

# VIII WORKSHOP ON PROBABILITY AND STATISTICAL METHODS

Feb 12, 13, 14 - 2020 - UFSCar, São Carlos, Brazil

## PROGRAM BOOK



Organized by Joint Graduate Program in Statistics PIPGES -  
UFSCar/USP

# 8th Workshop on Probabilistic and Statistical Methods

February 12–14, 2020

UFSCar, São Carlos, SP, Brazil

## PROGRAM

ICMC/USP - DEs/UFSCar



## About the 8WPSM

UFSCar, São Carlos, February 12–14, 2020

The Workshop on Probabilistic and Statistical Methods is an activity of the Joint Graduate Program in Statistics UFSCar/USP (PIPGEs), which brings together the research groups of probability and statistics working at ICMC-USP and UFSCar, in São Carlos/SP, Brazil.

The meeting intends to discuss new developments in statistics, probability and their applications. Activities include conferences and invited speaker sessions, contributed talks, poster sessions and a short course devoted to undergraduate/graduate students. The presentations of this new edition are related to probability and stochastic processes, statistical inference, regression models, survival analysis and related topics.

## Organizing Committee

Jorge Luís Bazan, ICMC-USP  
Mariana Curi, ICMC-USP(Chair)  
Rafael Izbicki, DEs-UFSCar(Chair)  
Renato Jacob Gava, DEs-UFSCar  
Vera Tomazella, DEs-UFSCar

## Scientific Committee

Adriano Polpo de Campos, The University of Western Australia  
Carlos Alberto de Bragança Pereira, IME/USP  
Francisco Louzada Neto, ICMC/USP  
Mário de Castro Andrade Filho, ICMC/USP  
Oswaldo Anacleto, ICMC/USP  
Vera Tomazella, UFSCar

## Support Committee

(students from PIPGEs)  
Ana Carolina do Couto Andrade  
Alex Leal Mota  
Bruna Luiza de Faria Rezende  
Caio Moura Quina  
Camila Sgarioni Ozelame  
Carlos Franklin Taco Pedraza  
Djidenou Hans Amos Montcho  
Gustavo Alexis Sabillón  
Jessica Suzana Barragan Alves  
Marcos Jardel Henriques  
Marco Inacio  
Marina Gandolfi  
Milena Nascimento Lima  
Naiara Caroline Aparecido dos Santos  
Osafu Augustine Egbon  
Patrícia Stülp  
Renata Cristina Carregari

# Invited Speakers

## Conferences

André Ponce de Leon F. de Carvalho, ICMC-USP  
Artur Lemonte, UFRN  
Dylan Molenaar, University of Amsterdam  
Gabriela Cybis, UFRGS  
Ido Ben-Ari, University of Connecticut  
Larissa Avila Matos, UNICAMP  
Mário de Castro, ICMC-USP  
Mauricio Sadinle, University of Washington

## Mini-Conferences

Andressa Cerqueira, UFSCar  
Guilherme Ludwig, UNICAMP  
Marcelo Andrade, ICMC-USP/UFSCar  
Michel H. Montoril, UFSCar

## Short Course

Fernanda de Bastiani, UFPE

# Special Sessions

## Machine Learning

*(Chair: R. B. Stern)*

Fabio Gagliardi Cozman, POLI-USP  
João Ricardo Sato, UFABC  
Mauricio Sadinle, University of Washington

## Probability

*(Chair: R. Gava)*

Cristian Favio Coletti, UFABC  
Sandro Gallo, UFSCar  
Miguel Abadi, IME-USP



# 8th Workshop on Probabilistic and Statistical Methods

February 12–14, 2020

UFSCar, São Carlos, SP, Brazil

## SCHEDULE

ICMC/USP - DEs/UFSCar



---

**WEDNESDAY 12 FEB**

**08h00 - 08h30** : Registration

**08h30 - 09h00** : Opening

**09h00 - 10h00** : Mauricio Sadinle, University of Washington

**10h00 - 10h30** : Coffee Break

**10h30 - 11h30** : André Ponce de Leon F. de Carvalho, ICMC/USP

**11h30 - 12h00** : Guilherme Ludwig, UNICAMP

**12h00 - 14h00** : Lunch

**14h00 - 15h00** : Mário de Castro, ICMC-USP

**15h00 - 16h00** : Oral Communications

- 15h00 - 15h20: *Andreza Aparecida Palma*, Diego Ferreira  
UFSCar - Campus Sorocaba / UFPR
- 15h20 - 15h40: *Rodrigo Malavazi Corder*, Gilberto A. Paula, Anaclara Pincelli  
and Marcelo U. Ferreira  
Biomedical Science Institute - University of São Paulo
- 15h40 - 16h00: *Ronaldo Dias*, Julian Collazos and Adriano Z. Zambom  
UNICAMP

**16h00 - 17h00** : Coffee Break / Poster Session 1

**17h00 - 17h30** : Michel H. Montoril, UFSCar

**17h30 - 18h00** : Andressa Cerqueira, UFSCar

## THURSDAY 13 FEB

**08h00 - 10h00** : Short Course by Fernanda de Bastiani, UFPE

**10h00 - 10h30** : Coffee Break

**10h30 - 11h30** : Iddo Ben-Ari, University of Connecticut

**11h30 - 12h00** : Marcelo Andrade, USP/UFSCar

**12h00 - 14h00** : Lunch

**14h00 - 15h00** : Gabriela Cybis, UFRGS

**15h00 - 16h00** : Oral Communications

- 15h00 - 15h20: *Caio Lucidius Naberezny Azevedo*, Jean-Paul Fox and Dalton Francisco de Andrade  
State University of Campinas
- 15h20 - 15h40: *Uriel Moreira Silva*, Felipe Carvalho Álvares da Silva, Luiz Henrique Duczmal, Denise Bulgarelli Duczmal  
UFMG
- 15h40 - 16h00: Boa Vista

**16h00 - 17h00** : Coffee Break / Poster Session 2

**17h00 - 18h00** : Larissa Avila Matos, UNICAMP

## FRIDAY 14 FEB

**08h00 - 10h00** : Short Course by Fernanda de Bastiani, UFPE

**10h00 - 10h30** : Coffee Break

**10h30 - 11h30** : Artur Lemonte, UFRN

**11h30 - 12h30** : Dylan Molenaar, University of Amsterdam

**12h30 - 14h30** : Lunch

**14h30 - 17h30** : Special Sessions

- *Machine Learning (Chair: R. B. Stern):*
  - 14h30 - 15h30: Fabio Gagliardi Cozman, POLI-USP
  - 15h30 - 16h30: João Ricardo Sato, UFABC
  - 16h30 - 17h30: Mauricio Sadinle, University of Washington
- *Probability (Chair: R. Gava):*
  - 14h30 - 15h30: Cristian Favio Coletti, UFABC
  - 15h30 - 16h30: Sandro Gallo, UFSCar
  - 16h30 - 17h30: Miguel Abadi, IME-USP



8th Workshop on Probabilistic and Statistical Methods

February 12–14, 2020

UFSCar, São Carlos, SP, Brazil

## ABSTRACTS

ICMC/USP - DEs/UFSCar



## Conferences

### **André Ponce de Leon F. de Carvalho (ICMC/USP, Brazil)**

From Big Data to Data Science

*Abstract: With to the recent expansion in data generation and the growing importance of exploring the knowledge contained in these data, Data Science is one of the fastest growing area of Exact Sciences. Large companies, like Amazon, Apple, Disney, Facebook, Google and Microsoft are hiring a large number of scientists, engineers and statisticians to work in this area. The ability to acquire, store and transmit data from the most diverse human activities, in the public and private sectors, has grown exponentially. This is generating massive volumes of data. These massive volumes of data, known as Big Data, come from a variety of sources and therefore have a wide variety of structures, ranging from traditional attribute-value tables to videos and messages on social networks. Analyzing these massive volumes of data can generate valuable information for decision making, enabling the extraction of new and useful knowledge. The difficulty of this analysis by traditional data analysis techniques has led to the development of new techniques, expanding the area of Data Science. This talk will present the main aspects, challenges and applications of Big Data and Data Science.*

### **Artur Lemonte Jorge (UFRN, Brazil)**

Poder local dos testes da razão de verossimilhanças, Wald, escore e gradiente sob ortogonalidade

*Abstract: Os poderes locais dos testes da razão de verossimilhanças, Wald, escore de Rao e gradiente sob a presença de um vetor de parâmetros,  $\omega$ , que é ortogonal aos parâmetros restantes são considerados nesta apresentação. Será mostrado que alguns dos coeficientes que definem os poderes locais destes testes ficam inalterados independentemente se  $\omega$  é conhecido ou precisa ser estimado, enquanto que os outros coeficientes podem ser expressados como a soma de dois termos, o primeiro deles corresponde ao termo que é obtido como se  $\omega$  fosse conhecido, e o segundo, um termo adicional produzido pelo fato de  $\omega$  ser desconhecido. Esse resultado será aplicado na classe de modelos de regressão não lineares mistos e os poderes locais dos testes serão comparados.*

**Dylan Molenaar (University of Amsterdam)**

Studying Variability of Measurement Model Parameters across Continuous Background Variables

*Abstract: Multi-group latent variable modeling approaches to study variability of measurement model parameters across categorical background variables like gender, cohort, and experimental conditions have been well established (e.g., Jöreskog, 1971; Meredith, 1993; Mellenbergh, 1989; Millsap, 2012). However, in many applications, the background variable is a continuous variable, for instance, age, socio economic status, or IQ. To study parameter variability in such cases, one can pragmatically choose to categorize the continuous background variable and apply a traditional multi-group model. However, this approach is suboptimal for various reasons. In the present talk, parametric and non-parametric alternatives to study parameter variability across continuous background variables are presented including moderated latent variable models, locally weighted latent variable models, and mixture factor models. The statistical properties of the models are studied, and the models are illustrated on real datasets. References: Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, 36, 409- 426 Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525-543. Mellenbergh, G. J. (1989). Item bias and item response theory. *International journal of educational research*, 13(2), 127-143. Millsap, R. E. (2012). *Statistical approaches to measurement invariance*. Routledge.*

**Gabriela Cybis (UFRGS, Brazil)**

Inference for clustering and classification of high dimensional small sample size data with r-package Uclust

*Abstract: High dimension small sample size datasets are common in fields like genetics, biological imaging and econometrics. These datasets present a significant challenge for statistical testing in clustering problems. We present a nonparametric U-statistic based approach for clustering and classification tailored for these types of problems. The methods require few assumptions about the data generating processes and have more statistical power than competing alternatives. We also present the r-package Uclust for clustering and classification under this framework. To showcase the methods, we present an application to classifying ECG data based frequency spectrums, and an application to hand movement classification based on a concept of the union of homogeneous subgroups.*

**Ido Ben-Ari (University of Connecticut, USA)**

A random walk with catastrophes

*Abstract: Random population dynamics with catastrophes (events pertaining to possible elimination of a large portion of the population) has a long history in the mathematical literature. In this paper we study an ergodic model for random population dynamics with linear growth and binomial catastrophes: in a catastrophe, each individual survives with some fixed probability, independently of the rest. Through a coupling construction, we obtain sharp two-sided bounds for the rate of convergence to stationarity which are applied to show that the model exhibits a cutoff phenomenon.*

**Larissa Avila Matos (UNICAMP, Brazil)**

A semiparametric mixed-effects model for censored longitudinal data

*Abstract: In longitudinal studies involving laboratory-based outcomes, repeated measurements can be censored due to assay detection limits. Linear mixed-effects (LME) models are a powerful tool to model the relationship between a response variable and covariates in longitudinal studies. However, the linear parametric form of LME models is often too restrictive to characterize the complex relationship between a response variable and covariates. More general and robust modeling tools, such as nonparametric and semiparametric regression models, have become increasingly popular in the last decade. In this article, we use semiparametric mixed models to analyze censored longitudinal data with irregularly observed repeated measures. The proposed model extends the censored LME model and provides more flexible modeling schemes by allowing the time effect to vary nonparametrically over time. We develop an EM algorithm for maximum penalized likelihood (MPL) estimation of model parameters and the nonparametric component. Further, as a byproduct of the EM algorithm, the smoothing parameter is estimated using a modified LME model, which is faster than alternative methods such as the restricted maximum likelihood (REML) approach. Finally, the performance of the proposed approaches is evaluated through extensive simulation studies as well as applications to datasets from AIDS studies.*

**Mário de Castro (ICMC/USP, Brazil)**

## Generalized fiducial inference for the Grubbs model

*Abstract: Generalized fiducial inference for the precision of a measuring instrument without available replications on the observations is our main interest. In this talk, we discuss two new estimation procedures for the precision parameters and the product variability under the Grubbs model considering the two-instrument case. One method is based on a fiducial generalized pivotal quantity and the other one is built on the method of the generalized fiducial distribution. The behavior of the point and interval estimators is assessed numerically through Monte Carlo simulation studies. The methodology is applied in the analysis of a data set from a methods comparison study. This is joint work with Lorena C. Tomaya (UFAC).*

**Mauricio Sadinle (University of Washington, USA)**

## Nonparametric Identified Methods to Handle Nonignorable Missing Data

*Abstract: There has recently been a lot of interest in developing approaches to handle missing data that go beyond the traditional assumptions of the missing data being missing at random and the nonresponse mechanism being ignorable. Of particular interest are approaches that have the property of being nonparametric identified, because these approaches do not impose parametric restrictions on the observed-data distribution (what we can estimate from the observed data) while allowing estimation under a full-data distribution. When comparing inferences obtained from different nonparametric identified approaches, we can be sure that any discrepancies are the result of the different identifying assumptions imposed on the parts of the full-data distribution that cannot be estimated from the observed data, and consequently these approaches are especially useful for sensitivity analyses. In this talk I will present some recent developments in this area of research and discuss current challenges.*

