

1.1 Comunicações Orais 1 - CO1

1.1.1 CO1.1 - A new flexible gamma generalized model with properties, applications and Bayesian approach.

Authors: *Fábio Prata Vieira; Gauss M. Cordeiro; Adriano K. Suzuki; Edwin M. M. Ortega.*

Abstract: We propose a new lifetime model called the odd log-logistic generalized gamma distribution that can be easily interpreted. Some of its special models are discussed. We obtain general mathematical properties of this distribution including the ordinary moments, and quantile functions. We discuss parameter estimation by the maximum likelihood method and a Bayesian approach, where Gibbs algorithms along with metropolis steps are used to obtain the posterior summaries of interest for survival data with right censoring. Further, for different parameter settings, sample sizes and censoring percentages, we perform various simulations and evaluate the behavior of the estimators. The potentiality of the new distribution is proved by means of two real data sets. In fact, the new distribution can produce better fits than some well-known distributions.

1.1.2 CO1.2 - Prior Specifications to Handle the Monotone Likelihood Issue in the Cox Regression Model.

Authors: *Frederico M. Almeida; Enrico A. Colosimo e Vinícius D. Mayrink.*

Abstract: The phenomenon of monotone likelihood is observed in the fitting process of a Cox model when the likelihood converges to a finite value while at least one parameter estimate diverges to infinity. Monotone likelihood primarily occurs in samples with substantial censoring of survival times and associated to categorical covariates. In particular and more frequent, it occurs when one level of a categorical covariate has not experienced any failure. A solution suggested by Heinze and Schemper (2001) is an adaptation of a procedure by Firth (1993) originally developed to reduce the bias of maximum likelihood estimates. The method leads to finite parameter estimates by means of penalized maximum likelihood estimation. In this case, the penalty might be interpreted as a Jeffreys type of prior well known in Bayesian inference. However, this approach has some drawbacks, especially biased estimators and high standard errors. In this paper, we explore other penalties for the partial likelihood function in the flavor of Bayesian prior distributions. An empirical study of the suggested procedures confirms satisfactory performance of both estimation and inference. We also explore a real analysis related to a melanoma skin data set to evaluate the impact of the different prior distributions as penalizations.

1.1.3 CO1.3 - Bayesian Approach for the Zero-Modified Poisson-Lindley Regression Model.

Authors: *Wesley Bertoli da Silva; Katiane S. Conceição; Marinho G. Andrade; Francisco Louzada Neto.*

Abstract: In this paper we propose the zero-modified Poisson-Lindley regression model as an alternative to model overdispersed count data exhibiting inflation or deflation of zeros in the presence of covariates. It will be shown that the zero modification can be incorporated by using the zero-

truncated Poisson-Lindley distribution. A simple reparametrization of the probability function will allow us to represent the zero-modified Poisson-Lindley distribution as a hurdle model. This trick leads to the fact that proposed model can be fitted without any previously information about the zero modification present in a given dataset. Bayesian procedures will be considered for estimation and inference concerns. A sensitivity study to detect points which can influence the parameter estimates will be performed based on the Kullback-Leibler divergence. A simulation study is presented in order to illustrate the performance of the developed methodology. The usefulness of the proposed model will be evaluated using a real dataset of faecal egg counts of gastrointestinal nematodes from 282 cows. Standard comparison respect to the Poisson, Poisson-Lindley and zero-modified Poisson regression models is also provided.

1.2 Comunicações Orais 2 - CO2

1.2.1 CO2.1 - Finite Mixture of Regression Models for Censored Data Based on Scale Mixtures of Normal Distributions.

Authors: *Camila Borelli Zeller; Celso Rômulo Barbosa Cabral; Víctor Hugo Lachos.*

Abstract: In statistical analysis, particularly in econometrics, the finite mixture of regression models based on the normality assumption is routinely used to analyze censored data. In this work, an extension of this model is proposed by considering scale mixtures of normal distributions (SMSN). This approach allows us to model data with great flexibility, accommodating multimodality and heavy tails at the same time. The main virtue of considering the finite mixture of regression models for censored data under the SMSN class is that this class of models has a nice hierarchical representation which allows easy implementation of inference. We develop a simple EM-type algorithm to perform maximum likelihood (ML) inference of the parameters of the proposed model. To examine the performance of the proposed method, we present some simulation studies and analyze a real dataset.

