

# VIII ENCONTRO DOS ALUNOS

Estatística e Experimentação Agronômica



**Anais do VIII Encontro dos Alunos em  
Estatística e Experimentação Agronômica**

**Escola Superior de Agricultura Luiz de Queiroz  
Piracicaba, 23 de novembro de 2018**

**Realização:**



Departamento de  
Ciências Exatas  
da ESALQ/USP



Programa de Pós-  
graduação em Estatística e  
Experimentação Agronômica



## **COMISSÃO ORGANIZADORA**

(ORGANIZING COMMITTEE)

- Eduardo Elias Ribeiro Junior<sup>1</sup> (ESALQ/USP);
- Pórtia Piscitelli Cavalcanti (ESALQ/USP);
- Welinton Yoshio Hirai (ESALQ/USP);
- Clarice Garcia Borges Demétrio (ESALQ/USP).

## **COMITÊ CIENTÍFICO**

(SCIENTIFIC COMMITTEE)

- Clarice Garcia Borges Demétrio (ESALQ/USP);
- Idemauro Antonio Rodrigues de Lara (ESALQ/USP);
- Rafael de Andrade Moral (Maynooth University);
- Renata Alcarde (ESALQ/USP);
- Thiago Oliveira de Paula (ESALQ/USP);
- Walmes Marques Zeviani (LEG/UFPR).

## **SÍTIOS ELETRÔNICOS**

(WEB PAGES)

- VIII Encontro dos Alunos  
<https://esalq-pggee.github.io/encontro2018/>;
- Departamento de Ciências Exatas  
<http://www.lce.esalq.usp.br/>;
- Escola Superior de Agricultura Luiz de Queiroz  
<http://www4.esalq.usp.br/>.

---

<sup>1</sup>E-mail: [jreduardo@usp.br](mailto:jreduardo@usp.br)



# VIII Encontro dos Alunos

## em Estatística e Experimentação Agronômica

<b>1 MINICURSO</b>	<b>1</b>
Modelos de Regressão Não Linear	
<i>Prof. Dr. Walmes Marques Zeviani . . . . .</i>	1
<b>2 CONFERÊNCIAS</b>	<b>1</b>
An Extended Random-effects Approach to Modeling Repeated, Overdispersed Count Data	
<i>Profa. Dra. Clarice Garcia Borges Demétrio . . . . .</i>	1
Análise de Dados Poisson Composto Longitudinais Multivariado	
<i>Prof. Dr. Afrânio Marcio Corrêa Vieira . . . . .</i>	1
Planejamento para o ajuste de curvas flexíveis	
<i>Profa. Dra. Luzia Aparecida Trinca . . . . .</i>	2
Alternative methods for modeling of the cure rate in survival studies	
<i>Profa. Dra. Vera Lúcia Damasceno Tomazella . . . . .</i>	2
<b>3 COMUNICAÇÕES ORAIS</b>	<b>3</b>
Delineamentos ótimos para experimentos com cana-de-açúcar	
<i>Short-Authors . . . . .</i>	3
Superdispersão em Dados Categorizados Multinomiais: uma Aplicação em Ciências Agrárias	
<i>Short-Authors . . . . .</i>	3
<b>4 PARTICIPANTES</b>	<b>4</b>



# 1 MINICURSO

## Modelos de regressão não linear: teoria e aplicações

*Prof. Dr. Walmes Marques Zeviani (LEG/UFPR)*

Em modelos regressão não-linear dados observados de uma variável resposta são descritos por uma função de uma ou mais variáveis explicativas que é não linear seus parâmetros. Assim como nos modelos lineares o objetivo é identificar e estabelecer a relação entre variáveis explicativas e resposta. Entretanto, enquanto os modelos lineares definem, em geral, relações empíricas, os modelos não-lineares são, em grande parte das vezes, motivados pelo conhecimento do tipo de relação entre as variáveis. Desta forma, as aplicações surgem nas diversas áreas onde relações físicas, biológicas, cinéticas, químicas, fisiológicas, dentre outras, são estabelecidas por funções não lineares que devem ter coeficientes (parâmetros) identificados (estimados) a partir de dados observados ou experimentais.