## Mini-Conferences

### **Andressa Cerqueira (UFSCar, Brazil)**

Learning communities in weighted networks

*Abstract: Network models have received an increasing attention from the statistical community, in particular in the context of analyzing and describing the interactions of complex random systems. In this context, community structures can be observed in many networks where the nodes are clustered in groups with the same connection patterns. In this talk, we will address the community detection problem for weighted networks in the case where, conditionally on the node labels, the edge weights are drawn independently from a Gaussian random variable with mean and variance depending on the community labels of the edge endpoints. We will present a fast and tractable EM algorithm to recover the community labels that achieves the optimal error rate.*

### **Guilherme Ludwig (UNICAMP, Brazil)**

Scalable modeling of nonstationary covariance functions with regularized B-spline deformations

*Abstract: We propose a semiparametric method for nonstationary covariance function modeling, based on the spatial deformation method of Sampson and Guttorp (1992), but using a low-rank, scalable, regularized deformation function. We show that fine tuning of the regularization parameters can ensure that the deformation does not fold in on itself, and therefore yields proper covariance function estimates. An application to rainfall data illustrates the method.*

### **Marcelo Andrade (ICMC-USP/UFSCar, Brazil)**

Um método de busca de matriz  $Q$  em modelos da TRI multidimensionais

*Abstract: Recentemente, a matriz  $Q$ , que é um elemento responsável por capturar a relação entre itens e traços latentes e está presente na grande maioria dos modelos de diagnóstico cognitivo (MDC), foi incorporada na formulação dos modelos da teoria da resposta ao item multidimensionais (TRIM). Na prática, para utilizar um modelo com matriz  $Q$ , seja ele um MDC ou um modelo da TRIM, é necessário primeiro estabelecer a relação entre itens e traços latentes construindo uma matriz  $Q$  apropriada. Embora este processo de construção da matriz  $Q$  seja tipicamente feita por especialistas no tema dos itens, ele é um processo subjetivo e, portanto, pode implicar em equívocos, resultando em importantes problemas práticos, como, por ex-*

*emplo, influenciar a estimação dos parâmetros do modelo. Neste trabalho, propomos um método de busca de matriz  $Q$  em modelos da TRIM baseado em critérios projetados para fornecer informação máxima sobre alguma propriedade de interesse. Para isso, utilizamos o algoritmo de troca, um método eficiente e sistemático para a busca de matrizes, visando encontrar uma matriz  $Q$  que maximize alguma propriedade de interesse. Além disso, o método fornece informações de eficiência que podem ser úteis na reavaliação de uma matriz  $Q$ . Um estudo de simulação foi realizado para analisar o desempenho do método proposto utilizando dois diferentes cenários construídos a partir da variação do número de itens e do número de dimensões do traço latente. Para ilustrar o uso do método de busca de matriz  $Q$  em dados reais, utilizamos um conjunto de dados com respostas de 1.111 estudantes a uma versão do BDI com 21 itens.*

**Michel H. Montoril (UFSCar, Brazil)**

Wavelet denoising of broadband pulsed signals via empirical Bayes thresholding

*Abstract: Cetacean bioacoustic research is rapidly increasing in importance to answer questions related to the human contribution on acoustic environmental changes. Acoustic towed array is a technique using sometimes opportunistic research vessel platforms in collaboration with other oceanographic investigations. This situation generates additional noise sources that affect or partially mask the signal of interest. This compromises the correct detection and possible classification of the species. In this work, combined wavelets and empirical Bayes thresholding are used to denoise the audio recording of odontocetes echolocation clicks. This method was able to denoise clicks without affecting their features.*

## Short-Course

**Fernanda de Bastiani (UFPE, Brazil)**

Distributional regression approach using GAMLSS

*Abstract: The generalized additive models for location, scale and shape, GAMLSS, are univariate distributional regression models, where all the parameters of the assumed distribution for the response variable can be modelled as additive functions of the explanatory variables. GAMLSS address the problem of choosing an appropriate distribution for the response variable, and models how the distribution parameters vary with changes in the explanatory variables. GAMLSS, its statistical modelling philosophy and its implementation in the software R will be introduced. The different distributions for modelling the response variable, and their properties will be described. These distributions include continuous (positively or negatively skewed and with high or low kurtosis), discrete and mixed distributions. Different additive terms for modelling the parameters of the distribution such as linear, non-parametric smoothing and random effects terms will be shown. Also different modelling selection techniques and diagnostics for checking the model adequacy will be addressed. All example given are real data examples.*

## Special Sessions

### Machine Learning

**Fabio Gagliardi Cozman (Poli/USP, Brazil)**

Explaining Machine Learning

*Abstract: There is now wide interest in techniques that explain automatic decisions generated through machine learning. And there are several ways to generate such interpretations and to explain what happens inside a machine learning "black-box" such as a large random forest or a deep neural network. This talk will discuss various perspectives on how best to explain machine learning techniques, focusing both on textual explanations and on explanations for large-scale embeddings.*

**João Ricardo Sato (UFABC, Brazil)**

Applications of Computational Statistics and Machine Learning in Brain Imaging Data

*Abstract: In this presentation, we will introduce how computational statistics and machine learning methods can be useful to analyze brain imaging data. This is an interdisciplinary research which requires state-of-the-art knowledge from both Neuroscience and machine learning to be successful. Brain imaging data will be introduced to the audience, with the main concern on data structure and quantitative features. Moreover, the mindset of how this combination may be approached in order to raise new questions will be discussed. The presentation will also focus on illustrations of these applications in neurodevelopment, brain disorders, brain-computer interfaces, and Education. Finally, the main challenges and perspectives will be discussed.*

**Mauricio Sadinle (University of Washington, USA)**

Least Ambiguous Set-Valued Classifiers with Bounded Error Levels

*Abstract: In most classification tasks there are observations that are ambiguous and therefore difficult to correctly label. Set-valued classifiers output sets of plausible labels rather than a single label, thereby giving a more appropriate and informative treatment to the labeling of ambiguous instances. We introduce a framework for multiclass set-valued classification, where the classifiers guarantee user-defined levels of coverage or confidence (the probability that the true label is contained in the set) while minimizing the ambiguity (the expected size of the output). We first derive oracle clas-*

sifiers assuming the true distribution to be known. We show that the oracle classifiers are obtained from level sets of the functions that define the conditional probability of each class. Then we develop estimators with good asymptotic and finite sample properties. The proposed estimators build on existing single-label classifiers. The optimal classifier can sometimes output the empty set, but we provide two solutions to fix this issue that are suitable for various practical needs.

## Probability

### **Cristian Favio Coletti (UFABC, Brazil)**

Convergence to a transformation of the Brownian Web for a family of Poissonian trees

*Abstract:* We introduce a system of one-dimensional coalescing random paths starting at the space-time points of a homogeneous Poisson point process in  $\mathbb{R} \times \{0\}$  which are constructed as a function of a family  $(\Lambda_n)_{n \in \mathbb{N}}$  of Poisson point processes. We show that under diffusive scaling this system converges in distribution to a continuous mapping of the Brownian Web.

*Joint work with Leon A. Valencia Henao (UdeA, Colombia).*

### **Sandro Gallo (UFSCar, Brazil)**

Optimal Concentration bounds for stochastic chains

*Abstract:* Concentration bounds usually give upper bounds for the probability that a given statistic of the process deviates from its mean, based on a finite sample. In view of the Central Limit Theorem, which is an asymptotic result, a natural question is whether we can get "Gaussian concentration bounds". This problem is quite well-studied in the context of sequence of i.i.d. random variables, but much less is known in the dependent case. Here we will present gaussian concentration bounds in the context of stochastic chains with countable alphabet, enjoying some kind of weak dependence. We will illustrate our results by examples and counter-examples, showing in particular in which sense our bounds are optimal.

### **Miguel Abadi (IME/USP, Brazil)**

Variations on Kac's Theme

*Abstract:* We presented several variations to the well known Kac's Lemma which are useful to compute size of clustering phenomena that apply still just under stationarity and ergodicity.



## Oral Communications

**Andreza Aparecida Palma (UFSCar - Campus Sorocaba, Brazil), Diego Ferreira (UFPR, Brazil)**

On the subprime crisis and the Latin American financial markets: A regime switching skew-normal approach

*Abstract: This paper analyzes the potential occurrence of financial contagion in Latin American markets from the recent US subprime crisis. Distinctively from the usual empirical approaches for contagion analyses, a regime-switching skew-normal (RSSN) model is implemented in order to assess both correlation and coskewness contagion as well as investigate the occurrence of structural breaks in the moments of the mean, variance and skewness. Even though correlation contagion was observed in all selected Latin American markets, coskewness contagion was only detected in Brazil. Variance structural breaks were found in all financial markets while structural breaks in the mean were only detected in Argentinian, Mexican and the US markets. Yet, joint contagion and structural break tests suggested the occurrence of these phenomena in all considered markets.*

### **Boa Vista**

MLOps – Do Desenvolvimento à Implantação de Modelos.

Desafios do Cadastro Positivo

*Abstract: Os modelos de credit scoring são modelos preditivos que auxiliam na tomada da decisão de aceitação, ou não, de proponentes ao crédito. Desde a década de 40, diversas metodologias foram propostas para a resolução deste problema e, no Brasil, estes modelos eram baseados em informações cadastrais, de consultas e de negativação dos indivíduos. Com a aprovação da lei do Cadastro Positivo, estes modelos passam a ser baseados em todo o histórico de crédito dos proponentes, possibilitando a criação de modelos mais assertivos. No entanto, novos desafios são enfrentados na criação dos modelos como, por exemplo, a manipulação de bases de dados extremamente grandes que passam de 16.000.000.000 de linhas. Abordaremos os desafios práticos do dia a dia da equipe de analytics da Boa Vista desde o desenvolvimento à implantação de modelos de crédito neste universo big data do Cadastro Positivo.*

**Caio Lucidius Naberezny Azevedo, Jean-Paul Fox, Dalton Francisco de Andrade (State University of Campinas)**

Time series and multilevel modeling for longitudinal item response theory data

*Abstract: Longitudinal Item Response Theory (IRT) data occurs when experimental units are submitted to measurement instruments (e.g., cognitive test, psychiatric questionnaires, biological essays among others) along different assessment conditions, as different time points. Very often, in this kind of study, we are interested in the so-called latent variables (or latent traits) and their behavior along these conditions, including the modeling of their inter-dependency structure. In this work we use some stationary and nonstationary time series and multilevel models to represent longitudinal IRT data. More specifically, we consider first order autoregressive (AR(1)), first order moving average (MA(1)), first order autoregressive moving average (ARMA(1,1)), antedependence (AD) time series models as well as the Uniform and Hankel dependency structures, induced by appropriate multilevel models. These structures are studied under a time-homocedastic and time-heteroscedastic fashions. We developed a Bayesian inference framework, which includes parameter estimation, model fit assessment and model comparison, through MCMC algorithms. Simulation studies are conducted in order to measure the parameter recovery and model comparison tools. A real data analysis, concerning a longitudinal cognitive study of Mathematics achievement, conducted by the Federal Brazilian government, is performed. All computational implementations are made through the WinBUGS program, using the R2WinBUGS package, from R program.*

**Rodrigo Malavazi Corder, Gilberto A. Paula, Anaclara Pincelli and Marcelo U. Ferreira (Biomedical Science Institute - University of São Paulo)**

Statistical modeling of surveillance data to identify correlates of urban malaria risk: a population-based study in the Amazon Basin

*Abstract: Despite the recent malaria burden reduction in the Americas, focal transmission persists across the Amazon Basin. Timely analysis of surveillance data is crucial to characterize high-risk individuals and households for better targeting of regional elimination efforts. Here we analyzed 5,480 records of laboratory-confirmed clinical malaria episodes combined with demographic and socioeconomic information to identify risk factors for elevated malaria incidence in Mâncio Lima, the main urban transmission hotspot of Brazil. Overdispersed malaria count data clustered into households were fitted with random-effects zero-inflated negative binomial regression models. Random-effect predictors were used to characterize the spatial heterogeneity in malaria risk at the household level. Adult males were identified as the population stratum at greatest risk, likely due to increased occupational exposure away of the town. However, poor housing and residence in the less urbanized periphery of the*

town were also found to be key predictors of malaria risk, consistent with a substantial local transmission. Two thirds of the 8,878 urban residents remained uninfected after 23,975 person-years of follow-up. Importantly, we estimated that nearly 14% of them, mostly children and older adults living in the central urban hub, were free of malaria risk, being either unexposed, naturally unsusceptible, or immune to infection. We conclude that statistical modeling of routinely collected, but often neglected, malaria surveillance data can be explored to characterize drivers of transmission heterogeneity at the community level and provide evidence for the rational deployment of control interventions.