1.2.2 CO2.2 - A multivariate linear regression analysis of censored data using the finite mixture of multivariate Student-t distribution.

Authors: *Victor H. Lachos, Edgar J. L. Moreno, Celso R. B. Cabral.*

Abstract: Mixture multivariate regression models are based on the assumption of normality (symmetry) and thus are sensitive to outliers, heavy-tailed and skewness. Besides this, these kind of data can be subjected to some upper and/or lower detection limits because of the restriction of experimental apparatus. For such data structures, we propose an extension of the classical normal model based on finite mixtures of multivariate Student- t distributions that deals with these issues simultaneously. This approach allows us to model data with great flexibility, accommodating multimodality, heavy tails and also skewness depending on the structure of the mixture components. We develop an analytically simple yet efficient EM-type algorithm for conducting maximum likelihood estimation of the parameters. The algorithm has closed-form expressions at the E-step, that rely on formulas for the mean and variance of the multivariate truncated Student- t distributions. Further, a general information-based method for approximating the asymptotic covariance matrix of the estimators is also presented. Results obtained from the analysis of both simulated and real data sets are reported to demonstrate the effectiveness of the proposed methodology. The proposed algorithm and methods are implemented in the new R package CensMixReg.

1.2.3 CO2.3 - Robust Regression Modeling for Censored Data based on Mixtures of Student-t Distributions.

Authors: *Víctor H. Lachos; Luis Benites; Celso R. Barbosa Cabral; Dipak K. Dey.*

Abstract: In the framework of censored regression models, the distribution of the error terms departs significantly from normality, for instance, in the presence of heavy tails, skewness and/or atypical observation. In this paper we extend the censored linear regression model with normal errors to the case where the random errors follow a finite mixture of Student-t distributions. This approach allows us to model data with great flexibility, accommodating multimodality, heavy tails and also skewness depending on the structure of the mixture components. We develop an analytically tractable and efficient EM-type algorithm for iteratively computing maximum likelihood estimates of the parameters, with standard errors as a by-product. The algorithm has closed-form expressions at the E-step, that rely on formulas for the mean and variance of the truncated Student-t distributions. The efficacy of the method is verified through the analysis of simulated and real datasets. The proposed algorithm and methods are implemented in the new R package CensMixReg.

1.3 Comunicações Oraís 3 - CO3

1.3.1 CO3.1 - Multistate Model Survival Model to Monitoring Action on Labour.

Authors: *Gleici da Silva Castro Perdoná; Alexandre Cristovão Maiorano; Hayala C. Cavenague de Souza; Francisco Louzada Neto; João Paulo Dias de Souza.*

Abstract: In this work we presents a propose methods statistical to solve problems in area of Reproductive Health. The project namely, SELMA, is part of a major project called BOLD (Better Outcomes in Labour Difficulty). The objective is the development of the Simplified, Effective, Labor Monitoring to Action (SELMA) Tool. SELMA is an algorithm to guide the health professionals, offering a solution to avert intrapartum-related fetal death of babies (stillbirth) and neonatal death or severe morbidity. The version of SELMA tool, which is presented here, is supplied by multistate survival and longitudinal models that provide the probabilities of occurrence of poor labour-related outcomes. The procedure is illustrated on data extracted from a prospective cohort study conducted in eight health facilities in Nigeria and Uganda.

1.3.2 CO3.2 - The Poisson-Weibull Regression Model.

Authors: *Valdemiro Piedade Vidas, Giovana Oliveira Silva, Francisco Louzada Neto.*

Abstract: In this paper, we proposed a regression model in the form location-scale based on the Poisson-Weibull distribution. This distribution arises on the competitive risk scenario. The new regression model was proposed for modeling data which has a increasing, decreasing and unimodal failure rate, and presented as particular cases the new exponential-Poisson regression model and the Weibull regression model. Assuming censored data, we considered the maximum likelihood approach for parameters estimation. For different parameter values, sample sizes and censoring percentages, various simulation studies were performed to study the means, bias relative and mean square error of the maximum likelihood estimative, and to compare the performance of the Poisson-Weibull regression model and their particular cases. The selection criteria AIC and likelihood ratio test were used for selection of regression model. Besides, we used the sensitivity analysis to detect in uential or outlying observations and residual analysis was used to check assumptions in the model. The relevance of the approach was illustrated with a data set.

