

XV MGEST

Departamento de
Estatística - UFMG



XV MGEST

ENCONTRO MINEIRO
DE ESTATÍSTICA
BELO HORIZONTE

5 e 6 de Outubro de 2017



ANAIS DO XV MGEST

Prédio da Escola de Engenharia
UFMG - Av. Antônio Carlos, 6627
Pampulha - Belo Horizonte - MG



Departamento de Estatística
Universidade Federal de Minas Gerais



XV MGEST

Encontro Mineiro de Estatística

**ANAIS DO
XV MGEST**

Departamento de Estatística

Universidade Federal de Minas Gerais

Belo Horizonte, 05 e 06 de outubro de 2017

COMISSÃO ORGANIZADORA

Coordenação Geral:

Profa. Glaura da Conceição Franco

Comissão Organizadora Local (UFMG):

Profa. Magda Carvalho Pires (Presidente)

Prof. Aloisio Joaquim Freitas Ribeiro

Profa. Denise Duarte

Profa. Lourdes Coral Contreras Montenegro

Prof. Marcos Antônio da Cunha Santos

Roger William Câmara Silva

Profa. Rosangela Helena Loschi

Prof. Thiago Rezende dos Santos

Prof. Vinícius Diniz Mayrink

Prof. Wagner Barreto de Souza

Comitê Científico

Prof. Marcos Oliveira Prates (Presidente – UFMG)

Prof. Clécio Ferreira (UFJF)

Prof. Emerson Cotta Bodevan (UFVJM)

Prof. Enrico Antonio Colosimo (UFMG)

Prof. Joel Augusto Muniz (UFLA)

Prof. Leandro Alves Pereira (UFU)

Profa. Letícia Milani Rodrigues (UNIFAL)

Prof. Luiz Alexandre Peternelli (UFV)

Prof. Ricardo Tavares (UFOP)

Equipe de apoio de alunos do DEST-UFMG

César Macieira

Guilherme Lopes de Oliveira

Isabela de Lima Severino

Isabella Amaro de Oliveira Pereira

Janete Furtado Ribeiro Aguirre

Larissa Natany Almeida Martins

Rafael Hilario Cruz Peixoto

Stella Cheng Guimaraes

Editoração

Profa. Magda Carvalho Pires

Isabella Amaro de Oliveira Pereira

Rafael Hilario Cruz Peixoto

Stella Cheng Guimaraes

ÍNDICE

CONFERÊNCIAS	5
Como ser um estatístico aplicado.....	5
Modelos de covariância baseados em covariáveis.....	6
Modelagem espacial na análise de experimentos agrônomo com parcelas espacialmente correlacionadas.....	7
Complex sampling issues in the analysis of longitudinal survey data.....	8
De Fisher até o "Big Data": continuidades e descontinuidades.....	9
Adaptive significance level: blending bayesian and classical concepts.....	10
MINICONFERÊNCIAS	11
Non-parametric statistical method applied to genomic selection.....	11
Modelos semi-paramétricos para a análise de dados de degradação.....	12
Visual interpretation of F test results based on effect size.....	13
Sequências de Bernoulli com propriedade de lapso de memória.....	14
SESSÕES TEMÁTICAS	15
ST1: Inferência em Processos Estocásticos.....	15
ST2: Educação Estatística.....	16
MINICURSOS	17
Teoria da Resposta ao Item - Modelos não dicotômicos com uso de software livre.....	17
Modelando dados amostrais complexos.....	18
Como publicar artigos de qualidade em Estatística e Estatística Aplicada.....	19
COMUNICAÇÕES ORAIS	20
SESSÃO PÔSTER	32

CONFERÊNCIAS

Conferência de Abertura

Como ser um estatístico aplicado

Dani Gamerman

DME – UFRJ

A palestra visa ilustrar formas de atuação de um estatístico aplicado. Usando técnicas modernas, a palestra apresentará aplicações em áreas como Epidemiologia, Finanças, Medicina e Meio Ambiente. Através desses exemplos, vai se procurar mostrar como um estatístico aplicado pode e deve atuar. A apresentação se concentrará nos resultados obtidos nas análises, em detrimento da teoria desenvolvida para elas. Assim, a apresentação será acessível para profissionais e alunos de todos os níveis e de todas as áreas.

Conferência 1

Modelos de covariância baseados em covariáveis

Paulo Justiniano
UFPR

Modelos espaciais definem modelam a dependência a partir de estruturas de vizinhança e/ou de distância entre observações que são espacialmente indexadas. Tal procedimento segue a intuição comum de que observações mais geograficamente mais próximas tendem a ser mais parecidas entre si. Por outro lado, informações auxiliares, na forma de covariáveis podem ser induídas na modelagem, tipicamente na estrutura de médias do modelo como em problemas usuais de regressão. Neste trabalho explora-se a ideia de que a proximidade pode ser definida não apenas no espaço geográfico mas também em espaço definido pelas covariáveis. Diversas estratégias de definição de tal espaço são apresentadas e avaliadas. São mostrados e comparados resultados obtidos com estruturas de vizinhança espaciais e definidas sobre as covariáveis individualmente bem como considerando combinações destas. A proposta expande a possibilidade de definição de modelos utilizados na modelagem de dados espaciais e não espaciais.

Conferência 2

Modelagem espacial na análise de experimentos agrônomico com parcelas espacialmente correlacionadas

Renato Ribeiro de Lima
DES-UFLA

A influência da heterogeneidade do solo, bem como de outras variáveis ambientais, na estimação e avaliação das diferenças entre tratamentos deve ser considerada no planejamento e nas análises estatísticas de experimentos agrônomicos. Existem muitos experimentos conduzidos em condições heterogêneas e nos quais as parcelas ou unidades experimentais são espacialmente correlacionadas. Nesses casos, é necessário que sejam utilizadas metodologias de análises estatísticas adequadas e que contemplem o uso da informação espacial e, conseqüentemente, da correlação existente entre as parcelas. Existem diferentes metodologias que consideram a dependência espacial entre as parcelas nas análises estatísticas, mais especificamente na análise de variância e nos procedimentos de comparações múltiplas. Dentre essas metodologias têm-se o uso da modelagem geoestatística, dos modelos autoregressivos espaciais (Spatial Autoregressive - SAR) e dos modelos autoregressivos condicionais (Conditional Autoregressive - CAR). Alguns estudos têm sido realizados com o objetivo de aplicar essas metodologias que consideram a dependência espacial nas análises de dados experimentais. Esses estudos apontam para um ganho em precisão na comparação de diferentes tratamentos, levando à detecção de diferenças cada vez menores entre os mesmos. Nesta conferência serão descritas algumas metodologias de análise de dados experimentais correlacionados espacialmente, com ênfase em Análise de Variância e testes para a comparação de médias, com resultados de aplicações a diferentes conjuntos de dados.

Conferência 3

Complex sampling issues in the analysis of longitudinal survey data

Marcel de Toledo Vieira
UFJF

Cross-sectional surveys offer a picture of the population at a particular time and are mainly interested in macro-level processes. Longitudinal studies are more similar to a movie and have the following features: (a) are concerned with micro-level processes, (b) provide a fuller and more detailed view of structural processes, (c) allow direct study of change and separation of age and cohort effects, (d) allow modelling of transitions into and out of particular states and what predicts these transitions. Moreover, because of the temporal dimension, we are better able to make causal inferences with longitudinal data. However, several challenges and issues arise when analysing longitudinal survey data, including: (i) panel attrition, (ii) wave nonresponse, (iii) item nonresponse, (iv) panel conditioning or practice effects, (v) complex data structure, (vi) complex survey design, and (vii) correlated observations for the same individual over time. In this talk we shall give special attention to feature (b) and issues (i), (vi) and (vii), listed above. Results of an empirical study using data from the British Household Panel Survey and of a simulation study will be presented and discussed. Methodological issues will be illustrated with empirical evidence from the analysis of longitudinal data on life satisfaction and household income.

Conferência 4

De Fisher até o "Big Data": continuidades e discontinuidades

Renato Martins Assunção
DCC - UFMG

Em dois grandes trabalhos, em 1922 e 1925, Fisher introduziu muitas das ideias (parâmetros, suficiência, eficiência, máxima verossimilhança) que, aliadas à teoria da decisão de Wald em 1950, sus tentaram a estrutura da estatística até os anos 80. A partir daí, tivemos o início da introdução de computadores para armazenar e analisar "grandes dados". Os conceitos da era Fisheriana evoluíram em resposta à este novo ambiente. "Suficiência" evoluiu para a "compressão de dados", o conceito de "eficiência" teve de incluir considerações computacionais e questões de escala, "parâmetros" e "máxima verossimilhança" foram atualizados para o contexto de modelos semi e não paramétricos, modelos maiores que os anteriores e com a preocupação de ter robustez. O aumento constante da capacidade computacional nos últimos 30 - 40 anos permitiu a implementação prática do ponto de vista Bayesiano. Tivermos o uso disseminado de métodos intensivos de computador como o "bootstrap" de Efron, bem como a introdução do aprendizado de máquina do ponto de vista e com métodos de ciência da computação. Após uma visão geral desse desenvolvimento histórico, vou apresentar alguns exemplos de meu trabalho recente.

Conferência de Encerramento

Adaptive significance level: blending bayesian and classical concepts

Carlos Alberto de Bragança Pereira
IME - USP

The main objective of this paper is to find a close link between the adaptive level of significance, presented here, and the sample size. We, statisticians, know of the inconsistency, or paradox, in the current tests of significance that are based on p-value statistics that is compared to the canonical significance levels (10%, 5%, and 1%): "Raise the sample to reject the null hypothesis" is the recommendation of some ill-advised scientists! This paper will show that it is possible to eliminate this problem of significance tests. Lindley's paradox – "increase the sample to accept the hypothesis" – also disappears. Obviously, we present here only the beginning of a possible interesting research. The intention is to extend its use to more complex applications such as survival analysis, reliability tests, and other areas. The main tools used here are the Bayes Factor and the extended Newman-Pearson Lemma, discussed by DeGroot (1986).

MINICONFERÊNCIAS

Non-parametric statistical method applied to genomic selection

Camila Azevedo

UFV

Genome wide selection (GWS) is based on a large number of molecular markers widely distributed over the genome that are assumed as independent variables under a multiple regression viewpoint. These markers enables to capture genes effects on traits of interest through their linkage disequilibrium (LD) with the quantitative trait loci (QTL) affecting these traits. In this context, GWS provide reliable genetic value estimation for individuals without phenotypes measures. However, the number of markers is generally much larger than the number of phenotyped individuals; and these markers are also highly correlated. Under this condition, statistical methods that takes into account dimensional reduction and multicollinearity modeling are required. A suitable statistical method for GWS should contemplate three mains attributes: (i) to accommodate the genetic architecture of trait in terms of genes of small and large effects; (ii) to regularize the estimation process in the presence of multicollinearity and large number of markers; and (iii) to perform efficient selection of markers that affect the traits. New classes of statistical-genomic methods approaching simultaneously all these mentioned attributes have been proposed. Among them highlights the non-parametric Delta-p method, which is based on the genetic distance between subpopulations. This method is based on the allele frequency changes due to selection and on the genetic gain concepts. Additionally, Delta-p enable to exploit the genetic values estimated from others methods under a selection index framework.

Modelos semi-paramétricos para a análise de dados de degradação

Cristiano de Carvalho Santos
UFMG

Estudos sobre tempo de vida são muito utilizados em diversas áreas do conhecimento, entre elas a área industrial. O objetivo, em geral, está em avaliar a confiabilidade de produtos ou peças a partir da distribuição de probabilidade do tempo até a falha. Neste contexto, os modelos de degradação surgem como uma alternativa para contornar o problema de dados com grande proporção de censura, como é o caso de estudos envolvendo produtos com alta confiabilidade. Estes modelos são motivados pelo fato de que grande parte das falhas estão associadas a uma característica do produto que se degrada ao longo do tempo. Desta forma, a modelagem é feita sob uma medida de degradação avaliada ao longo do tempo e a inferência sob o tempo de falha é realizada a partir da especificação de um nível crítico de degradação. O modelo misto, em que os efeitos aleatórios estão relacionados com a taxa de degradação, é uma das alternativas para modelar dados de degradação. Sob o paradigma Bayesiano, os efeitos aleatórios são tratados como parâmetros, o que simplifica a implementação de modelos flexíveis. A abordagem Bayesiana não-paramétrica pode ser usada para especificação de modelos ainda mais flexíveis e realísticos. Neste trabalho, propomos modelar o comportamento dos efeitos aleatórios através de misturas via processo de Dirichlet, considerando como núcleo da mistura as distribuições Normal e Skew-Normal. Em um estudo com dados simulados, mostramos que os modelos propostos são capazes de acomodar assimetria, caudas pesadas e multimodalidade no comportamento dos efeitos aleatórios. Os modelos propostos também são aplicados a dados reais sobre emissores de lasers, entre outros.

Visual interpretation of F test results based on effect size

Isabel de Souza Amorim
UFLA

The analysis of variance (ANOVA) is one of the most often employed statistical methods to study differences between treatments. However, the ANOVA often focus just on the p-values. Therefore, it would be valuable to supplement the F-testing with some good measures of overall effect size (ES). In this talk we will present a visual tool based on effect size measures to improve research interpretation of mixed model ANOVA results. The basic and straightforward idea is to interpret effects relative to the residual error and to choose the proper effect size measure. The close link between Cohen's d , the effect size in an ANOVA framework, and the Thurstonian (Signal detection) d -prime are used to suggest the delta-tilde barplot as a better visual tool to interpret mixed model results. A generic implementation of the method is available on the SensMixed R-package. The method is illustrated on a multifactorial sensory profile data set and compared to actual d -prime calculations based on Thurstonian regression modelling through the ordinal R-package. The use of the delta-tilde barplot can be viewed as good and relevant additional tool for interpretation of the ANOVA table, particularly in situations with more than a single factor and with several attributes. Although the delta-tilde barplot suggested here cannot substitute a good post-hoc analysis, they are valuable additional tools for a relevant interpretation of the ANOVA results. In addition, the delta-tilde barplot can help to move the focus a bit away from purely looking at p-values but rather focusing on the size of the effect.

Sequências de Bernoulli com propriedade de lapso de memória

Rodrigo Lambert
UFU

Para sequências binárias cuja distribuição depende do número de sucessos do passado, introduzimos o conceito de "Lapso de memória", que nada mais é do que uma chave que liga e desliga a dependência que o processo tem no passado. Para uma classe de sequências binárias com lapso de memória, provamos três teoremas clássicos de convergência: Lei forte dos grandes números, Teorema central do limite, e Lei do logaritmo iterado.

SESSÕES TEMÁTICAS

ST1: Inferência em Processos Estocásticos

Coordenador: Ivair Ramos Silva (UFOP)

Palestrantes:

Ivair Ramos Silva (UFOP) - Optimal Alpha Spending for Sequential Analysis With Binary Data

For sequential analysis hypothesis testing, we derive the optimal alpha spending function for continuous and for group sequential analysis of binomial processes. The method favors to construct designs that are uniformly better than any other method, such as the classical Wald, Pocock or O'Brien-Fleming alpha spending. These results are obtained through analytical derivations and based on exact calculations by means of linear programming.

Luiz Henrique Duczmall (UFMG) - A spatial scan statistic for beta regression models

Spatial Scan Statistics have been developed for geographical cluster detection in different types of models, for example, Bernoulli, multinomial, Poisson, exponential, Weibull and normal. However, some data are continuous in the interval $[0,1]$ such as rates, proportions and indices, or are limited in the interval (a,b) . In this work, we propose a spatial scan statistic for a Beta regression model. The test statistic is based on a likelihood ratio test and evaluated using bootstrap p-value. The proposed method is illustrated using index of basic education and infant mortality in the Brazilian Amazon. The statistical power, sensitivity and positive predicted value of the test are examined through a simulation study.

Reinaldo Antônio Gomes Marques (UNIFAL) - Monte Carlo and interacting particle strategies for stochastic process

In this talk we will review the recent Monte Carlo methodologies to deal with complex inference in stochastic process. It will be given an attention in the discrete process where some unknown quantity or the posterior distribution can not be computed in the close-form. We also will show the gain of using the particle interacting approximations to deal with online inference. Finally, some insights about how we can efficiently update information over time to perform statistical inference in the real-time process, it will be presented.

ST2: Educação Estatística

Coordenador: Marcos Nascimento Magalhães (IME-USP)

Palestrantes:

Marcos Nascimento Magalhães (IME-USP) - Priorizar ideias é o desafio

Várias carreiras têm disciplinas iniciais cobrindo itens básicos de Probabilidade e Estatística e, em muitas situações, há uma pressão para a apresentação de vários conteúdos e técnicas. Pretende-se, em um semestre, formar “estatísticos” que dominem inúmeras técnicas, além de operar com desenvoltura os software estatísticos. Com algumas exceções, isto redundava em aprendizado superficial de qualidade duvidosa. É preciso se contrapor e buscar uma discussão mais aprofundada de conceitos. Dessa forma, criaremos senso crítico e autonomia, elementos essenciais na formação do estudante. Nessa apresentação vamos discutir algumas iniciativas para enfrentar o desafio de priorizar ideias nas disciplinas básicas de Estatística.

José Francisco Soares (ABAVE) - Sucesso pessoal e contribuição social com a Estatística

Nesta apresentação, primeiramente, reflito, a partir do sucesso profissional de vários estagiários de Laboratório de Medidas Educacionais do Departamento de Estatística, sobre o que deve saber um estatístico para ter uma carreira profissional e socialmente relevante. Argumento que é a formação adquirida na UFMG que explica o seu sucesso, mais do que suas personalidades e origens sociais. Em um segundo momento, através de exemplos, mostro como a Estatística é essencial para o bom planejamento da educação básica brasileira, área onde estou envolvido há muitos anos.

Leandro de Oliveira Souza (Facip/UFU) - Possibilidades de insubordinação criativa no ensino de estatística

As ações de insubordinação criativa são atos políticos que ocorrem quando se rompe com regras pré-estabelecidas visando progresso e mudanças na sociedade com foco na justiça social. Visamos relatar um projeto de pesquisa que buscou o desenvolvimento de novas abordagens para o ensino de Estatística, por meio de atividades de investigação, exploração, tratamento e análise de dados. Objetivou-se ensinar estudantes a selecionar gráficos e tabelas adequadas para reportar informações a partir da construção de conceitos de amostra, população, indivíduo e variável. A questão que conduziu a pesquisa foi: como poderíamos preparar estudantes para que pudessem compreender conceitos estatísticos e produzir informações ao invés de apenas consumi-las? Dois estudantes de ensino médio protagonizam, a partir de uma problemática na sua escola, uma investigação que nos ajudarão a refletir sobre o ensino de estatística na educação básica.

MINICURSOS

Teoria da Resposta ao Item - Modelos não dicotômicos com uso de software livre

Marcos Antônio C. Santos
UFMG

Este é um breve minicurso de caráter aplicado dirigido aos interessados no tema, com ênfase na utilização das implementações no software R e outros softwares livres para análise de dados de vários modelos da TRI. Serão apresentadas uma rápida descrição dos principais modelos da TRI para respostas dicotômicas e não dicotômicas, como os modelos Resposta Gradual, Resposta Contínua e GGUM e descrição de implementações disponíveis. A construção de algumas funções úteis na análise TRI no ambiente R, como obtenção de matriz de correlação policórica e outras serão apresentadas.

Modelando dados amostrais complexos

Leonardo Bastos

FIOCRUZ

Nesse minicurso será apresentado uma alternativa para se ajustar modelos estatísticos dados amostrais complexos sem ignorar o desenho amostral. O ajuste de modelos será baseado na pseudo verossimilhança, sob o ponto de vista clássico. Algumas bases de dados abertas e scripts em R para analisá las serão apresentados. Também será apresentado uma breve discussão de como incluir o desenho amostral sob a ótica bayesiana.

Como publicar artigos de qualidade em Estatística e Estatística Aplicada

Eric Batista Ferreira
UNIFAL

A ciência contemporânea é difundida via artigos científicos, publicados em revistas e periódicos – impressos ou eletrônicos - publicações contínuas e regulares e que permitem que a comunidade científica fique a par do estado da arte e da evolução da ciência. Grande parte desses artigos é oriunda de Programas de Pós-graduação; produzida por discentes, docentes e muitas vezes em associação com estudantes de graduação. É por meio dessas pesquisas e descobertas que existe avanço em cada área do conhecimento. Segundo Pinto e Fausto (2012), esses trabalhos geralmente são publicados em periódicos que podem ser impressos ou eletrônicos. Os periódicos eletrônicos são uma tendência, pois são mais baratos e podem atingir um número enorme de leitores, com mais facilidade e rapidez, já que estão disponíveis via internet. Diante desse cenário, o que significa publicar com qualidade? Quais são as métricas valorizadas atualmente e quais devemos seguir? Essas e outras questões provocativas serão o centro da discussão travada neste minicurso.

COMUNICAÇÕES ORAIS

Comunicações Orais 1

Non-Gaussian stochastic volatility model with jumps via Gibbs Sampler	21
Teste T sequencial e a decisão sobre o coeficiente de determinação (R^2).....	22
Cadeias de Markov ocultas aplicadas a notificações de dengue no município de Lavras-MG ...	23
Indicadores educacionais e contexto escolar: uma análise das metas do IDEB.....	24
Uma abordagem bayesiana generalizada para a teoria de resposta ao item através de DIF e distribuições do traço latente via mistura.....	25
On the statistical challenge of correcting underreporting bias: is it possible without informative priors?.....	26

Comunicações Orais 2

Analysis of a longitudinal split-plot with subsampling experiment using GAMLSS	27
Análise bayesiana do fenômeno hot hand em basquete via modelos hierárquicos	28
Aplicação dos modelos de regressão censurados em estatísticas públicas.....	29
Secagem de banana d'água nas temperaturas de 40 e 70°C descrita pelo modelo de Lewis via inferência bayesiana.....	30
Social representations and cognitive networks: a graph-based approach.....	31

Comunicações Orais 1

Non-Gaussian stochastic volatility model with jumps via Gibbs Sampler

Arthur Tarso Rego^{*}, Thiago Rezende dos Santos

Understand the behavior of asset prices is essential for capital allocation decisions between the available investment option. Such decision depends on what one thinks about risks and returns associated with these investment options. On the literature there are models focused on estimating financial assets risk, however, most of them depend on MCMC methods based on Metropolis algorithms, which makes them computationally expensive. This work presents an alternative model capable of estimating the volatility without appeal to intensive computational methods. An analysis with real data is performed and results indicate that the proposed model was capable of estimating volatility, in a computationally efficient and automated way.