**Ronaldo Dias, Julian Collazos and Adriano Z. Zambom (UNICAMP)**  
Selection of the Number of Clusters in Functional Data Analysis

*Abstract: Identifying the number  $K$  of clusters in a dataset is one of the most difficult problems in clustering analysis. A choice of  $K$  that correctly characterizes the features of the data is essential for building meaningful clusters. In this paper we tackle the problem of estimating the number of clusters in functional data analysis by proposing a new measure that can be used with different procedures in selecting the optimal  $K$ . A combination of two test statistics is used, which measure the lack of parallelism and the mean distance between curves. Simulations demonstrate the efficacy of the proposed approach in comparison to existing methods in the literature and results suggest that the proposed method yields stable results and agreement of the procedures considered. The application of the proposed method with several real datasets is investigated.*

**Uriel Moreira Silva, Felipe Carvalho Álvares da Silva, Luiz Henrique Duczmal, Denise Bulgarelli Duczmal (Universidade Federal de Minas Gerais - UFMG)**

Optimized particle-based algorithms for state filtering and parameter learning

*Abstract: In this work we present optimal implementation guidelines for sequential Monte Carlo methods in state-space models with the presence of unknown static parameters. We study in detail typical examples of degeneration and propose computationally efficient solutions capable of improving the performance of the filtering devices. First, an alternative interpretation of the artificial rejuvenation concept explored by most traditional approaches is provided. We then explore the concept of fertility factors in this context and introduce dynamic renewal and regularized resampling schemes, which reduce the selective pressure on the particle set and also the number of resampling steps necessary in the filtering algorithm (thus reducing the set's*

*susceptibility to degeneracy and increasing efficiency). Finally, we provide empirical and simulation-based evidence that simple particle approximations built upon these concepts can satisfactorily deal with complex parameter estimation problems such as relative abundance ecological models.*

## Poster Session 1

**Adalto Speroto, Pablo Martin Rodriguez (PIPGES, ICMC-USP/UFSCar, Brazil)**

Aplicações do Problema do Colecionador de Cupons ao Modelo de Maki-Thompson

*Abstract: Neste trabalho, iremos estudar propriedades de um variável que surge do problema do colecionador de cupons e que é de fundamental importância para compreendermos o comportamento do modelo de Maki-Thompon em árvores homogêneas.*

**Adalto Speroto, Pablo Martin Rodriguez (PIPGES, ICMC-USP/UFSCar, Brazil)**

Conexões entre problemas de cupons e rumores estocástico

*Abstract: Este trabalho tem como objetivo estudar propriedades das variáveis aleatórias que envolvem o problema do colecionador de cupons. Além disso, evidenciamos aplicações dessas variáveis para compreender problemas de propagação de rumores estocásticos.*

**Alaor Cervati Neto (Universidade Federal de São Carlos, Brazil)**

Exploring information theory-based measures for non-linear dimensionality reduction in manifold learning

*Abstract: The processing of large amounts of data to extract useful information is one of the main issues that may be approached using machine learning. One way to obtain this information is by grouping data according to their common features. In very complex data sets, this task may be accomplished by finding simpler ways of representing the relations between this data, called manifold. There are many methods to find the data groups in a set automatically. However, finding a way to minimize the complexity of this data without losing relevant content is a computationally costly process. An alternative to that is treating these data sets as probability distributions of random variables and using concepts and measures from information theory to find their relations more efficiently. This work describes some manifold learning methods and information theory measures and proposes to join them in order to attain better results.*

**Alex de la Cruz Huayanay, Jorge L. Bazán (USP/UFSCAR, Brazil)**

## Longitudinal data analysis for binary response using alternative links

*Abstract: In this work we presented a broad class of asymmetric link functions, called power and reciprocal power, as an alternative to common regression analysis models for longitudinal data with binary response variable. Based on the results of a simulation study, the performance of the models with asymmetric bonding was shown. Parameter estimation was performed under Bayesian approach considering Markov Chain Monte Carlo (MCMC) methods through the No-U-Turn Sampler algorithm. Model comparisons were made using different criterias. Finally, an application of researched methodology was presented using a real data set to show the usefulness of the asymmetric link functions. In this application, the best model among several proposed models were identified.*

**Alex Leal Mota, Pedro L. Ramos, Paulo H. Ferreira, Vera L.D. Tomazella, Francisco Louzada (UFSCar-USP, Brazil)**

## Modeling water flow using the inverse generalized gamma distribution

*Abstract: The control of the water flow is often essential for safety reasons. Water flow rates either above or below the desired limit can affect both the population and the ecosystems. Thus, estimating well, the flow of water is important for the definition of environmental planning and public policies. In this work, we apply the inverse generalized gamma (IGG) distribution to five real data sets related to the average flows of water (in  $m^3/s$ ), from July to November, in the Piracicaba River, Brazil. For that to be possible, a classical inference via maximum likelihood estimators and a Bayesian inference considering objective priors for the IGG distribution parameters, are presented. Furthermore, extensive numerical simulations are conducted to investigate and compare the performances of the proposed estimation methods.*

**Amanda Morales Eudes D'Andrea, Vera Lucia Damasceno Tomazella (UFSCar/USP, Brazil)**

## Multiple repairable systems with a frailty term

*Abstract: In practice, many repairable systems, although apparently identical, show great heterogeneity among themselves and, often, this heterogeneity cannot be described by observed covariates. In the survival analysis literature, this unobservable heterogeneity is called frailty, and in survival models, frailty is modeled through multiplication (called the multiplicative frailty model) or by the addition (called the additive frailty model) of the function of risk by a positive random variable with mean one and, therefore, heterogeneity is described by its variance. In this paper we analyze data with multiple repairable systems assuming that there are unobservable factors that cause*

*heterogeneity between the systems. In modeling we use a random term (frailty variable) in the intensity function in a multiplicative way and we consider the gamma distribution for the frailty variable.*

**Ana Carolina do Couto Andrade, Gustavo Henrique de Araujo Pereira (Ufscar - Usp, Brazil)**

Extensão do resíduo quantílico em modelos de regressão von Mises

*Abstract: A regressão von Mises é frequentemente utilizada no âmbito dos dados circulares. Uma etapa importante para a validação do modelo é a análise de diagnóstico, que tem por objetivo identificar outliers, pontos influentes e verificar possíveis afastamentos das suposições feitas para o ajuste do modelo. O primeiro e o terceiro objetivo podem ser atingidos a partir da análise de resíduos. Diante do exposto, este trabalho teve por finalidade propor uma extensão do resíduo quantílico, adequando-o para o modelo de médias von Mises; bem como, estudar o seu comportamento em comparação com outros três resíduos. Usando simulações de Monte Carlo, os resultados sugerem que a distribuição de probabilidade do resíduo quantílico pode ser aproximada pela distribuição normal padrão, considerando um tamanho de amostra pelo menos moderado. A vantagem deste resíduo sobre os demais se dá, principalmente, pela invariabilidade de sua expressão algébrica ao considerarmos outras classes de modelos de regressão para dados circulares.*

**Átila Prates Correia (USP - UFSCar, Brazil)**

A new method for constructing bivariate symmetric copulas based on straight line homotopies

*Abstract: In this presentation, it is going to be exposed a new method for constructing families of copulas based on pre-existing ones. Although it has been demonstrated its validity to particular cases (such as AMH and FGM copulas), it still remains to study which additional restrictions over the argument copulas that make such methodology applicable. Once such conditions are determined, extend these results to higher dimensional copulas represents a natural approach to be considered.*

**Breno Gabriel da Silva, Terezinha Aparecida Guedes, Yana Miranda Borges, Aline Maria Orbolato Gonçalves-Zuliane (Universidade Estadual de Maringá, Brazil)**

Análise da Incidência no Câncer Cítrico via Distribuição Beta Inflacionada de Zeros

*Abstract: O porta-enxerto pode influenciar a tolerância da variedade copa a fitopatógenos, geralmente porta-enxertos que induzem menor vigor a copa, demonstram menores taxas de infecção do patógeno, comparado aos mais vigorosos. O objetivo foi avaliar a influência exercida pelos porta-enxertos limão cravo (*Citrus limonia*), tangerina sunki (*Citrus sunki*), tangerina cleópatra (*Citrus reshni*) e laranja caipira (*C. sinensis*) na resistência ao cancro cítrico, em nove genótipos de laranja doce (*Citrus sinensis*) variedade Pêra. Considerando que há situações que o modelo de regressão linear não pode ser utilizado devido à característica de uma determinada variável, torna-se necessário a utilização de técnicas apropriadas, ainda mais quando estamos lidando com taxas, frações e proporções. Diante disso, neste trabalho será utilizado o modelo de regressão Beta inflacionado, pois a variável resposta, incidência de cancro cítrico assume valores no intervalo  $[0, 1)$ .*

**Breno Gabriel da Silva, Yana Miranda Borges, Sandra Milena Diaz Puentes, Naiara Caroline Aparecido dos Santos, Talita Evelin Nabarrete Tristão de Moraes (Universidade Estadual de Maringá, Brazil)**

Regressão Multinomial Aplicada em um Estudo Referente ao Grau de Motilidade Espermática de Zangões

*Abstract: Neste estudo, tem-se que a variável resposta  $Y$  (grau de motilidade) é politômica e apresenta cinco categorias (Motilidade-1, Motilidade-2, Motilidade-3, Motilidade-4, Motilidade-5) referentes à escala de movimentação de espermatozoides de zangões. Tendo em vista a natureza da variável resposta, será ajustado o modelo logitos categoria de referência para investigar se o grau de motilidade dos espermatozoides estaria associado com a genética e o efeito de bloco. Assim, considerando o grau de Motilidade-5 como categoria de referência, os quatro logitos, para um dado vetor  $x = (x_1, x_2)$  das covariáveis  $X_1$  (Genética (1 - Um alelo de baixa produção e 2 - Dois alelos de alta produção)) e  $X_2$  (Bloco). Em relação à análise diagnóstica do modelo utilizou-se o gráfico dos resíduos de Pearson. Os dados foram analisados no software R versão 3.6.0.*

**Camila Bertini Martins, Adriano Polpo e Carlos A. B. Pereira (Universidade Federal de São Paulo, Brazil)**

Bayesian Meta-analytic Measure for Proportion

*Abstract: Meta-analysis is a method that summarizes results from different studies of the same subject. The Bayesian meta-analytic measure is based on a combination of the posterior density functions obtained in each of the studies. The measure preserves both the heterogeneity between and within the studies, and it is assumed that the whole data from each study are available. In this paper we present the Bayesian meta-analytic measure for proportion.*

**Carlos Aparecido dos Santos, Victor Battalini de Deus da Chagas, Larissa Reis dos Santos e Daniele Cristina Tita Granzotto (Universidade Estadual de Maringá, Brazil)**

Modelos de cópulas bivariados: uma aplicação ao estudo do tempo de falha em esteiras transportadoras de grãos com abordagem Bayesiana

*Abstract: Dados de sobrevivência bivariados e longitudinais são comuns em muitos estudos em diversas áreas, como medicina, biológicas ou engenharias. Usualmente, temos medidas repetidas para o mesmo indivíduo ou unidade, caracterizando uma possível associação entre essas medidas. Quando essas medidas apresentam alguma estrutura de associação entre elas, uma alternativa para capturá-la é através das funções cópulas ou modelos exponenciais bivariados generalizados. Propomos estudar algumas diferentes funções cópulas, como alternativa para derivar distribuições dos tempos de sobrevivência multivariados. Visando comparar as cópulas obtidas, propomos o estudo de alguns modelos exponenciais generalizados. Os resultados teóricos foram estudados numericamente, a partir de uma aplicação real e via estudo de simulação. Para a aplicação consideramos os tempos de parada de grandes transportadoras de grãos, devido à quebra ou desgaste de pequenas partes, tal como os mancais que sustentam estas plataformas.*

**Carlos Franklin Taco Pedraza, Dr. Vicente Garibay Cancho (UFSCar-USP, Brazil)**

Modelo de Regressão Potência-Normal Logística para Resposta no Intervalo Unitário

*Abstract: Neste trabalho é proposto uma nova classe de modelos estatístico para modelar dados limitados no intervalo contínuo  $(0,1)$ . O novo modelo é a composição da distribuição potência-normal e o quantil de outra família de distribuições com o mesmo suporte. A nova proposta é uma generalização do modelo proposto por Johnson. Além disso, a nova família é estendida para modelos de regressão como uma*

*alternativa aos modelos de regressão com resposta no intervalo unitário. Também para a classe de modelos proposto, desenvolvemos procedimentos inferenciais sob a perspectiva clássica, baseada na teoria de máxima verossimilhança.*

**Caroline Amantéa Stella, Adriano Kamimura Suzuki (USP/UFSCar, Brazil)**  
Modelo Logarítmico Inversa Gaussiana com Fração de Cura

*Abstract: Modelos de análise de sobrevivência com proporção de cura têm sido muito estudados. Atualmente uma ramificação desses modelos considera que os indivíduos da população tem heterogeneidade não observada. Para a consideração da heterogeneidade não observada inclui-se uma variável aleatória conhecida como fragilidade (efeito aleatório) sendo introduzida na função de risco. Nestes modelos que consideram a heterogeneidade não observada e indivíduos que nunca atingem a falha, a fragilidade segue uma distribuição discreta começando em zero. O modelo proposto considera a fragilidade seguindo distribuição Logarítmica e a distribuição basal seguindo a distribuição Inversa Gaussiana. A estimação dos parâmetros foi feita por método de máxima verossimilhança. As aplicações a dados artificiais e reais serão apresentadas.*

**Christian Eduardo Galarza Morales, Larissa Ávila Matos, Víctor Hugo Lachos (Universidade Estadual de Campinas, Brazil)**

Momentos das distribuições de seleção elíptica multivariadas duplamente truncadas com ênfase em modelos censurados

*Abstract: Neste trabalho, nós momentos duplamente truncados para uma classe geral de distribuições assimétricas chamada a família de seleção elíptica multivariada. Nesta grande família de distribuições pertencem versões complexas assimétricas multivariadas de distribuições elípticas conhecidas como as distribuições normal,  $t$  de Student, exponencial potência, hiperbólica, Slash, Pearson tipo II, normal contaminada, entre outras.*

*Num ponto de vista frequentista, momentos duplamente truncados tem muita importância no contexto de modelos censurados intervalares, quando consideramos a resposta  $Y_i$ ,  $i = 1, \dots, n$ , sendo uma amostra i.i.d. desta família, por exemplo, seguindo uma distribuição Normal ou  $t$  assimétrica.*

*Dada a flexibilidade desta família, nos podemos modelar características como assimetria, caudas pesadas e multimodalidade, enquanto a censura intervalar permite considerar adicionalmente às censuras intervalares, missing values e censuras à esquerda para respostas estritamente positivas ao considerar intervalos da forma  $(-\infty, \infty)$  e  $(0, c)$  respectivamente. Os resultados gerais para a família de seleção*

*elíptica, envolvem todos os momentos anteriormente usados na literatura de modelos censurados num ponto de vista frequentista. Os momentos encontram-se disponíveis ao usuário no pacote MomTrunc no R.*

### **Cristel Ecaterin Vera Tapia, Sandro Gallo (UFSCar/USP, Brazil)**

Estimation of the number of communities in a random multigraph model

*Abstract: Cerqueira and Leonardi (2018) introduced the Krichevsky-Trofimov estimator for the number of communities in the classic stochastic block model (falls in the general class of random graphs), where each entry of the adjacency matrix  $X$ , follows a Bernoulli distribution. Similar to Karrer and Newman (2011), we consider a multigraph version of the stochastic block model and replace the Bernoulli distribution by the Poisson distribution, in this way, we present in this work an extension of results in Cerqueira and Leonardi (2018).*