1.3.3 CO3.3 - On the Comparison of Risk of Death According to Different Stages of Breast Cancer via the Long-Term Exponentiated Weibull Hazard Model.

Authors: *Hayala C. Cavenague de Souza; Gleici da Silva Castro Perdoná; Francisco Louzada Neto.*

Abstract: Long-term survivors models have been extensively used for modelling time-to-event data with a significant proportion of patients whose do not experience poor outcome. In this paper we propose a new long-term survivor hazard model, which accommodates comprehensive families of cure rate models as particular cases, including Modified Weibull, Exponentiated Weibull, Weibull, Exponential and Rayleigh distribution, among others. The maximum likelihood estimation procedure is presented. A simulation study evaluates bias and mean square error of the considered estimation procedure as well as the coverage probabilities of the parameters asymptotic and bootstrap confidence intervals. A real Brazilian dataset on breast cancer illustrates the methodology. From the practical point of view, under our modelling, we provide a parameter that works as a metric to quantify and compare the risk between different stages of the disease. We emphasize that, we developed an online platform for oncologists to calculate the probability of survival of patients diagnosed with breast cancer according to the stage of the disease in real time.

1.4 Comunicações Orais 4 - CO4

1.4.1 CO4.1 - Negative Binomial Kumaraswamy-G Cure Rate Regression Model.

Authors: *Amanda Morales Eudes D'Andrea; Ricardo Rocha; Vera Lucia Damasceno Tomazella.*

Abstract: In survival analysis the presence of elements not susceptible to the event of interest are very common. These elements lead to what is called a fraction cure, or cure rate, or even long-term survivors. In this paper, we propose a unified approach using the negative binomial distribution for modeling cure rates under the Kumaraswamy family of distributions. The estimate is made by maximum likelihood. Furthermore, we propose an approach to the Negative Binomial Kumaraswamy-G regression model. Finally, we illustrate the distributions proposed using a real data set related to a melanoma study.

1.4.2 CO4.2 - Defective models induced by Gamma frailty term for survival data with cured fraction.

Authors: *Juliana Scudilio; Vinicius Calsavara; Ricardo Rocha; Francisco Louzada; Vera Tomazella; Agatha Sacramento.*

Abstract: In this paper we propose a defective model induced by a frailty term to fit survival data with cure fraction. Defective models have the advantage of modeling the proportion of cured without adding any extra parameters in the model, the opposite to what happens in the most models from the literature. Models with frailty terms incorporate an unobserved heterogeneity between individuals and this incorporation brings advantages for the estimated model. Because describe the influence of unobserved covariates in a proportional hazard model. We propose and discuss the defective Gamma-Gompertz and Gamma-inverse Gaussian regression models. Where the frailty term is assumed a Gamma distribution and baseline rate is the Gompertz defective distribution or Gaussian inverse defective distribution. Simulation studies are performed to verify the asymptotic properties of the maximum likelihood estimator. Finally, we apply the proposed models to two real data sets to illustrate the proposed methodology, in which one of them is newly added to the literature.

1.4.3 CO4.3 - Defective Models for Cure Rate Modeling with Interval-Censored Data.

Authors: *Vinicius F. Calsavara; Agatha S. Rodrigues; Ricardo Rocha; Vera Tomazella.*

Abstract: The regression models in survival analysis are most commonly applied for right-censored survival data. Here, we present a defective model for cure rate modeling for interval-censored event-time data. Defective distributions are characterized by having density function that integrate to values less than one when the domain of their parameter is different from the usual one. We consider the Gompertz and inverse Gaussian defective distributions which allow to model data containing cured elements, that is, a cure fraction. The parameter estimation is reached by maximum likelihood estimation procedure and Monte Carlo simulation studies are considered in order to evaluate the proposed models performance. We analyze two real data set from the medical area to illustrate the practice relevance of the proposed methodology.

1.5 Comunicações Oraís 5 - CO5

1.5.1 CO5.1 - Optimal burn-in policy based on a collection of cutoff points using mixture Inverse Gaussian degradation process and Copula theory.