^{*} arthur.tarso@outlook.com, Universidade Federal de Minas Gerais.

Teste T sequencial e a decisão sobre o coeficiente de determinação (R^2)

Cássia de Souza Santos^{*}; Eric Batista Ferreira; Davi Butturi-Gomes

Os testes sequenciais foram desenvolvidos por Wald (1948), dentre eles destaca-se o teste t sequencial para casos em que há normalidade. Ainda, os testes sequenciais diferenciam-se dos demais testes presentes na inferência estatística por utilizar amostras de tamanho variável com a garantia de ser mínima e possibilita ao pesquisador determinar as taxas de erro tipo I e tipo II. Rushton (1950) propôs uma versão para o teste t sequencial, porém Gouindarajulu e Haward (1989) em um estudo sobre expansões assintóticas uniformes, indicou uma modificação para este teste. Diante disso, o trabalho consiste em avaliar o comportamento do teste t sequencial de Rushton com a modificação na tomada de decisão e no tamanho amostral para amostras simuladas de R^2 ao longo de diferentes médias e desvios-padrão. Justifica-se a utilização do coeficiente de determinação (R^2), pois essa variável em um contexto de regressão pode ser modelada como tendo distribuição normal truncada de 0 a 1. Os resultados evidenciaram a coerência do teste, em que foi verificado ser menor que 20 quando o verdadeiro valor da média se distanciava de H_0 e tendia a 300 quando o valor médio hipotético era igual ao simulado. Sendo assim, o estudo permitiu conhecer o comportamento sequencial do coeficiente de determinação frente à tomada de decisão e ao tamanho amostral.

^{*} cassiaunifal@yahoo.com.br, Universidade Federal de Alfenas

Cadeias de Markov ocultas aplicadas a notificações de dengue no município de Lavras-MG

Deive Ciro de Oliveira*, Rodrigo Ferreira Abreu,
Rafael Agostinho Ferreira, João Domingos Scalon

A dengue é uma importante enfermidade e exige por parte de agentes públicos atuação para minimizar seus efeitos na população. Neste sentido, uma das ações de diagnósticos é determinar períodos temporais de maior gravidade da epidemia para intensificação de prevenção, maior contingente de medicamentos, médicos etc. Considerada a observação de séries de notificação de casos de dengue (anualidade), as cadeias de markov com estados ocultos permitem fazer inferência sobre janelas de maior ocorrência de notificações. Neste trabalho aplicou-se cadeia de markov ocultas a dados de notificação de casos de dengue no município de Lavras-MG em 2015. São apresentados resultados de estimação dos parâmetros além da discussão da estimação das janelas temporais de maior intensidade.

* deive.oliveira@gmail.com, Universidade Federal de Alfenas

Indicadores educacionais e contexto escolar: uma análise das metas do IDEB

Erica Castilho Rodrigues^{*}; Daniel Abud Seabra Matos

Buscamos relacionar características das escolas que estejam associadas à probabilidade de a escola atingir a meta do Índice de Desenvolvimento da Educação Básica (Ideb) 2013. Os objetivos são: analisar a associação da meta do Ideb com outros indicadores educacionais; comparar os anos iniciais e finais do ensino fundamental. Usamos métodos quantitativos: estatística descritiva, Teoria da Resposta ao Item e regressão logística (N anos iniciais= 36.236 escolas e N anos finais= 29.095 escolas). Nos anos iniciais, a variável de maior impacto sobre a probabilidade de a escola atingir a meta do Ideb é a infraestrutura. Nos anos finais, as variáveis que mais impactam são o nível socioeconômico e a infraestrutura. Existe um funcionamento diferencial entre os anos iniciais e finais (por ex.: o nível socioeconômico apresenta impacto positivo nos anos iniciais e negativo nos anos finais). Algumas relações entre o Ideb e outros indicadores educacionais são diferentes quando o fenômeno em questão é a meta do Ideb.

^{*} ericacastilrodrigues@gmail.com, Universidade Federal de Ouro Preto

Uma abordagem bayesiana generalizada para a teoria de resposta ao item através de DIF e distribuições do traço latente via mistura

Gabriel Oliveira Assunção* ; Flávio Bambirra Gonçalves

Atualmente, a Teoria de Resposta ao Item (TRI) vem sendo utilizada em testes para a avaliação psicológica, de conhecimento e até de qualidade de um funcionário. Com o foco na sua atuação para a avaliação de conhecimento, a TRI apresenta três parâmetros para avaliar um item, sendo eles: discriminação, dificuldade e acerto ao acaso. Com estes parâmetros, é possível verificar quais questões são boas para o estudo e assim poder melhorar o questionário e obter melhores resultados.

Pesquisas mostram que, em uma mesma população, o teste pode ser mais difícil para um grupo; para isso, é utilizado o Funcionamento Diferencial do Item (DIF), onde os parâmetros serão diferentes para cada um desses grupos. O intuito é utilizar o DIF para o parâmetro de dificuldade, a fim de perceber em quais grupos o item fica mais difícil e, assim, avaliar melhor a população. Com o intuito de melhor representação da população, os escores, traço latente, obtidos no exame serão calculados através de uma mistura de distribuições para, assim, modelar melhor características de multimodalidade, assimetria e caudas pesadas.

Será utilizada uma abordagem Bayesiana para se fazer inferência, através de um algoritmo de MCMC que foi implementado na linguagem Ox.

* gabrieloliveira1995@gmail.com, Universidade Federal de Minas Gerais.

On the statistical challenge of correcting underreporting bias: is it possible without informative priors?

Guilherme L. de Oliveira^{*}, Rosangela H. Loschi, Renato M. Assunção

A major challenge when monitoring risks in socially deprived areas of under developed countries is that economic, epidemiological and social data are typically underreported. Thus, statistical models that do not take the data quality into account will produce biased estimates for the relative risks. Building models able to provide reliable estimates for these quantities is the great challenge in this context. We present different Bayesian approaches for mapping risks when available data sets are subjected to underreporting. The usual practices to overcome this problem is to consider either models for censored data or marginalization under the true but unknown number of cases. In both cases, special attention must be given to the construction of prior distributions for the data reporting process. In fact, such prior information have shown to be quite important for achieving good posterior estimates for the relative risks and we discuss about that in this work. We present different models applying them to account for potential underreporting of early neonatal mortality counts in Minas Gerais State, Brazil, where data quality is known to be poor.

^{*} guilopes2110@gmail.com, Universidade Federal de Minas Gerais.

Comunicações Orais 2

Analysis of a longitudinal split-plot with subsampling experiment using GAMLSS

Gustavo Thomas* ; Clarice Garcia Borges Demétrio

Generalized additive models for location, scale and shape (GAMLSS) can be used for mixed modeling purposes, because they can incorporate random effects in their formulation. Besides, nonparametric functions of the available covariates might also be used, so that experiments with random and nonlinear effects can be modeled with this framework. Such an analysis is described in this work, highlighting the need to model both the mean and dispersion parameters of the assumed response distribution.

* gustavothomas17@yahoo.com.br, ESALQ/USP

Análise bayesiana do fenômeno hot hand em basquete via modelos hierárquicos

Lucas José Gonçalves Freitas^{*}; Renata Souza Bueno; Gustavo Henrique Mitraud Assis Rocha

O fenômeno Hot Hand surgiu nos esportes coletivos, especificamente no Baseball, e agora habita várias áreas do conhecimento, como economia e psicologia. Trata-se da crença de que é mais provável ter êxito em eventos com registro de sucessos imediatamente anteriores. Entre os fãs, profissionais e patrocinadores de Basquete, a crença em Hot Hand representa maiores investimentos em treinamento, novas apostas esportivas e grandes audiências de televisão. Este trabalho se propõe a analisar dados da NBA para validar ou não, de acordo com as dificuldades impostas por cada jogo, as hipóteses de Hot Hand em arremessos sucessivos de um mesmo atleta, utilizando para este objetivo modelos hierárquicos e métodos bayesianos de inferência.

^{*} matlucas13@yahoo.com.br, Escola Nacional de Ciências Estatísticas – ENCE.

Aplicação dos modelos de regressão censurados em estatísticas públicas

Sofia Elizabeth Pignataro de Lima^{*}; Gustavo Henrique Mitraud Assis Rocha;
Alinne de Carvalho Veiga

Este projeto de iniciação científica aplicou modelos de regressão censurados em dados de estatísticas públicas brasileiras através de métodos bayesianos. Para a distribuição dos erros dos modelos, assumiu-se distribuições tanto simétricas (normal e t-Student) quanto assimétricas (skew-normal e skew-t). Para estimar os parâmetros dos modelos, foram utilizadas, como principais ferramentas, as técnicas de MCMC Gibbs Sampling, Metropolis com passeio aleatório e ARMS. A partir dos resultados obtidos, foi feita uma comparação entre as estimativas de cada modelo. As estatísticas públicas escolhidas como fonte de dados foram as da Pesquisa de Orçamentos Familiares (POF), de 2008-2009, realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE).

^{*} sofia.pignataro@gmail.com, ENCE - Escola Nacional de Ciências Estatísticas

Secagem de banana d'água nas temperaturas de 40 e 70°C descrita pelo modelo de Lewis via inferência bayesiana

Thais Destefani Ribeiro* ; Joel Augusto Muniz

A banana é o fruto mais consumido mundialmente, e o Brasil é o quarto maior produtor mundial da cultura estando esta maior concentrada nos estados da Bahia, Pará, São Paulo e Minas Gerais. Este é um fruto de sabor agradável que está presente dentre as diversas camadas da população, no entanto é altamente perecível, razão pela qual sua comercialização deve ser rápida, racional e feita com uma série de cuidados para que não haja perdas expressivas e o fruto chegue ao seu destino em boas condições. Sendo assim a secagem da banana surge como uma saída para se reduzir estas perdas. Este trabalho teve como objetivo utilizar o modelo de Lewis na descrição da cinética de secagem de banana D'água, nas temperaturas de 40 e 70°C via inferência bayesiana. Foram comparadas as diferenças em médias das taxas de secagem e das razões de umidade ao final do processo. O software Open Bugs foi usado para realizar as análises e o R para verificar a convergência das cadeias. Avaliou-se a qualidade no ajuste do modelo com o Desvio Padrão Residual, Erro de Monte Carlo e o Intervalo de Credibilidade a 95%. De modo geral a metodologia de inferência bayesiana foi eficiente no ajuste dos dados de cinética de secagem de banana, e o modelo de Lewis ajustou-se bem aos dados em ambas temperaturas. Com base nas diferenças calculadas observou-se a perda de umidade de 0,5 ao final do processo entre as temperatura de 70°C e 40°C no mesmo período de avaliação, com variação na taxa de secagem de 0,2263 g de H₂O/min.

* t.destefani.ribeiro@gmail.com, Universidade Federal de Lavras

Social representations and cognitive networks: a graph-based approach

Wesley Pereira* ; Gilvan Guedes; Denise Duarte; Rodrigo Ribeiro; Marina Amorim

Introdução: A análise das representações sociais pode ser entendida como uma análise de conceitos que são compartilhados socialmente. Esses conceitos sociais dependem de um processo de negociação coletiva, de significação, e acima de tudo de efetivos canais de comunicação para que esse significado atinja e circule entre os atores sociais, expondo-os ao objeto de significação de forma ampla. Alguns autores destacam o importante papel da comunicabilidade na formação de uma representação social em determinado contexto, mas pouco se explora esse aspecto explicitamente espacial na literatura das Representações Sociais. Objetivos: Este estudo propõe duas fórmulas para transformar dados sobre representações sociais em dados relacionais apropriados para a criação de grafos. Discutem-se também os potenciais do conceito de comunicabilidade para difusão social de conceitos coletivamente significados quando explicitamente modelados espacialmente. Metodologia: Para atender ao objetivo proposto, será apresentada uma proposta metodológica de análise espacial das Representações Sociais utilizando-se uma combinação de Análise de Redes Sociais para as Representações Sociais e o georreferenciamento dos atores (egos). Os dados são obtidos por meio da Técnica de Livre Associação de Palavras e analisados com a fórmula de afinidade entre pessoas proposta por Pereira (2015). Cada ego é, então, georreferenciado e após a obtenção de clusters de egos com pensamentos comuns (utilizando-se de um algoritmo de agrupamento baseado na maximização da modularidade), esses subgrafos são mapeados espacialmente. Como, em geral, os dados são obtidos através de amostra, a representação espacial desses conceitos apresenta descontinuidade na superfície, exigindo alguma técnica de interpolação espacial. Neste trabalho, propõe-se a utilização do interpolador espacial IDW (Inverse Distance Weight), adotando-se como o critério de interpolação o valor da afinidade de Pereira (2017) para ponderação de áreas em redes de pessoas que compartilham significados comuns. A superfície espacial suavizada, então, permite observar áreas "quentes" e "frias" no espaço, identificando se e onde há concentração espacial de atores que compartilham significados comuns. Aplicou-se a metodologia acima descrita para dados coletados para 115 indivíduos com diferentes níveis de exposição ao Zika Vírus em Governador Valadares, Minas Gerais, entre Agosto e Novembro de 2016. Foram considerados como critérios de estratificação: nível de exposição (indivíduos nunca contaminados por nenhuma doença transmitida pelo *Aedes Aegypti*; infectados por Dengue ou Chykungunya, mas não por Zika; infectados por Zika), sexo, status socioeconômico. Os dados foram tratados no R para análise das redes de pessoas (utilizando-se os pacotes *wordvok*, *igraph* e *fmsb*), e a análise espacial foi feita no ArcGis 10.2. Resultados: Foram encontrados 5 grupos principais de atores sociais (egos) que compartilham significados comuns (Epidemia, Consequências, Caos, Falta de Comprometimento e Sintomas). Aparentemente a análise de pontos sugere difusão espacial desses grupos, embora o suavizador IDW utilizando a fórmula de afinidade ponderada de Pereira revela que alguns atores são responsáveis isolados pela difusão desses conceitos. Conclusão: A concentração espacial observada sugere que alguns atores são mais populares na difusão espacial do conceito social, e devem, portanto, ser o foco de estudo para o desenvolvimento de campanhas com o objetivo de promover comportamentos preventivos em relação ao Zika Vírus naquelas comunidades.

* wesleyhpereira@gmail.com, Universidade Federal de Minas Gerais.

SESSÃO PÔSTER

Sessão Pôster 1

O algoritmo Wang-Landau nos grafos aleatórios	38
Curva de crescimento de equinos mangalarga marchador utilizando modelos de regressão não lineares.....	39
Beef cattle body weight prediction using time series	40
O desempenho em matemática influencia na escolha da profissão?.....	41
Função cointegração aplicada às séries índice de atividade econômica e população economicamente ativa	42
Novo índice de vulnerabilidade juvenil: violência e desigualdade socioeconômica segundo a desigualdade racial – Brasil, 2010	43
Impactos sobre o programa Bolsa Família: um estudo nas maiores cidades do sul de Minas Gerais.....	44
Agrupamento dos municípios do Triângulo Mineiro e Alto Paranaíba em relação às suas características socioeconômicas.....	45
Modelos de riscos proporcionais de Weibull com erros de classificação e censura intervalar.....	46
Seleção de variáveis e comparação de modelos em regressão logística	47
Exportação de café verde no Brasil: uma análise utilizando o modelo de Holt-Winters.....	48
Perdas na esperança de vida ao nascer decorrentes das principais causas de morte em Minas Gerais, 2015	49
Aplicação do modelo quase-binomial no processo de incubação de ovos.....	50
Análise de agrupamentos no estudo dos indicadores de mortalidade no Brasil	51
Comparação de modelos não lineares na descrição da altura de tronco de árvores.....	52
Modelo HMM para classificação de condições de mercado.....	53
Métodos estatísticos de proteção de dados confidenciais sob a condição de differential privacy.....	54
Análise temporal da exportação de barris de petróleo no Brasil	55
Previsões sobre a taxa de admissão de emprego no Brasil	56
Partially linear models and their applications to change point detection of chemical process data	57
Uma rotina em R para o teste Scheffé em experimentos em fatorial triplo.....	58
Estudo do número de óbitos no Brasil no período de 2000 a 2015 utilizando o modelo de series temporais.....	59
Probabilidade conjunta do gráfico de controle de Shewhart	60
Análise estatística da aceitabilidade do brigadeiro sem lactose com adição de biomassa de banana verde.....	61

Distribuição exponencial ajustada com estimadores equivariantes em eventos de precipitação máxima	62
Aplicação inovadora na análise de padrão de pontos marcados	63
Geodesign: pacote no R para planejamento amostral ótimo em geoestatística	64
Perfil de uma unidade de dor torácica em um hospital universitário quanto ao tipo de cuidado de enfermagem.....	65
Adequabilidade dos métodos numéricos no ajuste da distribuição Gumbel	66
Aplicação de teoria de resposta ao item na avaliação de questões em estatística aplicada.....	67
Análise do número de caso de tuberculose via modelos de séries temporais no estado do Ceará.....	68
Função verossimilhança e estimadores de máxima verossimilhança.....	69
Comparação de modelos de regressão não linear na descrição da liberação de CO ₂ de resíduo orgânico no solo.....	70
Perfil dos periódicos científicos publicados pela Universidade Federal de Minas Gerais (UFMG).....	71
Escolha entre o modelo cúbico com platô e o Von Bertalanffy que melhor descreve número médio de tubérculo por planta de batata	72
Aplicação da geoestatística espaço-temporal na distribuição das chuvas na Paraíba, Brasil.....	73
Daltonismo e acessibilidade em representações gráficas no Brasil	74
Comportamento espacial e temporal do IDEB-MG dos anos finais do ensino fundamental	75
Análise do sistema de medição em uma linha de classificação de frutos de melão amarelo	76
Teste Scheffé em fatorial duplo cruzado: uma rotina em R.....	77
Comparação entre os modelos Gompertz e logístico para avaliar o crescimento em matéria fresca da batata.....	78
Modelos lineares generalizados: aplicações às Ciências Atuariais.....	79
Utilização da biblioteca JavaScript D3 para a construção de gráficos estudados no ambiente web.....	80
Prior specifications to handle monotone likelihood in the Cox regression model.....	81
Redes neurais artificiais como alternativa à seleção genômica ampla para a predição de características complexas	82
A utilização da teoria de resposta ao item e modelagem log-normal do tempo para classificação de palavras.....	83
Previsão do perfil das instituições envolvidas em estratégias de F&A do setor bancário brasileiro	84
Regressão quantílica na estimação da eficiência técnica da agricultura familiar em Minas Gerais.....	85
Detection of inconsistencies in geospatial data with geostatistics.....	86
Regressão logística semi-paramétrica.....	87
Teste de independência usando amostras pequenas	88

Testes de hipótese para monitoramento do vetor de médias de processos multivariados sob amostragem sequencial: extensão para o caso de dados auto correlacionados.....	89
Gama exponenciada Poisson truncada no zero.....	90
O modelo linear de degradação sob uma perspectiva bayesiana: aplicação em emissores de laser.....	91
Modelo de regressão logística em dados de redações do consumidor	92
Análise espaço – tempo e intensidade de proliferação da doença ferrugem asiática da soja nos municípios de Chapada dos Guimarães e Tangará da Serra no estado do Mato Grosso.....	93
Construção de algoritmos para estimação em modelos GLARMA	94
A distribuição beta logística	95
Avaliação da gestão ambiental municipal do estado de Minas Gerais	96
Um modelo de séries temporais para previsão do consumo de energia elétrica da indústria no Brasil.....	97
Descrição de crescimento do acesso CNPAE-102 de pinhão manso utilizando modelo não linear.....	98
Regressão via componentes independentes parciais aplicada à seleção genômica	99
Modelando o tempo de concessão de patentes brasileiras usando análise de sobrevivência.....	100
Perfil dos estudantes de graduação da UFMG	101
Correlação entre área segura e valor arrecado da produção de café no estado de MG: uma análise espacial.....	102
Avaliação dos conhecimentos em doença falciforme no curso de educação à distância em Minas Gerais: aplicação do modelo de regressão de Poisson univariado.....	103
Análise temporal do índice de expectativas das famílias via metodologia de Box and Jenkins.....	104
Erros do modelo de regressão linear simples: e se a suposição de normalidade dos erros não forem válidas?	105
Avaliação do efeito alelopático de Xanatse (<i>Colophospermum Mopane</i>) na germinação e crescimento inicial de abóbora (<i>Cucurbita Máxima</i>).....	106
Mineração de dados no esporte: um estudo sobre o campeonato brasileiro de futebol Série A 2016.....	107
Modelo de previsão da série de índice de atividade econômica do Banco Central (IBC-BR)...	108
Produção de cana-de-açúcar nas regiões Centro-Oeste e Nordeste do Brasil descrita por modelos mistos.....	109
Levantamento da visão política-social dos alunos dos cursos de humanas da UFMG	110
Análise de satisfação dos alunos do Centro de Ciências Agrárias da Universidade Federal do Espírito Santo.....	111
Representação da estrutura do pensamento coletivo sobre as enchentes do Rio Doce: conectando indivíduos afins através da teoria dos grafos.....	112
Poder exato do teste sequencial condicional da razão de verossimilhanças	113