### **Daniel Camilo Fuentes Guzman, Clécio da Silva Ferreira, Camila Borelli Zeller (UFScar-USP, Brazil)**

Modelos de Regressão para Dados Censurados sob a Classe de Distribuições de Misturas de Escala Normal Assimétricas

*Abstract: Um problema frequente na análise de regressão é quando a observação da variável resposta é censurada para alguns indivíduos. Isto ocorre em várias situações práticas, por razões como limitações do equipamento de medição ou do desenho experimental. Estes fenômenos podem ser modelados mediante modelos estatísticos e matemáticos. No âmbito dos modelos de regressão censurados, os erros aleatórios são rotineiramente considerados como tendo uma distribuição normal, principalmente por conveniência matemática. No entanto, este método tem sido criticado na literatura por causa de sua sensibilidade a desvios da suposição de normalidade. Nesse trabalho, primeiro estabelecemos uma nova ponte entre o modelo de regressão censurado e a classe de distribuições assimétricas estudadas por Ferreira et al.. As misturas de escala assimétricas das distribuições normais são frequentemente utilizadas para procedimentos estatísticos que envolvem dados assimétricos e caudas pesadas. A principal virtude dos membros dessa família de distribuições é que eles são fáceis de serem simulados e também fornecem algoritmos tipo Esperança-Maximização (EM) para a estimativa de máxima verosimilhança. Neste trabalho, estendemos o algoritmo EM para o algoritmo MCEM para modelos de regressão lineares censurados. O algoritmo do tipo EM foi discutido com ênfase nas distribuições Normal Assimétrica,  $t$ -Student Assimétrica, Slash Assimétrica e Normal-Contaminada Assimétrica. Os métodos propostos são verificados através da análise de vários estudos de simulação*

*e aplicação em conjuntos de dados reais.*

**Davi Keglevich Neiva, Ricardo Blanco, Felipe Araujo, Mariana Lopes, Felipe Souza (Serasa Experian, Brazil)**

Modelos de Machine Learning, cuidados e boas práticas

*Abstract: Nos últimos anos muito se tem falado sobre novas técnicas de modelagem, principalmente aquelas referentes a machine learning, que vêm apresentando possibilidade de melhor precisão quando comparadas com técnicas mais tradicionais. No entanto, essas técnicas muitas vezes criam modelos de difícil interpretação e que se ajustam bastante aos dados, o que pode dificultar consideravelmente a avaliação da eficiência do modelo na predição de novas observações.*

*Para evitar obter modelos de machine learning que são facilmente descalibrados e que apresentam baixa performance em um médio ou até mesmo no curto prazo é preciso ter atenção redobrada em vários pontos do processo de modelagem. Nesses modelos desvios da amostra e mudanças na população impactam muito mais as predições e avaliar formas de desenvolver modelos mais robustos demanda habilidade técnica, experiência na construção dos modelos e principalmente conhecimento dos dados envolvidos.*

*Neste trabalho, são apresentadas boas práticas utilizadas no desenvolvimento de modelos de Machine Learning aplicados à modelagem de crédito e algumas das análises que devem ser realizadas visando o desenvolvimento de um modelo com alto grau de robustez.*

**Deborah Bassi Stern, Rafael Izbicki (UFSCar/USP, Brazil)**

Vector representation of texts applied to prediction models

*Abstract: Natural Language Processing has gone through substantial changes over time. It was only recently that statistical approaches started receiving attention. The Word2Vec model is one of these. It is a shallow neural network designed to fit vectorial representations of words according to their syntactic and semantic values. The word embeddings acquired by this method are state-of-art. This method has many uses, one of which is the fitting of prediction models based on texts. It is common in the literature for a text to be represented as the mean of its word embeddings. The resulting vector is then used in the predictive model as an explanatory variables. In this dissertation, we propose getting more information of text by adding other summary statistics besides the mean, such as other moments and quantiles. The improvement of the prediction models is studied in real datasets. Keywords: Natural language processing, Neural networks, WordVectors, Prediction models.*

**Diego Carvalho do Nascimento, Bruno Pimentel, Renata Souza, João P. Leite, Dylan J. Edwards, Taiza E.G. Santos and Francisco Louzada (USP/UFSCar)**

Dynamic Time Series Smoothing for Symbolic Interval Data applied to Neuroscience

*Abstract: This work aimed to appraise a multivariate time series, high-dimensionality data-set, presented as intervals using a Symbolic Data Analysis (SDA) approach. SDA reduces data dimensionality, considering the complexity of the model information through a set-valued (interval or multi-valued). Additionally, Dynamic Linear Models (DLM) are distinguished by modeling univariate or multivariate time series in the presence of non-stationarity, structural changes and irregular patterns. We considered neurophysiological (EEG) data associated with experimental manipulation of verticality perception in humans, using transcranial electrical stimulation. The innovation of the present work is centered on use of a dynamic linear model with SDA methodology, and SDA applications for analyzing EEG data.*

**Diogo Barboza Moreira, Luís Ernesto Bueno Salasar (UFSCar, Brazil)**

Uma proposta de teste de hipóteses bayesiano não paramétrico com aplicação em dados funcionais

*Abstract: Este trabalho apresenta uma abordagem bayesiana não paramétrica para testar a igualdade de distribuição de duas amostras independentes de dados funcionais. Dados são considerados de natureza funcional quando a cada unidade experimental é associada a realização de um processo estocástico. Usualmente, os dados funcionais são representados a partir da consideração de uma sistema ortogonal de funções. Neste trabalho, consideramos o sistema ortogonal de Fourier para representação dos dados funcionais e, sob uma perspectiva bayesiana não paramétrica, desenvolvemos um procedimento de teste considerando diferentes noções de distância entre funções. O método proposto é ilustrado em um conjunto de dados reais de temperaturas diárias obtidas ao longo de um ano em diferentes regiões climáticas do Canadá e comparado com diferentes métodos não-paramétricos da literatura a partir de um estudo de simulação para avaliação do poder.*

**Éder Silva de Brito, Vera Lucia Damasceno Tomazella; Daniele da Silva Baratela Martins Neto (USP/UFSCar, Brazil)**

Inferência para distribuição Weibull baseada em Censura Tipo II Progressiva

*Abstract: Many statistical analysis require the ability to test whether a sample comes from a given distribution, a goodness-of-fit test. Although much work exists*

*on performing this test under a frequentist framework, there are almost no attempts for tackling this problem under a Bayesian perspective. The reason is that the null hypothesis is precise (sharp), which created difficulties for most Bayesian methods. In this work we show how the Full Bayesian Significance Test, a framework for testing parametric sharp hypothesis, can be extended to nonparametric models by using pseudo-densities. We compare our method to the classical  $t$ -test, Kolmogorov-Smirnov, Anderson-Darling, Cramer-Von Mises tests in simulated datasets, and show that our method performs better and equally well to some of them.*

**Éder Silva de Brito, Vera Lucia Damasceno Tomazella; Daniele da Silva Baratela Martins Neto (USP/UFSCar, Brazil)**

Inferência para distribuição Weibull baseada em Censura Tipo II Progressiva

*Abstract: Testes de confiabilidade realizados em amostras com censura são recorrentes em diversas áreas, justificados pelo interesse em redução de tempo e custos destas observações. Dentre os esquemas de censura apresentados na literatura, tem-se ampliado estudos acerca de amostras com Censura do Tipo II Progressiva, uma generalização da censura do Tipo II clássica. Neste modelo deseja-se observar certa quantidade de falhas  $e$ , a cada falha ocorrida, outras unidades em funcionamento são retiradas aleatoriamente do experimento (caracterizando o modelo progressivo de censura), até que se obtenha um número pré-determinado de falhas observadas. Neste trabalho apresentamos métodos inferenciais para distribuição Weibull baseados em amostras na presença de censura tipo II progressiva. Neste contexto consideramos dois métodos para obtenção dos estimadores, o de máxima verossimilhança e o algoritmo EM, bem como suas medidas de variação assintótica. Um estudo de simulação foi realizado com o objetivo de verificar a qualidade dos estimadores obtidos pelos métodos apresentados, sob diferentes cenários para os tamanhos de amostras e esquemas de censura. Por fim, a metodologia foi aplicada considerando exemplos com dados reais.*

**Elizbeth Chipa Bedia, Vicente Garibay Cancho; Dipankar Bandyopadhyay (ICMC-USP/DEs-UFSCar, Brazil)**

A frailty model for multistate semi-competing risk data with applications to colon cancer

*Abstract: Semi-competing risks are a generalization of the competing risks scenario, where a subject may experience both terminal and non-terminal events, usually dependent, and the event time to the intermediate non-terminal event (say, tumor recurrence in cancer studies) is subject to dependent censoring by the terminal event*

(say, death), but not vice-versa. As an alternative to the latent failure time formulation of semi-competing risks with joint survival functions here, we consider an illness-death (multistate) shared frailty framework, where the dependency between the terminal and non-terminal failure times is incorporated via the power variance frailty between the conditional transition rates that are assumed Markov. Inference is conducted via maximum likelihood. A simulation study is conducted to evaluate the finite sample performance of the model parameters. Finally, we compare and contrast our power variance frailty proposal to known alternatives by application to a colon cancer dataset.

**Fabiano Rodrigues Coelho, Jorge Luis Bazán e Cibele Maria Russo Novelli (USP e UFSCar, Brazil)**

Bayesian skew-probit regression model with random effects: An application to longitudinal data

*Abstract: Longitudinal Binary response data are commonly analyzed using logistic mixed regression models, however when data are unbalanced this model could be inadequate. Motivated by application data where the response variable is unbalanced, we propose the use of the asymmetric probit link under Bayesian approach. We perform the Bayesian estimation process through the NUTS algorithm. For the application, we use the know Madras schizophrenia data where the binary variable  $y$  indicates the presence or absence of psychiatric symptoms over  $t = 0, \dots, 11$ . months during the first year after hospitalization for schizophrenia for patients  $i = 1, \dots, 86$  and other variables such as age and gender were considered. Results using model comparison methods and residual analysis considering standardized quantile residual to show that the proposed model is more suitable than the traditional model. Additionally, interpretations of the model in the application are provided.*

**Fátima Aparecida Kian (UFABC, Brazil)**

Estudos sobre a Linguagem Estatística/Probabilística no Ensino Fundamental

*Abstract: Neste estudo nos concentramos em descrever e analisar como surgem elementos linguísticos no processo de ensino e aprendizagem de estatística e probabilística no ensino fundamental, entendida como uma linguagem precisa e especializada. Para isso, escolheu-se realizar um estudo exploratório de uma de processo de instrução com alunos do quarto ano do ensino fundamental (cerca de 10 anos) que tenham recebido algum tipo de instrução prévia sobre o assunto. Especificamente, a multiplicidade de termos expressões orais e escritas, símbolos e representações utilizadas com a finalidade de que os alunos aprendam o conceito de forma gradual e*

*adquiram os respectivos conceitos básicos em estatística que coadune com nossa fundamentação teórica. Espera-se que os resultados atualmente parciais possam mostrar palavras e expressões verbais da língua comum relacionado principalmente ao significado intuitivo de probabilidade e estatística.*

*Palavras-chave: matemática estatística, linguagem*

**Felipe Hernandez Bisca, Rafael Izbicki, Murilo Cantoni (UFSCar, Brazil)**  
Multivariate conditional probability density estimation with copulas

*Abstract: Most machine learning regression models only yield single point estimations for the label of a new observation. However, when dealing with multi-modal or asymmetric distributions, a single point estimate is not enough to summarize the full uncertainty over such label. One solution for this case is to estimate the full conditional density function of the label given the features, which is more informative. For instance, this density can be used to compute probability regions rather than single point estimates. Conditional densities become especially useful when modelling multivariate responses, which is often the case in fields such as cosmology. Most well known conditional density estimators are too slow to be computed or do not generalize to multivariate-response settings. To minimize such problems, our method estimates multivariate densities using copula to aggregate estimates of univariate conditional densities given by the recent-developed FlexCode. We show that this solution leads to improved results when compared to other state-of-the-art techniques.*

**Francisco José dos Santos, Olímpio José dos Santos, Jorge Henrique e Silva Júnior, William Figueredo Cruz, Claudimeire Lima de Sousa (IFPI, Brazil)**

Análise da disponibilidade dos recursos hídricos da Cidade de Picos-Piauí de 2017 a 2040

*Abstract: As águas subterrâneas representam maior fonte para abastecimento de algumas regiões do Brasil. A presente pesquisa foi desenvolvida no município de Picos que está localizado na região Sudeste do Piauí. Inicialmente foi feito uma estimativa matemática com base nos dados dos dois últimos censos do Instituto Brasileiro de Geografia e Estatística. Os dados foram tabulados com a ajuda de algumas equações para chegar aos resultados. Foi feito a projeção populacional e uma avaliação de disponibilidade hídrica do aquífero Serra Grande da zona urbana do município, baseados em dados existentes fornecidos por Vidal (2003) e pelos os últimos censos 2000 e 2010. Analisou-se o consumo diário per capita de água, a projeção da população até 2040, o consumo diário e o consumo anual de 2017 a 2040. Neste sentido a recarga anual*

é de  $1,84 \times 10^8$ , e com base nesses dados estima-se que a disponibilidade de água no aquífero serra grande para a população urbana de Picos em 2040 é de aproximadamente 3.244.000.000 m<sup>3</sup>, e o consumo de 2017 a 2040 pode chegar a 110.519.863,56 m<sup>3</sup>. Analisando a discrepância da disponibilidade de água em 2040 com o consumo anual dos próximos 23 anos, chega a um resultado significativo em relação à população existente o que compromete o manancial do aquífero podendo não suportar a demanda da população. Desse modo importante que a população tenha consciência da possível escassez de água no futuro e uma alternativa é que sejam implantadas políticas de conscientização específicas para tentar minimizar o futuro problema. Palavras-chave: águas subterrâneas; aquífero; população.