Authors: *Lia Hanna Martins Morita; Vera Lucia Tomazella; Narayanaswamy Balakrishnan.*

Abstract: Burn-in tests have been discussed in reliability literature, wherein we operate those items until observe high degradation values that could possibly separate the weak items from the normal items, before being delivered to consumers. This concept is often called screening procedure in the degradation literature and is subjected to misclassification errors, which are of much importance in burn-in studies. Commonly, the underlying degradation process is assumed to be a Wiener process or a Gamma process and several papers concerning optimal burn-in policies have been discussed in the literature assuming that the underlying degradation path follows a Wiener process. However in this paper we propose the mixture inverse Gaussian process which can deal with monotone paths and has interest properties. We present a decision rule for classifying an item as normal or weak, based on burn-in time and a collection of cutoff points. Then, an economic cost model is used to determine the optimal burn-in time and the optimal cutoff points, whose estimation is based on analytical method or approximated method involving Copula theory. Finally an example of real LASER data set in the reliability literature is presented to illustrate the proposed procedure.

1.5.2 CO5.2 - Where geography lives? A projection approach for spatial confounding.

Authors: *Renato Martins Assunção; Marcos Oliveira Prates; Erica Castilho Rodrigues.*

Abstract: Spatial confounding between the spatial random effects and fixed effects covariates has been recently discovered and showed that it may bring misleading interpretation to the model results. Solutions to alleviate this problem are based on decomposing the spatial random effect and fitting a restricted spatial regression. In this paper, we propose a different approach: a transformation of the geographic space to ensure that the unobserved spatial random effect added to the regression is orthogonal to the fixed effects covariates. Our approach, named SPOCK, has the additional benefit of providing a fast and simple computational method to estimate the parameters. Furthermore, it does not constrain the distribution class assumed for the spatial error term. A simulation study and a real data analysis are presented to better understand the advantages of the new method in comparison with the existing ones.

1.5.3 CO5.3 - Avaliação de produtos através da verificação da presença de $m \geq 1$ quesitos.

Authors: Démerson André Polli; Carlos Alberto de Ribeiro Diniz.

Resumo: Este trabalho apresenta dois modelos simples, construídos a partir de misturas de distribuições discretas, para a modelagem de avaliações de produtos ou serviços através de uma nota associada a $m \geq 1$ quesitos que podem ser julgados satisfatórios ou não. A nota final atribuída pelos avaliadores é o número de quesitos (dentre m) julgados como satisfatórios. Os modelos apresentados baseiam-se na premissa que as avaliações podem ser feitas de forma consciente, avaliando os quesitos, ou aleatoriamente, inclusive com inflação em uma determinada nota $0 \leq N \leq m$.

1.6 Comunicações Orais 6 - CO6

1.6.1 CO6.1 - The Gamma Burr XII Distribution: Theory and Practice.

Authors: Edleide de Brito; Giovana O. Silva; Gauss M. Cordeiro; Clarice G. B. Demétrio; Maristela D. Oliveira.

Abstract: We propose a new four-parameter distribution called the gamma Burr XII distribution. The new model contains as special cases some well-known distributions discussed in the literature such as the logistic and Burr XII distributions, among several others. We obtain the moments, incomplete moments, generating and quantile functions, mean deviations, Bonferroni and Lorenz curves and reliability. Besides, we introduced a regression model considering the new gamma Burr XII distribution. The method of maximum likelihood is used for estimating the model parameters. We determine the observed information matrix. Finally, the usefulness and flexibility of the family is illustrated using three real data sets.

1.6.2 CO6.2 - Improved hypothesis testing in a general multivariate elliptical model.

Authors: Tatiane F. N. Melo; Silvia L. P. Ferrari e Alexandre G. Patriota.

Abstract: This paper investigates improved testing inferences under a general multivariate elliptical regression model. The model is very flexible in terms of the specification of the mean vector and the dispersion matrix, and of the choice of the error distribution. The error terms are allowed to follow a multivariate distribution in the class of the elliptical distributions, which has the multivariate normal and Student-t distributions as special cases. We obtain Skovgaard's adjusted likelihood ratio statistics and Barndorff-Nielsen's adjusted signed likelihood ratio statistics and we compare the methods through simulations. The simulations suggest that the proposed tests display superior finite sample behavior as compared to the standard tests. Two applications are presented in order to illustrate the methods.

