**Seminários do PPG em Estatística e Experimentação Agronômica**

**Segundo Semestre de 2022**

**Coordenadora: Clarice G.B. Demétrio**

|  |  |  |
| --- | --- | --- |
| Data/Horário | Palestrante | Título |
| 01/09, 14h  Presencial | Clarice G.B. Demétrio (ESALQ/USP)  (Lab A) | Collaboration: a way of changing lives by sharing minds |
| 08/09, 14h  Presencial | Idemauro A. R. de Lara (ESALQ/USP)  (Lab A) | Experimental Design, Categorical Data and Models |
| 15/09, 14h  Online | Rafael de A. Moral (University of Maynooth, Ireland) | Profiling Television Watching Behaviour Using Bayesian Hierarchical Models |
| 20/09, 8h  Online | Chris Brien (University of Adelaide, Australia) | [The design and analysis of a two-phase experiments involving human subjects: a case study](https://briencj.github.io/webcjb/multitier/TwoPhaseHumanExpt.v1.ppsx) |
| 22/09, 14h  Online | Elias T. Krainski (KAUST) | Implementing a class of non-stationary and non-separable spatio-temporal models |
| 30/09, 10h  Presencial | Gustovo Yun, Cristian Marcelo Villegas Lobos  (Lab A) | Tutorial sobre mapas do Brasil em R. Aplicação conjunta dos pacotes *geobr*, *ggplot2* e *tidyverse* |
| 06/10, 10h | ??????????????  (Lab A) | Semana Luiz de Queiroz |
| 14/10, 10h | ?????????????  (Lab A) |  |
| 21/10, 10h | Daiane Zuanetti (UFSCar)  (Lab A) | Data-driven reversible jump algorithms for variable selection |
| 28/10, 10h  Presencial | Rodrigo Domiciano Marques (ESALQ/USP)(Lab A) | Introdução à criação de pacotes em R |
| 04/11, 10h | Breno, Gabriela, Rodrigo, Talita  (Lab A) | Introdução ao Latex (????) |
| 10/11, 14h  Online | Geert Molenberghs (University of Hasselt, Bélgica) | Handling negative correlation and/or underdispersion in Gaussian and Non-Gaussian hierarchical data. |
| 18/11, 10h | livre | Semana RBRas |
| 25/11, 10h | John Hinde  (Lab A) | Random Effect models for grouped toxicological data: modelling variation and induced clustering |
| 01/12, 10h  Presencial | Clarice G.B. Demétrio (ESALQ/USP) |  |
| 09/12, 10h  Presencial | Deoclécio Jardim Amorim (ESALQ/USP)  (Lab A) |  |

**Data/Horário:** 01/09, 14h – Presencial (português)

**Palestrante:** Clarice G.B. Demétrio (ESALQ/USP)

**Title:** Collaboration: a way of changing lives by sharing minds

**Summary:**

It shows how collaboration can change an academic life.

**Data/Horário:** 08/09, 14h – Presencial (português)

**Palestrante:** Idemauro A. R. de Lara (ESALQ/USP)

**Title:** Experimental Design, categorical data and models

**Summary:**

Experimental designs with response variable categorical and polytomous (nominal or ordinal) are common in agrarian sciences. The aim of this seminar is to present a motivational introduction to categorical data analysis, whose procedures are an integral part of multivariate methods. We present some motivational studies related to sensorial analysis, animal behaviour, citriculture and entomology and, additionally, we discuss  sampling schemes and probability distributions for the response variable. Regression models for categorical data are also presented, in cross-sectional and longitudinal contexts, including recent research results. Finally, topics for future research in this area are presented.

**Data/Horário:** 15/09, 14h – Online (português)

**Palestrante:** Rafael de Andrade Moral (University of Maynooth, Ireland)

**Title:** Profiling Television Watching Behaviour Using Bayesian Hierarchical Models

**Summary:**

Customer churn prediction is a valuable task in many industries. In telecommunications it presents great challenges, given the high dimensionality of the data, and how difficult it is to identify underlying frustration signatures, which may represent an important driver regarding future churn behaviour. Here, we propose a novel Bayesian hierarchical joint model for count and time-to-event data, that is able to characterise customer profiles based on how many events take place within different television watching journeys, and how long it takes between events. The model drastically reduces the dimensionality of the data from thousands of observations per customer to 11 customer-level parameter estimates and random effects. We test our methodology using data from 40 BT customers (20 active and 20 who eventually cancelled their subscription) whose TV watching behaviours were recorded from October to December 2019, totalling more than 400,000 observations.