**Gabriela Maria Leme Trivellato, Gabriel Adrián Sarriés, Gustavo Nazato Furlan (ESALQ/CENA/USP, Brazil)**

Machine Learning para análise do banco de dados do "Sistema de avaliação ponderada da multifuncionalidade da agricultura (MFA)"

*Abstract: Este resumo propõe-se a apresentar o emprego de Machine Learning para análise do banco de dados da pesquisa de Mestrado "Sistema de avaliação ponderada da multifuncionalidade da agricultura (MFA)", em andamento no Programa de Pós-Graduação Interunidades em Ecologia Aplicada, da ESALQ/CENA/USP. Esta pesquisa centra-se no desenvolvimento de um índice/sistema de avaliação da MFA, inspirado principalmente em três índices/indicadores de sustentabilidade ambiental, da tendência "Tableaux De Bord" Via Indicadores: 1. o APOIA - Novo Rural (Avaliação Ponderada de Impacto Ambiental de Atividades do Novo Rural); 2. o Sistema de Avaliação de Impactos de Inovações Tecnológicas Agropecuárias (Ambitec-Agro); 3. o método francês IDEA (Indicateurs de Durabilité des Exploitations Agricoles). O índice criado funda-se: em matrizes de ponderação; na Escala de Likert; nos critérios da Matriz de Eisenhower e; no princípio da média ponderada. Possui quatro dimensões, cada uma composta por cinco indicadores e; cada indicador, em média, por cinco subindicadores. O índice foi testado para os 26 estados brasileiros e o Distrito Federal (DF), utilizando os resultados preliminares do Censo Agropecuário 2017, do IBGE. Foi construído um banco de dados para ser analisado por Machine Learning, nos softwares JMP, Weka e SAS. Esta pesquisa encontra-se na fase de análise de dados. Até o momento, uma PCA (Principal Components Analysis) com Biplot realizada no software JMP, revelou: os estados e DF diferenciam-se em seus desempenhos nas quatro dimensões consideradas (D1, D2, D3, D4) e valor final do índice (Ind). Uma Discriminant Analysis linear (no software JMP) revelou que as regiões brasileiras caracterizam-se por desempenhos distintos nas quatro dimensões. As referidas dimensões tratam-se das quatro principais funções atribuídas à MFA na realidade rural brasileira, definidas por Maria José Carneiro e Renato Maluf: a) produção socioeconômica das famílias rurais; b) promoção da segurança alimentar das*

*próprias famílias rurais e da sociedade; c) manutenção do tecido social e cultural; d) preservação dos recursos naturais e da paisagem rural. Nosso objetivo final é verificar a pertinência do índice desenvolvido por meio da identificação de suas fragilidades. O uso de Machine Learning para análise do banco de dados da pesquisa tem-se mostrado útil, tendo revelado a aplicabilidade do índice para diferenciar o desempenho de estados e regiões em termos de MFA. 1 M. J. Carneiro, R. S. Maluf. Para além da produção: multifuncionalidade e agricultura familiar. Mauad Editora Ltda, Rio de Janeiro, 2003.*

### **Gabriela Massoni, Rafael Bassi Stern (Universidade Federal de São Carlos, Brazil)**

Análise de textos por meio de processos estocásticos na representação word2vec

*Abstract: A problemática da análise de dados textuais passa pelo desafio de transformar uma observação do seu conjunto de dados, que está no formato de texto, em valores numéricos. Em geral, a metodologia escolhida para executar tarefas deste tipo é a bag-of-words (Matsubara et al., 2003). Trata-se essencialmente de codificar o aparecimento de todas as palavras em cada observação. Ou seja, esse método utiliza como covariáveis binárias que indicam se a palavra aparece no texto ou não. Existem variações de como utilizar o bag-of-words de forma inteligente para resolver alguns problemas. Entretanto, esse método possui limitações, como não conseguir capturar relações semânticas entre as palavras ou ainda não capturar a inversão de sentido causado por palavras como não.*

*A metodologia word2vec (Goldberg and Levy, 2014) é um procedimento baseado em redes neurais artificiais, que surgiu como alternativa para a captura de relações semânticas entre as palavras. Ela não considera apenas as palavras ( $w$ ) de uma sentença, mas também o seu contexto ( $c$ ) de ordem  $L$ , ou seja, as  $2L$  palavras mais próximas à  $w$ . Nesta abordagem, cada palavra é representada por um vetor numérico de ordem  $d$ . Na modelagem Skip-gram (SG), utiliza-se como entrada de uma rede neural as palavras da base de dados, para prever os contextos (saída). Como o word2vec utiliza redes neurais, o poder preditivo é substancialmente aumentado quando comparado ao bag-of-words.*

*Ao aplicar o modelo word2vec, obtém-se uma matriz de valores. Desta forma, para o ajuste de modelos estatísticos, faz-se necessário transformar os valores desta matriz em valores para cada observação do banco. Pode-se pensar em métricas a partir dos valores para alcançar este objetivo. Usualmente, o que se faz é tomar uma média dos valores dos vetores de cada palavra que está presente na observação. Contudo, tomar a média descarta informação sobre o texto, como a ordem em que as palavras aparecem. Uma alternativa é considerar que o texto é um processo estocástico em que, dado cada palavra do texto, temos uma probabilidade de que a próxima seja outra palavra. Assim é possível capturar uma dependência temporal no texto, melhorando*

as predições do modelo.

**Gilberto Pereira de Alcântara Junior, Gustavo Henrique de Araujo Pereira (PIPGES, Brazil)**

Performance preditiva do lasso, stepwise e máxima verossimilhança em modelos de regressão logística

*Abstract: Duas etapas fundamentais na construção de um modelo de regressão são a estimação dos parâmetros e a seleção de covariáveis. A estimação dos parâmetros por máxima verossimilhança combinada com o método de seleção de variáveis por stepwise é muito utilizada. Entretanto o stepwise possui um custo computacional muito grande e em alta dimensionalidade o mesmo pode se tornar inviável de ser utilizado. O uso do método de penalização lasso pode ser uma solução em alta dimensionalidade e o mesmo é capaz de selecionar covariáveis e estimar os parâmetros do modelo. Como a estimação por máxima verossimilhança não tem alto custo computacional, surge o questionamento do desempenho conjunto do método lasso para selecionar covariáveis e o método de máxima verossimilhança para estimação dos parâmetros. O modelo de regressão logística é muito utilizado em diversas áreas porque é comum termos variáveis resposta binárias e por possuir parâmetros de fácil interpretação. O presente trabalho tem como objetivo estudar o desempenho preditivo da combinação dos métodos de estimação via máxima verossimilhança e seleção via lasso, contra lasso, stepwise com máxima verossimilhança e lasso com stepwise no modelo de regressão logística. Estudos de simulação foram realizados sob diferentes cenários e seis bases de dados foram utilizadas para aplicação.*

**Giovanni Moura de Holanda, Jorge Moreira de Souza (FITec - Inovações Tecnológicas, Brazil)**

Capability based control charts: an experimental approach to manufacturing process

*Abstract: Control charts are widely used to evaluate the control acceptance limits to monitor the variability throughout the industrial manufacturing process to minimize the production of nonconforming items. Acceptance limits are very dependent on the estimated standard deviation that quantifies the variability around the mean.*

*Before starting production, an important activity is to measure and calibrate each item of the process chain. Each item has a set of parameters whose calibration is based on a subgroup of observations aiming at reducing the item variability. The size of the subgroups is generally small in order to optimize cost and duration of the test. The short-term standard deviation is a measure of the variability within subgroups.*

*In a lot production many factors can influence the variability as raw material,*

item fatigue, etc. This is detected by the long-term standard deviation that measure the variability of the samples produced.

Process capability indices quantify the short-term,  $C_p$ , and long-term,  $P_p$ , variations; however, they are not a direct quantification of the percentage of nonconforming items. This is expressed generally by the ppm, parts per million, of the items produced that are outside the specification. For example, a target  $C_p > 3$  implies a severe control of identifying and correcting special causes (out-of-control points) that will lead to a 16 ppm and, consequently, more costly control activities. However, if a limit of 160 ppm is accepted, the control activities are less costly (less out-of-control points) and lower specified,  $C_p > 1.5$ . In other words, the cost of identifying and correcting special causes should be compatible with the cost of off-target products.

The present study addresses the use of modified control chart to derive the ppm and the acceptance limits given the target  $C_p$ , the probability of reject a good item, the corresponding upper and lower specification limits. This approach is being carried out to be specifically applied for monitoring and controlling the equipment set-up of a semiconductor manufacturing company.

## **Giovanni Pastori Piccirilli, Jorge Luis Bazán Guzmán (PIPGEs UFSCar-USP, Brazil)**

Alternatives mixed regression models for bounded response variables

*Abstract: Regression models for bounded response variables play a important role in many application that the response are percentages, proportions or rates. The main models aims model the relationship between one or more covariates with the conditional mean of the response variable and are based in the beta distribution. The present work aimed to present two quantile mixed regression models for limited responses based on the Logistic-Gumbel distribution and the Johnson SB distribution for limited responses. The mixed models allows a flexibility for model in situations that exists a within-subject correlation. For classical Inference it was used the RS algorithm to obtain the estimates of penalized maximum likelihood. The bayesian approach was also adopted using MCMC methods using NUTS algorithm, an extension of HMC algorithm. Model selection criteria were also presented. The results of an analysis of data regarding the percentage of votes of political Party in the state of Sergipe over time are presented. It has been shown that the quantile models presented here are an alternative to the Beta mixed regression model.*

**Gustavo Alexis Sabillón Lee, Daiane A. Zuanetti (Universidade de São Paulo / Universidade Federal de São Carlos, Brazil)**

Estimação para Modelos Markovianos Ocultos Não-Homogêneos

*Abstract: Os modelos Markovianos ocultos são paradigmas estatísticos que podem ser utilizados para modelar processos estocásticos onde valores observáveis dependem diretamente de uma sequência de variáveis aleatórias não observáveis. No modelo Markoviano oculto o sistema que está sendo modelado é considerado um processo de Markov com estados não observáveis (isto é, ocultos) e em cada estado oculto temos a emissão de um valor observável. Os modelos Markovianos ocultos podem ser homogêneos ou não homogêneos. O foco principal deste trabalho, serão os modelos Markovianos não-homogêneos. Neste trabalho, apresentaremos alguns procedimentos de estimação utilizados com modelos Markovianos. Faremos uma comparação entre estes procedimentos utilizando o Bayes Information Criteria para os procedimentos frequentistas que serão estudados, e o Deviance Information Criteria para o método Bayesiano que será estudado. A estimação dos parâmetros é realizada sob abordagem Bayesiana e clássica, fazendo uma comparação da performance dos mesmos usando alguns indicadores como ser o erro quadrático médio e o viés. O método de estimação Monte Carlo em cadeias de Markov especificamente Metropolis-within-Gibbs e as comparações dos modelos são desenvolvidas utilizando diferentes critérios para comparação de modelos.*

## Poster Session 2

**Isaac Esteban Cortés Olmos, Mário de Castro Andrade Filho (Universidade de São Paulo, Brazil)**

An application of nonparametric mixed models to longitudinal data under elliptical distributions

*Abstract: In recent decades, there has been a strong demand in the development of nonparametric models because they can account for flexible functional forms from the data to capture possibly complicated relationships between longitudinal outcomes and covariates. The objective of this work is to present an illustrative example of the process of estimation and inference of nonparametric mixed models under elliptical errors for a longitudinal data structure. To carry out this, penalized likelihood equations are applied to allow deriving maximum penalized likelihood estimates which are robust against outliers in the sense of Mahalanobis distance.*

**Isis Fernanda Mascarin, Katiane Silva Conceição (USP/UFSCar, Brazil)**  
Distribuições discretas zero-modificadas para modelar dados de contagem zeros faltantes

*Abstract: A análise de dados de contagem ocupa um importante lugar dentro da estatística aplicada, uma vez que muitos problemas reais são expressos em termos de enumerações. Frequentemente, conjuntos de dados de contagem apresentam discrepâncias na frequência da observação zero, que pode ser alta ou baixa, e assim refere-se ao conjunto de dados como zero-inflacionado ou zero-deflacionado, respectivamente. Além disso, existem situações onde a observação zero não ocorre nos conjuntos de dados e, muitas vezes, modelos zero-truncados são inadequadamente considerados, visto que há uma probabilidade positiva (e não nula) para ocorrência de tal evento, embora este não tenha ocorrido. Este trabalho tem como objetivo apresentar o procedimento de estimação dos parâmetros das distribuições zero-modificadas em situações em que a frequência da observação zero nos conjuntos de dados é nula e a probabilidade de ocorrência de tal valor é positiva (zero-deflacionada).*

**Izabella Gonçalves Rendeiro da Silva, Profa. Dra. Samira Maria Leão de Carvalho; Enio Laubyer Ferreira e Silva (Universidade Federal do Pará, Brazil)**

Study of Flocculants Consumption in Bayer Process

*Abstract: In this work, was studied the consumption of flocculant in sedimentary tanks in the Bayer Process in a period of 12 months in relation to the values specified in project. And, through the quality tool SPC (Statistical Process Control) 'Histogram, Control charts and boxplot', using 355 data, was determined the second and third quarter, in addition to the last month of the year, as an abrupt period of variation. To evaluate the deviation of consumption by sedimentation tank, around 1,532 data were used in the construction of the Waterfall Chart where distributions of flocculants flows that are fed in sedimentation tanks have been shown. Subsequently, using 18,400 data, graphical analysis of the behavior parameters of the tanks was performed which presented the highest values of flocculant consumption in the period studied. According to the analyses carried out in the 2nd and 3rd quarter of the year of the study, in 4 sedimentation tanks, Tanks 1 and 2 presented consumption values below the specified value, while Tanks 3 and 4 presented high consumption in the same period. When evaluating flocculant consumption from the 2nd for the 3rd quarter of Tanks 3 and 4, a reduction of about 51% was observed in the period, attributed to the operation output of Tank 3. Therefore, the increase in flocculant consumption in the study period is due to tank 3 operation.*