Sessão Pôster 2

Uma análise utilizando a metodologia Box & Jenkins para o índice de atividade econômica brasileira.....	114
Preço dos imóveis sob a perspectiva da regressão linear – Windsor, Canadá 1987	115
Determinantes do endividamento de curto prazo das resseguradoras atuantes no Brasil	116
Fundos de investimento remunerados pelo risco: qual escolher?	117
Teste Scheffé no R para experimentos em parcela subdividida.....	118
Variáveis associadas à probabilidade de sobrevivência até os 60 anos na população da região metropolitana de Belo Horizonte	119
Análise de concordância no descarte de pintos de corte pelo critério de refugagem	120
Gerenciador de métricas de publicação (GMP).....	121
O uso das redes neurais artificiais para prever o padrão de resposta da resistência a ferrugem em coffea arábica	122
Perfil econômico-financeiro das instituições bancárias adquiridas em processos de fusões e aquisições (F&A).....	123
Avaliação do impacto da não - normalidade multivariada no teste de hipótese de razão de verossimilhanças para o vetor de médias de processos multivariados sob amostragem sequencial.....	124
Ajuste de modelos geoestatísticos espaço-temporal para avaliar a variação diária de temperatura em Lavras – MG.....	125
Métodos bayesianos aplicados na estimação do mérito genético do comprimento de carcaça em suínos	126
Imputação múltipla em regressão logística aplicada a dados de efetividade da terapia antirretroviral	127
Modelo multivariado para taxas e proporções espacialmente dependentes	128
Acidentes de trânsito e condições de mobilidade nas capitais do Norte, Sul e Centro-Oeste do Brasil.....	129
Análise discriminante e análise de cluster aplicadas em dados sobre Pokémon.....	130
Desempenho do gráfico de controle T^2 de Hotelling	131
Zero-inflated mixed Poisson regression models.....	132
Regressão quantílica na avaliação da adaptabilidade fenotípica de cultivares de feijão.....	133
Modelo de regressão lognormal com censura intervalar e erros de classificação.....	134
Comparação do desempenho de modelos da teoria de resposta ao item para itens dicotômicos: um estudo de simulação	135
Uma análise da inadimplência ocorrida nas operações de crédito para pessoas jurídicas.....	136
Modelos GLARMA para séries temporais	137
Métodos de simulação de dados geográficos sintéticos para bases confidenciais.....	138

Estimação teórica do tamanho crítico amostral para variáveis com dependência espacial	139
Análise bayesiana na previsão de ventos extremos em Sorocaba-SP.....	140
Modelos de séries temporais na análise da série de produção de automóveis no Brasil	141
Fatores associados à expectativa de vida ao nascer no Japão	142
Consumo de energia elétrica na região Sudeste do Brasil: uma aplicação do modelo SARIMA	143
Inferential aspects for the INAR(1) process with zero-inflated Poisson innovations.....	144
Extending JAGS for spatial data	145
Erros do modelo de regressão linear simples: e se a suposição variância constante não for válida?.....	146
Análise da volatilidade da taxa de retorno das ações da vale na Bovespa utilizando o modelo APARCH.....	147
Predição na presença de diferentes níveis de multicolinearidade por meio de métodos de aprendizado estatístico.....	148
Análise de série temporal da produção de petróleo no Brasil de 2000 a 2016	149
Comparação de métodos para seleção de preditores	150
Avaliação do risco de investimento em grandes empresas brasileiras: uma aplicação do modelo GARCH	151
Germinação e crescimento inicial do feijão vulgar (<i>Vigna Unguiculata</i>), sob efeito extrato aquoso de Mecrusse/Cimbirre (<i>Androstachys Johnsonii</i>)	152
Avaliação do efeito espacial no mapeamento de mortalidade neonatal precoce em Minas Gerais nos triênios de 2000-02 a 2012-14 na presença de dados sub-registrados.....	153
Identificação de múltiplos pontos de mudança em sequências multivariadas.....	154
Modelos de probabilidades da família Tweedie para o consumo de energia elétrica (per-capto) da região sudeste do Brasil.....	155
Caracterização do perfil dos registros de violência sexual contra mulheres mineiras de 2010 a 2014	156
Comparação de medidas de distância em análise de clusters para determinação de autoria em textos literários	157
Estudo sobre a evasão discente de uma universidade de Minas Gerais via regressão logística	158
O uso do Shiny para o ensino de matemática e estatística no ensino médio.....	159
Aplicativo pedagógico interativo de regressão linear	160
Uma análise exploratória sobre saúde e alimentação	161
Estudo da rentabilidade do Itaú Unibanco através de regressão linear	162
Estudo comparativo entre a faixa original do NIR de bancada com a faixa reduzida para dados de teor de sacarose aparente.....	163
Comportamento temporal do consumo residencial de energia elétrica nas regiões Sudeste e Nordeste do Brasil	164

Bandas de confiança para quantis usando núdeo-estimador.....	165
Modelo de previsão para a série do custo da cesta básica em São João del-Rei.....	166
Investigação com estudantes do ensino médio sobre o entendimento das questões relacionadas à realidade LGBT+	167
Partição de dados para uso em validação cruzada e sua influência sobre a qualidade de ajuste e seleção de variáveis em modelos de calibração	168
Validação do inventário de motivações para uso de jogos online a partir de estudo exploratório com usuários do Pokémon Go.....	169
Análise do risco de evasão no curso de estatística UFOP considerando dados censurados	170
Estimação bayesiana de intensidade de tráfego em filas multiservidoras.....	171
Modelos lineares generalizados e suas aplicações em Ciências Atuariais sob a visão bayesiana e frequentista da inferência	172
Análise de variância multivariada aplicada a bebidas energéticas consumidas por não atletas	173
Análise de desempenho de operadoras de saúde utilizando análise multivariada.....	174
Mortalidade por doença de Chagas no Brasil: análise e predições bayesianas dos efeitos de idade, período e coorte de nascimento	175
Infecção pelo vírus HPV: um modelo logístico com abordagem bayesiana	176
Análise da confiabilidade de um questionário para determinar a satisfação dos alunos utilizando o software R.....	177
Uma aplicação de modelo linear generalizado sob enfoque bayesiano para dados aumentados.....	178
Aprendizagem bayesiana e identificabilidade em modelos de regressão com erros nas covariáveis.....	179
Um estudo da tendência temporal da taxa suicida no Brasil.....	180
Uma análise espacial da dinâmica da desigualdade de renda em São Paulo nos anos de 1991, 2000 e 2010	181
Uma análise do programa Bolsa Família utilizando pareamento por escore de propensão	182
Biblioteca 'Wordevok' do software R.....	183
Desenvolvendo uma TIC para o ensino de probabilidade	184
Distribuições de probabilidade Fuzzy para severidade de sinistros	185
Confecção de pictogramas no R.....	186

Sessão Pôster 1

O algoritmo Wang-Landau nos grafos aleatórios

Adrian Hinojosa*

Na modelagem estatística clássica ou Bayesiana de Grafos Aleatórios um problema recorrente no uso do MCMC e do algoritmo de Metropolis-Hastings é a presença de misturas na distribuição das estatísticas sumárias. Esta situação é chamada de “degeneração do modelo”, isto é distribuições com grafos densos e esparsos. Este problema tem sido explicado recentemente numa série de trabalhos importantíssimos na área de limites de grafos aleatórios, pelo fato do modelo apresentar “transição de fase”, especialmente na região onde as arestas têm baixa densidade e o número de triângulos ser alto. O que implica que os algoritmos demoram em convergir ou tem tendência.

Uma inovadora solução que é usada na simulação de distribuições de Gibbs na mecânica estatística é o uso do algoritmo de Wang-Landau, que é uma extensão do algoritmo de Metropolis-Hastings. Neste trabalho apresentamos um algoritmo baseado no Wang-Landau para simular grafos do modelo exponencial de grafos aleatórios (ERGM) na região do espaço paramétrico onde temos degeneração.

* adrian@est.ufmg.br, Universidade Federal de Minas Gerais.

Curva de crescimento de equinos mangalarga marchador utilizando modelos de regressão não lineares

Rafaela Aparecida Ribeiro; Adriele Aparecida Pereira* ; Felipe Amorim Caetano de Souza;
Joel Augusto Muniz; Tales Jesus Fernandes

Os equinos da raça Mangalarga Marchador são muito difundidos pelo país devido à rusticidade e aplicabilidade no trabalho diário das propriedades rurais. Estudos sobre o crescimento e proporções morfométricas têm demonstrado o quanto é importante conhecer as características físicas do animal, pois são determinantes instrumentos na movimentação de um cavalo. No entanto, muitos criadores e profissionais possuem dúvidas sobre a confiabilidade de alguns métodos de estimação. O uso de modelos de regressão pode ser uma alternativa mais confiável para a análise de crescimento, pois estabelecem uma relação entre as variáveis associadas com o crescimento e sua idade, possibilitando a estimação das medidas do animal em qualquer idade. Estes modelos permitem resumir as informações contidas em um conjunto de dados em apenas alguns parâmetros com interpretações práticas diretas. Assim, este trabalho teve o objetivo de avaliar o ajuste dos modelos de regressão não lineares Logístico e o platô de resposta quadrática e indicar o mais adequado para descrever o crescimento de equinos da raça Mangalarga Marchador utilizando a variável comprimento corporal, em função da idade. Os dados analisados foram de 228 eqüinos da raça Mangalarga Marchador dos seis meses até aos 176 meses de idade que foram agrupados em sexos e divididos em 16 classes de idade. Os modelos estudados foram comparados segundo os avaliadores de qualidade: coeficiente de determinação ajustado e desvio padrão residual (DPR). Toda a parte computacional envolvida na elaboração deste trabalho foi realizado utilizando-se o software estatístico R (R DEVELOPMENT CORE TEAM, 2013). Embora o comportamento dos dados tenha característica de crescimento platô, os avaliadores de qualidade indicaram que o modelo Logístico é mais adequado para descrever a curva de crescimento da variável comprimento corporal. O modelo platô tende a superestimar o comprimento adulto dos animais em ambos os sexos.

* adrieleapvga@yahoo.com.br, Universidade Federal de Alfenas

Beef cattle body weight prediction using time series

Giovana Roseiro; Hiron Pereira Farias; Adriele Giaretta Biase^{*}; Danieli Perez; Tiago Zanett Abertini; Marcelo Neves Ribas; Luigi Cavalcanti; Júnior Fernandes; Carlos Tadeu dos Santos Dias; Dante Pazzanese Duarte Lanna

Cattle daily body weight and growth series, identifying their components and getting predict models of daily body weight. A total of 231 Senepol (*Bos taurus* genotype) heifers were evaluated, with a mean initial weight of 400 ± 56.7 kg. The animals were evaluated in feedlot for a maximum of 93 days in a collective pen equipped with an Intergado[®] automatic weighing system at Grama Farm, Pirajuí - São Paulo State. The diet was with 14.9% Crude Protein (CP) and 67.5% Total Digestible Nutrients (TDN, Dry Matter basis) was offered ad libitum. Predictions for the future body weight (BW) at 7, 14 and 21 days were performed and computed the percentage of models that best fit based on Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC) and Mean Square Prediction Error (MSPE). The ARIMA model (p, d, q) represented 100% of the best animal growth fit, based on BIC. The ARIMA (0, 1, 1) presented more than 55% of the fit analysis considering forecasts of 7, 14 and 21 days. The ARIMA models presented forecasts relatively close to those observed, detecting the trend in the series. Among the adjusted models, the ARIMA model (0, 1, 1) was the one that presented the best results to predict the average daily beef cattle Body Weight.

^{*} adrielegbiase@gmail.com, @Tech

O desempenho em matemática influencia na escolha da profissão?

Alice Silva Duarte^{*}; Cláudia Adam Ramos

É cultural associar o desempenho em matemática à escolha profissional. Diversos testes de aptidão utilizam a facilidade em resolver problemas matemáticos e o gosto pela disciplina como forma de direcionamento para o mercado de trabalho. Assim, o objetivo deste trabalho é checar se de fato o desempenho dos alunos em matemática está associado à escolha da sua profissão. Para esta análise foram considerados três cursos cujos perfis exibem alguma semelhança, mas cujas matrizes curriculares são muito diferentes. O estudo foi feito considerando o desempenho dos alunos de Administração Pública, Ciências Econômicas e Ciências Atuariais em Matemática. Com base na análise quantitativa dos dados das 11 turmas de cada curso investigado, e das notas destes discentes em matemática, observou-se que, os alunos de Ciências Atuariais foram os que mais aprovaram em matemática (45%), seguidos dos alunos de Ciências Econômicas (35%) e de Administração Pública (24%). Estes resultados parecem indicar, que, de fato, existe uma relação entre escolha profissional e gosto pela matemática.

^{*} alice_sduarte@hotmail.com, Universidade Federal de Alfenas

Função cointegração aplicada às séries índice de atividade econômica e população economicamente ativa

Aline Caroline Rodrigues* ; Letícia Lima Milani

Resumo: De acordo com Gujarati e Porter (2011) duas variáveis serão cointegradas se existir uma relação de longo prazo, ou de equilíbrio entre elas. Busca-se, portanto, neste trabalho identificar a existência, ou não, de uma relação de longo prazo entre as séries índice de atividade econômica e população economicamente ativa (%), tendo em vista a importância econômica de ambas para diagnosticar o bom funcionamento da economia. Os dados são provenientes do Ipeadata sendo de abrangência mensal de 2006 até 2016. Para verificar a estacionariedade das séries utilizou-se do teste Dickey–Fuller, em uma primeira análise, verificou-se que as séries são não estacionárias, por apresentarem raiz unitária. Desta forma, aplicou-se a primeira diferença, e em seguida o mesmo teste apontou que a série tornou-se estacionária. Como a série é $I(1)$, integrada de ordem um, foi possível aplicar o teste de cointegração de Engle e Granger, o qual a um nível de significância de 5% indicou que as séries índice de atividade econômica e população economicamente ativa apresentam cointegração. Verificou-se também que o valor da estatística Durbin-Watson, foi maior do que o coeficiente de determinação R^2 sendo, portanto, uma regressão não espúria como explanado por Gujarati e Porter (2011). Comprovou-se a confiabilidade da cointegração pela análise dos resíduos, em que o teste Dickey–Fuller apresentou estacionariedade dos mesmos, ou seja, são ruído branco. Portanto, após o tratamento estatístico dos dados verificou-se a existência de cointegração entre as séries indicando uma associação positiva de longo prazo entre elas. Tratando-se de economia brasileira tal associação é relevante para tomada de decisões econômicas (políticas expansionistas ou contractionistas) que afetam a população ativa e a condução da economia brasileira em seus processos produtivos.

* alineacr25@gmail.com, Universidade Federal de Alfenas

Novo índice de vulnerabilidade juvenil: violência e desigualdade socioeconômica segundo a desigualdade racial – Brasil, 2010

Allan Martins Coelho de Oliveira^{*}; Aline Bezerra da Silva; Waldir Jesus de Araújo Lobão.

O presente trabalho analisa a vulnerabilidade juvenil no Brasil, considerando a predisposição da juventude às respostas e consequências negativas que o envolvem em diversas dimensões da vida: a educação, o trabalho, a renda etc., privilegiando o recorte da desigualdade racial. Para alcançar esse objetivo, recomenda-se um novo índice composto que representa a abrangência e a intensidade das diferentes situações de violência e de desigualdade socioeconômica dos indivíduos segundo a desigualdade racial. O novo índice combina indicadores de violência (três variáveis) e desigualdade socioeconômica (quatro variáveis). Os resultados evidenciam a amplitude e a persistência da desigualdade racial existente no Brasil.

^{*} allan.m.coelho@gmail.com, Escola Nacional de Ciências Estatísticas

Impactos sobre o programa Bolsa Família: um estudo nas maiores cidades do sul de Minas Gerais

Amanda Mambeli Lopes^{*}; Carlos Guilherme de Souza; Gabriel Rodrigo Gomes Pessanha

De acordo com a Constituição Federal de 1988, art. 227, o Estado, em união da família e da sociedade, deve garantir à criança, ao adolescente e ao jovem, o direito a vida, por meio dos cuidados “à saúde, à alimentação, à educação, ao lazer, à profissionalização, à cultura, à dignidade, ao respeito, à liberdade e à convivência familiar e comunitária”. Nesse sentido, o Programa Bolsa Família (PBF) é uma Política Pública social pensada como um meio para atingir tal dever do Estado, da sociedade e da família, na erradicação da pobreza e da vulnerabilidade social. Visto a importância dessa política na garantia dos direitos dos cidadãos em situação de pobreza, é que foi elaborado o presente constructo. O objetivo desse estudo é utilizar o método quantitativo de regressão linear múltipla para compreender quais variáveis explicativas selecionadas impactam no número de beneficiários e no número de famílias contempladas pelo PBF, tendo como objeto empírico os maiores municípios do sul de Minas Gerais, em termos populacionais. Para isso, realizou-se a regressão através dos mínimos quadrados ordinários (MQO), com séries temporais anuais de 2004 a 2012, por meio do software livre, GRETL. As variáveis independentes foram encontradas no portal eletrônico IMRS - Índice Mineiro de Responsabilidade Social. Como resultado, obteve-se o R^2 de 0,476887, o que significa dizer que em 47,6887% das variações no número de famílias beneficiárias do PBF são explicadas pelas variações na taxa bruta de mortalidade padronizada por mil habitantes, gasto per capita com atividades de educação, gasto per capita com desenvolvimento econômico, gasto per capita com saneamento, gasto per capita com esporte e lazer, taxa de emprego no setor formal e rendimento per capita no setor formal. O P-valor (F) obtido foi de $1,21 \times 10^{-10}$, que resulta em uma significância de R^2 ao nível de 1%. Com a análise dos resíduos, foi verificada a robustez do modelo, com os resultados superiores a 5%. Dessa forma, o trabalho não se prende a esvaziar os estudos na área, mas ampliá-los para outras localidades e com outras questões referentes ao Programa, visto a sua importância para o âmbito social do país.

^{*} amandamambelli@hotmail.com, Universidade Federal de Alfenas - Campus Varginha

Agrupamento dos municípios do Triângulo Mineiro e Alto Paranaíba em relação às suas características socioeconômicas

Ana Carolina de Oliveira* ; Leonardo Henrique de Andrade Bento do Nascimento;
Patrícia de Siqueira Ramos

Este trabalho consiste em agrupar os municípios da mesorregião Triângulo Mineiro e Alto Paranaíba (TMAP) utilizando a técnica multivariada de análise de agrupamento. A mesorregião TMAP foi escolhida considerando sua importância econômica para o estado. Assim, o objetivo é identificar os grupos de municípios mais desenvolvidos com base em variáveis socioeconômicas e no método não hierárquico das k-médias. Os dados utilizados são provenientes do Censo Demográfico 2010 realizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE), consultados a partir do Índice Mineiro de Responsabilidade Social (IMRS). Esses dados são referentes aos 66 municípios presentes nessa região e foram analisados a partir de 9 variáveis socioeconômicas. O método de Ward foi aplicado e, visualmente, definiu-se o número de grupos através do ponto de corte de maior distância do dendrograma. Posteriormente, o método não hierárquico das k-médias alocou os municípios em relação às características socioeconômicas semelhantes em grupos, sendo eles: grupo 1 (G1) - 21 municípios; grupo 2 (G2) - 9 municípios; grupo 3 (G3) - 19 municípios; e o grupo 4 (G4) - 17 municípios. Observado o mapa da região, construído com relação aos grupos, percebe-se que os municípios alocados no G4 são, em grande parte, os de maior área territorial e contingente populacional. Além disso, há uma concentração dos municípios desse grupo em uma região específica, entretanto não é possível comprovar uma relação espacial a partir da metodologia adotada no estudo. A respeito do coeficiente de variação, observa-se que o G4 apresenta uma menor variabilidade relativa dos dados, o que aponta para uma semelhança no comportamento das variáveis nesses municípios. Após a análise de agrupamento, sugere-se que o G4 possui um desenvolvimento socioeconômico superior visto que dentre os 10 melhores PIB da mesorregião, 8 estão presentes nesse grupo.

* ana_coliveira@outlook.com, Universidade Federal de Alfenas

Modelos de riscos proporcionais de Weibull com erros de classificação e censura intervalar

Ana Carolina Fernandes Dias* ; Magda Carvalho Pires .

A análise de sobrevivência tem como principal objetivo estimar a distribuição do tempo até a ocorrência do evento de interesse. Este tempo é denominado tempo de falha, podendo ser, por exemplo, o tempo até a morte de um paciente. Usualmente os estudos de análise de sobrevivência são aplicados à área médica onde a detecção do evento de interesse é obtida através de exames de diagnóstico, exames que por sua vez estão sujeitos a erros de classificação. Sendo assim, se esses erros são ignorados o processo de inferência sobre a distribuição dos tempos de falha será comprometido e produzirá estimativas viciadas, comprometendo o resultado do estudo. Apresentamos assim um modelo capaz de incorporar os erros de classificação para dados de censura intervalar, considerando o caso em que os indivíduos são avaliados em diferentes instantes. Para obter as estimativas de máxima verossimilhança, utilizamos o pacote 'optim' do software R. Estudos de simulação demonstraram que a metodologia proposta apresentou resultados satisfatórios quando comparados ao modelo convencional (que desconsidera os erros de classificação), produzindo estimativas menos viciadas.

* karol.fdias@hotmail.com, Universidade Federal de Minas Gerais

Seleção de variáveis e comparação de modelos em regressão logística

Ana Júlia Alves Câmara^{*}; Rosangela Helena Loschi

Esse estudo tem como objetivo avaliar e aplicar métodos de seleção de variáveis e comparação de modelos em regressão logística sob uma abordagem bayesiana. São introduzidos conceitos básicos sobre o modelo do ponto de vista clássico e bayesiano, além do método de seleção avaliado, o SSVS (Stochastic Search Variable Selection) e conceitos sobre o DIC e a CPO, os quais são métodos utilizados na avaliação da qualidade de ajuste dos modelos.

Além de um estudo com dados simulados, foi realizado um estudo de caso com os dados do vestibular da UFMG, a fim de identificar os fatores responsáveis por influenciar o RSG médio dos alunos do ICEX. Através desse estudo, observou-se que poucas variáveis relativas à vida pré-acadêmica influenciam de fato no rendimento do estudante dentro da universidade. No entanto, dentre os fatores mais relevantes tem-se o tipo de escola do ensino médio, a opção de língua estrangeira, a ocupação e escolaridade da mãe.