1.6.3 CO6.3 - Métodos de Regressão para o Problema de Quantificação.

Authors: Afonso Vaz; Rafael Izbicki; Rafael Stern.

Abstract: Em várias aplicações de aprendizagem estatística, utilizamos classificadores como uma maneira de melhorar a estimativa da proporção de ocorrência da classe positiva em uma determinada população, da qual não temos um número elevado de amostras rotuladas. Essa tarefa é denominada Quantificação. Por outro lado, considerando o mesmo cenário, podemos estar interessados em avaliar a associação entre a proporção e uma determinada covariável. Por exemplo: Uma empresa pode estar interessada em avaliar como a proporção de resenhas positivas de uma página do Facebook/Twitter, sobre um de seus produtos, evolui ao longo do tempo. Uma estratégia para avaliar esse

comportamento é considerar um modelo de regressão entre os dados provenientes de um estudo de quantificação e o tempo. Neste trabalho, apresentamos os elementos relacionados ao problema de regressão para quantificação. Descrevemos, brevemente, dois métodos comumente utilizados para contornar a situação, bem como, propomos dois métodos inéditos. Apresentamos simulações para comparar os métodos, assim como os analisamos sob o ponto de vista teórico.

1.7 Comunicações Orais 7 - CO7

1.7.1 CO7.1 - Penalized Maximum Likelihood Estimation in the Modified Extended Weibull Distribution.

Authors: *Verônica Maria Cadena Lima; Francisco Cribari Neto.*

Abstract: We address the issue of performing inference on the parameters that index the modified extended Weibull (MEW) distribution. We show that numerical maximization of the the MEW log-likelihood function can be problematic. It is even possible to encounter maximum likelihood estimates that are not finite, i.e., it is possible to encounter monotonic likelihood functions. We consider different penalization schemes to improve maximum likelihood point estimation. A penalization scheme based on the Jeffreys invariant prior is shown to be particularly useful. Simulation results on point estimation, interval estimation and hypothesis testing inference are presented. Two empirical applications are presented and discussed.

1.7.2 CO7.2 - Linear Regression Models with Finite Mixtures of Skew Heavy-Tailed Errors.

Authors: *Luis Benites; Rocío Maehara; Victor H. Lachos.*

Abstract: We consider estimation of regression models whose error terms follow a finite mixture of scale mixtures of skew-normal (SMSN) distributions, a rich class of distributions that contains the skew-normal, skew-t, skew-slash and skew-contaminated normal distributions as proper elements. This approach allows us to model data with great flexibility, accommodating simultaneously multimodality, skewness and heavy tails. We developed a simple EM-type algorithm to perform maximum likelihood (ML) inference of the parameters of the proposed model with closed form expressions for both E- and M-steps. Furthermore, The empirical information matrix is derived analytically to account for standard errors. The practical utility of the new method is illustrated with the analysis of a real dataset and several simulation studies. The proposed algorithm and methods are implemented in the R package `FMsmnReg()`.

1.7.3 CO7.3 - Regressão quantílica logística robusta para respostas intervalares.

Authors: *Christian E. Galarza; Victor H. Lachos.*

Abstract: Nos modelos de regressão, muitas vezes a variável de resposta é uma variável que apenas toma valores num intervalo contínuo limitado, de modo que os métodos tradicionais de estimação, tais como regressão por mínimos quadrados, modelos de efeitos mistos, modelos não paramétricos, entre outros, podem ser inadequados. Desconsiderar esta característica da resposta na maioria dos casos conduz a estimações fora dos possíveis valores que podem ser observados. A regressão quantílica (RQ) logística constitui um método eficaz para preencher esta lacuna. Robusta contra os outliers, sua inferência é válida sem importar da distribuição do erro, e as predições sempre estarão restritas ao intervalo da resposta, caracterizando completamente a distribuição. Os métodos de regressão quantílica paramétrica (frequentistas e bayesianos) são baseados na distribuição Laplace assimétrica (DLA), porém este é um suposto muito forte que raramente é satisfeito. Neste trabalho

propomos um modelo de predição robusta para respostas intervalares usando uma função simples de ligação que permite uma fácil interpretação dos parâmetros e supondo o erro com distribuição pertencente a uma nova família de distribuições assimétricas principalmente de caudas pesadas. Uma aplicação será desenvolvida para ilustrar o modelo, usando o pacote lqr do software R.