**Data/Horário:** 20/09, 8h – Online (inglês)

**Palestrante:** Chris Brien (University of Adelaide, Australia)

**Title:** The design and analysis of a two-phase experiments involving human subjects: a case study

**Summary**

Two-phase experiments were introduced in 1952 by the Australian statistician George McIntyre. Their use has been most prevalent in agriculture experiments, especially plant breeding experiments. However, there is the potential for their application to be much more widespread. This potential is not being realized because of a lack of awareness of multiphase experiments within the statistical community. A potted history of two-phase experiments will be given and an explanation of what constitutes a two-phase experiment provided, using a scenario in a sports Science context as an introductory example. The use of the anatomy of a design for understanding the confounding in the experiment will be outlined. The case study involves a pain-rating experiment reported in a 1997 paper by Solomon, Prkachin, &amp; Farewell. The data from a subset of the experiment was analyzed by Farewell and Herberg (2003) and by Jarrett, Farewell and Herzberg (2020) using analyses-of-variance and linear mixed models. A further re-analysis using linear mixed models that has been described by Brien (2022) will be outlined. Suggested improvements for the design of future similar experiments, based on the results of the re-analysis, will also be presented.

**References**

Brien, C. J. (2022). Designing, understanding and modelling two-phase experiments with human subjects. Statistical Methods in Medical Research, 31(4), 626-645.

Farewell, V. T.; Herzberg, A. M. (2003). Plaid designs for the evaluation of training for medical practitioners. Journal of Applied Statistics, 30(9), 957-965.

Jarrett, R. G., Farewell, V. T.; Herzberg, A. M. (2020). Random effects models for complex designs. Statistical Methods in Medical Research, 29(12), 3695-3706.

Solomon, P. E., Prkachin, K. M.; Farewell, V. (1997). Enhancing sensitivity to facial expression of pain. Pain, 71(3), 279-284.

**Data/Horário:** 22/09, 14h – Online (português)

**Palestrante:** Elias T. Krainski (King Abdullah University of Science and Technology, KAUST, Arábia Saudita)

**Title:** Implementing a class of non-stationary and non-separable spatio-temporal models

**Summary:**

In this talk we start introducing the building block framework commonly considered in the statistical analysis of real data. The modeling framework to build models considering the Stochastic Partial Differential Equations (SPDE) approach is briefly introduced. We then consider a class of non-separable spatio-temporal SPDE models and a simple way to introduce non-stationarity. Some of the computational implementation details are considered and illustrated. A real dataset is considered to illustrate the methodology.

**Data/Horário:** 30/09, 10h – Presencial (português)

**Palestrantes:** Gustavo Yun, Cristian Marcelo Villegas Lobos

**Title:** Tutorial sobre mapas do Brasil em R. Aplicação conjunta dos pacotes *geobr*, *ggplot2* e *tidyverse*

**Summary:**

O R é uma linguagem de programação gratuita e de código aberto, que se baseia na atuação direta de seus usuários para a promoção de facilidades, melhorias e inovações acessíveis a partir de pacotes. Essa linguagem é muito utilizada para análise de dados, seja para importação, manipulação, análise estatística ou divulgação dos resultados. Uma das possibilidades da linguagem R é a elaboração de mapas. Neste tutorial, propõe-se construir, de maneira participativa, mapas do Brasil no âmbito nacional, estadual e municipal. Para isso, será abordado o uso dos pacotes *geobr* (PEREIRA; GONÇALVES, 2019) para acessar dados geoespaciais de diferentes escalas geográficas do Brasil, o *ggplot2* (WICKHAM, 2016) como ferramenta de confecção dos mapas e o *tidyverse* (WICKHAM et al., 2019) para manipulação das bases de dados.

**Referências**

PEREIRA, R.H.M.; GONÇALVES, C.N.. **geobr:** Loads Shapefiles of Official Spatial Data Sets of Brazil, 2019. Disponível em: https://github.com/ipeaGIT/geobr.

WICKHAM, H.. ggplot2: Elegant Graphics for Data Analysis. **Springer-Verlag**, New York, 2016. Disponível em: https://ggplot2.tidyverse.org.

WICKHAM, H. et al.. Welcome to the tidyverse. **Journal of Open Source Software**, 4(43), 1686, 2019. doi:10.21105/joss.01686.

**Data/Horário:** 21/10, 10h – Online (português)

**Palestrante:** Daiane Zuanetti (UFSCar)

**Title:** Data-driven reversible jump algorithms for variable selection

**Summary:**

We propose data-driven reversible jump algorithms for selecting variables in a regression model. Our main motivation is to identify relevant SNPs for a specific phenotype, estimate their additive, dominance and epistasis (interaction) effects in genetic dataset that usually is a problem involving high-dimension data. We compare the DDRJ’s performance with other traditional methods of variable selection, using real and simulated data sets, and it shows good performance. Joint work with Prof. Dr. Luis Milan.