^{*} anajulia.camara@gmail.com, Universidade Federal de Minas Gerais

Exportação de café verde no Brasil: uma análise utilizando o modelo de Holt-Winters

Ana Mara da Silva^{*}; Letícia Lima Milani Rodrigues

O café, nas suas diferentes formas, é um dos principais commodities exportados pelo Brasil desde os tempos de sua colonização. O produto foi inserido no território brasileiro em 1727 oriundo na Guiana Francesa, país sul-americano. O Brasil é a principal referência no mundo em exportações cafeeiras desde 1771, quando chegava a Portugal, país colonizador do Brasil, a primeira mostra de café provindo de terras brasileiras (FREITAS, 2006). No século XIX, a produção de café, que outrora era explorada apenas em nível nacional, foi inserida no mercado mundial tornando o Brasil importante e reconhecido internacionalmente na indústria cafeeira. A produção de café concentra-se atualmente na região sudeste do país, principalmente nos estados de São Paulo e Minas Gerais, devido principalmente as suas condições geográficas propícias ao cultivo. No presente trabalho, observou-se a série de exportações de café verde, também chamado de café em grãos, no período de janeiro de 2001 a julho de 2016. A partir da análise dos dados, foram feitos os seguintes testes: Teste de Sinal, para apurar a presença de tendência na série; Teste de Fisher, para a apuração de sazonalidade. Feitos os testes, pode-se perceber que as exportações de café apresentam tendência e sazonalidade, sendo necessária a adoção de um modelo apropriado para estudar séries com tais características. O modelo de Holt-Winters foi o escolhido para a análise, modelo esse que é baseado em equações constantes associadas a cada uma das componentes do padrão da série: nível, tendência e sazonalidade (MORETTIN; TOLOI, 2006). A abordagem determinada é totalmente adequada para o estudo de séries com tendência e sazonalidade, que é o caso das exportações cafeeiras do Brasil. As Medidas de Acurácia, MAPE, MAD E MSD, importantes para demonstrar os erros e ajudar na decisão se o modelo é bem ajustado aos dados, mostram-se boas, com percentuais de erros aceitáveis. A previsão obtida através do modelo pode ser considerada correta, pois os valores reais e previstos se mostraram bem próximos.

^{*} anamarasilva737@yahoo.com.br, Universidade Federal de Alfenas

Perdas na esperança de vida ao nascer decorrentes das principais causas de morte em Minas Gerais, 2015

Ana Paula Rocha Rosa^{*}; Larissa Gonçalves Souza

Nas últimas décadas, o Brasil tem experimentado um processo de transformação no seu perfil epidemiológico, caracterizado pela gradual substituição de óbitos por doenças infecciosas e parasitárias para aqueles ocasionados por doenças crônico-degenerativas. Nesse contexto, o objetivo deste trabalho foi mensurar o ganho potencial de esperança de vida ao nascer de homens e mulheres de Minas Gerais, em 2015, caso os óbitos por doenças do aparelho circulatório, do aparelho respiratório, neoplasias e causas externas fossem eliminados. Os dados coletados no DATASUS foram o número total de óbitos por todas as causas, número total de óbitos por cada causa selecionada e população no meio do ano de referência, estratificados por sexo e grupos de idade. Foram construídas tábuas de mortalidade ordinárias e tábuas associadas a um único decremento para cada causa de morte específica. A abordagem utilizada para construção das tábuas de um único decremento assume que a função da força de mortalidade pela causa i é proporcional a função da força de decremento por todas as causas combinadas no intervalo etário x a $x+n$. Os resultados indicam que a eliminação dos óbitos por doenças do aparelho circulatório levam aos maiores ganhos potenciais de esperança de vida ao nascer tanto para os homens (4,86 anos), quanto para as mulheres (5,79 anos). Por outro lado, o menor ganho potencial de esperança de vida ao nascer para as mulheres está associado à eliminação dos óbitos por causas externas (0,95 anos) e para os homens às doenças do aparelho respiratório (2,38 anos). De forma geral, os resultados permitem analisar a importância de cada uma dessas causas de óbito no perfil de morbimortalidade do estado, contribuindo para o planejamento adequado de políticas públicas de saúde.

^{*} anapaularocharosa@yahoo.com, Universidade Federal de Alfenas

Aplicação do modelo quase-binomial no processo de incubação de ovos

Andréia da Silva Meyer^{*}; Nelson José Peruzzi; Amanda Liz Pacífico Manfrim Peticarrari;
Gladys Dorotea Cacsire Barriga; Joseane Padilha da Silva.

O ovo é um sistema biológico complexo que, em condições normais, garante a sobrevivência e desenvolvimento de um embrião até a eclosão, proporcionando-lhe proteção e subsistência, porém qualquer variação severa pode acarretar instabilidade ao sistema e levar à morte embrionária. Estudos experimentais têm demonstrado que a taxa de eclosão depende, entre outros fatores, das características físicas dos ovos (Narushin and Romanov, 2002b). O presente trabalho tem como objetivo estudar a relação entre a taxa de eclosão e quatro parâmetros físicos dos ovos: peso do ovo (X_1); Espessura da casca do ovo (X_2); Esfericidade do ovo (X_3) e a relação entre gema e albúmen (X_4). Foi utilizada a teoria de modelos lineares generalizados, Nelder e Wedderburn (1972), para análise dos dados e atribuída a distribuição binomial para a variável resposta. Para corrigir a superdispersão detectada adotou-se o método da quase-verossimilhança na estimação do fator de heterogeneidade e consequente adição do parâmetro de dispersão no modelo. Para verificar o ajuste do modelo utilizou-se o gráfico normal de probabilidades com envelope simulado. Todas as análises foram realizadas no ambiente R: A language and Environment for Statistical Computing (2011). O modelo quase-binomial ajustou-se bem aos dados, mostrando uma boa ferramenta para análise de dados dessa natureza.

^{*} andreia.meyer@feb.unesp.br, FEB/UNESP

Análise de agrupamentos no estudo dos indicadores de mortalidade no Brasil

Anísio Pereira dos Santos Júnior* ; Priscila Neves Faria

O Sistema de Informações sobre Nascidos Vivos (SINASC) e o Sistema de Informações sobre Mortalidade (SIM) são as fontes de informação do Ministério da Saúde que possibilitam o monitoramento dos eventos vitais no Brasil e permitem a construção de indicadores de saúde de forma contínua para todo o país. Assim, objetivou-se com o presente trabalho analisar as informações de óbitos e aplicar metodologias de Estatística Multivariada na análise dos dados do Ministério da Saúde e do SIM a fim de obter respostas quanto à semelhança entre Estados da Federação no que diz respeito à mortalidade no ano de 2011. Para isso, foi utilizada a distância de Mahalanobis para a obtenção da matriz de distâncias e, para a formação dos grupos, o método aglomerativo do Centroid. O método de otimização utilizado para determinar o número ótimo de grupos no dendrograma foi o de Tocher original, que resultou na escolha de três grupos, onde dois grupos apresentaram estados que ficaram isolados, formando um grupo: São Paulo formou o primeiro grupo e o estado do Rio de Janeiro outro, possivelmente devido aos elevados números de óbitos em ambos os estados. Cerca de 32% dos óbitos nos dois estados estão relacionadas a causa doenças do aparelho circulatório. Tais motivos dessa elevada taxa estão associados principalmente aos fatores de risco como a obesidade, o fumo, hipertensão, diabetes, hipercolesterolemia, sedentarismo e estresse. Outro fator dessa porcentagem elevada pode estar relacionado à qualidade da assistência médica disponível.

* anisio4322@hotmail.com, Universidade Federal de Uberlândia

Comparação de modelos não lineares na descrição da altura de tronco de árvores

Ariana Campos Frühauf*, Jaqueline Gonçalves Fernandes, Joel Augusto Muniz,
Ana Carolina Maioli Campos Barbosa, Gabriel de Assis Pereira

Estudar as características de crescimento de árvores é importante para sua preservação e assim manutenção de seus benefícios para nosso planeta em geral. Este trabalho tem como objetivo avaliar o ajuste dos modelos Gompertz e Logístico aos dados de altura do tronco de árvores, comparando-se os modelos com base no valor do coeficiente de determinação ajustado e do critério de informação de Akaike.

* arianafruhauf@gmail.com, Universidade Federal de Lavras

Modelo HMM para classificação de condições de mercado

Arthur Tarso Rego^{*}

Conhecer as condições de mercado afeta a performance de estratégias de alocação de capital e pode ter impactos significativos nos retornos. Certos tipos de estratégia são mais bem sucedidas em períodos de alta volatilidade do mercado, como por exemplo a negociação via High-Frequency Trading (HFT). Outras, como a buy-and-hold são mais eficazes em momentos de baixa volatilidade, e beneficiam-se com o retorno dos ativos através de dividendos.

O objetivo deste trabalho é apresentar o método HMM, usando o pacote depmixS4 do R, como forma de classificar as condições de mercado com base em dados de retornos e volatilidade do índice Ibovespa, em detrimento das formas tradicionais de clustering.

^{*} arthur.tarso@outlook.com, Universidade Federal de Minas Gerais

Métodos estatísticos de proteção de dados confidenciais sob a condição de differential privacy

Augusto Felix Marcolin^{*}; Thais Paiva Galletti

O trabalho trata do assunto de privacidade em bases de dados. Apresentamos técnicas para anonimização dos indivíduos, assim como, uma garantia de confidencialidade aos presentes na bases de dados.

^{*} augustomarcolin@gmail.com, Universidade Federal de Minas Gerais

Análise temporal da exportação de barris de petróleo no Brasil

Bárbara Alves Lopes* ; Marcela de Marillac Carvalho; Manoel Vítor de Souza Veloso;
Letícia Lima Milani Rodrigues;

Ao longo dos anos, o Brasil vem ganhando destaque no que diz respeito à produção e exportação de petróleo. Houve uma melhora do ritmo de produção e exportação de barris de petróleo processados no país, devido a fatores como abertura do setor de exploração de petróleo no final da década de 1990. O presente trabalho analisa a série temporal referente a quantidade exportada de petróleo pelo país, em barris, no período de janeiro de 2001 a dezembro de 2016, obtendo previsões para o período de janeiro a março de 2017, por meio da metodologia Box-Jenkins. Um modelo SARIMA(1,1,0)(0,1,1)₁₂ com a inserção dos pontos de intervenção, referentes aos períodos de julho de 2005, abril de 2008 e dezembro de 2010, dadas por dummies temporais, foi o mais apropriado. Verificou-se a evolução da quantidade exportada pelo país, no período analisado, sendo influenciada por eventos econômicos. O modelo estimado conseguiu detectar os efeitos da descoberta do pré-sal, os impactos da crise do subprime e as implicações da variação da moeda nacional e das políticas econômicas anticíclicas adotadas pelo governo, no pós-crise, e produziu previsões próximas a realidade, confirmando a validade das intervenções incluídas no modelo.

* barbaraalvesl@yahoo.com.br, Universidade Federal de Alfenas

Previsões sobre a taxa de admissão de emprego no Brasil

Bruna da Costa Silva* ; Luiz Otávio Pala; Gislene Araujo Pereira

O presente trabalho analisa o número total de admissões de empregados no Brasil, sob o regime da Consolidação das Leis de Trabalho, que é um importante indicador econômico, por estar diretamente ligado ao desenvolvimento industrial, pobreza e desigualdade social (BARROS et al.). Os dados utilizados são provindos do Instituto de Pesquisa Econômica Aplicada (IPEA), dispostos sobre forma de série temporal mensal, correspondentes ao período de janeiro de 2000 a abril de 2017. Neste sentido, com o objetivo de avaliar o número de empregados, foi utilizado da metodologia de Box e Jenkins, com auxílio dos Softwares estatísticos R e Gretl. Após a análise do comportamento temporal e da adequação de um modelo que comporte tal objetivo, foram realizadas previsões para o nível de admissão de emprego nos meses entre maio e junho de 2017. Para identificação das componentes da série temporal tendência e sazonalidade, foram utilizados os testes de raiz unitária (Dickey-Fuller) e de sazonalidade (Fisher), respectivamente. No entanto, a sazonalidade não foi eliminada mesmo após a realização da diferenciação da série. Sendo assim, foi necessário a utilização de um modelo para análise de séries temporais que considere a sazonalidade. Foram considerados três modelos SARIMA(p,d,q)(P,D,Q), estes apresentaram significância considerável em conjunto a ruído branco. Desta forma, foi escolhido o modelo que apresentou menor medida de acurácia, de modo a retornar a melhor previsão da série possível, dentre os modelos apontados, expresso de forma SARIMA(1, 1, 2)(1, 1, 1). Por fim, foram realizadas as previsões para os meses do ano de 2017, apontando o desenvolvimento temporal do número de admissões no Brasil, sendo justificado pelo cenário econômico dos últimos anos.

* brunasilva208@hotmail.com, Universidade Federal de Alfenas - Campus Varginha

Partially linear models and their applications to change point detection of chemical process data

Clécio S. Ferreira; Camila B. Zeller^{*}; Aparecida M. Mimura; Júlio C. Silva

In many chemical data sets, the amount of radiation absorbed (absorbance) is related to the concentration of the element in the sample by Lambert–Beer’s law. However, this relation changes abruptly when the variable concentration reaches an unknown threshold level, the so-called change point. In the context of analytical chemistry, there are many methods that describe the relationship between absorbance and concentration, but none of them provide inferential procedures to detect change points. In this work, we propose partially linear models with a change point separating the parametric and nonparametric components. The Schwarz information criterion is used to locate a change point. A back-fitting algorithm is presented to obtain parameter estimates and the penalized Fisher information matrix is obtained to calculate the standard errors of the parameter estimates. To examine the proposed method, we present a simulation study. Finally, we apply the method to data sets from the chemistry area. The partially linear models with a change point developed in this work are useful supplements to other methods of absorbance–concentration analysis in chemical studies, for example, and in many other practical applications.

^{*} camila.zeller@ufjf.edu.br, Universidade Federal de Juiz de Fora

Uma rotina em R para o teste Scheffé em experimentos em fatorial triplo

Beatriz de Oliveira Rodrigues^{*}; Felipe de Melo Taveira; Flavio Bittencourt; Adriana Dias

Em experimentos em estrutura fatorial são aqueles em que os tratamentos são formados pelas combinações de dois ou mais fatores (duas ou mais variáveis independentes). São usados em praticamente todas as áreas de pesquisa. Uma das vantagens dos experimentos em estrutura fatorial é que eles permitem ao pesquisador observar vários fatores e estimar simultaneamente os vários efeitos principais e interação, permitindo a obtenção de conclusões mais amplas sobre o experimento. Entretanto, a aplicação de um teste de médias, como o teste Scheffé é difícil de ser executado manualmente e poucos programas de análises estatísticas fazem o teste para o caso de um fatorial triplo. Desta forma, este trabalho tem por objetivo apresentar uma rotina computacional que realiza o teste Scheffé em experimentos em estrutura fatorial tripla, considerando todos os casos, com interação tripla e/ou dupla significativa ou somente com os fatores principais significativos.

^{*} a14006@bcc.unifal-mg.edu.br, Universidade Federal de Alfenas

Estudo do número de óbitos no Brasil no período de 2000 a 2015 utilizando o modelo de series temporais

Camila Moreira Ribeiro* ; Mariana de Souza Gonçalves;
Bruna Leandra Zeferino Silva; Letícia Lima Milani

A mortalidade, tratada neste trabalho em sua forma geral, abrange todos os óbitos ocorridos no Brasil ao longo dos anos. Sua contabilização é de extrema importância, já que é um instrumento utilizado para a monitoração das condições de saúde populacional. Através de indicadores como o de mortalidade geral, é viável avaliar o desenvolvimento de distintas regiões, encontrando deficiências e problemas e propondo possíveis soluções aos mesmos. Em geral, a mortalidade tem demonstrado tendência decrescente em todo o mundo. Escolheu-se trabalhar com números de todo o país, para que se possa encontrar um panorama amplo de saúde nacional ao decorrer dos anos. Dessa maneira, neste trabalho, propõe-se observar o número de óbitos brasileiros, utilizando uma série de dados extraída do Ministério da Saúde, disponibilizada através do Sistema de Informações de Mortalidade (SIM). Para tanto, foi feita uma análise de série temporal, utilizando o modelo ajustado por meio da metodologia SARIMA (p,d,q)(P,D,Q). Foram coletados dados mensais, de janeiro de 2000 a dezembro de 2015, totalizando 180 observações. Portanto, por meio da análise do comportamento desta série temporal, buscase estudar o Brasil e seus níveis de mortalidade, verificando efeitos da sazonalidade e tendência. Por conseguinte, foi feita a modelagem, para que a previsão pudesse ser realizada. Dessa forma, realizou-se a previsão para os três meses posteriores. Finalmente, pretende-se também, reafirmar a importância da contabilização de indicadores de saúde, visto suas aplicações em diversas áreas de estudo, além de serem parâmetros indispensáveis para calcular o desenvolvimento e condições de saúde da população. Conclui-se, através dos resultados, que o modelo SARIMA (1,1,1)(1,1,1)₁₂ foi o que obteve melhores previsões, visto que essas acompanharam as oscilações da série.

* camilam.ribeiro@yahoo.com.br, Universidade Federal de Alfenas

Probabilidade conjunta do gráfico de controle de Shewhart

Camila Rafaela Gomes Dias^{*}; Juliana Vieira Gomes; José Ivo Ribeiro Júnior.

O presente trabalho buscou avaliar o desempenho do monitoramento de duas variáveis (Y_1 e Y_2), de forma univariada, pelo gráfico de controle de Shewhart da medida individual, em termos das probabilidades conjuntas dos alarmes falso (α_c) e verdadeiro (P_d), verificando os efeitos das correlações (ρ) 0; 0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8; 0,9; 1 entre Y_1 e Y_2 .

^{*} camilarafaelagomesdias@gmail.com, Universidade Federal de Viçosa

Análise estatística da aceitabilidade do brigadeiro sem lactose com adição de biomassa de banana verde

Camilla Stéfane Tolentino^{*}; Fabrícia de Matos Oliveira

Resumo: A banana verde é apropriada para o preparo de subprodutos, como a biomassa, isso se deve ao alto conteúdo de amido resistente (AR) e fibras assim melhorando o seu valor nutricional. A biomassa não altera o sabor dos alimentos em que está sendo adicionada, apenas aumenta o volume com a singularidade de acrescentar vitaminas e sais minerais, podendo ser agregada à maioria das receitas, tomando os alimentos funcionais do tipo prebióticos. Já os alimentos sem lactose devem apresentar a ausência da açúcar do leite e seus derivados, uma pessoa que tem intolerância à lactose não consegue processar no seu organismo o açúcar do leite e dividi-lo em glicose e galactose, havendo então a importância de ampliar cada vez mais ofertas de alimentos para intolerantes à lactose. Este trabalho teve como objetivo elaborar e verificar a aceitabilidade do brigadeiro sem lactose com adição de biomassa banana verde. Analisou-se o índice de aceitação do consumidor de produtos formulados a partir da polpa da banana verde e sem lactose por meio da análise sensorial, assim contribuindo com os subprodutos da banana verde auxiliando na qualidade de vida e ampliando a oferta de produtos para intolerantes a lactose. Para o presente estudo sensorial foi utilizadas amostras de brigadeiro feito em laboratório de análise sensorial, onde foram avaliadas os atributos: aroma, cor, sabor e textura na escala hedônica. Foi feita análise estatística descritiva dos dados e também análise de componentes principais. Na análise descritiva verificou-se que todos os atributos obtiveram notas altas, indicando a boa aceitação do produtos, na análise de componentes principais verificou-se que os atributos que mais influenciaram na aceitabilidade do brigadeiro foi textura e sabor e cor.

^{*} camillastefane@hotmail.com, Universidade Federal Uberlândia -UFU, Campus Patos de Minas.

Distribuição exponencial ajustada com estimadores equivariantes em eventos de precipitação máxima

Carlos José dos Reis* ; Lucas Monteiro Chaves; Eleanderson Campos Eugênio Filho;
Laís de Souza Monteiro.

A teoria da estimação equivariante, para famílias de locação, escala e de locação-escala, iniciou-se com os trabalhos de Edwin James George Pitman em 1939 e, a partir da década de 1950, uma discussão mais geral sobre esse assunto passou a receber contribuições de vários autores. Os estimadores equivariantes surgem das ações de um grupo sobre o espaço paramétrico, o espaço amostral e o espaço das estimativas. Devido à carência de trabalhos sobre esse assunto na literatura estatística brasileira, o objetivo geral neste trabalho foi estudar o ajuste da distribuição Exponencial de dois parâmetros, estimados com estimadores equivariantes, aos dados de precipitação máxima de Piracicaba-SP. Dessa maneira, os objetivos específicos foram: i) Verificar o ajuste da distribuição Exponencial aos dados de precipitação máxima nos meses de dezembro a março e ii) calcular as precipitações máximas esperadas para os tempos de retorno de 10, 30, 50, 70 e 100 anos no município. As séries históricas de precipitação máxima (1917-2016) foram obtidas junto à estação convencional do posto agrometeorológico da Escola Superior de Agricultura “Luiz de Queiroz” (ESALQ/USP). Verificou-se que os dados de precipitação máxima em todos os meses estudados satisfizeram a pressuposição de independência, conforme os resultados do teste de Ljung-Box ($\alpha = 0,05$). Verificou-se também que a distribuição Exponencial ajustou-se aos dados em todos os períodos mensais, com base no teste de kolmogorov-Smirnov ($\alpha = 0,05$). As maiores precipitações máximas para Piracicaba-SP, calculadas utilizando-se a distribuição Exponencial, são esperadas para os meses de fevereiro e dezembro.

* carlosjreis17@yahoo.com.br, Universidade Federal de Lavras

Aplicação inovadora na análise de padrão de pontos marcados

Carolina Cristina Bicalho^{*}; Raul A. Viana Madeira; Joelma Pereira; João Domingos Scalon

O objetivo básico da análise de processos pontuais marcados é verificar se existe independência entre marcas e pontos. Uma das ferramentas mais utilizadas neste tipo de análise é a função de correlação marcada. Nesse trabalho são utilizados métodos de Monte Carlo aplicados na função de correlação marcada para testar a hipótese de independência entre marcas e pontos. Esses métodos são aplicados a dados da distribuição de alvéolos em fatias de pão francês. Os resultados mostram que os métodos de Monte Carlo, associados a função de correlação marcada, podem ser instrumentos úteis para a análise da relação existente entre a localização e o tamanho dos alvéolos em pão francês e, conseqüentemente, auxiliar e inovar o controle da qualidade de produtos de panificação.

^{*} carolinabicalho@gmail.com, Universidade Federal de Lavras

Geodesign: pacote no R para planejamento amostral ótimo em geoestatística

Catarina Dall Agnol Zidde^{*}; Gustavo da Silva Ferreira

Neste trabalho é apresentado o pacote do R "geodesign", criado para realizar o processo de planejamento amostral ótimo em Geoestatística nos casos em que se deseja aumentar uma amostra pré-existente via maximização de funções utilidade.