**Jacques Henrique Bessa Araújo, Paulo César Cortez (Federal University of Ceará, Brazil)**

Exploratory Analysis of Deep Neural Networks Applied to Speech Enhancement

*Abstract: Automatic speech recognition consists of converting speech in a word sequence through algorithms in embedded systems. The accuracy has a high correlation with signal processing algorithms to improve quality and intelligibility. A manner to reduce the word error rate is through noise attenuation algorithms. In this paper, we explore alternative setups to predict clean speech from noisy speech through nonlinear regression using deep neural networks. We established a multi-condition training and testing database, with three noise types and four main values of signal-to-noise ratios (SNRs). Furthermore, we also trained and tested signals with SNR values generated randomly in a range between 0 to 15 dB to increase the generalization capacity of the network. We use three objective metrics for performance evaluation, and we compare the results with classical algorithms such as shallow neural network, Wiener filtering, and spectral subtraction. The results show that the deep neural network (DNN) method outperforms the classical algorithms in all scenarios of SNR values.*

**Jessica Suzana Barragan Alves, Breno Gabriel da Silva, Naiara Caroline Aparecido dos Santos, Patrícia Stülp, Talita Evelin Nabarrete Tristão de Moraes, Yana Miranda Borges (USP/UFSCAR, Brazil)**

Avaliação da atividade antimicrobiana do extrato hidroalcoólico de *Baccharis dracunculifolia* liofilizado contra as cepas de bactérias Gram-negativas

*Abstract: Nos dias atuais, várias pesquisas têm seus estudos voltados para obtenção de compostos que possam substituir os compostos sintéticos e antibióticos. Uma alternativa para o combate dos microrganismos são os extratos vegetais com ação antimicrobiana, devido à resistência destes a múltiplas drogas. Dessa maneira tem-se realizado buscas contínuas de novos produtos com propriedades antimicrobianas eficazes e econômicos para combater a resistência desses microrganismos patogênicos. Ao se pensar em compostos naturais, surge o interesse em verificar a ação antimicrobiana da planta brasileira do grupo Asteraceae, *Baccharis dracunculifolia*, a qual é rica em compostos antioxidantes. Com isto, o presente trabalho tem como objetivo avaliar a atividade antimicrobiana in vitro do extrato hidroalcoólico de *Baccharis dracunculifolia* liofilizado contra as cepas de seis bactérias Gram-negativas e verificar se há diferença entre os níveis de concentrações inibitórias mínima (MIC) do extrato utilizado, apropriando-se de métodos não paramétricos. Os microrganismos testados nos ensaios incluem as bactérias Gram negativas: *Escherichia coli*, *Salmonella enteritides*, *Pseudomonas aeruginosa*, *Aeromonas hydrophila*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*.*

**Jonathan Kevin Jordan Vasquez, Josemar Rodrigues (USP-UFSCAR, Brazil)**

Aplicação da distribuição Waring Generalizada para investigar a superdispersão dos fatores de risco em Análise de Sobrevivência de longa duração em dois estágios

*Abstract: No modelo de longa duração em dois estágios o tempo de vida de um paciente depende do número de fatores de risco responsáveis pela sua sobrevivência ou cura. O número de fatores de risco é um efeito aleatório latente que expressa o comportamento heterogêneo dos pacientes em relação ao risco básico da população conhecido na literatura recente como fragilidade discreta. Esta fragilidade está diretamente conectada ao fenômeno de superdispersão dos fatores de risco e o mecanismo destrutivo ou de reparo dos fatores de risco. Várias distribuições discretas com caudas pesadas ("J-shaped") têm sido utilizadas para explicar o excesso de variabilidade mas sem sucesso para separar o efeito individual ( predisposição: "proneness") dos efeitos externos não aleatórios e latentes do paciente ( submissão : "liability"). Os efeitos latentes interferem na fragilidade onde o efeito predisposição ou insegurança do paciente é o mais importante na cura, planejamento preventivo ou tratamento de câncer em medicina de precisão. A distribuição Binomial Negativa (BN) é a mais utilizada mas o seu mecanismo de destruição ou reparo não é flexível o suficiente para permitir*

*uma destruição ou não reparo interno sem ruídos externos. A distribuição Waring Generalizada (WG) é mais apropriada para o modelo de cura em dois estágios devido a existência de um mecanismo de defesa inteligente que depende exclusivamente do paciente sem ruídos externos e uma taxa de cura totalmente personalizada úteis em tratamento de câncer por imunoterapia. A distribuição de WG identifica na superdispersão três fontes de variabilidade que impactam a taxa de cura e a sobrevivência dos pacientes: aleatório, submissão e predisposição. Por simplicidade computacional na solução dos problemas inferenciais somente uma versão especial da distribuição WG (distribuição Waring) será utilizada nas simulações e aplicações via algoritmo EM.*

**Josué dos Santos da Conceição (UNICAMP, Brazil), Dr. Diógenes Ferreira Filho (UFRRJ, Brazil)**

Previsão de desempenho de estudantes usando um modelo de regressão com variáveis Dummy

*Abstract: O presente trabalho procurou investigar a relação entre as notas obtidas por alunos nos períodos 2017-2 e 2017-3 (conhecido como “período extraordinário”), a fim de procurar obter a nota mínima necessária que um aluno deveria obter para se estimar, com probabilidade de erro pequena, que ele seria aprovado no período extraordinário. Usando dados de acesso restrito, cujo uso foi devidamente autorizado pelo órgão competente, foram usadas as notas obtidas pelos alunos nos períodos 2017-2, 2017-3 e o CRA (Coeficiente de Rendimento Acumulado) como variáveis. O conjunto de dados utilizado contém notas referentes a duas disciplinas (Cálculo III e Cálculo IV) e originou-se na Universidade Federal Rural do Rio de Janeiro (UFRRJ). Após tentativas de ajuste de modelos de ordens 1, 2, 3 e 4 e com interação entre as variáveis, verificou-se que o modelo mais adequado para explicar a nota obtida pelo aluno em 2017-3 deveria levar em consideração se o aluno cursou a disciplina Cálculo III ou Cálculo IV. Desta forma, foi obtido um modelo com a inclusão de uma variável categórica ou “Dummy”. Foi possível obter notas mínimas que os alunos deveriam alcançar para serem aprovados nas disciplinas Cálculo III ou Cálculo IV separadamente. Os resultados obtidos indicaram a possibilidade de se estabelecer um critério para aceitação dos alunos nas disciplinas estudadas, de acordo com a nota obtida e estabelecer uma ordem de prioridade na alocação dos alunos nas turmas. Estes critérios podem diminuir o número de alunos inscritos nas disciplinas, gerando economia de espaço físico (salas de aula) e pessoal (professores e funcionários) para a universidade. Adicionalmente, foram analisadas as suposições necessárias para a validade do modelo, tanto sob a perspectiva gráfica quanto a de testes, como os de Durbin-Watson, Shapiro-Wilk e Kolmogorov-Smirnov. São também discutidos os perigos relacionados com a adoção de critérios isolados na avaliação da validade de modelos de regressão. O trabalho baseia-se na pesquisa realizada para a elaboração do Trabalho de Conclusão de Curso de graduação do autor.*

**Kally Chung, Kally Chung, Paulo Justiniano Ribeiro Junior, Wagner Hugo Bonat (UFPR, Brazil)**

Modelos Lineares Generalizados para Dados Espaciais

*Abstract: Modelos e métodos geoestatísticos usualmente são bem definidos e razoavelmente consensuais para tratar dados gaussianos, isto é, dados contínuos e de distribuição simétrica. A utilização de variogramas para a estimação dos parâmetros que descrevem padrões espaciais e a Krigagem para a predição inicialmente propostos se consolidaram na prática em diversas áreas, sendo posteriormente associados à geoestatística baseada em modelos com inferência formal e predição baseadas na distribuição normal multivariada. No entanto, para dados binários ou potencialmente com excesso de zeros, de contagem, contínuos assimétricos e outros mais dados não gaussianos, há poucas e não consensuais abordagens apropriadas. Logo, o propósito deste trabalho é a apresentação de uma classe de modelos denominado modelo linear generalizado para dados espaciais (MLGDE) que lidam com dados não gaussianos, tendo casos gaussianos e independentes como particulares. A estruturação do modelo é baseada nos 1o. e 2o. momentos através de funções de ligação tradicionais do modelo linear generalizado (MLG) e de matriz de covariância definida por funções de variância da distribuição da família Tweedie e funções de correlação usuais em geoestatística. Adota-se o algoritmo Chaser para a estimação dos parâmetros e o preditor simples de Krigagem para a predição. As análises de dados realizadas com o conjunto de dados CTC e Rongelap mostram que o MLGDE é versátil, robusto e eficiente para trabalhar com dados não gaussianos, gaussianos e também com dados independentes.*

**Katy Rocio Cruz Molina, Vera Tomazella, Vinicius Calsavara (UFSCar/USP, Brazil)**

Modelo de Sobrevivência Induzida por Fragilidade Poisson Zero-Modificada

*Abstract: Modelos de sobrevivência com um termo de fragilidade, são apresentados como uma extensão do modelo de risco proporcional de Cox (Cox, 1972) em que um efeito aleatório, denominado fragilidade, é introduzido na função de risco de forma multiplicativa com o objetivo de modelar a heterogeneidade não observada das unidades em estudo. A distribuição para a variável de fragilidade é assumida ser contínua e não negativa. Entretanto, há algumas situações nas quais é apropriado considerar a fragilidade distribuída discretamente, por exemplo, quando a heterogeneidade dos tempos de sobrevivência surge por causa da presença de um número aleatório de falhas por unidade ou pela causa da exposição a danos em um número aleatório de ocasiões. Neste trabalho consideramos o modelo de fragilidade com distribuição disc-*

*reta Poisson Zero-Modificada, a qual tem vários casos particulares. Neste contexto podemos observar a possibilidade de indivíduos com fragilidade zero que corresponde a um modelo de falha limitado que contém uma proporção de unidades que nunca falham (sobreviventes de longa duração) ou modelo de fração de cura. O modelo proposto é aplicado a um conjunto de dados reais de melanoma.*

**Leonardo Biazoli, Patrícia de Siqueira Ramos, Lincoln Frias (Universidade Federal de Alfenas, Brazil)**

**The Spatial Distribution of Rural Insurance in Brazil**

*Abstract: This paper analyzes the spatial distribution of rural insurance policies in Brazilian municipalities in 2017. To this end, it uses data from the Ministério da Agricultura, Pecuária e Abastecimento and, as a first step, performs an exploratory analysis. The southern states were the ones that demanded rural insurance the most, representing about 70% of the total amount. The most insured crops were soybean, second-crop corn, grape, wheat and rice. To verify the existence of spatial autocorrelation two analyses were conducted: one including all Brazilian municipalities and another one limited to the country's southern region (since this is the region where most policies are located). Spatial autocorrelation, measured by Moran's global index  $I$ , indicates the kind and strength of the correlation between the municipalities: if they are surrounded by units in the same situation, High-High (HH) or Low-Low (LL) spatial clusters are formed; on the other hand, if the surrounding municipalities are in the opposite situation, High-Low (HL) and Low-High (LH) outliers are formed. It was observed that rural insurance policies are not randomly distributed throughout the Brazilian territory, with the detection of groups, mainly spatial clusters. They indicate a greater similarity between geographically close municipalities and it was possible to observe a concentration of municipalities with HH behavior in the southern region of Brazil, whereas in the northern and northeastern regions municipalities with LL dominated. Similarly, considering only the southern region, spatial dependence can also be observed, with a concentration of municipalities with HH behavior in the northern and southeastern portions of this region and LL clusters being found in its metropolitan and coastal regions.*

**Lucas Moura Faria e Silva, Jony Arrais Pinto Junior (Universidade Federal Fluminense, Brazil)**

**Quantificando o número de armas de fogo nas microrregiões brasileiras: uma abordagem com efeitos espaciais correlacionados**

*Abstract: Existe um consenso na literatura que a proporção de suicídios por arma*

*de fogo (PSAF) é a melhor medida indireta de prevalência de arma de fogo. Entretanto, esta proxy parece não ser uma boa medida em locais com baixa densidade populacional, visto que suicídios são eventos raros. Recentemente, Cerqueira et al. (2018) propuseram três medidas indiretas que levam em consideração características pessoais das vítimas e um conjunto de efeitos espaciais como novas alternativas para a prevalência de armas de fogo. As medidas propostas mostraram boa correlação com a PSAF, mas apresentaram diferenças que os autores acreditam indicar uma medida indireta mais acurada. Entretanto, as novas proxies consideram efeitos espaciais independentes, o que não parece apropriado dado o contexto espacial inerente. Em geral, dados subdivididos espacialmente apresentam dependência espacial, isto é, dados de uma determinada região assemelham-se mais às regiões vizinhas do que às regiões mais distantes. Assim, a proposta deste trabalho foi criar proxies para prevalência de armas de fogo, seguindo uma metodologia semelhante à proposta de Cerqueira et al. (2018), modelando a probabilidade de um indivíduo cometer suicídio por arma de fogo, considerando variáveis explicativas e os efeitos espaciais correlacionados das microrregiões brasileiras. Sob uma perspectiva completamente bayesiana, foram comparados modelos ajustados com diferentes estruturas de dependência no espaço, considerando distribuições a priori condicionais autorregressivas (CAR). Além disso, os resultados obtidos foram comparados com os resultados de Cerqueira et al. (2018) referente a proxy que não utiliza algum método de regularização.*

**Luiz Henrique Marra da Silva Ribeiro, Vinícius Fonseca Dal Poggetto, Danilo Beli, Adriano Todorovic Fabro, José Roberto de França Arruda (Unicamp, Brazil)**

Bayesian inference on plane frame mechanical properties for predicting the robustness of wave propagation bandgaps

*Abstract: Phononic crystals are periodic structures that exhibit frequency ranges where elastic waves cannot propagate, called Bragg-scattering bandgaps. This study aims at simulating the plane wave propagation in a frame structure with two-dimensional periodicity using Bayesian statistics to understand the bandgap behavior with respect to variability in the Young's modulus and mass density of the frame material. The spectral element method is used to compute Frequency Response Functions (FRF) that are used as numerical experiments. Six random samples taken from a Gamma distribution are used to model the varying mechanical properties. From the numerically computed FRFs, the equivalent wavenumbers are estimated via a Prony method. A Markov chain Monte Carlo (MCMC) algorithm that considers an error between the simulated and observed wavenumbers is used to simulate the posterior distribution of the plane frame mechanical properties. Stochastic dispersion diagrams are simulated via Monte Carlo from these posterior distributions. The MCMC algorithm, by using the additional information related to the wavenumber, presented a more precise distri-*

