

# 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>
Visualização de dados utilizando ggplot2	
<i>Short-Authors</i> . . . . .	3
A distribuição odd log-logística gama generalizada e modelo de regressão associado para dados censurados	
<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, Overdispersed 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

#### Visualização de dados utilizando ggplot2

*Authors*

Resumo: O ggplot2 foi criado em 2005 por Hadley Wickham, é um pacote de visualização de dados para a linguagem de programação estatística R. A partir da definição de gráficos estabelecida em The Grammar of Graphics livro escrito por Leland Wilkinson em 2005, Hadley escreveu um artigo em 2010 com o seguinte título: A Layered Grammar of Graphics, mostrando que podem ser divididas em camadas as características principais de um gráfico, o que é a essência do ggplot2. Este pacote tem como vantagens: obter gráficos mais elegantes, facilidade na criação destes gráficos, estrutura padronizada das funções deixando assim o processo ainda mais intuitivo, além da possibilidade de criar uma grande quantidade de gráficos com poucas linhas de código. Serão apresentados diversos gráficos sendo eles: gráfico de pontos, barras, dispersão e boxplot, além de suas variações. Serão utilizados para ilustrar a metodologia, os dados diários de precipitação pluviométrica da estação convencional do posto meteorológico da ESALQ (<http://www.esalq.usp.br/departamentos/leb/anos.html>).

#### Data visualization using ggplot2

Abstract: The ggplot2 was created in 2005 by Hadley Wickham, is a data visualization package for the statistical programming language R. From the definition of graphics established in 'The Grammar of Graphics' book written by Leland Wilkinson in 2005, Hadley wrote a article in 2010 with the following title: 'A Layered Grammar of Graphics', showing that the main characteristics of a graphic can be divided into layers, which is the essence of ggplot2. This package has the advantages of: obtaining more elegant graphics, easy creation of these graphics, standardized structure of functions, making the process even more intuitive, and the possibility of creating a large number of graphics with few lines of code. There will be several graphs being shown: dotcharts, barcharts, scatterplot and boxplot, besides its variations. It will be used to illustrate the methodology, the daily rainfall data of the conventional station of the meteorological station of the ESALQ (<http://www.esalq.usp.br/departamentos/leb/anos.html>).

#### A distribuição odd log-logística gama generalizada e modelo de regressão associado para dados censurados

*Authors*

Resumo: Neste trabalho, propomos uma extensão da distribuição gama generalizada (GG) chamada de distribuição odd log-logística gama generalizada (OLLGG), que tem flexibilidade para acomodar várias formas da função de taxa de risco (hrf), tais como crescente, decrescente, em forma de U e unimodal, e que também é apropriado para discriminar entre alguns modelos probabilísticos alternativos, como as distribuições GG, Weibull e exponencial, comumente usados na análise do tempo de sobrevivência. Além disso, é comum, em situações práticas, que o tempo de sobrevivência esteja relacionado a covariáveis que ajudam a explicar a variabilidade deste.

Portanto, outro objetivo deste trabalho é propor modelos de regressão, na forma de locação e escala, denominado modelo de regressão log-odd log-logística gama generalizada (LOLLGG), para dados censurados. Ilustramos a potencialidade do novo modelo de regressão por meio de um conjunto de dados reais do nível de tensão e finalizamos com algumas considerações finais.

The odd log-logistic generalized gamma distribution and its associated regression model for censored data

**Abstract:** In this paper, we propose an extension of the generalized gamma (GG) distribution called the odd log-logistic generalized gamma (OLLGG) distribution, which has flexibility to accommodate various forms of the hazard rate function (hrf), such as increasing, decreasing, U-shaped and unimodal, and that is also appropriate to discriminate between some alternative probabilistic models such as the GG, Weibull and exponential distributions, commonly used in analyzing survival time. Besides this, it is common in practical situations for the survival time to be related to covariables that help explain the variability this time. Therefore, another objective of this work is to propose regression models, in the form of location and scale, called the log-odd log-logistic generalized gamma (LOLLGG) regression model, with censored data. We illustrate the potentiality of the new regression model by means of a real data set of the voltage level and ends with some concluding remarks.

## 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)