^{*} catarinazidde@gmail.com, ENCE / IBGE

Perfil de uma unidade de dor torácica em um hospital universitário quanto ao tipo de cuidado de enfermagem

Clesnan Mendes-Rodrigues* ; Arthur Velloso Antunes; Guilherme Silva Mendonça; Iolanda Alves Braga; Fabíola Alves Gomes; Karla Fabiana Begosso Sampaio da Fonseca Carbonari

A utilização de sistemas de classificação de pacientes para delimitação da carga de trabalho em Enfermagem tem crescido como ferramenta administrativa. Apesar disto, o perfil de algumas unidades especializadas ainda é desconhecido. O objetivo deste estudo foi avaliar o perfil e a carga de trabalho em enfermagem em uma Unidade de Dor Torácica de um Hospital Universitário. Foi utilizado o sistema de classificação de pacientes adulto de Santos e classificados 66 paciente-dia. As avaliações foram nos turnos da manhã, tarde e noite. Os dados obtidos para o estudo são dados secundários e seu uso foi aprovado institucionalmente. O escore médio foi de 26,15 (mediana: 26,5; desvio padrão: 4,32; mínimo: 14; máximo: 43). Os tipos de cuidado não foram igualmente distribuídos ($p < 0,0001$), e o mais comum foi alta dependência (63,64%), seguido de semiintensivo (24,24%), sendo que o tipo de cuidado também não mostrou dependência do turno de avaliação ($p = 0,7551$). As áreas de cuidado que obtiveram maior pontuação quatro foram à deambulação e cuidado corporal. Para o estado mental, a motilidade e a alimentação predominou a pontuação um. Os dados demonstram a dependência dos pacientes quanto aos cuidados de enfermagem, principalmente a restrições de locomoção e repouso e cuidados restritos ao leito. Como predominou o tipo de cuidado alta dependência, os pacientes desse tipo de unidade depende de 10 horas de cuidados de enfermagem por dia. A unidade apresentou uma sub-dimensionamento independente dos parâmetros utilizados no calculo. Essas informações e indicadores podem permitir um cuidar mais planejado e um melhor dimensionamento de pessoal em unidades de dor torácica.

* clesnan@bol.com.br, Universidade Federal de Uberlândia

Adequabilidade dos métodos numéricos no ajuste da distribuição Gumbel

Daiane de Oliveira Gonçalves* ; Luiz Alberto Beijo; Fabricio Goecking Avelar

Devido ao aumento da ocorrência e da amplitude de eventos extremos como, por exemplo, chuvas torrenciais seguidas ou não de grandes períodos de estiagem, a previsão dos seus comportamentos torna-se de suma importância para governos e populações em geral. Um modelo comumente utilizado para descrever o comportamento das observações extremas é a distribuição Gumbel. O Método da Máxima Verossimilhança é o mais utilizado na estimação dos parâmetros da distribuição Gumbel. Como não há solução analítica para os Estimadores de Máxima Verossimilhança de todos os parâmetros da distribuição Gumbel, é necessário o uso de métodos numéricos para se aproximar os valores das estimativas desses parâmetros. Métodos numéricos diferentes produzem valores diferentes para as estimativas dos parâmetros. Como pequenas diferenças nas estimativas dos parâmetros podem causar grandes diferenças na previsão dos eventos extremos, é preciso determinar qual método numérico produz estimativas mais precisas e acuradas. Nesse sentido, o presente trabalho propôs a comparação da acurácia e da precisão das estimativas de máxima verossimilhança dos parâmetros da distribuição Gumbel obtidas por dois métodos numéricos iterativos muito utilizados para esse fim: o Método de Newton-Raphson e o Método de Nelder-Mead. Para isso, foram simulados dados da distribuição Gumbel com os seguintes cenários: parâmetros posição com os valores 10, 20, 30, 50 e 100; parâmetro escala iguais a 20%, 40%, 60%, 80%, 100% e 120% do valor do parâmetro posição e tamanhos amostrais 20, 30, 40, 50, 60 e 100. Para cada um desses cenários foram simuladas 2.000 séries de dados. A acurácia e a precisão foram avaliadas por meio do erro quadrático médio e do viés médio relativo. Concluiu-se que o Método de Newton-Raphson e o Método de Nelder-Mead apresentaram precisão e acurácia iguais tanto para a estimativa do parâmetro escala quanto para a do parâmetro posição.

* daiane_og@hotmail.com, Universidade Federal de Alfenas

Aplicação de teoria de resposta ao item na avaliação de questões em estatística aplicada

Deive Ciro de Oliveira*, Júlio Sílvio de Sousa Bueno Filho

O processo de avaliação é uma tarefa de suma importância em qualquer projeto de ensino aprendizagem. Assim a construção/aplicação de métodos para avaliar questões presentes em avaliações pode melhorar sua elaboração. Neste trabalho descreve-se a construção inicial de um banco de questões em estatística aplicada e sua avaliação utilizando-se Teoria de Resposta ao Item (IRT:Item Response Theory). Dois resultados são apresentados: o primeiro descreve uma ferramenta operacional para gerenciamento de avaliações facilitando a extrapolação da escala (aplicação de muitas avaliações). O segundo resultado apresenta a estimação do modelo TRI de 2 parâmetros (τ : chute inicial (fixo=0.5), d : intercepto e α : inclinação) para respostas dicotômicas. Neste modelo o chute inicial τ relaciona-se a chance de acerto da questão por mero acaso, o parâmetro d intercepto relaciona-se a dificuldade da questão e α informa sobre grau de discriminação do item. O conjunto de dados em análise engloba uma prova de 33 questões aplicada a um conjunto de 35 alunos. Baseado nos resultados de estimação discute-se as questões quanto a sua qualidade enquanto instrumento de avaliação. Ressalta-se que análise é um mero indicativo da qualidade da questão servindo de indício para uma análise empírica mais detalhada. Ademais, deve-se considerar o número reduzido de amostras (35 alunos) o que afeta a análise. Foram identificadas 11 questões como não adequadas. Destas 5 se caracterizam pelo $\hat{\alpha} < 0$ que pode representar uma anomalia no método de ensino (erro do docente, má compreensão por parte do aluno, etc). Os 6 itens restantes apresentaram valores constantes em α o que pode ser um indicativo destes não apresentarem poder de discriminação entre: alunos com habilidade exigida e alunos sem habilidade exigida. As questões restantes se mostraram efetivas quanto a tarefa de discriminação, variando segundo α e d .

* deive.oliveira@gmail.com, Unifal-MG/UFLA

Análise do número de caso de tuberculose via modelos de séries temporais no estado do Ceará.

Denize Palmito dos Santos^{*}; Kelly Pereira de Lima; Allana Livia Beserra Paulino;
Pablo Lourenço Ribeiro de Almeida; Augusto Ramalho de Moraes

A tuberculose (TB) é uma doença infectocontagiosa, no caso, estima-se que um terço da população mundial está infectada e sendo que grande parte irá desenvolver e transmitir a doença (Kritski et al, 2007). O aumento dos casos de tuberculose está relacionado diretamente ao aumento da pobreza, à má distribuição de renda e ao avanço da urbanização. A Organização Mundial de Saúde estima que anualmente sejam notificados cerca de seis milhões de novos casos em todo o mundo, levando mais de um milhão de pessoas a óbito. Além disso, a OMS prevê que se não houver a reversão da gravidade desse quadro de mortalidade, até 2020 um bilhão de pessoas sejam infectadas, e no caso, 35 milhões possam morrer em decorrência da TB. O Brasil ocupa o 17º lugar é responsável, por cerca de 80% do total de casos de tuberculose no mundo. Desta maneira, no Brasil, a TB é a nona causa de hospitalização e a quarta em mortalidade por doenças infecciosas. Com isso, esse estudo tem como objetivo compreender o comportamento dos casos notificados de tuberculose no estado do Ceará, por meio de séries temporais, além disso, comparar os modelos de suavização, os modelos de espaço de estados (ets) e o modelo SARIMA, com relação à previsão. Assim, as informações obtidas nesse trabalho podem subsidiar a criação de estratégias e intervenções locais, ou mesmo, o melhoramento da política de saúde pública no Ceará. Os dados para análise foram obtidos no Sistema Nacional de Agravos e Notificações (SINAN), durante o período mensal de 2001 a 2015. Ao analisar os dados verificou-se que não houve a necessidade de transformação dos dados. Além disso, usou-se o teste de Cox-Stuart e verificou-se que a série não possui o componente de tendência. Para verificar a sazonalidade foi usando o teste não paramétrico de Kruskal-Wallis constatou-se a existência do componente de sazonalidade. No caso, ajustou-se um modelo de suavização exponencial de Holt-Winters (HW), os modelos de espaço de estados (ets) e o modelo SARIMA₁₂(3,0,1)x(0,1,1) em termos de previsão. Os modelos obtiveram estimativas bem próximas, mas ao usa o critério de seleção do modelo (MAPE) averiguou-se que o SARIMA teve o menor valor de critério, em relação ao demais. Com isso, conclui-se que o modelo SARIMA₁₂(3,0,1)(0,1,1) foi considerado o mais adequado para descrever o comportamento da série, bem como ser utilizado para fazer as previsões, pois comparado aos outros modelos, o mesmo possuiu o menor valor de MAPE.

^{*} denizepalmito@gmail.com, UFLA- Universidade Federal de Lavras

Função verossimilhança e estimadores de máxima verossimilhança

Desirée Thiely Alves Ramalho Costa *

Estudo feito para obter e comparar os estimadores não viciados (ENV) e os estimadores de máxima verossimilhança (EMV) para os parâmetros do modelo normal com média e variância desconhecidos e para os parâmetros do modelo de regressão linear simples.

* desiree_fly@hotmail.com, Universidade Federal de Minas Gerais

Comparação de modelos de regressão não linear na descrição da liberação de CO₂ de resíduo orgânico no solo

Edilson Marcelino Silva^{*}; Jaqueline Gonçalves Fernandes; Joel Augusto Muniz;

A criação de animais em confinamento gera grandes quantidade de resíduos. É de fundamental importância conhecer taxa de mineralização do carbono na dinâmica de decomposição. O objetivo do trabalho foi comparar modelos não lineares que descrevem a mineralização do carbono no solo. Com base nos avaliadores de qualidade do ajuste foram obtido bons ajustes pelos modelos. Com base no AIC o modelo que melhor descreveu a liberação de CO₂ dos dejetos de suínos foi o Stanford & Smith, seguido dos modelos de Marion e Juma, respectivamente.

^{*} edilsonmg3@hotmail.com, Universidade Federal de Lavras

Perfil dos periódicos científicos publicados pela Universidade Federal de Minas Gerais (UFMG)

Edilson Messias Gonçalves Júnior* ; Ronaldo André Lopes; Eric Batista Ferreira

O conhecimento científico que é produzido pela humanidade é registrado e divulgado principalmente por periódicos científicos e, cada vez mais, o formato eletrônico vem sendo preferido em detrimento do impresso, devido a praticidade e economia de recursos e tempo.

A Universidade Federal de Minas Gerais (UFMG) publica revistas desde 1894, quando se inicia a sua Revista da Faculdade de Direito e, atualmente, conta com 62 publicações eletrônicas e impressas em seu portal.

Com a finalidade de traçar um perfil quali e quantitativo dos periódicos científicos da UFMG, este trabalho considerou dois períodos de tempo, um triênio (2010 a 2012) e um quadriênio (2013 a 2016). Neles, foi investigado o índice Qualis das revistas, os índices h e g, o número de citações por artigo, o número de artigos publicados, o número de autores por artigo, o índice hi-norm e o fator de impacto. Foi utilizado o software Harzing's publish or perish para obter as métricas supracitadas e o portal Web Qualis. O índice de correlação de Spearman, assim como o teste de significância para o mesmo, foi utilizado para medir o grau da associação do Qualis e as outras variáveis.

Pôde-se constatar que 66% das revistas não possuem DOI, todas possuem ISSN, 58% foram criadas após o ano 2000, 55% tem periodicidade semestral, 90% não cobram do autor, o número médio de citações por artigo era 1,95 de 2010-2012 e foi para 3,10 de 2013 a 2016, o número de artigos publicados era de 90,55 de 2010-2012 e passou para 152,42 de 2013-2016, o número de autores por artigo era de 2,01 de 2010-2012 e passou para 2,22 de 2013-2016, o índice g era de 7,76 de 2010-2012 e passou para 5,58 de 2013-2016, o índice h era de 5,15 de 2010-2012 e passou para 4,00 de 2013-2016, o fator de impacto médio é de 0,23 nos últimos dois anos (2015-2016) e o qualis modal era B5 de 2010-2012 e permaneceu sendo de 2013-2016.

* juninho.goncalves780@gmail.com, Universidade federal de Alfenas - UNIFAL

Escolha entre o modelo cúbico com platô e o Von Bertalanffy que melhor descreve número médio de tubérculo por planta de batata

Édipo Menezes da Silva* ; Victor Ferreira da Silva; Jaqueline Gonçalves Fernandes;
Tales Jesus Fernandes; Joel Augusto Muniz

A cultura da batata (*Solanum tuberosum* L.) originou na Cordilheiras dos Andes, próximo ao Lago Titicaca e foi levada para outras regiões do mundo pelos colonizadores europeus, se tornando importante fonte de alimento para população mundial. A China é o maior produtor juntamente com a Índia, sendo que ambos representam mais de um terço da produção mundial. Informações da Embrapa (EMBRAPA, 2017) destacam que a produção anual do Brasil é próxima de 3,5 milhões de toneladas, envolvendo cerca de 5000 produtores concentrados em sete estados (MG, SP, PR, RS, SC, GO e BA). Da produção brasileira, 90% é destinada ao consumo in natura e 10% para a indústria, destacando-se entre os cultivares mais plantada a Ágata. No estudo do desenvolvimento da planta da batata ao longo do tempo a principal variável é o número médio de tubérculos por planta quem tem um comportamento sigmoidal bem descrito por modelos não lineares. O trabalho teve por objetivo comparar o ajuste do modelo cúbico com platô e o de Von Bertalanffy em estudo com número médio de tubérculo por planta ao longo do tempo, em dias após o plantio. Os dados foram extraídos de um experimento de Fernandes et al. (2010), realizado em Itaís, SP, com a cultivar Ágata, onde a variável resposta foi avaliada aos 28, 35, 42, 49, 56, 63, 70, 77, 84, 91 e 97 dias após o plantio. A estimação dos parâmetros foram feitas pelo método de Gauss-Newton, utilizando-se rotina do software R versão 3.4.0. O ajuste dos modelos foi comparado com base nos valores de R^2 ajustado. Foi feito o diagnóstico sobre os resíduos, avaliando-se a normalidade, homocedasticidade e a independência com base nos teste de Shapiro-Wilk, Breusch-Pagan e Durbin-Watson. Os resultados indicaram o R^2 ajustado para o modelo com platô e Von Bertalanffy, respectivamente, de 0,92 e 0,99, mostrando que o modelo de Von Bertalanffy é o mais adequado para descrever a variação do número médio de tubérculo por planta da batata da cultivar Ágata.

* ediposvm01@gmail.com, Universidade Federal de Lavras

Aplicação da geoestatística espaço-temporal na distribuição das chuvas na Paraíba, Brasil

Elias Silva de Medeiros^{*}; Kelly Pereira de Lima; Henrique José de Paula Alves;
Prof. Dr. Renato Ribeiro de Lima; Prof. Dr. Ricardo Alves de Olinda

A modelagem espaço-temporal vem sendo abordada em trabalhos teóricos, bem como em estudos que utilizam para aplicação na resolução de diversos fenômenos, e este tipo de abordagem tem ganhado visibilidade nos últimos anos, uma vez que, fenômenos indexados tanto no espaço como no tempo precisam ser compreendidos. Assim, o interesse em analisar a dinâmica espaço-temporal tem proporcionado o desenvolvimento de modelos espaço-temporais, em áreas como epidemiologia, demografia, ciências ambientais, hidrologia, entre outras. A metodologia geoestatística espaço-temporal será aplicada aos dados da precipitação total mensal do estado da Paraíba, localizado na região nordeste do Brasil. O principal objetivo desse trabalho foi ajustar diferentes estruturas de covariâncias espaço-temporal da classe produto-soma generalizada. Após os ajustes dessas estruturas foram comparadas as diferenças entre o semivariograma amostral e o ajustado dessa classe. Os resultados indicaram que a distribuição das chuvas no estado da Paraíba está correlacionada tanto espacialmente e quanto temporalmente. Entretanto, análises mais detalhadas devem ser utilizadas para determinar a escolha de uma estrutura de covariância espaço-temporal.

^{*} eliasestatistica@gmail.com, Universidade Federal de Lavras

Daltonismo e acessibilidade em representações gráficas no Brasil

Elizabeth Duane Santos da Costa* ; Lucas Ramos Dias França; Luís Fernando Israel
Assunção; Mariana de Castro Pasqualini

No olho humano, encontram-se os cones, células fotossensíveis responsáveis pela identificação das frequências das cores azul, verde e vermelho, contidas no espectro de luz visível. Um ou mais cones podem apresentar um mau funcionamento devido à condições genéticas, caracterizando o distúrbio visual conhecido com Daltonismo ou Discromatopsia, dividido em três tipos: Deuteranopia, Protanopia e Tritanopia. Pelo menos 8% dos homens e 0,5% das mulheres são afetados por algum tipo, sendo 8,7 milhões de pessoas apenas no Brasil. Dessa forma, a presente pesquisa busca compreender a acessibilidade dos gráficos produzidos pelos institutos IBGE (Instituto Brasileiro de Geografia e Estatística) e IBOPE (Instituto Brasileiro de Opinião Pública e Estatística) no período entre maio de 2016 e maio de 2017. Para isso, foram exduídas variações na intensidade, combinações entre tipos e monocromatia, com o intuito de evitar possíveis redundâncias. No total, 992 gráficos foram analisados e 520 foram encaixados na categoria de gráficos potencialmente problemáticos, ou seja, os que contêm duas ou mais cores e que dependem exclusivamente delas para a identificação das variáveis. A partir disso, os gráficos foram transformados usando o plugin Color Vision Deficiency Simulator do software GIMP para simular a visão deficiente e, em seguida, foi feita a análise da identificação de cores nos gráficos modificados. Com essa análise, verificou-se que 14% dos gráficos produzidos pelos dois institutos no intervalo de um ano são inacessíveis para 8,7 milhões de daltônicos, o que indica um problema na elaboração dessas imagens, que têm por objetivo traduzir informações estatísticas de modo compreensível.

* duanecosta@gmail.com, Universidade Federal de Minas Gerais

Comportamento espacial e temporal do IDEB-MG dos anos finais do ensino fundamental

Erika Eunizia Nunes^{*}; Prof. Dr. Ednaldo Carvalho Guimarães- Orientador

O presente trabalho teve como objetivo analisar a dependência espacial do IDEB (Índice de Desenvolvimento da Educação Básica) do 9º ano do ensino fundamental e verificar, descritivamente, se houve melhorias entre os índices do IDEB no período de 2007 a 2013. Utilizou-se uma amostra dos índices do IDEB de 302 cidades de Minas Gerais, dos anos 2007, 2009, 2011 e 2013. Com base nas metas para as escolas, estipuladas pelo INEP, os índices do IDEB foram codificados no valor 1 para as observações abaixo da meta e 0 caso contrário. A variável dicotômica foi submetida à análise geoestatística, sendo utilizada a ferramenta de ajuste de semivariogramas para a descrição e determinação da dependência espacial e a krigagem indicadora para a interpolação e mapeamento das probabilidades de não se atingir as metas do IDEB. O modelo de semivariograma ajustado foi o exponencial para todos os anos, com alcances da dependência espacial variando de 1,17 a 2,34 graus. Verificou-se que, ao longo do tempo, houve uma tendência de redução das áreas com maior probabilidade de não se atingir as metas do IDEB.

^{*} erikayurischulz@hotmail.com, Universidade Federal de Uberlândia

Análise do sistema de medição em uma linha de classificação de frutos de melão amarelo

Fabrício Pelizer de Almeida* ; Renner Ribeiro de Oliveira

Os dados de uma determinada medição são utilizados com frequência em manufatura de modo comparativo aos limites de controle previamente estabelecidos, como parâmetro importante para a avaliação e diagnóstico do processo. Outro aspecto relevante da análise do sistema de medição (SMA) é a possibilidade de se determinar a associação entre as variáveis que incidem no processo de estudo, com o intuito de se verificar o status da padronização e a capacidade de atendimento das especificações de produto e processo (MONTGOMERY, 2004). O presente artigo analisou o sistema de medição das características físicas de frutos de melão amarelo in natura, no município de Sobral (CE). Os dados foram coletados em uma linha de produção no período de 03 à 24 de abril de 2015, em uma empresa de beneficiamento de frutos de melão para exportação, especificamente quanto à característica de peso (g). Foram considerados na pesquisa os três operadores responsáveis, dez peças escolhidas aleatoriamente e três repetições sob o mesmo critério e o atendimento das especificações para cada parâmetro analisado. Os dados foram tabulados e submetidos à análise estatística utilizando-se o pacote Quality Tool e a função Gage-Study R&R (ANOVA) do software MINITAB (v. 17). O modelo ANOVA, com e sem interação entre frutos, operadores e o sistema mostrou-se bastante consistente na avaliação das medições durante o processo. No teste de F ($p < 0,005$) apenas os frutos isoladamente, e com interação frutos-operadores foram significativos ($p < 0,001$). A variação fruto-à-fruto corresponde à 82,8% das variações totais do sistema enquanto que a interação fruto-operadores representa 8,57% do total de medidas. A desuniformidade dos lotes de frutos de melão (cerca de 41,4%, ou seja, a variação total do calibre peça-à-peça) é uma das causas evidentes de variação acima do limite aceitável no processo, afetando a avaliação pelos operadores que não conseguem reproduzir o padrão operacional, por esse motivo. Portanto, caberia o enfoque de padronização de frutos e pré-seleção desses lotes na empresa, diminuindo a variabilidade de medidas peça-à-peça e corroborando com a proposta de treinamento técnico e acuracidade operacional na linha de produção agroindustrial.