*bution inference and wavenumber dispersion estimation. Although the 95% confidence interval of the probability distributions used to sample the mechanical properties numerical experiments are inside the 95% credible interval of the posterior distributions obtained via the software OpenBUGS®, and via the used MCMC algorithm, the latter yields much narrower posterior distributions. Analyzing the frequency range DC-50 kHz, it is possible to verify that there are two band gaps around the frequencies of 20 and 35 kHz that are robust with respect to the proposed variation of the mechanical properties. These investigations are relevant for the practical use of periodic frames built using additive manufacturing technologies.*

**Márcio Luis Lanfredi Viola, Graziela de Fátima Valero Minezildo (Universidade Federal de São Carlos, Brazil)**

Measuring Complexity in Epidemic Dengue Periods

*Abstract: Dengue is a worrying infectious disease for Brazil, because of the epidemics of this disease that the country has been experiencing. In this work, we apply the complex measures due to Lopez-Mancini-Calbet (LMC) and to Shiner-Davison-Landsberg (SDL) on the incidence of dengue in Sao Paulo city. José R. C. Piqueira (2017) used the LMC and SDL complexity measures to investigate how macroeconomic variables are related with crisis events, specifically, the measures were applied in dollar-real parity series.*

**Mariane dos Santos Bispo, Camila Borelli Zeller (Universidade Federal de Juiz de Fora, Brazil)**

Double Reweighted estimator for the parameters of the multivariate SMSN distributions

*Abstract: The scale mixtures of skew-normal (SMSN) distributions have many useful applications in robust statistical analysis. The parameter estimation of the SMSN-distribution is carried out using maximum likelihood (ML) estimation method, and the ML estimates are obtained via the ExpectationMaximization (EM) algorithm. In this article, we will use the maximum Lq-likelihood (MLq) estimation method introduced by Ferrari and Yang (2010), Do?ru , Bulutb, and Arslan (2018) to estimate all the parameters of the multivariate SMSN-distribution. We modify the EM algorithm to obtain the MLq estimates. We provide a simulation study and a real data example to illustrate the performance of the MLq estimators over the ML estimators.*



**Marília Gabriela Rocha, Caio Lucidius Naberezny Azevedo (UNICAMP, Brazil)**

Contributions to the study of the main unfolding models of item response theory with applications in psychology

*Abstract: It is of great interest in the field of Psychometrics the use of latent variable models. Throughout the project, a methodological review of the main parametric and one-dimensional unfolding models of Item Response Theory (IRT) was performed, discussing aspects such as interpretation, applicability and estimations. GGUM was applied in the analysis of the Brazilian version of the Observer Alexithymia Scale (OAS), considering the adequacy that unfolding models have in relation to measurement instruments related to attitude.*

**Naiara Caroline Aparecido dos Santos, Jorge Luis Bazán Guzmán (USP/UFSCar, Brazil)**

Bayesian Approach to the Rasch Poisson Model with Application in Attention Data

*Abstract: The Rasch Poisson Count Model (RPCM) is designed for testing with items where the answer is error counts (or successes in various tasks). The RPCM is an item Response Theory model in which we want to estimate a one-dimensional latent trait and a parameter associated with the items. The objective of this work is to develop the estimation of the model under Bayesian approach considering Integrated Nested Laplace Approximation (INLA). The proposed methodology is applied to data from the attention test developed by Beyzaee (2017), which has 20 items and answers from 238 individuals. The results show that the Bayesian estimation is efficient and close to the estimation using glmer, function under approach of mixed generalized linear models using approximate marginal maximum likelihood method. Additionally, the general checks of the model indicated that the RPCM model fits the studied attention data.*

**Patricia Stulp, Carlos Aparecido dos Santos, Willian Luís de Oliveira, Jessica Suzana Barragan Alvez, Naiara Caroline Aparecido dos Santos (USP-UFSCar, Brazil)**

Modelo Bivariado Poisson-Exponencial

*Abstract: Existem muitas maneiras de se introduzir um modelo bivariado, entre as quais destacam-se os modelos em que uma variável resposta é discreta e a outra contínua. Utilizando o critério da fatoração, segue uma proposta de modelo bivariado com resposta mista, com dependência entre as respostas, cuja variável resposta discreta segue uma distribuição Poisson e a variável contínua, condicionada à var-*

*iável resposta discreta, segue uma distribuição Exponencial. A fim de se verificar a qualidade do modelo, isto é, se as propriedades dos estimadores de máxima verossimilhança são verificadas, foi realizado um estudo de simulação com vários cenários.*

**Paulo Victor do Espirito Santo da Silva, José Gracildo de Carvalho Júnior (UFPA, Brazil)**

Statistical Methods Applied to Public Safety

*Abstract: Several areas of human knowledge study the phenomenon of criminality in order to identify the behavioral pattern, in relation to the criminal practice employed in common areas of living or even in isolated places. The objective of this study was to statistically model the records of injury in the State of Pará, from 2010 to 2017. For this purpose, the univariate time series analysis methodology was used to identify the model that best captured the variability of the records. of injury evaluated in this study, considering the use of the script of a statistical software with proven reliability in the implementation of this scientific methodology. The results obtained with tests of statistical significance at the 5% level returned satisfactory diagnoses (independence, normality and constant variance) of the forecast errors, even in distinct subgroups, leading to the unambiguous presence of white noise in the identified model. The results highlight victims in the totality of females; mostly single, between 18 and 34 years old, incomplete elementary school. The highest occurrence of crimes was on weekends, from 20:00 to 00:00 hours, beaten without using an instrument, in private homes and in the municipality of Belém. By the SARIMA model (1,1,1) (1,1,1) 12 identified it became possible to predict and project records of the crime of personal injury in the State of Pará, within a confidence interval of 95% probability.*

**Rafael Rocha de Oliveira Garcia, Cibele Maria Ruddo Novelli (USP/UFSCar, Brazil)**

Joint procedure for variable selection for fixed and random effects in linear mixed-effects models

*Abstract: Linear mixed-effects (LME) models are mainly used to describe clustered data. As an example, one can fit a LME model using data from repeated measures designs, as those constitute clustered data. The problem of selecting predictors in a model is well known to statisticians and it can be used in different scenarios. The selection may be used to perform a dimension reduction or to improve prediction, as less predictors may imply in decreasing the variance. In the context of linear mixed-effects models, variable selection may be carried out by setting a set of predictors, say the fixed effects, as observed and performing selection for the random effects, then*

setting this subset as observed and select the fixed effects predictors. To improve the method for selecting variable in LME, a lasso-type joint procedure is used, in which one can select predictors for both fixed and random effects at the same time. By analyzing a data set of total concentration of nitrate in the atmosphere, we notice that there is residual autocorrelation. We then propose to include this autocorrelation in the procedure, using an autoregressive (AR) process in order to improve the variable selection method. We propose a constrained EM algorithm in which we incorporate the AR process in the estimation procedure.

### **Ricardo Potozky de Oliveira (UFPR, Brazil)**

Hidden Markov Models in Linguistic Typology

*Abstract: Markov stochastic models are used in a wide variety of fields - from economics to genetics - and a Markov model for an autonomous system with partially observable system states is called Hidden Markov Model (HMM). Linguistic Typology, as a branch of comparative linguistics, aims at the description of the structural diversity of natural languages. This field, still giving its first steps among brazilian researchers, can provide insights on the better understanding of the unique status of human language. The main objective of this work is to present applications of HMMs in the research horizon of the linguistic typology. This work is an indirect result of the author's participation at the Linguistic Typology Laboratory (LATIP/UFBA), under the guidance of Prof. Dr. João Paulo Lazzarini Cyrino, and his collaboration with an online platform of statistical analysis of linguistic data.*

### **Richardson Costa de Carvalho, José Gracildo de Carvalho Júnior (Universidade Federal do Pará, Brazil)**

Statistical Methods Applied to Public Security, A Fuzzy Approach

*Abstract: The public security is a duty of the state that aims among another activities, the achievement of acts to nullify or contain the criminality. Technical-scientific procedures from the probability theory and statistics are of utmost importance to help with public security issues. Due to the need to collect, analyze, interpret and present results related to public security, this study is fully justified by the use of time series analysis combined to fuzzy set theory, given to improving of the accuracy of forecasting models of criminal occurrences, among another numerical characteristics. The results of this study indicated that the crime of robbery was the offense most registered in the Metropolitan Region of Belém (32.28%), besides that the identified model got capture with great statistical significance the inherent variability to the data evaluated in this study. The goal of stabilish a forecasting model by fuzzy time series*

*was successfully achieved, given that, was possible accomplish forecast to the related data to the registered cases of offense in the Metropolitan Region of Belém, besides having been possible the projection in this kind of crime with reasonable reliability, being able to contribute to the operational planning and to effective combat actions to the criminality by the public security agencies in this region.*

**Solange Ferreira Silvino, Yana Miranda Borges, Breno Gabriel da Silva, Sandra Milena Diaz Puentes (Faculdade Unyleya)**

Investigando a Relação do Caráter Genético de Zangões no Número de Espermatozoides

*Abstract: Na maioria dos países da América do Sul a apicultura ainda é trabalhada praticamente de maneira extrativista sem um conhecimento técnico, o que a torna menos competitiva com as produções de países como Alemanha e Argentina que implementam sistemas de melhoramento genético para manter as características desejáveis de suas colônias. No entanto, a criação extensiva é praticada pela pouca disponibilidade de material genético avaliado para ofertar no meio apícola. Assim, neste trabalho pretende-se verificar se existem efeitos significativos do caráter genético na variável resposta número de espermatozoides. Inicialmente utilizou-se a análise descritiva por meio de tabelas e box-plot para extrair as informações iniciais dos dados. Para verificar indícios de correlação entre a variável resposta número de espermatozoides e a variável caráter genético utilizou-se o gráfico de correlações. Quanto a suposição de normalidade da variável resposta, utilizou-se o teste de Shapiro-Wilk, considerando o nível de significância a 5%. A fim de verificar se existem efeitos significativos do caráter genético na variável resposta número de espermatozoides utilizou-se o modelo de efeitos. Os dados foram analisados no software R versão 3.6.0.*

**Solange Ferreira Silvino, Yana Miranda Borges, Breno Gabriel da Silva, Brian Alvarez Ribeiro de Melo (Faculdade Unyleya)**

Modelos Probabilísticos na Análise de Séries de Temperatura Mínima Média em Manaus - AM

*Abstract: A relevância em estudar fenômenos climatológicos baseia-se na influência que variáveis dessa natureza exercem no mundo. Entre as variáveis mais observadas, destaca-se a temperatura, cujo efeito de sua variação pode vir a causar impactos significativos, como na proliferação de espécies biológicas, produção agrícola, saúde da população, etc. Distribuições de probabilidade vêm sendo estudadas para verificar o melhor ajuste para descrever e/ou prever o comportamento de variáveis climáticas e, sob esse contexto, o presente estudo avaliou, dentre seis distribuições de probabilidade, a de melhor ajuste para descrever uma série histórica de temperatura mínima média mensal. As séries utilizadas neste estudo englobam um período de 38*

*anos (1980 a 2018) separados por mês, da estação meteorológica da estação MANAUS - AM (OMM: 82331) obtidas no INMET, totalizando 459 observações. Foram utilizados os testes Difference-Sign e Turning Point para verificar independência dos dados e o método da máxima verossimilhança para estimar os parâmetros. Para selecionar a distribuição de melhor ajuste foram utilizados os testes de Kolmogorov-Smirnov, Anderson-Darling, Cramér-von Mises, Critério de Informação de Akaike e gráficos quantil-quantil. Foram avaliadas as distribuições Log-Normal, Gama, Weibull, Gumbel tipo II, Benini e Rice, destacando-se as distribuições Rice, Log-Normal e Gumbel II como as de melhor desempenho.*

**Talita Evelin Nabarrete Tristão de Moraes, Naiara Caroline Aparecido dos Santos, Breno Gabriel da Silva, Yana Miranda Borges (UEM, Brazil)**  
Inferências Bayesiana em Modelo com Fração de Cura para Dados de Câncer de Pele

*Abstract: O câncer de pele é o tipo de câncer mais frequente na população brasileira (INCA, 2009), correspondendo a cerca de 30% de todos os tumores malignos registrados no país, sendo a região Sul do Brasil a que apresenta maiores índices de desenvolvimento da doença devido à alta concentração de pessoas com pele mais clara nessa população. Sendo este dividido em melanoma e não melanoma, em que o melanoma é o tipo mais grave. Pelo fato do câncer de pele apresentar altos percentuais de cura, se detectado e tratado precocemente, sua taxa de mortalidade é uma das mais baixas. Para este trabalho, utilizou-se os dados fornecidos pelo Instituto Nacional de Câncer (INCA), em que selecionou-se pacientes diagnosticados com câncer de pele entre os anos de 2000 e 2005, acompanhados até 2013 no estado do Paraná, tendo como objetivo analisar o modelo de sobrevivência estimando a proporção de indivíduos curados. Neste cenário, ajustaram-se modelos de mistura, onde a população é dividida em duas partes: uma que representa o tempo de sobrevivência para os indivíduos suscetíveis (não curados) e a outra com uma distribuição que permite calcular o tempo de sobrevivência dos imunes (curados), com distribuição de sobrevivência Weibull na presença de fração de cura sob o enfoque bayesiano, por meio do método de aproximações de Laplace encaixadas e integradas, INLA (Integrated Nested Laplace Approximation)*

**Vitor Gustavo de Amorim, Sandro Gallo (PhD supervisor) (ICMC-USP/UFSCar, Brazil)**

Approximate Distributions of the Time Elapsed Until the Occurrence of a Pattern

*Abstract: Consider a stochastic process taking values in a finite set  $\mathcal{S}$ . We are interested in the time elapsed until the process hits for the first time a given pattern (finite string of symbols)  $A$ . This random variable is called hitting time and denoted*

by  $\tau_A$ . If the process starts with  $A$ , we call this variable return time and denote by  $\tau_{A|A}$ . We are interested in approximated distributions for both variables. This problem occurs naturally in search of rare patterns in biological sequences modelled by a Markov chain or in a sequence of bits (0's and 1's) modelled by an i.i.d. process. For these processes, we know that return and hitting time have approximately exponential distribution. We present two results that establish an explicit parameter for the exponential distributions and give precise error terms for the approximation. We present also applications in mentioned processes (i.i.d. and Markov chain) and we calculate their respective approximate distributions, error terms and expected value for hitting and return times.

