisfatória a um estudo de caso de uma empresa que produz itens de segurança para automóveis na região de São João del Rei-MG.

Palavras-Chave: *Estatística T^2 ; Limite superior de controle; Controle estatístico de qualidade; Controle estatístico de processos; Gráfico de controle multivariado.*

Sessão: Probabilidade e Processos Estocásticos

Título: A Hubbert Diffusion Process

Autores: Istoni da Luz Sant'Ana, Patricia Román Román, Francisco Torres Ruiz

Resumo: A problem of great current interest is to chart accurately the progress of oil production. It is well known that oil exploration is conducted in cycle and, in fact, after the oil production reaches its peak in a specific system, a fatal decline will begin. In this work, we propose a stochastic model, based on the theory of diffusion process, associated with the Hubbert curve. With this model in mind, we intend to give a probabilistic treatment to the one cycle oil production, including the forecasting of its peak and peak-time. After building the model, comprehensive study is presented, including its main characteristics and a simulation of sample paths. The inferential study is carried out on the basis of discrete sampling and includes numerical aspects related to the search of initial solutions to be taken into account for solving the likelihood equations. The possibilities of the new process are illustrated by means of simulated data.

Palavras-Chave: *Hubbert curve; Diffusion process; Peak oil*

Título: Ajuste da Distribuição de Gumbel para a análise de dados de precipitação do município de São João do Cariri-PB

Autores: Manoel Rivelino Gomes de Oliveira, Moacyr Cunha Filho, Ana Patrícia Siqueira Tavares Falcão, Ewerton Pereira de Oliveira, Maria Das Vitórias Alexandre Serafim

Resumo: A precipitação é um integrador complexo das relações naturais entre o tipo climático, temperatura, umidade relativa do ar, velocidade e direção dos ventos e vegetação. Dessa forma, a busca de um ajuste da série de precipitação do município de São João do Cariri-PB, à Distribuição de valor extremo ou de Gumbel, permitirá entender as características dessas chuvas em regiões semiáridas. Para tal foram utilizados dados de precipitação pluviométrica, coletados em uma estação meteorológica. A análise, dos resultados obtidos revela que a Distribuição de Gumbel ajustada aos dados de precipitação representa adequadamente as característica e variabilidade da série em estudo.

Palavras-Chave: *Distribuição de Gumbel; ajuste; precipitação.*

Título: Continuidade das probabilidades de transição de processos de renovação obtidos por agregação de estados

Autores: Walter Augusto Fonseca de Carvalho

Resumo: Este trabalho é dedicada ao estudo dos processos de renovação binários obtidos como agregação de estados a partir de processos Markovianos com alfabeto finito. Utilizamos uma abordagem matricial para obter condições sob as quais o processo agregado pertence a cada uma das seguintes classes: (1) Markoviano de ordem finita, (2) Processo de ordem infinita com probabilidades de transição contínuas.

Palavras-Chave: *Transformação Agregante; Processo Markoviano Oculto; Processos de Renovação; Agregabilidade.*

Título: New Stochastic Orders based on the Inactivity Time

Autores: S. Yasaei Sekeh, G.R. Mohtashami Borzadaran, A. H. Rezaei Roknabadi

Resumo: Measure of uncertainty in past lifetime distribution is particularly suitable measure to describe the information in problems related to ageing properties of reliability theory based on distribution of components or systems. This measure has been defined by Ruiz and Navarro (1996) on using the physical significance. In this paper, we introduce a new measure related to the moment orders and define new stochastic orders based on that. We also provide some stochastic comparisons with other certain well-known ageing stochastic orders. Finally, a few properties for series systems and mixture with respect to defined measures are discussed.

Palavras-Chave: *Inactivity time; Stochastic orders; Moment inactivity time; Series system; Mixture.*

Sessão: Séries Temporais e Econometria

Título: Financial Crisis of 2007-2009 and Stress Test with Structural Changes for the Brazilian Financial Market

Autores: Aléssio Tony Cavalcanti de Almeida; Bruno Ferreira Frascaroli

Resumo: Due to the financial crisis of 2007-2009 and the serious problems still faced in particular by some countries in the Eurozone, we can check the current inseparability between financial and productive system. This context indicates the importance of analysis committed to measuring risk involving financial activities. Thus, the central objective of this work is to assess the stress effects of a financial distress in daily return series of the domestic market index on the daily return series of the major companies' assets in the Brazilian financial market, considering the period of international economic crisis and possible structural changes in the stress parameters. We use the methodology CoVaR, described by Adrian and Brunnermeier (2011), and analysis of structural changes in Quantile Regression (QR), highlighted by Qu (2008) and Oka and Qu (2011), to estimate the modifications in the stress test before and after of the last economic crisis. The main innovation of this study is to consider structural changes in the stress test, obtained by CoVaR for the quantiles 0.01, 0.05 and 0.10. The results show that some of the reaction coefficient of the returns of firms operating in the Brazilian financial market had structural changes, as well as the value at risk of the market index decreased between the two periods. In general, the marginal contribution of financial distress in the market index, independently of the quantile evaluated, generated a lower stress on the returns of firms after the international crisis than before of this adverse macroeconomic context.