# 2 CONFERÊNCIAS

## An Extended Random-effects Approach to Modeling Repeated, Over-dispersed Count Data

*Profa. Dra. Clarice Garcia Borges Demétrio (ESALQ/USP)*

Non-Gaussian outcomes are often modeled using members of the so-called exponential family. The Poisson model for count data falls within this tradition. The family in general, and the Poisson model in particular, are at the same time convenient since mathematically elegant, but in need of extension since often somewhat restrictive. Two of the main rationales for existing extensions are (1) the occurrence of overdispersion (Hinde and Demétrio 1998, Computational Statistics and Data Analysis 27, 151-170), in the sense that the variability in the data is not adequately captured by the models prescribed mean-variance link, and (2) the accommodation of data hierarchies owing to, for example, repeatedly measuring the outcome on the same subject (Molenberghs and Verbeke 2005, Models for Discrete Longitudinal Data, Springer), recording information from various members of the same family, etc. There is a variety of overdispersion models for count data, such as, for example, the negative-binomial model. Hierarchies are often accommodated through the inclusion of subject-specific, random effects. Though not always, one conventionally assumes such random effects to be normally distributed. While both of these issues may occur simultaneously, models accommodating them at once are less than common. This paper proposes a generalized linear model, accommodating overdispersion and clustering through two separate sets of random effects, of gamma and normal type, respectively (Molenberghs, Verbeke and Demétrio 2007, LIDA, 13, 513-531, Molenberghs et al, 2010, Statistical Science, 25: 325-347, Vangeneugden et al, 2011, Journal of Applied Statistics, 38: 215-232, Molenberghs, Verbeke and Demétrio 2017, SORT, 41, 3-54). This is in line with the proposal by Booth, Casella, Friedl and Hobert (2003, Statistical Modelling 3, 179-181). The model extends both classical overdispersion models for count data (Breslow 1984, Applied Statistics 33, 38-44), in particular the negative binomial model, as well as the generalized linear mixed model (Breslow and Clayton 1993, JASA 88, 9-25). Apart from model formulation, we briefly discuss several estimation options, and then settle for maximum likelihood estimation with both fully analytic integration as well as hybrid between analytic and numerical integration. The latter is implemented in the SAS procedure NLMIXED. The methodology is applied to data from a study in epileptic seizures.

## **Análise de Dados Poisson Composto Longitudinais Multivariado**

*Prof. Dr. Afrânio Marcio Corrêa Vieira (UFSCar)*

Distribuição Poisson Composta é uma distribuição contínua assimétrica, com massa de probabilidade positiva em  $Y = 0$ . Registros pluviométricos, valores pagos para apólices de seguros, dentre outras situações apresentam dados com este comportamento. Apresentaremos um problema em que múltiplas expressões bioquímicas de variedades do algodão foram mensuradas ao longo do tempo, sob um delineamento experimental planejado. Na análise, uma estratégia utilizando modelos lineares generalizados misto permite a análise multivariada das expressões bioquímicas, levando em consideração a não-normalidade, dependência temporal e estrutura do delineamento experimental.

## **Planejamento para o ajuste de curvas flexíveis**

*Profa. Dra. Luzia Aparecida Trinca (UNESP/Botucatu)*

O ajuste de curvas ou superfícies sempre faz parte da análise de resultados experimentais, nos quais procura-se estabelecer relações entre a variável resposta e os vários fatores quantitativos. Os polinômios de segunda ordem são largamente empregados e suas limitações frequentes, devido a simetria imposta, não raramente levam à falta de ajuste e ao uso de modelos de alta ordem nem sempre interpretáveis ou parcimoniosos. Para curvas ou superfícies assimétricas, inclusive com assíntotas, na década de 1990, foram sugeridos os polinômios fracionários, inspirados na família de transformações Box-Tidwell, para análise de dados observacionais. Vários trabalhos mostraram que polinômios fracionários (PF) de até segunda ordem podem gerar uma grande variedade de curvas úteis para modelar as relações de interesse prático. Em princípio, os PF podem também resolver os problemas de falta de ajuste dos modelos de primeira e segunda ordem na análise de dados experimentais. No entanto, quando tentamos ajustar um PF aos dados de um experimento, esbarramos em, pelo menos, dois problemas. O primeiro é que o PF de segunda ordem, como definido originalmente, inclui dois parâmetros para cada fator, as potências, além dos coeficientes de regressão. O segundo é que o delineamento clássico apresenta pontos esparsos e simétricos na região experimental, resultando em pouca informação para estimação dos parâmetros extras do polinômio. Nesse trabalho propomos uma versão de PF de segunda ordem que restringe a estimação de uma única potência para cada fator. A ideia é que a potência determina a transformação apropriada aos níveis do fator para que o polinômio de segunda ordem seja uma boa aproximação para a relação subjacente. Sob esse modelo mais parcimonioso, estudamos o comportamento de delineamentos eficientes para estimar todos os parâmetros. Como o modelo é não linear precisamos incorporar informação a priori para a construção dos delineamentos. Resultados mostram que o delineamento resultante para o PF pode ser bem diferente do delineamento clássico, indicando que o problema de estimação das potências deve ser considerado no planejamento do experimento. O método pode ser estendido para os modelos lineares generalizados nas situações em que seja apropriado especificar o preditor linear por uma relação curva assimétrica.