* fabricio.almeida@uniube.br, Universidade de Uberaba

Teste Scheffé em fatorial duplo cruzado: uma rotina em R

Felipe de Melo Taveira^{*}; Beatriz de Oliveira Rodrigues; Flavio Bittencourt; Adriana Dias

Experimentos em que os tratamentos são formados pela combinação de dois fatores (variáveis independentes) são denominados experimentos em estrutura bifatorial. Experimentos em que os tratamentos apresentam esta estrutura podem ser instalados em delineamentos como o inteiramente ao acaso, em blocos completos, em quadrados latinos, etc. Conclusões mais amplas podem ser obtidas em experimentos deste tipo por permitirem avaliar os efeitos principais e os efeitos da interação dos fatores em estudo. Quando os fatores em estudo forem qualitativos (ou pelo menos um deles) e o teste F da análise de variância for significativo, tanto a interação ou para um dos fatores, pode-se utilizar o teste de comparação múltipla de Scheffé. Este é um teste que permite ao pesquisador realizar variadas comparações entre os efeitos do modelo que estão em teste, para isto, basta definir o(s) contraste(s) de interesse. Este trabalho teve como objetivo apresentar uma rotina em linguagem R que permitirá ao usuário realizar o teste de Scheffé para experimentos bifatoriais com interação significativa ou quando se a interação não for significativa, nos casos em que pelo menos um dos fatores qualitativos sejam significativos.

^{*} felipe.melo.taveira@gmail.com, Universidade Federal de Alfenas

Comparação entre os modelos Gompertz e logístico para avaliar o crescimento em matéria fresca da batata

Felipe Augusto Fernandes* ; Édipo Menezes da Silva; Victor Ferreira da Silva;
Tales Jesus Fernandes;Joel Augusto Muniz.

A batata (*Solanum tuberosum* L.) é considerada a terceira fonte alimentar da humanidade, sendo suplantada pelo arroz e trigo (Silva, G. O. et al., 2015). A cultivar de batata Asterix possui uma película rosada, é uma das mais cultivadas no Brasil apresentando rendimento e produtividade elevados. Conforme Fernandes et al. (2010), modelar o crescimento, pode facilitar a tomada de decisões relativas ao manejo da cultura. Como esse crescimento apresenta um padrão não linear, uma boa tentativa de ajuste são os modelos não lineares. O objetivo deste trabalho foi comparar o ajuste dos modelos não lineares Gompertz e Logístico na descrição do crescimento em matéria fresca do tubérculo da batata. Os dados analisados foram retirados de Fernandes et al. (2010) e são referentes a produtividade de matéria fresca de tubérculo da batata cultivar Asterix, considerando os dias após o plantio. Todas as análises estatísticas foram desenvolvidas no software estatístico R versão 3.4.0. A qualidade de ajuste fornecidas pelos modelos foi comparada utilizando os seguintes avaliadores: Critério de informação de Akaike (AIC); soma de quadrados de erro (SQE), e o coeficiente de determinação (R^2). Nos dois modelos, os parâmetros foram significativos, segundo o teste t , ao nível de 1%. Ao realizar os testes Shapiro-Wilk, Durbin-Watson e Breusch-Pagan, verificou-se que os resíduos são normais, independentes e homocedásticos ao nível de 1% de significância. O modelo que melhor se ajustou ao crescimento da produtividade foi o de Gompertz, pois apresentou valores menores de AIC e SQE, além de um maior coeficiente de determinação.

* fernandesfelipepest@gmail.com, Universidade Federal de Lavras

Modelos lineares generalizados: aplicações às Ciências Atuariais

Fernanda Venturato Roquim* ; Danilo Machado Pires

Um modelo de regressão consiste em relacionar uma variável resposta a uma ou mais variáveis explicativas a fim de se entender algum fenômeno. Em um modelo linear clássico, a variável resposta precisa seguir uma distribuição normal e, por isso, é necessário buscar uma solução para atender às demais distribuições. Nelder e Wedderburn (1972), propuseram como alternativa os modelos lineares generalizados (MLG) que agregam todas as distribuições que pertençam à família exponencial, abrindo um leque de opções para a distribuição da variável resposta. Em especial, uma área que beneficia-se bastante dos MLG é a área das Ciências Atuariais que lida com dados de diversas naturezas, incluindo técnicas específicas de análise de riscos e expectativas dentro do ramo segurador e previdenciário. A atuária é uma ciência relativamente nova, sendo escassas as bibliografias em português sobre o tema, fato que conflita diretamente com a situação de expansão do mercado segurador nacional. O objetivo deste trabalho foi estudar os modelos lineares generalizados implementados às ciências atuariais, principalmente em precificação de seguros, contribuindo para o avanço da ciência na região. Esta pesquisa demandou levantamento de bibliografia selecionada sobre o tema. Após o estudo dos métodos, foram ajustados três modelos à dados da atuária, Poisson, Gama e Normal Inversa através de recursos computacionais. Por fim espera-se fortalecer esta linha de pesquisa no meio acadêmico, podendo os resultados obtidos serem divulgados à comunidade e os modelos encontrados explorados pelas seguradoras.

* fer_venturato@yahoo.com.br, Universidade Federal de Lavras

Utilização da biblioteca JavaScript D3 para a construção de gráficos estudados no ambiente web

Elizabeth Duane Santos da Costa; Francisco José de Sousa Junqueira* ; Marina Alves Amorim; Rafael Tavares Silva; William Giani Duani Martins

Estudo e utilização da biblioteca javascript D3, na recriação de gráficos estudados na disciplina DAD, em formato de página web, explorando características interativas e de personalização visual.

* franjobh@gmail.com, Universidade Federal de Minas Gerais

Prior specifications to handle monotone likelihood in the Cox regression model

Frederico M. Almeida^{*}, Enrico A. Colosimo, Vinícius D. Mayrink

The monotone likelihood is a phenomenon that may affect the fitting process of well-established regression models such as the Cox proportional hazards model. In short, the problem occurs when the likelihood converges to a finite value, while at least one parameter estimate diverges to $+\infty$. In survival analysis, monotone likelihood primarily appears in samples with substantial censored times and containing many categorical covariates; it is often observed when one level of a categorical covariate has not experienced any failure. A solution suggested in the literature (known as Firth correction) is an adaptation of a method originally created to reduce the bias of maximum likelihood estimates. The method leads to a finite estimate by means of a penalized maximum likelihood procedure. In this case, the penalty might be interpreted as a Jeffreys type of prior widely used in the Bayesian context. 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. A simulated study is developed, based on Monte Carlo replications and distinct sample sizes, to evaluate the impact of the suggested priors in terms of inference. A real application is also presented to illustrate the analysis using a melanoma skin data set.

^{*} falmeida856@gmail.com, DEST-UFMG

Redes neurais artificiais como alternativa à seleção genômica ampla para a predição de características complexas

Gabi Nunes Silva^{*}; Isabela de Castro Sant'Anna; Ithalo Coelho de Sousa;
Cosme Damião Cruz; Moysés Nascimento

No melhoramento genético o grande desafio é recomendar genótipos superiores para aumentar a produtividade e a qualidade da produção. Para tanto, a seleção genômica ampla (GWS), tem sido adotada agregando informações moleculares aos valores fenotípicos mensurados. No entanto, problemas estatísticos relacionados à dimensionalidade e multicolinearidade e problemas genéticos relacionados a não inclusão de efeitos de dominância e epistasia nas modelagens adotadas prejudicam a acurácia da GWS. Com base no exposto, propõe-se o uso das Redes Neurais Artificiais (RNA) na predição de valores genéticos para duas características quantitativas, em detrimento ao modelo RR-BLUP para GWS. No estudo foram utilizados 500 indivíduos provenientes de uma população simulada F1, oriunda do cruzamento de duas populações contrastantes, também obtidas via simulação. Os 500 indivíduos foram genotipados com 1000 marcadores. Procedeu-se então à redução de dimensionalidade das matrizes de marcadores via Regressão Stepwise. Assim, para as análises com RR-BLUP\GWS e RNA foram utilizados os arquivos reduzidos, com 500 indivíduos e 150 marcas selecionadas. Os resultados demonstraram que a redução de dimensionalidade aumenta em até duas vezes a acurácia de ambas as metodologias. Demonstraram ainda que as RNAs propiciaram predições mais acuradas que o RR-BLUP (REQMv de 106.4712 e 3.6312 para o RR-BLUP e RNA, respectivamente) no cenário mais simples e que presença de epistasia e dominância não foram fatores limitantes para as metodologias propostas, tendo em vista o baixo valor de REQM alcançado (20.06761) para o cenário mais complexo avaliado após a redução. Esse trabalho demonstra que o uso das Redes Neurais Artificiais no melhoramento constitui uma promissora ferramenta para fins de predição.

^{*} gabi.nunessilva@yahoo.com.br, Universidade Federal de Viçosa

A utilização da teoria de resposta ao item e modelagem log-normal do tempo para classificação de palavras

Gabriel Oliveira Assunção^{*}; Marcos Antônio da Cunha Santos

O estudo apresenta um método para o aperfeiçoamento na avaliação da medição de Dislexia do Desenvolvimento. Foi utilizado como meio de medição a aplicação de um teste de 323 palavras com vista a medir a precisão de acertos e o tempo utilizado para a leitura. Foram utilizadas no aperfeiçoamento: a Teoria de Resposta ao Item (TRI) para a precisão e, para o tempo de leitura, uma modelagem através da distribuição Log-Normal. O aperfeiçoamento foi realizado através de uma reclassificação das palavras com os resultados apresentados nas técnicas citadas anteriormente. Para isto, foi utilizado o método K-means de agrupamento, mas o tempo de resposta não mostrou significante nesta tentativa. Foi proposto um agrupamento utilizando todos os resultados anteriores para, assim, haver uma nova classificação das palavras e uma busca de melhora no teste. Os resultados obtidos foram que o método K-means agrupa utilizando apenas o parâmetro de dificuldade do item para classificar os itens como difíceis ou fáceis, e que a correlação entre a capacidade de leitura e o tempo de leitura era baixa. Foi concluído, então, que o método proposto para condensar informações sobre a precisão e para o tempo é bom, apenas um parâmetro seria necessário para classificar as palavras e que o fato da correlação baixa entre a capacidade de leitura e o tempo de leitura pode ser significativo para este resultado.

^{*} gabrieloliveira1995@gmail.com, Universidade Federal de Minas Gerais

Previsão do perfil das instituições envolvidas em estratégias de F&A do setor bancário brasileiro

Gabriel Rodrigo Gomes Pessanha*, Poliana Maria Benelli, Matheus Saraiva Alcino,
Matheus Henrique Ribeiro

O objetivo deste trabalho é identificar o poder de determinação das características da firma, especialmente, as econômicas-financeiras, no perfil assumido (adquirida ou adquirente) pelas instituições em processos de Fusões e Aquisições (F&A). O estudo pode ser classificado como descritivo-exploratório com abordagem quantitativa. Os dados utilizados no decorrer do trabalho se referem ao setor bancário brasileiro e estão compreendidos entre os anos de 1994 e 2015 e foram obtidos por meio da base de dados do Economática e do Banco Central. Os modelos utilizados no decorrer deste trabalho foram: análise discriminante linear, regressão logística e redes neurais artificiais. Os resultados indicaram que as características econômico-financeiras da firma são decisivas para a ocorrência de tais estratégias. Ressalta-se que os resultados encontrados pelas redes neurais foram mais assertivos que as demais metodologias empregadas, todavia, na tentativa de sanar uma das limitações do referido método, propôs-se a utilização de uma modelagem híbrida com o ajuste a priori de outras metodologias, pois assim é possível fazer um estudo prévio do relacionamento entre as variáveis induzidas no modelo. Os resultados foram, na sua maioria, coerentes com a teoria e com as expectativas iniciais. Por fim, ressaltam-se a importância e a relevância dos resultados encontrados, dado o potencial de contribuição com a escassa literatura sobre F&A em mercados emergentes e para o atual debate acerca da concentração no setor bancário e dos seus efeitos sobre os retornos dos bancos adquirentes, mediante teste empírico das teorias já desenvolvidas e promoção de um patamar para novas pesquisas.

* gabrielrgp@yahoo.com.br, Universidade Federal de Alfenas - UNIFAL

Regressão quantílica na estimação da eficiência técnica da agricultura familiar em Minas Gerais

Gabriela França Oliveira* ; Raimundo Cardoso de Oliveira Neto;
Ana Carolina Campana Nascimento; Moysés Nascimento; Camila Ferreira Azevedo.

A agricultura familiar é bastante presente no Brasil, bem como Minas Gerais, e constitui a base da economia de muitos municípios no país. Por este motivo, é uma modalidade cuja eficiência influencia os índices de desenvolvimento municipal. Neste sentido, estudos têm sido realizados para avaliar a eficiência destes produtores. Estes estudos baseiam-se em técnicas como Análise Envoltória de Dados (DEA) e Fronteira Estocástica (SFA). No entanto, estes modelos podem resultar em estimativas inconsistentes da eficiência, uma vez que esta modalidade de agricultura é bastante heterogênea. Visando contornar esta limitação, este trabalho propõe estimar a eficiência técnica da agricultura familiar em Minas Gerais, a nível municipal, por meio do ajuste de regressão quantílica, utilizando dados do Censo Agropecuário de 2006. Para tanto, considerou-se como insumos a área utilizada para agricultura familiar no município, a quantidade máquinas e mão-de-obra, além dos gastos com demais despesas e, como produto, o valor total da produção (em mil reais). Como resultados, verificou-se que a tecnologia de produção utilizada na agricultura familiar é diferente entre os municípios. De modo que, a área utilizada influenciou negativamente a produção em municípios dos quantis 0,15 ao 0,55; a quantidade de maquinário apresentou efeito positivo sobre o total produzido apenas para municípios nos extremos da distribuição da produção; os gastos com demais despesas influenciaram positivamente todos os municípios, entretanto, a influência foi maior em municípios do segundo quartil e ainda; a mão-de-obra influenciou igualmente a produção de todos os municípios. Além disso, observou-se uma média de eficiência no estado foi de 41,7%. Desta forma, a produção da agricultura familiar no estado de MG pode ser aumentada em até 58,3% sem a necessidade de aumentar a quantidade de insumos utilizados.

* gabriela.franca@ufv.br, Universidade Federal de Viçosa

Detection of inconsistencies in geospatial data with geostatistics

Adriana Maria Rocha Trancoso Santos; Gerson Rodrigues dos Santos^{*}; Paulo César Emiliano; Nilcilene das Graças Medeiros; Amy L. Kaleita; Lígia de Oliveira Serrano Pruski.

Almost every researcher has come through observations that “drift” from the rest of the sample, suggesting some inconsistency. The aim of this paper is to propose a new inconsistent data detection method for continuous geospatial data based in Geostatistics, independently from the generative cause (measuring and execution errors and inherent variability data). The choice of Geostatistics is based in its ideal characteristics, as avoiding systematic errors, for example. Comparing it to a most utilized detections method, BoxPlot, the importance and functionality of the new method was verified, since the BoxPlot did not detect any data classified as discrepant. The proposed method pointed that, in average, 1,2% of the data of possible regionalized inferior outliers and, in average, 1,4% of possible regionalized superior outliers, in relation to the set of data used in the study.

^{*} prof.gersonrodrigues@gmail.com, UFV

Regressão logística semi-paramétrica

Gisele de Oliveira Maia^{*}; Clécio da Silva Ferreira

Os modelos semi-paramétricos possuem a seguinte forma $Y=X\beta+Ng+\epsilon$ e em que o erro segue uma distribuição normal com média zero e variância igual a σ^2 , assim $Y \sim N(X\beta+Ng,\sigma^2)$. $X\beta$ é a parte paramétrica do modelo e Ng a parte não-paramétrica. O vetor β dos parâmetros paramétricos são usadas para interpretar a relação existente entre as variáveis explicativas e Y (variável resposta). O objetivo principal nos modelos paramétricos é a estimação do vetor β , no entanto no modelo não-paramétrico é a estimação da curva g .

A matriz N é denotada por matriz incidente, ela está relacionada aos componentes do vetor t ($g(t_i)$) que serão selecionados para a estimação da curva g , estes valores de t selecionados são os nós. Para a estimação da curva g usa-se os splines cúbicos, que através de cada ponto(nó) escolhido do vetor t fornece uma estimativa para a g , assim juntando cada estimativa de g em cada ponto temos a curva g estimada. Trabalhamos com três métodos para estimar a curva g , o método de Green&Silverman(1993), em que os nós são selecionados como sendo todos os valores distintos do vetor t , então todos os nós terão diferentes valores. Também vimos o método de Eilers&Marx(1996) que seleciona os nós de acordo com um certo intervalo determinado a priori, assim ele divide o vetor t em intervalos. E propusermos uma generalização do método de Eilers&Marx(1996), dando a opção do usuário entrar com a quantidade de nós e seus valores.

O intuito principal do trabalho foi estudar a Regressão Logística Semi-Paramétrica (com função de ligação logit) que é um Modelo Linear Generalizado(MLG) possuindo as seguintes características, $E(Y)=\text{logit}^{-1}(\mu/(1-\mu))=X\beta+Ng$, em que Y é a proporção de sucessos em n tentativas independentes, cada um com probabilidade de ocorrência μ e $n*Y \sim \text{Binomial}(n, \mu)$

Na estimação do vetor β e da curva g utilizamos o método iterativo de Newton-Raphson combinado com os fundamentos de estimação de Green&Silverman(1993) para a parte não-paramétrica. Obtemos excelentes estimações tanto para o vetor β como para a curva g .

As próximas metas são avaliar essas estimações e prosseguir estudando os outros MLGs para o caso semi-paramétrico. Agradecimento a FAPEMIG.

^{*} giseleemaia07@gmail.com, Universidade Federal de Juiz de Fora

Teste de independência usando amostras pequenas

Gizele Zeferino de Almeida*

Objetivo: O objetivo do trabalho é verificar desempenho de três testes estatísticos não paramétricos que testam a associação entre duas variáveis aleatórias quando se trata de amostras pequenas. Os testes são: Teste de Blum, Kiefer e Rosenblatt (1961), Teste de Hoeffding (1948) e o Teste de Spearman (1904).

Metodologia: Um código de programação foi desenvolvido e implementado no software R para testar o desempenho dos testes e o seu respectivo poder. Foram geradas amostras aleatórias independentes e aplicadas a cada teste. Esperava-se, ao nível de 5% de significância, que na maioria das vezes os testes não rejeitassem a hipótese de independência. Ao gerar amostras dependentes esperava-se que na maioria das vezes a hipótese de independência fosse rejeitada.

Resultados: Os testes de Spearman e o teste de Blum tiveram ótimos desempenhos e responderam de forma similar aos vários cenários de simulações. O poder dos testes já era alto a partir do tamanho amostral igual a 6. O teste de Hoeffding não se mostrou tão poderoso para amostras iguais a 6 e 7, mas à medida que aumenta o tamanho da amostra ele já se torna altamente poderoso.

Conclusão: Conclui-se que o teste de Spearman e o teste de Blum são testes eficientes e poderosos quando se trata de testar a independência de amostras de tamanho bem reduzido. O teste de Hoeffding também apresentou um bom desempenho, mas somente a partir do tamanho de amostra igual a 11.

* gizelegt1@hotmail.com, Universidade Federal de Minas Gerais

Testes de hipótese para monitoramento do vetor de médias de processos multivariados sob amostragem sequencial: extensão para o caso de dados auto correlacionados

Graziele A. Diniz^{*}; Sueli Aparecida Mingoti

O objetivo é apresentar os resultados dos testes de hipótese para o vetor de médias para o caso em que há violação da independência das observações, sob uma amostragem sequencial, que em algumas situações se torna mais vantajosa que a amostragem aleatória simples. Na prática é comum encontramos dados autocorrelacionados, estes por sua vez mudam a estrutura da matriz de covariâncias. Assim, para um melhor funcionamento dos testes estatísticos e a qualidade do monitoramento do processo, em controle de qualidade, é necessário considerar a autocorrelação dos vetores que contém as observações amostrais. Para isso, um modelo de séries temporais se torna apropriado para modelar o comportamento dos dados. O teste estatístico avaliado foi o da Razão de Verossimilhança sob uma amostragem sequencial dos dados. A avaliação do impacto da dependência dos dados no Erro Tipo I e no Poder do Teste foi realizada através da simulação Monte Carlo. Resultados numéricos dos vários cenários simulados são importantes para mostrar o impacto da estrutura de correlação dos dados, tendo-se em cada cenário estudado as estimativas do Poder do Teste e do Erro Tipo I bem como o tamanho de amostra necessário para tomar-se a decisão em favor ou contra a hipótese nula, além de outras informações.

^{*} grazi_alexandrina@hotmail.com, Universidade Federal de Minas Gerais (UFMG)

Gama exponenciada Poisson truncada no zero

Guilherme Aparecido Santos Aguiar^{*}; Fernando Moala

There will be presented a new probability distribution, the Zero Truncated Poisson Exponential Gamma Distribution (ZTPEG). The distribution Exponentiated Gamma that was introduced by (Gupta, 1998). This distribution has the flexibility to model monotonous and not monotone failure rates. Thus, using the composition method and the proposed distribution by Gupta, the model has been developed and studied. Often, obtaining a new distribution may be advantageous due to the different shapes of its hazard curve, and this new proposed model has a very flexible hazard rate, and just two parameters. With the help of the software Maple and R, it was possible to obtain some probabilistic results of the new distribution such as n - th time, stochastic ordering and stress-strength parameter. For the new distributions found, were also made simulations for different parameter values in order to find the best estimation method for each parameter. The methods used were: maximum likelihood, ordinary least-squares, weighted least-squares, Cramér-von-Mises, Anderson Darling, Anderson Darling - RT (right-tail), Anderson Darling - LT (left-tail), Anderson Darling - 2LT (left-tail second order), Kolmogorov and bayesian estimator with the prior Gamma. Some techniques to compare the estimators were used. Finally, an applications was also performed, where the adjustment of some proposed distributions in relation to the database was tested.