**Data/Horário:** 28/10, 10h – Presencial (português)

### Palestrante: Rodrigo Domiciano Marques (ESALQ/USP)

**Title:** Introdução à criação de pacotes em R

**Summary:**

O R é uma linguagem e ambiente para estatística computacional, de livre distribuição e código aberto para análise estatística e gráfica. Também dispõe de ferramentas mais amplas e versáteis que permitem a seus usuários a criação e compartilhamento de compilados de funções na forma de pacotes ou bibliotecas. Dentre as várias formas para se criar um pacote no R, será abordado o uso da interface R-Studio e dos pacotes devtools para criar, checar, documentar as funções as serem compiladas e importar funções de outros pacotes, e roxygen2 para edição dos arquivos de documentação e páginas de ajuda ao usuário.

**Referências**

MAYER, Fernando; ZEVIANI, Walmes. Pacotes R. Cursos LEG, 2016. Disponível em http://cursos.leg.ufpr.br/prr/capPacR.html Acesso em 31/08/2022.

RSTUDIO Cheatsheets. RStudio, 2022. Disponível em

http://www.rstudio.com/resources/cheatsheets; Acesso em 31/08/2022.

SILGE, Julia; NASH, John C.; GRAVES, Spencer. Navigating the R Package Universe. R Journal, v. 10, n. 2, p. 558, 2018.

WICKHAM, Hadley, BRYAN, Jenny. R packages: organize, test, document, and share your code. &quot; O&#39;Reilly Media, Inc.&quot;, 2022. Disponível em

<https://r-pkgs.org/> Acesso em 31/08/2022.

**Data/Horário:** 10/11, 14h – Online (inglês)

**Palestrante:** Geert Molenberghs (University of Hasselt, Bélgica)

**Title:** Handling negative correlation and/or underdispersion in Gaussian and Non-Gaussian hierarchical data

**Summary:**

The occurrence and interpretation of negative variance components in the context of linear mixed models is well understood at this point, even though the issue is surrounded by subtle issues for estimation and testing (Verbeke and Molenberghs 2003, Molenberghs and Verbeke 2007). Broadly, negative variance components often point to negative within-cluster correlation. It is even possible to give such linear mixed models a meaningful hierarchical interpretation (Molenberghs and Verbeke 2011). Matters are more complicated when the outcomes are non-Gaussian, either in the context of the generalized linear mixed model, or extensions thereof that allow for flexible modeling of both within-unit correlation as well as overdispersion (Molenberghs *et al.* 2010). An additional complication is that, in practice, not only negative variance components due to negative correlation, but also underdispersion instead of overdispersion can occur, sometimes even jointly. With focus on both continuous and count data, we describe how models can be made sufficiently flexible and, in a number of cases, interpreted hierarchically (Luyts *et al.* 2019).

**References**

Luyts, M., Molenberghs, G., Verbeke, G., Matthijs, K., Demétrio, C.G.B., and Hinde, J. (2019). A Weibull-count approach for handling under- and/or over-dispersed clustered data structures. *Statistical Modeling,***19,** 569-589.

Molenberghs, G. and Verbeke, G. (2007). Likelihood ratio, score, and Wald tests in a constrained parameter space. *The American Statistician*, **61**, 1-6.

Molenberghs, G. and Verbeke, G. (2011). A note on a hierarchical interpretation for negative variance components. *Statistical Modeling*, **11**, 389-408.

Molenberghs, G., Verbeke, G., Demétrio, C.G.B., and Vieira, A. (2010). A family of generalized linear models for repeated measures with normal and conjugate random effects. *Statistical Science*, **25**, 325-347.

Verbeke, G. and Molenberghs, G. (2003). The use of score tests for inference on variance components. *Biometrics*, **59**, 254-262.

**Data/Horário:** 25/11, 10h – Presencial (inglês)

**Palestrante:** John Hinde (University of Galway)

**Title:** Random Effect models for grouped toxicological data: modelling variation and induced clustering

**Summary:**

This talk will focus on the anlaysis of data from a toxicological study involving the use of a microbial control for termite that attacks sugar cane. The study involves different isolates of the control applied to groups of the termites in a completely randomised experiment with replicates. The insect mortatlity within the groups is measured daily for 8 days and can be treated as a multinomial reponse or a a form of grouped discrete survival data. We consider models from each each of these approaches and focus on assessing the efficacy and reprodicubility of the various isolates. This leads us to include random affects for the replicate variation and also to treat the isolates themselves as a random factor. This allows the classification of the isolates and, in particular, by using a finite mixture representation for the isolate random effect we are able to cluster them implicitly. The analyses will illustrate various different mixed model fitting approaches available in different R packages.

**Data/Horário:** 01/12, 10h – Presencial (português)

**Palestrante:** Clarice G.B. Demétrio (ESALQ/USP)

**Title:**

**Summary:**