## **Alternative methods for modeling of the cure rate in survival studies**

*Profa. Dra. Vera Lúcia Damasceno Tomazella (UFSCar)*

In medical studies, it is common that some units under study are not susceptible to the event of interest, called immune or cured elements. A class of models, referred to as cure rate models, considers these situations and has been studied by several authors in the recent years. The cure fraction is of interest to patients and a useful measure to monitor trends and differences



in survival of curable disease. In this presentation we discuss some alternative methods for modeling cure rate in particular the Defective models. Defective models have the advantage of modeling the proportion of cured without adding any extra parameters in the model, in contrast to the most models from the literature.

### 3 COMUNICAÇÕES ORAIS

#### **Delineamentos ótimos para experimentos com cana-de-açúcar**

*Authors*

Resumo: Um dos objetivos do programa de melhoramento genético é a seleção de novos melhores clones. Com delineamentos ótimos e modelos bem ajustados consegue-se melhorar a precisão dessa seleção, porém em fases iniciais do programa de melhoramento genético se tem uma limitação de material genético o que dificulta tal seleção devido a falta de repetição. No presente trabalho avaliamos experimentos que apresentam parte dos novos clones repetidos, ou em alguns casos, apenas uma repetição de cada, como também o critério adotado para se obter o delineamento ótimo.

#### **Optimal Designs with Sugarcane**

Abstract: One of the objectives of the breeding program is the best selection of new clones. With optimum designs and well-adjusted models, the accuracy of this selection can be improved, but in the early stages of the breeding program there is a limitation of genetic material which makes this selection difficult due to lack of repetition. In the present work we evaluated experiments that present part of the new clones repeated, or in some cases, only one repetition of each, as well as the criterion adopted to obtain the optimal design.

#### **Superdispersão em Dados Categorizados Multinomiais: uma Aplicação em Ciências Agrárias**

*Authors*

Resumo: Variáveis politômicas são comuns em experimentos agrônômicos, apresentando natureza natural ou ordinal. Voigt (2013), desenvolveu um experimento em casa de vegetação, cujo objetivo foi avaliar o florescimento de plantas adultas de laranjeira da variedade x11, enxertadas sobre o limão Cravo e sobre o citrumelo Swingle. Em cada planta avaliou-se o número de ramos novos, classificados como sendo: terminal, lateral, sem flor e abortada. O Modelo dos Logitos Generalizados é uma classe de modelos que pode ser empregada para a análise desses dados. Uma das características deste modelo é a pressuposição de que a variância é uma função conhecida da média e, espera-se, que a variância observada esteja próxima da variância pressuposta pelo modelo assumido. Contudo, há casos em que, os dados são mais heterogêneos do que a variância especificada pelo modelo, fenômeno este conhecido como superdispersão. Verifica-se na literatura que há uma abordagem de textos sobre superdispersão aplicados a dados de contagem e proporção para variáveis dicotômicas. Assim sendo, este trabalho tem como objetivo caracterizar o problema da superdispersão com dados multinomiais, assim como selecionar um modelo adequado para acomodar de forma satisfatória a superdispersão presente nos dados. Neste trabalho são ajustados os modelos Multinomial e Dirichlet-Multinomial e os resultados comparados, por meio da deviance e do gráfico half-normal plot. Avaliou-se forte evidências da presença de superdispersão para o modelo Multinomial. Então, realizou-se um teste de hipótese baseado na razão de verossimilhança, indicando que o modelo Dirichlet-Multinomial é o mais

adequado.