^{*} guiaguiar@hotmail.com, Universidade Federal de Minas Gerais

O modelo linear de degradação sob uma perspectiva bayesiana: aplicação em emissores de laser

Guilherme Augusto Veloso^{*}; Rosangela Helena Loschi

Grande parte da literatura em Confiabilidade lida com dados em que a medida observada é o tempo de falha (ou tempo de vida) das unidades em teste e o objetivo principal é caracterizar o quão confiáveis são os produtos. Entretanto, para produtos que possuem um alto grau de confiabilidade, é comum estar diante de uma situação na qual, ao final do ensaio, poucas ou até mesmo nenhuma falha é observada, resultando em um alto índice de censuras. Como uma alternativa aos dados de tempo até a falha, os ensaios de degradação tem cada vez mais conquistado o seu espaço dentro da área de Confiabilidade. A sua metodologia capaz de estudar os mais diversos perfis de degradação aliada aos métodos computacionais de simulação tem sido muito eficaz no objetivo de encontrar quantidades de interesse que são padrões em Confiabilidade. No presente trabalho, o modelo linear de degradação com efeitos aleatórios Weibull foi utilizado e o processo de estimação consistiu-se da implementação do algoritmo Metropolis Hastings e, com isso, se tornou possível controlar a convergência das cadeias bem como a autocorrelação das amostras $\{a_{posteriori}\}$ buscando a representatividade dessas amostras e a parcimônia do modelo, algo que é dificultoso em uma análise automatizada. Como uma aplicação prática, os dados de emissores de laser, discutido na literatura da área de Confiabilidade, foram reanalisados.

^{*} guilhermeaugustoveloso@yahoo.com.br, Universidade Federal de Minas Gerais

Modelo de regressão logística em dados de reclamações do consumidor

Gustavo Henrique O. Pereira* ; Maria Imaculada S. Silva; Nádia Giaretta Biase

A satisfação do consumidor é uma das condições mais buscadas pelas empresas em geral, em tempos de grande concorrência e períodos de recorrentes crises econômicas no mercado brasileiro. Neste sentido, algumas empresas no âmbito nacional desenvolveram em parceria, uma plataforma aberta (consumidor.gov.br) que, com base na avaliação dos consumidores, retorna índices de desempenho das empresas em relação às reclamações registradas na plataforma, como forma de resgatar a confiança do consumidor. A Regressão logística é um método estatístico que gera um modelo bastante flexível e de fácil interpretação para variáveis com respostas dicotômicas, e que pode ser usado para associar a resposta positiva do consumidor ou sua probabilidade com algumas variáveis explicativas, permitindo assim um planejamento adequado das empresas para atingir o objetivo de atrair e manter o consumidor. O objetivo deste trabalho foi utilizar os dados disponíveis na plataforma mencionada e o modelo de regressão logística para verificar quais variáveis podem contribuir para obter uma avaliação positiva do consumidor a uma reclamação. Das 14 variáveis disponíveis no site, 6 foram selecionadas e cada uma em duas categorias, para o ajuste do modelo, sendo que destas, apenas 3 apresentaram significância. Para a variável segmento de mercado, a estimativa da razão de chances (OR do inglês Odds Ratio) foi de 2,69, indicando que o segmento Operadoras de Telecomunicações tem cerca de 2,7 vezes mais chance de sucesso, ou seja, de obter uma avaliação positiva do consumidor, quando comparado com o segmento Bancos e Operadoras de cartão de crédito. Para o tipo de problema, a contestação ou cobrança tem 2,6 vezes mais chance de sucesso do que o problema vício de qualidade (OR = 2,6). Para a variável tempo de resposta, a estimativa da OR foi de 0,89, indicando que a chance de sucesso é aproximadamente 11% menor quando se aumenta uma unidade no tempo que a empresa demora para apresentar a resposta.

* gustavo.pererao13@hotmail.com, Universidade Federal de Uberlândia

Análise espaço – tempo e intensidade de proliferação da doença ferrugem asiática da soja nos municípios de Chapada dos Guimarães e Tangará da Serra no estado do Mato Grosso

Bruna Pontara Vilas Boas Ribeiro; Henrique José de Paula Alves^{*}; Marcelo Carvalho Alves;
Elisandra Batista Zambenedetti

A soja constitui uma das culturas de maior relevância para a economia brasileira e, possivelmente, a que apresenta crescimento mais expressivo no cultivo e no segmento agroindustrial nacional. De acordo com a Conab (2017) a estimativa da produção de grãos na safra 2016/2017 é de 745,6 milhões de toneladas e, destas, 535,3 milhões de toneladas referentes à cultura da soja. A ferrugem Asiática da Soja (FAS), causada pelo fungo (*Phakopsora pachyrhizi* H. Sydow & Sydow), é uma das doenças mais severas que incide na cultura da soja, com danos de perda da planta variando de 10% a 90% nas diversas regiões geográficas onde foi relatada. As primeiras epidemias severas da “FAZ” na América do Sul foram relatadas no Paraguai, na safra 2000/01, e no Brasil, na safra 2001/02, nas regiões sul do estado de Goiás, no Mato Grosso, norte do Mato Grosso do Sul e no Rio Grande do Sul. As condições climáticas exercem fundamental importância nas epidemias de FAS. O melhoramento foliar contínuo, promovido por orvalho ou pela chuva, sob condições ótimas de temperatura (18°C a 26,5°C) favorece o rápido desenvolvimento da doença, sendo o fator precipitação considerado o mais importante no progresso da doença nas condições de campo. Considerando a importância da cultura da soja para o agronegócio brasileiro e mundial, utilizou-se a krigagem, considerando a modelagem via funções de covariâncias estacionárias não-separáveis (Gneiting, 2002) para avaliar a evolução espaço-temporal da “FAS” e obter informações direcionadas quanto à severidade, avanço espaço-temporal dessa doença, para caracterização da área foliar sadia e a quantificação de danos.

^{*} jpahenrique@gmail.com, Universidade federal de Lavras

Construção de algoritmos para estimação em modelos GLARMA

Isabela de Lima Severino*

O trabalho tem como objetivo construir algoritmos para estimação em modelos GLARMA, focando em séries temporais de contagem, definindo sua função de verossimilhança para modelos onde os dados têm distribuição de Poisson e estimando os parâmetros deste modelo, através da construção de um algoritmo em linguagem R. Sendo desenvolvido durante Iniciação Científica com a orientação da Professora Glaura da Conceição Franco.

Utilizando a metodologia segundo o artigo Observation-driven models for Poisson counts dos autores Richard A. Davis, William T. M. Dunsmuir and Sarah B. Streett, publicado na revista Biometrika, Vol. 90, No. 4 (Dec., 2003), pp. 777-790, onde os modelos GLARMA aqui considerados são uma subclasse de modelos espaciais de estados Não-Gaussiana.

Como resultado foi o desenvolvimento de um algoritmo de estimação dos parâmetros do modelo GLARMA, com um bom desempenho em séries temporais de tamanhos variados, diferentes números de variáveis preditoras e combinações de ARMA(p,q) diversos.

O algoritmo gera estimativas muito próximas as estimativas do pacote “glarma”, que é um pacote muito difundido no meio acadêmico e empresarial. No entanto, o pacote “glarma” gera alguns erros de convergência para algumas combinações de ARMA(p,q), o que não acontece com o algoritmo desenvolvido nesse trabalho.

* ils_isabelalima@yahoo.com.br, Universidade Federal de Minas Gerais

A distribuição beta logística

Isabella Carolina Araujo*

A beta logística é uma distribuição pertencente a classe de distribuições beta generalizada. Esta classe contém distribuições que são geradas a partir do logito da distribuição beta. Apresentaremos aplicações da distribuição beta logística para demonstrar sua flexibilidade e compará-la com outras distribuições.

* isbellacarolinasl@gmail.com, Universidade Federal de Minas Gerais

Avaliação da gestão ambiental municipal do estado de Minas Gerais

Isabella Guimarães Gonçalves* ; Renato Ribeiro de Lima

Um dos maiores desafios da gestão ambiental municipal, enfrentado atualmente no Estado de Minas Gerais, é iniciar a gestão em municípios onde esta é praticamente inexistente, bem como torná-la eficiente nos municípios que já a têm. Nesse sentido, salienta-se a importância da existência de uma base cartográfica digitalizada e da utilização de ferramentas como os Sistemas de Informação Geográfica (SIGs) para que a gestão ambiental possa ser mais facilmente implantada e para que seja eficiente. Dessa forma, o objetivo deste trabalho foi avaliar os municípios do Estado de Minas Gerais quanto à gestão ambiental, utilizando-se como indicadores a realização de licenciamento ambiental e a implantação do Cadastro Ambiental Rural (CAR). Além de verificar se a gestão ambiental já é realizada, com base nesses dois indicadores, avaliou-se a sua eficiência, verificando a existência de base cartográfica digitalizada e a utilização de SIG. Foram considerados os dados da base municipal de 2015, disponíveis no site do IBGE. Inicialmente foram obtidos mapas temáticos de Minas Gerais com a identificação de cada município, para a visualização da distribuição espacial dos municípios que continham ou realizaram cada uma das variáveis avaliadas. Posteriormente foi realizada a contagem dos municípios que responderam afirmativamente a cada variável analisada, obtendo-se tabelas de contingência, com posterior aplicação do teste de qui-quadrado, para avaliar a associação entre as variáveis. Observou-se que uma minoria dos municípios de Minas Gerais respondeu afirmativamente à existência de base cartográfica digitalizada, existência de SIG, realização de licenciamento ambiental e implantação do CAR, sendo que apenas cinco dos 853 municípios possuíam todas as variáveis, enquanto que 422 não possuíam qualquer uma delas. Além disso, as variáveis que apresentaram associação foram: existência de base cartográfica digitalizada e SIG; existência de base cartográfica digitalizada e realização de licenciamento ambiental; e existência de SIG e realização de CAR. Assim, conclui-se que, de fato, a gestão ambiental no Estado de Minas Gerais, no ano de 2015, era precária, sendo necessária a avaliação de bases de dados mais recentes, as quais não estão disponíveis no site do IBGE. Porém, a utilização de ferramentas como a base cartográfica digitalizada e SIG potencializa a gestão, sendo a utilização dessas ferramentas pelos municípios um passo importante na busca pela eficiência na gestão ambiental.

* isag_gbel@hotmail.com, Universidade federal de Lavras

Um modelo de séries temporais para predição do consumo de energia elétrica da indústria no Brasil

Ithalo Coelho de Sousa^{*}; Valmaria Rocha da Silva Ferraz; Moysés Nascimento;
Filipe Ribeiro Formiga Teixeira; Laís Mayara Azevedo Barroso; Gabi Nunes Silva;
Eduardo Campana Barbosa; Mauricio Silva Lacerda

O consumo de energia elétrica no Brasil vem crescendo de uma maneira geral, sendo a indústria responsável pelo consumo de 39% da energia produzida no Brasil. Algumas alternativas são utilizadas, que podem alterar a produção de energia, uma delas é a modelagem do consumo de energia, visando adequar a demanda a oferta de energia uma vez que a energia precisa ser gerada instantaneamente de acordo com a necessidade do consumo, evitando assim o desperdício. O objetivo deste trabalho foi apresentar previsões para a série de consumo de energia elétrica na indústria do Brasil para o ano de 2017. Para a modelagem da série foi utilizado a série do consumo de energia elétrica na indústria brasileira entre janeiro de 1976 e dezembro de 2016. A série foi transformada para se obter uma periodicidade bimestral uma vez que utilizando os dados com periodicidade mensal a análise residual não foi satisfatória. De acordo com as características da série utilizamos o modelo de Holt-Winters (HW) aditivo e multiplicativo, ambos com e sem atenuação de tendência. Quando considerado os modelos de Box Jenkins, foi utilizada a transformação de Cox-Box para estabilizar a variância dos dados com o passar do tempo, em seguida foi utilizado o teste aumentado de Dickey-Fuller que indicou a não estacionariedade dos dados, tornando-se necessário tomar uma diferença para tornar a série estacionária, ainda assim, a função de autocorrelação (ACF) continuou com picos nas defasagens 6, 12 e 18, sugerindo uma sazonalidade de período 6, logo tomamos uma diferença sazonal de período 6. Após o tratamento dos dados selecionamos 4 modelos através dos critérios AIC, AICc e BIC e o comparamos através do Erro Quadrático Médio (EQM) e do Erro Absoluto Médio (EAM) sendo o SARIMA(0,1,0)x(2,0,0)₆ o melhor modelo. A análise residual do modelo não indicou nenhum problema sendo o modelo possível de ser utilizado para realizar previsões. Por fim, comparamos entre os modelos de HW e de Box Jenkins através do EQM e do EAM, sendo o SARIMA(0,1,0)x(2,0,0)₆ predominantemente melhor para fazer previsões.

^{*} ithalo.coelho@gmail.com, Universidade Federal de Viçosa

Descrição de crescimento do acesso CNPAE-102 de pinhão manso utilizando modelo não linear

Jaqueline Gonçalves Fernandes* ; Edilson Marcelino Silva; Édipo Menezes da Silva;
Kelly Pereira de Lima; Joel Augusto Muniz

O Pinhão manso (*Jatropha curcas* L.) pertence à família das Euforbiáceas. A espécie é nativa do Brasil e tem alto potencial para produzir óleo para fins energéticos. Descrever suas características é fundamental para identificar os principais acessos com as melhores características, como a maior altura ou a maior taxa de crescimento. O objetivo deste trabalho foi ajustar o modelo não linear Gompertz (1825) considerando a estrutura de erros independentes e de erros autoregressivos de primeira ordem (AR1), aplicado aos dados de altura da planta de pinhão manso, acesso CNPAE-102 (tóxico, susceptível a oídio). A comparação entre a qualidade do ajuste dos modelos foi feita com base nos valores do coeficiente de determinação ajustado e do critério de informação de Akaike. O modelo Gompertz é usado em estudos de crescimento por suas características em fornecer estimativas de parâmetros com interpretação prática (MAZZINI et al., 2003). Na estimação dos parâmetros utilizou-se o método de mínimos quadrados e o algoritmo de Gauss-Newton implementados em rotina do software R versão 3.3.0. Para verificar os pressupostos de normalidade, homogeneidade e independência dos resíduos foram utilizados os testes de Shapiro Wilk, Breusch Pagan e Durbin Watson respectivamente. Os dados foram provenientes de um experimento estabelecido na Embrapa Cerrados em Planaltina, Distrito Federal (GURGEL, 2011). As plantas foram medidas, em metros, a cada quinze dias, durante um ano (agosto de 2009 a julho de 2010). O modelo Gompertz sem AR1 apresentou o melhor ajuste. Condição-se que a altura média do CNPAE-102 após a maturação é de 2,9 m, o ponto de inflexão da curva é em 0,78 m e a taxa de crescimento é 0,007 m.

* jaqgonfer@gmail.com, Universidade Federal de Lavras

Regressão via componentes independentes parciais aplicada à seleção genômica

Jaquicele Aparecida da Costa^{*}; Leísa Pires Lima; Camila Ferreira Azevedo; Moysés Nascimento; Ana Carolina Campana Nascimento; Marcos Deon Vilela de Resende; Fabyano Fonseca e Silva

A seleção genômica ampla (GWS) visa o aumento da eficiência na estimação de valores genômicos (GEBV) e a rapidez em identificar indivíduos geneticamente superiores por meio da utilização de um grande número de marcadores moleculares distribuídos no genoma. A aplicação da GWS enfrenta desafios estatísticos relacionados à alta dimensionalidade e a multicolinearidade entre os marcadores. Dessa forma, métodos baseados em redução de dimensionalidade são necessários na GWS. Dentre estes destacam-se, sob enfoque univariado, a Regressão via Componentes Principais (PCR), a Regressão via Componentes Independentes (ICR) e os Quadrados Mínimos Parciais (PLS). No entanto, os valores genômicos dos indivíduos são estimados com maior acurácia, quando provenientes de análises que consideram simultaneamente diversas características de interesse de um programa de melhoramento, ou seja, análises sob um enfoque multivariado. Diante do exposto, o presente trabalho teve por objetivo propor e avaliar a ICR sob um enfoque multivariado, denominada Regressão via Componentes Independentes Parciais (ICP). O método ICP foi comparado aos métodos de redução de dimensionalidades já aplicados a GWS, ICR, PCR e PLS, utilizando-se sete características de produtividade em arroz. Os bancos de dados contemplam 370 indivíduos e 36.901 marcadores moleculares. Os resultados mostraram que o ICP é adequado e acurado para a estimação dos GEBVs, além de apresentar maiores similaridades com a ICR e PCR.

^{*} jaquicelecosta@gmail.com, Universidade Federal de Viçosa

Modelando o tempo de concessão de patentes brasileiras usando análise de sobrevivência

Jéssica Fernandes^{*}; Ludmilla Mattos; Rafael Morais

O tempo de concessão de patentes entre empresas que acessaram o crédito subsidiado do Banco de Desenvolvimento e as que não financiaram apresentam grande variabilidade. Para estudar essas diferenças aplicamos a técnica de Análise de Sobrevivência. Esse método consiste na análise do intervalo de tempo entre as datas de depósito e concessão de patentes considerando censura, ou seja, quando não temos o tempo final de aprovação da patente.

^{*} jessicafernandesa@hotmail.com, Instituto de Pesquisa Econômica Aplicada

Perfil dos estudantes de graduação da UFMG

Desirée Thiely Alves Ramalho Costa; Felipe Amorim Sousa; Gustavo Henrique Pereira Godinho; Isabella Carolina Araújo; Jéssica Luiza Ferreira Ramos*

O trabalho analisou e comparou as características dos alunos assistidos e não assistidos pela Fundação Mendes Pimentel (FUMP) e investigou a percepção que os estudantes têm dos gastos relacionados à manutenção na Universidade Federal de Minas Gerais através da aplicação de questionários a alunos de graduação no campus Pampulha. Através de análise descritiva e análise de correspondência percebeu-se que os critérios de classificação da FUMP discriminam bem alunos de baixa renda. Parte significativa dos alunos afirmou ter dificuldades e necessidade da FUMP para se manter na universidade, confirmando a importância da existência de assistência estudantil na universidade.

* jessica.luiza297@gmail.com, Universidade Federal de Minas Gerais

Correlação entre área segurada e valor arrecado da produção de café no estado de MG: uma análise espacial

Jorge Garcia Filho^{*}; Marcelo Silva de Oliveira

O estado de Minas Gerais é o maior produtor de café do Brasil, responsável por 50% da produção nacional, e sendo este produto um dos responsáveis para o desenvolvimento do produto interno bruto (PIB) do estado. Apesar de sua importância, o setor cafeeiro sofre com perdas devido aos riscos relacionados à atividade agrícola. Assim como qualquer setor de atividade econômica, a agricultura é uma atividade caracterizada pelo risco. E atualmente, a transferência de riscos se dá principalmente por meios de seguros. Com isso, este trabalho tem como objetivo analisar a correlação e a correlação espacial entre valor segurado da produção de café e valor arrecado com a produção, para os anos de 2014 e 2015. Sendo assim, verificar se a área segurada possui alguma relação com a arrecadação da produção do café, no estado de MG, e averiguar se tais variáveis se correlacionam espacialmente. A análise da correlação será realizada através do coeficiente da correlação de Pearson, e para a análise da correlação espacial entre as variáveis será utilizado o Índice Global de Moran.

^{*} garcia.jorge.j@gmail.com, Universidade Federal de Lavras

Avaliação dos conhecimentos em doença falciforme no curso de educação à distância em Minas Gerais: aplicação do modelo de regressão de Poisson univariado

Leonardo Gonçalves Pinheiro Júnior* ; Katy Karoline Santos Diniz; Heloísa de Carvalho Torres; Ilka Afonso Reis; Ana Paula Pinheiro Chagas Fernandes; Adriana Silvina Pagano⁴

Introdução: A doença falciforme é considerada um problema de saúde pública e apresenta manifestações clínicas nos primeiros anos de vida com repercussão na morbimortalidade. Nesse sentido, é importante que o profissional da área da saúde tenha o conhecimento da doença para o acompanhamento, diagnóstico, tratamento e prevenção de complicações causadas pela doença falciforme. O curso de educação à distância Doença Falciforme: Linha de Cuidados na Atenção Primária à Saúde vem sendo realizado pelo Centro de Educação e Apoio para Hemoglobinopatias (CEHMOB), desde 2010, com a finalidade de fortalecer a capacidade técnica e política em doença falciforme do profissional da área da Saúde. **Objetivo:** estudar os fatores associados ao nível de conhecimento dos grupos de profissionais da área da saúde que estavam inscritos no curso de educação à distância sobre a doença falciforme. **Métodos:** estudo transversal realizado em 2016. Cento e cinquenta e três (153) profissionais da área da Saúde aceitaram participar do estudo, sendo que 72 (47%) concluíram o curso de educação à distância (grupo de conduintes) e 81 (53%) não concluíram o curso (grupo de não conduintes). Os dados foram coletados por meio de um questionário contendo questões sobre variáveis sociodemográficas e o instrumento desenvolvido pela equipe de pesquisas para avaliação do conhecimento em doença falciforme, que continha 13 questões de múltipla escolha. Os testes Exato de Fisher e Qui-Quadrado foram utilizados para analisar a homogeneidade dos dois grupos em relação às variáveis sócio-demográficas. A análise univariada via regressão de Poisson possibilitou a seleção dos potenciais fatores de influência do escore de conhecimento. Na regressão de Poisson, utilizou-se um método de estimação robusto para que o modelo pudesse absorver o comportamento da estrutura de dispersão da variável resposta. Toda a análise foi realizada no programa estatístico R. **Resultados:** Os escores variaram entre um mínimo 0 e um máximo de 13 pontos. Os conduintes apresentaram escores médios e medianos maiores do que os não conduintes. O escore médio de acerto foi 9,76 (DP=2,62) para os profissionais conduintes e 6,54 (DP=3,22) para os não conduintes. Observou-se diferença significativa (valor-p = 0,000) entre os grupos conduinte e não conduinte quanto ao escore de conhecimento, sendo que, em média, os conduintes apresentaram escores 49,2% maiores. Houve diferença significativa nos escores médios de conhecimento quando comparados profissionais de níveis de formação diferentes e com diferentes tempos de trabalho na Rede de Atenção à Saúde. Não foi observada diferença significativa nos escores médios do conhecimento quando comparadas as categorias das variáveis gênero, categoria de profissional, tipo de serviço de saúde e se atendeu/e ou conheceu pessoas com a doença na área de abrangência do serviço de saúde. **Conclusão:** o curso tem contribuído para o aumento de conhecimento dos profissionais.