Palavras-Chave: *Stress Test; Quantile Regression; Structural Changes; Financial Crisis.*

Título: Generalized Normal ARMA Model

Autores: Eder Angelo Milani; Marinho G. Andrade; Carlos Alberto Ribeiro Diniz

Resumo: Using the generalized normal distribution, introduced by Nadarajah (2005), and the ideas of the GARMA model, introduced by Benjamin et al. (2003), we develop an autoregressive process called the generalized normal ARMA model (GN-ARMA). This new model is a alternative to model time series which have heavier or lighter tails than the normal distribution. Some properties of the processes are discussed, estimation of parameters is addressed and a method for forecasting a value of the dependent variable is proposed. Simulation studies are developed and a real dataset involving streamflow is analysed using both model, the usual ARMA model and the GN-ARMA model.

Palavras-Chave: *Time series; Generalized normal distribution; ARMA models; GARMA models.*

Título: Modeling reliability data using non-gaussian state space models

Autores: Thiago Rezende dos Santos; Dani Gamerman; Glaura C. Franco

Resumo: This paper presents a likelihood structure, encountered in several reliability models, that consists on a decomposition of data information in stages or times, thus leading to latent parameters. Piecewise exponential models, proportional hazards models and a few software reliability models are among the models in this unifying framework.

Latent states are related across stages through a non-Gaussian state space model framework. A distinctive feature of the models is their mathematical tractability, allowing for the exact computation of the marginal likelihood function, despite the non-Gaussianity of the state-space. Real-life examples illustrate the approach and reveal some advantages of our approach. Agradecemos à FAPEMIG pelo suporte financeiro.

Palavras-Chave: *Software reliability; state-space; Non-Gaussian filtering; Bayesian; Proportional hazards.*

Título: On the Performance of Wavelet-Based Long-Memory Parameter Estimation

Autores: Lucas Massaroppe, Flavio Caduda, Luiz A. Baccalá

Resumo: After reviewing wavelet techniques used to estimate the long memory parameter d used in ARFIMA models, Monte Carlo simulations are used to evaluate the performance of different discrete wavelet transforms under various mother wavelet choices. By comparing computed small sample bias, standard deviations and mean-square errors from the different methods, MODWPT's (Maximum Overlap Discrete Wavelet Packet Transform) is shown to outperform all other options under a minimum mean-square error criterion using the D(4) wavelet filter.

Palavras-Chave: *Fractionally integrated models; Long-memory; Wavelets.*

Sessão: Teoria da Resposta ao Item

Título: Avaliação da proficiência em inglês acadêmico através de um teste adaptativo informatizado

Autores: Vanessa Rufino da Silva; Mariana Cúri

Resumo: Este trabalho descreve as etapas de transformação de um exame de proficiência em inglês (EPI), utilizado no programa em Ciências da Computação e Matemática Computacional (CCMC) de pós-graduação do ICMC-USP, em um teste adaptativo informatizado (TAI) baseado em um modelo da Teoria de Resposta ao Item (TRI). O formato do EPI desde o ano de 2002, a estrutura dos dados coletados com suas aplicações semestrais e os resultados de uma simulação para verificar a qualidade de recuperação dos traços latentes em um TAI serão brevemente descritos para esclarecer sua utilização como alicerce do TAI implementado. O software TAI-PI (Teste Adaptativo Informatizado para Proficiência em Inglês) foi implementado em Java e o SQLite foi utilizado para armazenamento do banco de dados. A metodologia estatística adotada considerou o modelo de resposta gradual unidimensional, o critério de Kullback-Leibler para seleção de itens, o método de estimação da esperança a posteriori os traços latentes e a abordagem Shadow test para imposição de restrições (de conteúdo e tamanho da prova, por exemplo) na composição do teste de cada indivíduo. Os resultados de uma aplicação real do TAI-PI a alunos de mestrado do CCMC em novembro de 2013 são apresentados, evidenciando a boa qualidade da nova metodologia de avaliação informatizada.

Palavras-Chave: *teste adaptativo computadorizado; teoria de resposta ao item; shadow test.*

Título: Comparação empírica de modelos multidimensionais da teoria da resposta ao item

Autores: Rafael Tezza; Antonio Cezar Bornia; Débora Spenassato; Andréa Cristina Trierweiler

Resumo: Vários modelos para medir traços latentes multidimensionais têm sido desenvolvidos e aplicados com a finalidade de identificar o modelo mais adequado para as mais diversas configurações de traços latentes multidimensionais tanto hierárquica quanto não hierárquica. O presente artigo tem como objetivo comparar as estruturas hierárquicas e não hierárquicas de três modelos multidimensionais da teoria da resposta ao item, tendo como base dados reais referentes da qualidade de interfaces web de comércio eletrônico. Os modelos comparados foram o unidimensional múltiplo, multidimensional compensatório e o bi-fatorial.

Palavras-Chave: *teoria da resposta ao item multidimensional; comparação de modelos; bi-fatorial; qualidade; websites.*