## Overdispersion in Categorized Multinomial Data: an Application in Agrarian Sciences

**Abstract:** Polytomic variables are common in agronomic experiments, presenting natural or ordinal nature. Voigt (2013), developed a greenhouse experiment, whose objective was to evaluate the flowering of adult plants of orange variety 'x11', grafted on the Rangpur Lime and on the Citrumelo Swingle. In each plant, the number of new branches was evaluated, being classified as: terminal, lateral, without flower and aborted. The Generalized Logits Model is a class of models that can be used to analyze the data. One of the characteristics of this model is the assumption that the variance is a known function of the mean. It is expected that the observed variance be close to the variance assumed by the assumed model. However, there are cases in which, the data are more heterogeneous than the variance specified by the model, this phenomenon is known as overdispersion. It is verified in the literature that there is an approach of overdispersion texts applied to counting and proportion data for dichotomous variables. Thus, this work aims to characterize the problem of overdispersion with multinomial data, as well as to select a suitable model to satisfactorily accommodate the overdispersion present in the data. In this work, the Multinomial and Dirichlet-Multinomial models were adjusted, and the results are compared, through deviance and the half-normal graph. There was a strong evidence in presence of overdispersion for the Multinomial model evaluated. Then, a hypothesis test was performed based on the probability ratio, indicating that the Dirichlet-Multinomial model is the most adequate.

## 4 PARTICIPANTES

- Jhessica Letícia Kirch (ESALQ/USP)
- Roseli Aparecida Leandro (ESALQ/USP)
- Taciana Villela Savian (ESALQ/USP)
- Pollyane Vieira da Silva (ESALQ/USP)
- Janaína Marques e Melo (ESALQ/USP)
- Suelen Cristina Gasparetto (ESALQ/USP)
- Alessandra de Lima Goes (ESALQ/USP)
- Maria Letícia Salvador (ESALQ/USP)
- Cristiane Mariana Rodrigues da Silva (ESALQ/USP)
- Rita de Cássia de Lima Idalino (UNESP)
- Iuri Emmanuel de Paula Ferreira (UFSCAR - LAGOA DO SINO)
- Edijane Paredes Garcia (UNESP/BOTUCATU)
- Hercílio da Silva Freitas Júnior (ESALQ/USP)
- Fabrício Pilonetto (ESALQ/USP)
- Giovanni Casagrande Silvello (ESALQ/USP)
- Yuniel Tejeda Mazola (CENA/USP)
- Timóteo Herculino da Silva Barros (ESALQ/USP)
- Silvio Henrique Menezes Gomes (ESALQ/USP)

- Paula Midori Castelo (UNIFESP)
- Igor Engler Lima (ESALQ/USP)
- Cristian Marcelo Villegas Lobos (ESALQ/USP)
- Fábio Prata Vieira (ESALQ/USP)
- Mayara Aparecida Correa (ESALQ/USP)
- Humberto de Jesus Eufrade Junior (FCA/UNESP)
- Luana Tais de Freitas (ESALQ USP)
- Fagner Junior Gomes (ESALQ/USP)
- Welinton Yoshio Hirai (ESALQ/USP)
- Andréia Pereira Maria Hilário (ESALQ/USP)
- Alvaro Ramirez (ESALQ USP)
- Vivian Aparecida Brancaglioni (ESALQ/USP)
- Idemauro Antonio Rodrigues de Lara (ESALQ/USP)
- Maira Blumer Fatoletto (ESALQ/USP)
- Aldeir Ronaldo Silva (ESALQ/USP)
- Clarice Garcia Borges Demétrio (ESALQ/USP)
- Glória Cristina Vieira de Sousa (ESALQ/USP)
- Maria Beatriz Duarte Gavião (UNICAMP)
- Mariely Lopes dos Santos (UNIVERSIDADE DE SÃO PAULO ESCOLA SUPERIOR DE AGRICULTURA LUIZ DE QUEIROZ)
- Wagner Wolff (ESALQ/USP)
- Julio Cezar Souza Vasconcelos (ESALQ/USP)
- Roberto de Oliveira Rodrigues (ARYSTA (TAMBÉM ALUNO DE ENG. AGRONÔMICA NA FAESB))
- João Gabriel Ribeiro (ESALQ/USP)