### **Walef Machado de Mendonça, Leandro Ferreira (Universidade Federal de Alfenas, Brazil)**

Sistema baseado em regras fuzzy no estudo da probabilidade de ruína para indenizações com distribuição gama

*Abstract: Comumente denominado como probabilidade de ruína, o risco de uma seguradora ficar com reserva insuficiente para pagar suas indenizações, se apresenta como um grande desafio, uma vez que não existe uma expressão que possa ser aceita como ideal na comunidade atuarial. Nesse contexto, o objetivo deste trabalho foi construir um Sistema Baseado em Regras Fuzzy (SBRF) integrado ao modelo probabilístico de Cramér-Lundberg, para determinar a probabilidade de ruína de uma seguradora. A variável valor da indenização, gerada por simulação Monte Carlo, foi utilizada como variável de entrada do SBRF, considerando probabilidade de ruína como a variável de saída. Neste trabalho, foi analisado um cenário em que as indenizações seguem uma distribuição gama. SBRFs foram criados utilizando funções de pertinência gaussiana, método de inferência de Mandani e método de defuzzificação centro de gravidade. Foram construídos modelos com 5, 10 e 15 conjuntos fuzzy. Para a avaliação do modelo, os dados simulados foram divididos em dois conjuntos (treinamento e teste), sendo que para ambos os conjuntos de dados foram calculados o coeficiente de determinação e erro quadrático médio. Foi constatado que o coeficiente de determinação ficou acima de 90%, tanto para o conjunto de treinamento quanto para o conjunto de teste, nos modelos com 5, 10 e 15 conjuntos fuzzy, o que indica que os SBRFs foram bem ajustados e possuem boa capacidade para a estimação da probabilidade da ruína.*

**Walkiria Maria de Oliveira Macerau, Luis A. Milan (UFSCar/USP, Brazil)**

Estimação via Algoritmo EM de modelos de misturas de distribuições assimétricas e simétricas

*Abstract: Modelos de mistura de distribuições são ferramentas utilizadas para lidar populações heterogêneas. Entretanto, além da heterogeneidade, os dados podem apresentar assimetria dentro das subpopulações. Neste caso, é mais indicado trabalhar com misturas de distribuições assimétricas, pois essas distribuições conseguem captar a assimetria diretamente, evitando transformação nos dados. Neste trabalho, consideramos modelos de mistura finita de distribuições assimétricas e simétricas, por meio de uma abordagem clássica, baseada na estimação Via Algoritmo EM para estimar os parâmetros do modelo de mistura e os critérios de seleção de modelos AIC, BIC e ICL para selecionar o número de componentes de mistura. Para verificar a aplicabilidade do algoritmo, aplicamos a metodologia em um conjunto de dados reais da doença de Alzheimer. A implementação do método foi realizada utilizando o software R.*

**Yana Miranda Borges, Breno Gabriel da Silva, Luma Alana Vieira Henrique, Claudete Aparecida Mangolin (Universidade Estadual de Maringá, Brazil)**

Aplicação do Modelo Gama em Dados de Embriogênese Somática e Regeneração de Plantas a partir de Linhagens de Milho-Pipoca

*Abstract: O milho pipoca é um alimento muito consumido no Brasil, apesar de seu cultivo se restringir a pequenas áreas, devido à limitação de cultivares de alta qualidade e à tecnologia de produção inadequada. A transformação genética é uma opção para melhorar as variedades existentes que consegue ir além dos avanços obtidos pelo melhoramento tradicional. A realização dessa técnica só é possível através do desenvolvimento de um protocolo eficiente para a regeneração de plantas. Assim, neste trabalho inicialmente foi realizada a análise descritiva por meio de medidas resumo e gráficos. A fim de verificar se existem efeitos significativos dos meios (MS e N6), concentrações de 2,4-D (0;1,2;2,2 e 3,2), e os níveis de Sacarose (0%, 3% e 6%) na variável resposta percentual de calos embriogênicos utilizou-se o modelo de regressão Gama com função de ligação inversa. Para escolher o protocolo a ser utilizado, considerando os níveis de concentrações 2,4-D, os meios e os níveis de Sacarose utilizou-se o teste de Tukey para comparações múltiplas.*

**Yana Miranda Borges, Breno Gabriel da Silva, Sandra Milena Diaz Puentes, Naiara Caroline Aparecido dos Santos, Talita Evelin Nabarrete Tristão de Moraes (Universidade Estadual de Maringá, Brazil)**

Análise do Volume de Sêmen de Zangões de Abelhas *Apis mellífera* Africanizada via Modelo de Efeitos

*Abstract: A maioria das pesquisas feitas em torno da genética de abelhas tem o enfoque em abelhas europeias, por tanto, falta informação sobre a genética de abelhas africanizadas, presentes no Brasil. Por isso se faz necessário determinar a importância da diversidade genética dentro de um apiário, fomentando o uso da tecnologia. Para isso, a qualidade espermática dentro da espermateca quando o acasalamento é natural ou pela inseminação instrumental, precisa ser avaliada. Logo, este trabalho tem por objetivo avaliar se existem efeitos significativos do caráter genético na variável resposta volume (Quantidade de sêmen produzido por zangão). Empregou-se inicialmente a análise descritiva por meio de tabelas e box-plot para extrair as informações iniciais dos dados. Quanto a suposição de normalidade da variável resposta utilizou-se o teste de Shapiro-Wilk e para verificar a suposição de homogeneidade de variâncias utilizou-se o teste de Bartlett em que em ambos os testes considerou-se o nível de significância a 5%. A fim de verificar se existem efeitos significativos do caráter genético na variável resposta volume (Quantidade de sêmen produzido por zangão) utilizou-se o modelo de efeitos. A análise diagnóstica do modelo foi realizada por meio dos gráficos semi-normal de probabilidades com envelope simulado e qq-plot. Os dados foram analisados no software R versão 3.6.0.*

## Participants

Adalto Speroto  
 Adraiana S. Resende  
 Alaor Cervati Neto  
 Alex de la Cruz Huayanay  
 Alex Fernandes Mansano  
 Alex Leal Mota  
 Alex Pereira da Silva  
 Alexandre Alves Madi  
 Amanda Morales Eudes D'Andrea  
 Amanda Post da Silveira  
 Amanda Yumi Ambriola  
 Amandia de Oliveira Sá  
 Ana Carolina do Couto Andrade  
 Ana Paula Dos Anjos Vitorasse  
 André Luiz Grion  
 André Ponce de Leon F. de Carvalho  
 Angelina Carrijo de Oliveira Ganancin Faria  
 Anísia Maria da Rocha Nogueira  
 Átila Prates Correia  
 Barioni Júnior  
 Beatriz Rodrigues  
 Brenno Gustavi Barbosa  
 Breno Gabriel da Silva  
 Bruna Luiza de Faria Rezende  
 Brunna Quatrochi  
 Caio Lucidius Naberezny Azevedo  
 Caio Moura Quina  
 Camila Bertini Martins  
 Camila Sgarioni Ozelame  
 Carla Arnaud de Azevedo  
 Carlos Aparecido dos Santos  
 Carlos Franklin Taco Pedraza  
 Caroline Amantéa Stella  
 Christian Eduardo Galarza Morales  
 Cibele Maria Russo Novelli  
 Claudio Policastro  
 Cleide Mayra Menezes Lima  
 Conrado Paschoal Braga  
 Cristel Ecaterin Vera Tapia  
 Cristian Favio Coletti  
 Cristian Villegas  
 Cristina Ortiz Sobrinho Valete  
 Daiane Aparecida Zuanetti  
 Daiane de Souza Santos  
 Daniel Camilo Fuentes Guzman  
 Davi Keglevich Neiva  
 Débora Mayumi Rissato  
 Deborah Bassi Stern  
 Devis Ordoño Vilca  
 Diego Carvalho do Nascimento  
 Diogo Barboza Moreira  
 Djidenou Hans Amos Montcho  
 Douglas Decicino de Andrade  
 Douglas Vinícius Gonçalves Araújo  
 Éder Silva de Brito  
 Edson Hernandes Francelino  
 Eduardo Vargas Ferreira  
 Elias Silva de Oliveira  
 Elizabeth Chipa Bedia  
 Emily Sato  
 Erika Soares Machado  
 Evelyn Barros Bernardo  
 Fabiano Rodrigues Coelho  
 Fábio Pratavia  
 Fátima Aparecida Kian  
 Felipe Hernandez Bisca  
 Fernanda Yuka Ueno  
 Francisco José dos Santos  
 Francisco Rodrigues  
 Gabriel Silva Malta  
 Gabriela Maria Leme Trivellato  
 Gabriela Massoni  
 Gilberto Pereira de Alcântara Junior  
 Giovanni Moura de Holanda  
 Giovanni Pastori Piccirilli  
 Gladys Choque Ulloa  
 Guilherme Antonio Alves de Lima  
 Guilherme Mendonça Freire  
 Gustavo Alexis Sabillón Lee  
 Hugo Leonardo Salomão Monteiro  
 Igor Araújo dias santos  
 Igor de Souza Bezerra  
 Irineu Dinato Junior  
 Isabella Costa Rocha  
 Isabella Emanuele Corrêa  
 Isadora Souza Fumagali  
 Isis Fernanda Mascarin  
 Iuri Emmanuel de Paula Ferreira  
 Izabella Gonçalves Rendeiro  
 Jeremias Junior  
 Jessica Suzana Barragan Alves  
 Jonas Bodini Alonso  
 Jonathan Kevin Jordan Vasquez  
 Jony Arrais Pinto Junior  
 Jorge Luis Bazán Guzmán  
 Jorge Moreira de Souza  
 José Victor Saldanha Costa Silva Gomes  
 Josimara Tatiane da Silva  
 Josué dos Santos da Conceição  
 Josué dos Santos da Conceição

Juliana Buzinaro Andrikonis  
Juliana Cobre  
Juliana Shibaki Camargo  
Juliane Bertolo  
Kally Chung  
Karina Fernandes de Oliveira  
Katy Rocio Cruz Molina  
Ladislau de Oliveira dos Santos  
Laís Sebastiany de Souza Santos  
Lara Inacio de Moraes  
Lara Roberta da Silva Costa  
Leandro Resende Mundim  
Leonardo Biazoli  
Leonid Alemán Gonzales  
Lívia Karoline de Freitas Carlos  
Lorena Magalhães Drummond  
Luben Miguel Cruz Cabezas  
Lucas Moura Faria e Silva  
Luis Aparecido Milan  
Luis Ernesto Bueno Salazar  
Luís Felipe Barbosa Fernandes  
Luiz Antonio de Jesus Rocha  
Luiz Fernando Silva Eugenio dos Santos  
Luiz Gabriel Fernandes Cotrim  
Luiz Henrique Marra da Silva Ribeiro  
Marcelo Miky Mine  
Marcílio Ramos Pereira Cardial  
Márcio Luis Lanfredi Viola  
Marcos J Ribeiro  
Marcos Jardel Henriques  
Marcus Vinicius Teixeira Borba  
Maria da Conceição Pereira Saraiva  
Maria Fernanda Freitas Santos  
Maria Luiza Matos Silva  
Mariana Curi  
Mariana Vieira Ribeiro Lopes  
Mariane dos Santos Bispo  
Marília Gabriela Rocha  
Marina Gandolfi  
Marina Gonzaga de Oliveira  
Mário de Castro  
Mateus Borges Comito  
Matheus Ferreira Lino Chamorra  
Matheus Toshio Hisatugu  
Michele Maciel Sacramento  
Milena Nascimento Lima  
Milene Regina dos Santos  
Milton Miranda Neto  
Naiara Caroline Aparecido dos Santos  
Nancira Ribeiro Madi  
Natália da Silva Martins Fonseca  
Nayara Maria Barbosa de Sousa  
Oliviana Xavier do Nascimento  
Osafu Augustine Egbon  
Patrícia Stülp  
Patricia Viana da Silva  
Paulo Cesar Valet  
Paulo Henrique Ferreira da Silva  
Paulo Roberto de Lima Gianfelice  
Pedro Floriano Ribeiro  
Pedro Nasevicius Ramos  
Pedro Vinícius Chabbuh  
Rafael Izbicki  
Rafael Rocha de Oliveira Garcia  
Rafael Simões de Paula  
Rafaela Apreia Marques  
Renan Silva chun  
Renata Cristina Carregari  
Renato da Silva Fernandes  
Ricardo Ehlers  
Ricardo Felipe Ferreira  
Ricardo Potozky de Oliveira  
Ricardo R. da Silva  
Richardson Costa de Carvalho  
Roberto Shimizu  
Rodrigo Malavazi Corder  
Ronaldo Dias  
Roseli Aparecida Leandro  
Samuel Fontes Machado  
Sandra Cristina de Oliveira  
Sandra Valéria Coelho da Silva  
Sandro Gallo  
Silvia Jusnan Bravo Cedeno  
Solange Ferreira Silvino  
Tainá Santana Caldas  
Taís Roberta Ribeiro  
Talita Evelin Nabarrete Tristão de Moraes  
Tamiris Camargo dos Santos Mariano  
Uriel Moreira Silva  
Vera Tomazella  
Victor Cândido Reis  
Vinicius Hideki Yamada Santiago  
Vitor Gustavo de Amorim  
Walef Machado de Mendonça  
Walkiria Maria de Oliveira Macerau  
Welton Aparecido Teixeira  
Wesley Ricardo da Silva Satelis  
Wilbur Naike Chiuyari Veramendi  
Yana Miranda Borges





Organizers



Support

