

Universidade Federal do Rio Grande do Sul  
Instituto de Matemática e Estatística  
Departamento de Estatística



## Anais de Resumos

XI SEMANÍSTICA

XI Semana Acadêmica do Departamento de Estatística  
da UFRGS

<http://www.ufrgs.br/semanistica>

Porto Alegre - 27, 28 e 29 de outubro de 2021

## 6 Comunicações Orais da Pós-Graduação

### Mini Palestra 1:

#### Métodos Estatísticos para o Estudo da Dinâmica Migratória do Vírus da Gripe

Aline Foerster Grande

**Resumo:** O objetivo desse trabalho é modelar a incidência do vírus influenza no Brasil em um determinado mês  $t$ , a partir de dados de incidência e diversidade genética coletados em meses anteriores em outras localidades. Para tanto, são aplicados os modelos de Granger causalidade e de regressão com defasagens, que visam determinar a influência de determinadas variáveis na previsão de outras, bem como permitem uma modelagem fina de causalidade no contexto de séries temporais. Os modelos propostos podem ser utilizados na previsão da incidência da gripe no Brasil, conhecimento que pode ser estratégico para a implementação de políticas de vacinação pelo Governo, bem como desenvolvimento de estratégias ótimas de controle de epidemias desta doença. Tal análise mostrou-se fundamental neste último ano, uma vez que como os casos graves de influenza requerem hospitalização e cuidados intensivos, incluindo a necessidade de respiradores artificiais, e com a pandemia da COVID-19, esses bens preciosos tornaram-se escassos em muitos países. Assim, a previsão dos casos de influenza através desses modelos é essencial para orientar o sistema público de saúde na alocação de recursos e planejamento para a concomitância das duas doenças no inverno.

### Mini Palestra 2:

#### Quantile autoregressive distributed lag model global variable selection

Taís Loureiro Bellini, Eduardo Horta

**Resumo:** Quantile regression models the conditional distribution of a response variable  $Y$  on the vector of covariates  $X$  at different quantile levels offering a description for the whole conditional distribution (Koenker and Xiao (2006)). Using quantile regression, as opposed to traditional linear regression, provides several advantages, but can add complexity to certain operations. For example, when performing variable selection, there might be a different set of variables selected for each quantile. Frumento and Bottai (2016) propose a parametric modeling of quantile regression coefficient functions that allows us to estimate in one single minimization problem the coefficients for all quantiles in a grid.

Sottile et al. (2020) use this approach to perform LASSO variable selection using information on all quantiles simultaneously. In this work, we propose a global coefficient estimation and variable selection method based on the estimator presented in Sottile et al. (2020), introducing the group LASSO penalty, suggested in Yuan and Lin (2006), and applied in a Quantile Autoregressive Distributed Lag (QADL) model. Furthermore, since we are in a time series context, we also evaluate the variable selection penalization applying higher penalties to higher lags, as proposed in Konzen and Ziegelmann (2016). The results suggest that a weighted penalized approach can provide better results in selecting the variables as well as in estimating the coefficients. In particular, both LASSO and group LASSO penalization with higher weights for higher lags were the ones that had lower mean squared error to estimate most of the tested scenarios and set the zero coefficients correctly more often.

### Mini Palestra 3:

#### Correlation Selection in phylogenetic multivariate probit models

Felipe Grillo Pinheiro, Taiane Schaedler Prass e Gabirela Bettella Cybis

**Resumo:** The multivariate phylogenetic latent liability model, first proposed by Cybis et al. (2015), and the recent phylogenetic multivariate probit model (PMPM), developed by Zhang et al. (2021), are important tools for investigating the association structure between mixed-type biological traits controlling for the shared evolutionary history of related organisms. We model these associations through the correlation matrix,  $R$ , of the latent Gaussian variables in a multivariate Brownian diffusion process along a phylogenetic tree informed by molecular sequences. However, besides the well-known limitation of parameter identifiability in probit models, another difficulty lies in the arbitrary criteria used to determine significance of these associations. Correlations have been considered significant if a chosen percentage highest posterior density (HPD) interval does not contain zero. Estimating sparse correlation matrices provides both, a systematic solution for elimination of spurious correlations and parameter reduction, which is a major gain since the number of parameters scales quadratically in trait dimension. However, due to model assumptions or identifiability reasons, when the covariance matrix is assumed to be a correlation matrix, as in probit models, the options for prior distributions on  $R$  are limited, especially if one requires sparsity in addition. To bypass this limitation, Bayesian inference for probit models is usually performed using the data augmentation representation of Chib and Greenberg (1998), where the binary traits, that require unit variance to be identifiable, are rescaled. Consequently, the correlation matrix is expanded to a covariance matrix that can be modelled using standard conjugate priors