

REGRESSION MODELS FOR RESPONSES IN THE UNIT INTERVAL: SPECIFICATION, ESTIMATION AND COMPARISON

Wagner Hugo BONAT¹
Paulo Justiniano RIBEIRO JUNIOR¹
Walmes Marques ZEVIANI¹

- **ABSTRACT:** *Regression models are widely used on a diversity of application areas to describe associations between explanatory and response variables. The initially and most adopted form of Gaussian linear models was gradually extended to accommodate different kinds of response variables. Several of such models were described as particular cases of the class of the generalized linear models (GLM) which allows, under the same framework, a diversity of the formats for the response variable and functions linking the parameters of the distribution to a linear predictor. Since then GLM's structure became benchmark for several further extensions and developments in statistical modeling such as generalized additive models, overdispersion, among others. Response variables with values restricted to an interval most often (0, 1), are usual in social sciences, agronomy, psychometrics among other areas. The Beta or the Simplex distributions are often adopted although other options are mentioned in the literature. A set of regression models for restricted response variables which includes the usually adopted formats and also allows for a wider range of models are declared here under a general specification. Individual models are defined by the choices for the three components, the probability distribution for the response, the link function between a parameter of the distribution of choice and the linear predictor and the transformation function for the response. We report results of the analysis for four different datasets considering Beta, Simplex, Kumaraswamy and Gaussian distributions, and logit, probit, complementary log-log, log-log, Cauchit and Aranda-Ordaz as options for the link and transformation functions. Likelihood based analysis for model fitting, comparison and choice are carried out on a unified way and computer code is made available. Results shows there is no prominent model within the class illustrating the importance of tools for exploring a wide class of models at each analysis.*
- **KEYWORDS:** *Maximum likelihood; restricted variables; proportions; indexes; rates.*

¹ Universidade Federal do Paraná – UFPR, Laboratório de Estatística e Geoinformação – LEG, Departamento de Estatística – DEST, CEP: 81531-990, Curitiba, PR, Brasil. E-mail: wagner@leg.ufpr.br / paulojus@leg.ufpr.br / walmes@leg.ufpr.br