* leogpjr@gmail.com, Universidade Federal de Minas Gerais

Análise temporal do índice de expectativas das famílias via metodologia de Box and Jenkins

Luiz Otávio de Oliveira Pala* ; Gislene Araujo Pereira

O presente trabalho explora o comportamento temporal do índice de expectativas (IEC). Este, é um indicador que analisa a relação de confiança das famílias sobre as condições presentes e suas perspectivas futuras, dada a atual situação econômica do país (IPEA, 2017). É um indicador estimado mensalmente pela Federação do Comércio do Estado de São Paulo, e está contido no intervalo entre zero e duzentos pontos, onde quanto mais próximo de duzentos, melhor a relação de confiança e de expectativa das famílias. Desta forma, tem-se por objetivo analisar a evolução temporal do índice de expectativas, por meio da metodologia de Box and Jenkins, durante o período de janeiro de 1995 a maio de 2017, verificando o comportamento esperado deste indicador, dado o processo de retração econômica presenciado no Brasil nos últimos meses. Para a estimação do modelo proposto, verificou-se que a série temporal apresentou as componentes de tendência e sazonalidade, sendo necessário a realização das primeiras diferenças para obtenção de uma série estacionária, fundamentado pelo Teste de Dickey-Fuller, a um nível de significância de 5%. Por fim, a partir da análise das autocorrelações da série temporal, identificou-se a presença de um modelo ARIMA sazonal para tal objetivo, de forma SARIMA(0, 1, 0)(0, 1, 1), estimado utilizando o filtro de Kalman gerando um processo de ruído branco não Gaussiano. O modelo foi estimado pelo Software R, que possibilitou toda a análise e discussão dos resultados propostos. Assim, nota-se a importância da realização de estudos econométricos aplicados a indicadores econômicos, que possibilitem perspectivas futuras de aceleração econômica e das condições de expectativas dos agentes e dos indivíduos. Portanto, este trabalho contribui de forma a expandir a aplicação de estudos econométricos associados a indicadores de expectativas econômicas do país, visando projeções futuras do cenário macroeconômico do país.

* luizotavio.oliveira@gmail.com, Universidade Federal de Alfenas

Erros do modelo de regressão linear simples: e se a suposição de normalidade dos erros não forem válidas?

Maiara Fernanda Oliveira de Miranda^{*}; Ilka Afonso Reis

Esse estudo sobre Regressão Linear Simples trata da violação da suposição de Normalidade dos Erros e como isso afeta a estimação dos parâmetros. Para isso, simulamos populações de um milhão de pares (X, Y) , com X seguindo distribuição Uniforme entre 10 e 20, β_0 sendo igual a 5, β_1 igual a 0,5, $Y = \beta_0 + \beta_1 * X + \epsilon$. Quanto à distribuição dos erros ϵ , utilizamos a distribuição Normal padrão, a distribuição t-student com 10, 20 e 30 graus de liberdade e a distribuição LogNormal de três parâmetros com 25%, 50%, 75% e 99% dos valores abaixo de zero (parâmetros de forma e de escala iguais a 0 e 1, respectivamente). Retiramos, de cada uma dessas populações, 1000 amostras de tamanho $n=10, 50, 100, 250, 500$ e 1000 elementos. O modelo de regressão linear simples foi ajustado aos dados de cada uma das 1000 amostras e as estimativas dos parâmetros foram resumidas por meio de médias e desvio-padrões e visualizadas graficamente por meio de histogramas. Percebemos que a violação da suposição de Normalidade dos Erros atrapalha visivelmente a estimação de β_0 e, em certo momento, a variância do erro, mas não afeta as estimativas de β_1 . Para as distribuições t-student, o β_0 ficou bem estimado, com valores muito próximos ou igual ao real mesmo com amostras pequenas, enquanto, nas distribuições LogNormal, ele teve valores muito discrepantes, chegando até a ser negativo para LogNormal que deixa 99% dos valores abaixo de zero. A variância do erro fica um pouco subestimada apenas para distribuições Log-Normal com 50% dos valores abaixo de zero. Ou seja, quando a suposição da normalidade dos erros não é válida, haverá problemas principalmente na estimação do β_0 . Quanto ao teste F da tabela ANOVA, não foram encontrados problemas.

^{*} maiaraoliveira.m@gmail.com, Universidade Federal de Minas Gerais

Avaliação do efeito alelopático de Xanatse (*Colophospermum Mopane*) na germinação e crescimento inicial de abóbora (*Cucurbita Máxima*)

Mario Sebastião Tuzine* ; Rafael Estefane Ubisse; Eduardo Langane ;Natalino Calegario

Este trabalho teve como objetivo, determinar o potencial alelopático das folhas de Xanatse (*Colophospermum mopane*) na germinação e crescimento inicial das sementes e plântulas de abóboreiras (*Cucurbita máxima*), ensaio conduzido no Laboratório Regional Sul de Sementes de Lionde Distrito de Chokwe, província de gaza Moçambique.

Para a condução deste experimento, foram utilizadas sementes de abóbora adquiridas localmente com 95% de percentagem de germinação. Foram utilizadas folhas de plantas de Xanatse, coletadas no Distrito de Mabalane, província de gaza Moçambique. O experimento foi conduzido no substrato 750 gramas de areia, esterilizada a uma temperatura de 175°C durante 12 horas. Sendo conduzido com um foto período de 24 horas de luz fluorescente por um período de sete dias, a germinação foi observada de 12 em 12 horas. Foi usado o delineamento inteiramente casualizado com cinco tratamentos que consistiam em concentrações de 0% (água), 2,5%, 5,0%, 7,5% e 10% da diluição do extrato puro (100g de xanatse para 900 ml água) e cinco repetições de 250 sementes cada. As variáveis analisadas foram a Porcentagem de germinação, o Índice da velocidade de germinação (IVG). Outro bioensaio de crescimento foi conduzido com as mesmas proporções para avaliar a biomassa das componentes hipocótilo e raízes, Altura de plântulas, Comprimento da radícula. As médias de tratamentos foram comparadas pelo teste de Duncan a 5% de probabilidade no pacote estatístico The R Project for Statistical Computing.

Os extratos com xanatse apresentaram menores de germinação das sementes, da altura de plântulas e do comprimento da radícula. A semente de abóbora colocada em água destilada, apresentou uma germinação de 96%, enquanto que com a diluição de 10% do extratos puro de mopane obteve uma germinação de apenas 20%, a produção de matéria seca nos hipocótilo assim como na radícula foi estatisticamente inferior sendo 0,35g/planta no tratamento de 0% e variou 0,22 a 0,12g/planta nos demais tratamentos .

Concluindo que os extratos aquosos de folhas reduzem a porcentagem de germinação das sementes, crescimento altura e produção de biomassa das plântulas de abóbora indiciando efeito alelopático

* tuzine@yahoo.com.br, Instituto superior Politenico de Gaza -UFLA

Mineração de dados no esporte: um estudo sobre o campeonato brasileiro de futebol Série A 2016

Pedro Pires Costa^{*}

Trabalho desenvolvido através de Iniciação Científica, explorando a estatística no esporte.

^{*} pedro.pires97@hotmail.com, Universidade Federal de Minas Gerais

Modelo de previsão da série de índice de atividade econômica do Banco Central (IBC-BR)

Denise de Assis Paiva; Rejane Corrêa da Rocha^{*}; Luciane Teixeira Passos Giarola

Este trabalho teve como principais objetivos estudar o comportamento da série dessazonalizada do Índice de Atividade Econômica do Banco Central (IBC-Br) e fazer previsões, utilizando a metodologia de Box & Jenkins. O IBC-Br é um indicador utilizado com o intuito de antecipar o resultado do Produto Interno do Bruto (PIB) do país, constituindo-se de um parâmetro de avaliação do crescimento da economia brasileira ao longo dos meses. A série foi obtida no site do Banco Central do Brasil, Sistema Gerenciador de Séries Temporais (Bacen Outras/SGS) e é composta de 72 observações, correspondendo ao período de janeiro de 2003 a abril de 2017. Foram ajustados os modelos, ARIMA (p,q,d), utilizando o software Gretl. Para a escolha do melhor modelo foi utilizado o Critério de Informação de Akaike (AIC) e o erro percentual médio absoluto (MAPE). Desta forma, dentre os modelos ajustados, o que melhor se ajustou aos dados foi o ARIMA (2,1,0), sendo o AIC=481,2110 e MAPE=0,61%. Concluiu-se, que a metodologia de Box & Jenkins é útil para descrever e fazer previsões da série de Índice de Atividade Econômica do Banco Central (IBC-Br).

^{*} rejcrocha@gmail.com, UFSJ

Produção de cana-de-açúcar nas regiões Centro-Oeste e Nordeste do Brasil descrita por modelos mistos

Marina Rodrigues Maestre; Simone Daniela Sartorio de Medeiros^{*}; Sérgio Ricardo
Rodrigues de Medeiros

É de interesse estudar a produção da cana-de-açúcar, uma vez que essa é matéria-prima de uma variedade grande de produtos que são muito úteis nos dias de hoje, como por exemplo, o etanol. Avaliada ao longo dos anos, a partir de 2006 até 2010, nos estados das regiões Centro-Oeste (CO) e Nordeste (NE), essa variável é analisada utilizando modelos lineares mistos, sendo o tempo um fator longitudinal. Tal abordagem tem o atrativo de explicar de forma conveniente a variabilidade existente entre as regiões e as correlações dentro dessas regiões. O presente trabalho tem por objetivo utilizar a classe de modelos lineares mistos para ajustar um modelo a um conjunto de dados de produção de cana-de-açúcar, em toneladas, ao longo do tempo nos estados das regiões CO e NE. Para alcançar o melhor modelo, foram utilizados testes como o da razão de verossimilhanças, Wald, os critérios de informação AIC e BIC e gráficos de diagnóstico dos resíduos condicionais padronizados. Para a parte fixa do modelo resultante foram ajustadas duas retas, sendo uma para cada região. Estas retas diferiram quanto aos interceptos e quanto aos coeficientes angulares, ou seja, retas concorrentes. Foi significativa a inclusão de efeito aleatório somente no intercepto da região NE e nos coeficientes angulares das duas regiões. Por meio dos testes utilizados, houve heterogeneidade de variâncias para os níveis do fator Estado. Assim, pode-se concluir que em ambas as regiões, a produção de cana-de-açúcar tem tido um aumento significativo com o passar dos anos.

^{*} sisartorio@cca.ufscar.br, CCA/UFSCar

Levantamento da visão política-social dos alunos dos cursos de humanas da UFMG

Thaís Pacheco Menezes^{*}; Luiza Sette Camara; Leonardo Gonçalves; Laise Rodrigues; Luísa Barbosa

O objetivo do presente trabalho foi avaliar o perfil político e socioeconômico dos alunos regularmente matriculados nos cursos diurnos e vespertinos da área de Ciências Humanas da UFMG. Para alcançar tal objetivo, aplicou-se um questionário a uma amostra estratificada por curso e por gênero de 296 pessoas. De posse dos dados, foram usadas frequências absolutas e relativas para descrever as variáveis qualitativas e de medidas de tendências centrais para as quantitativas. Para verificar a associação entre as variáveis do estudo, foram usados os testes Qui-quadrado e Exato de Fisher. Nos casos de associação positiva, realizou-se a análise de correspondência. Ao final, conduziu-se que a maioria das respostas relacionadas à política foram centradas em pensamentos de esquerda sendo que a exigência de nível superior para candidatos a cargos políticos foi a que mais dividiu opiniões. No que tange à segurança, o percentual de alunos que sofreram algum delito no campus Pampulha ou em suas imediações foi pequeno apesar da classificação de segurança ter sido 2.98 em uma escala de 1 a 5. Por fim, percebeu-se que com o conturbado cenário político e econômico, a expectativa para o mercado de trabalho após a formatura para a grande parte dos alunos é baixa.

^{*} thaispacheco1207@gmail.com, Universidade Federal de Minas Gerais

Análise de satisfação dos alunos do Centro de Ciências Agrárias da Universidade Federal do Espírito Santo

Pimentel, U. * ; Gorini, E.

A realização de pesquisas de satisfação no âmbito acadêmico é de fundamental importância para avaliar as distintas dimensões que envolvem o ensino e a aprendizagem. O objetivo do presente estudo é analisar a percepção dos alunos em relação às condições físicas e pedagógicas oferecidas no Centro de Ciências Agrárias (CCA) da Universidade Federal do Espírito Santo (UFES). A pesquisa foi realizada com uma amostra de 394 alunos. Para a coleta dos dados utilizou-se um questionário dividido em 55 itens, esses foram subdivididos em quatro tópicos, sendo alguns de acordo com a escala Likert. Os resultados obtidos com essa análise indicaram que os acadêmicos estão satisfeitos com a qualidade do ensino no CCA, no entanto estão insatisfeitos com a estrutura física da faculdade e o sistema de transporte oferecido pela mesma.

* uillapimenteleng@gmail.com, UFES

Representação da estrutura do pensamento coletivo sobre as enchentes do Rio Doce: conectando indivíduos afins através da teoria dos grafos

Wesley Pereira^{*}; Denise Duarte; Gilvan Guedes; Rodrigo Ribeiro; Marina Amorim

Baseado no trabalho de Monografia apresentado em 2016/2 pelo primeiro autor, o trabalho busca apresentar a técnica utilizada e os resultados obtidos ao se utilizar da Teoria dos Grafos para decompor a estrutura do pensamento coletivo da população de Governador Valadares sobre as Enchentes do Rio Doce coletado sob Técnica de Associação Livre de Palavras, fundamentada na Teoria das Representações Sociais.

^{*} wesleyhpereira@gmail.com, Universidade Federal de Minas Gerais

Poder exato do teste sequencial condicional da razão de verossimilhanças

Wilson Manoel Lopes* ; Ivair Ramos Silva;

O objetivo da análise sequencial é identificar, o mais cedo possível, se hipóteses sobre o processo original serão confirmadas pelos dados que chegam sequencialmente ao longo do tempo. O cMaxSPRT (conditional maximized sequential probability ratio test) é derivado do teste MaxSPRT(maximized sequential probability ratio test) de Kuldorf (2011), que por sua vez é uma extensão do SPRT (sequential probability ratio test) de Wald (1945). O CMaxSPRT é mais adequado que os outros testes quando não é possível obter uma estimativa confiável da média histórica, e tem sido muito utilizado no monitoramento de vacinas pós-comercializadas nos EUA. Foi obtida uma expressão analítica para o poder do teste cMaxSPRT não sequencial, que pode ser utilizada como cota superior para o poder do teste cMaxSPRT sequencial. Através desta expressão foram geradas tabelas para as principais configurações do teste. A expressão analítica e as tabelas serão implementadas no pacote “sequential” do R, permitindo informar ao usuário quando estiver tentando configurar um teste com um dado poder alvo maior que a cota, evitando assim que o usuário espere horas por um resultado impossível de ser alcançado. Com as tabelas implementadas no pacote, os valores poderão ser acessados sem a necessidade de executar os cálculos, reduzindo tempo de processamento de funções de configuração do teste CMaxSPRT.

* wilsolopes@gmail.com, Universidade Federal de Ouro Preto

Sessão Pôster 2

Uma análise utilizando a metodologia Box & Jenkins para o índice de atividade econômica brasileira

Aline Caroline Rodrigues^{*}; Luiz Otávio de Oliveira Pala; Letícia Lima Milani

Resumo: Destaca-se a importância do índice de atividade econômica brasileira, o qual de acordo com o Banco Central do Brasil, atua como proxy para o desempenho dos três principais setores da economia: agropecuária, indústria e setor de serviços, além dos impostos sobre os produtos. Mediante tal importância, este trabalho analisou por meio da metodologia Box&Jenkins, a série temporal do índice de atividade econômica brasileira, sendo a abrangência mensal de janeiro de 2006 a dezembro de 2016, os dados são provenientes do Ipeadata e as análises se deram por meio dos softwares R e Gretl. Primeiramente, realizou-se uma análise visual da série e, em seguida, identificou-se a existência das componentes de tendência e sazonalidade, apontadas pelos testes de Cox-Stuart e Fisher, a um nível de 5% de significância, respectivamente. Os testes apontaram a presença das componentes de tendência e sazonalidade, sendo necessário a aplicação do operador de diferenças da série original, que se manteve sazonal mesmo após tal procedimento. Desta forma, foi necessário a utilização de um modelo que se ajuste ao comportamento sazonal, expresso na forma de SARIMA(p, d, q)(P, D, Q), com parâmetros avaliados pelas funções de autocorrelação (fac) e autocorrelação parcial (facp), que indicaram possíveis modelos para o estudo da série. Posteriormente, optou-se pela escolha do modelo SARIMA(2, 1, 2)(0, 1, 1), justificado pelo critério de seleção de modelos (AIC). Posteriormente, a análise dos resíduos apontou um processo de ruído branco, verificado pelo teste Ljung-Box. Por fim, foram realizadas previsões da série, indicando que os valores preditos pelo modelo apresentaram tendência em subestimar tal indicador.

^{*} alineacr25@gmail.com, Universidade Federal de Alfenas

Preço dos imóveis sob a perspectiva da regressão linear – Windsor, Canadá 1987

Allan Martins Coelho de Oliveira* ; Joana Guedes

Utilizar a teoria do modelo de regressão múltipla para entender o comportamento do preço das moradias na cidade de Windsor no Canadá em 1987, assim como conhecer a influência das variáveis físicas, estruturais e espaciais que o imóvel possua.

* allan.m.coelho@gmail.com, Escola Nacional de Ciências Estatísticas

Determinantes do endividamento de curto prazo das resseguradoras atuantes no Brasil

Ana Carolina de Oliveira* ; Leonardo Henrique de Andrade Bento do Nascimento;
Gabriel Rodrigo Gomes Pessanha

Este trabalho analisa a estrutura de capital das Resseguradoras atuantes no Brasil, mais especificamente os determinantes do Endividamento de Curto Prazo (ECP). Os dados utilizados foram extraídos dos demonstrativos financeiros de resseguradoras e disponibilizados pela Superintendência de Seguros Privados (SUSEP). Analisou-se, ao todo, 11 resseguradoras que atuaram no Brasil no período de 2013 a 2016. A abordagem é de natureza quantitativa com a aplicação da técnica estatística de Regressão Linear Múltipla. Os valores das variáveis contábeis foram ajustados aos efeitos inflacionários considerando-se do indexador IGP-DI. Os dados utilizados no decorrer deste estudo foram organizados em forma de painel balanceado de séries temporais empilhadas e analisados através do software GRETL. Os pressupostos da regressão linear múltipla foram testados e respeitados. Com relação ao modelo, foi constatado um baixo número de variáveis independentes significativas; entretanto as variáveis presentes no modelo final explicam, em média, 64,94% das variações observadas no ECP das Resseguradoras. Tal estatística foi confirmada pelo teste F que se mostrou significativo ao nível de 1%. Os resultados revelam que as variáveis que se mostraram significativas, ao nível de 5%, para o modelo ajustado, foram o índice Liquidez Corrente e a Lucratividade. Essas variáveis apresentaram um relacionamento inverso com a variável dependente. Além disso, esse comportamento indica que um aumento nessas variáveis gera, em média, uma diminuição na variável explicada. Ressalta-se que os fatores analisados neste trabalho explicam apenas parte das variações no ECP das empresas, o que sugere que outras variáveis também influenciam a forma como as resseguradoras se financiam no Brasil. Entre essas variáveis, podem existir fatores comportamentais específicos dos gestores e macroeconômicos não tratados no estudo. Considera-se que há um vasto campo a ser pesquisado no que diz respeito às Resseguradoras, pois, pela sua importância, existem poucos estudos relacionados a esse ramo no Brasil.

* ana_coliveira@outlook.com, UNIFAL-MG

Fundos de investimento remunerados pelo risco: qual escolher?

Ana Paula Rocha Rosa^{*}; Tatiane de Fátima Claudiano; Marçal Serafim Cândido

Fundos de Investimento tem, em sua essência, o objetivo de promover a captação e aplicação coletiva de recursos no mercado financeiro; aqueles que optam por empregar suas finanças em um fundo de investimento, encontram-se em busca de alcançar maiores rentabilidades de suas respectivas aplicações, aliadas a redução do risco que estão expostos, este que é sempre inerente a qualquer operação no âmbito das finanças. Nesse contexto, este trabalho tem como finalidade realizar análises para o retorno de três fundos de investimento: Referenciado DI (BB REF DI Plus Es), Ações Ibovespa Indexados (BB Ações Ibov Index) e Fundo Multimercado (BBMM Macro LP 200). Tais análises consistiram em mensurar o desempenho dos fundos de investimento escolhidos, quais fatores afetaram esse desempenho e qual seria a estimativa de perda máxima esperada através de medidas de risco. Para a análise do desempenho, foi utilizado a Regressão Linear Múltipla, aplicando-se o método dos Mínimos Quadrados Ordinários (MQO) através do software Gnu Regression Econometrics and Time-series Library (GRETSL); a base de dados referente as cotações diárias dos fundos foram retiradas do endereço eletrônico do Banco do Brasil. Para a análise dos fatores que afetam o desempenho dos fundos, foi utilizado um indicador que permite avaliar a relação risco e retorno de um investimento, denominado Índice de Sharpe, sendo os dados retirados do endereço eletrônico Uol Economia e pelo Instituto de Pesquisa Econômica Aplicada (IPEA DATA). Para analisar a perda máxima esperada em um fundo de investimento, foi utilizado o método que avalia o risco em operações financeiras, o VAR (Value at Risk), utilizando o software Excel. Tendo como base os fundos escolhidos, que são remunerados pelo risco, o objetivo desse trabalho foi classificar os fundos de investimento de acordo com os retornos esperados observados de cada um.

^{*} anapaularocharosa@yahoo.com, Universidade Federal de Alfenas

Teste Scheffé no R para experimentos em parcela subdividida

Beatriz de Oliveira Rodrigues^{*}; Felipe de Melo Taveira; Flavio Bittencourt; Adriana Dias

Em muitas situações é preferível instalar um experimento em parcela subdividida do que em fatorial cruzado, pois são mais fáceis de instalar e, também, porque em muitos casos a parcela (unidade experimental) é representada por “uma unidade física”, como um indivíduo ou um equipamento, o que permite receber os níveis do fator secundário dentro do fator principal, isto é, um indivíduo pode ser observado tantas vezes quanto o número de níveis do fator secundário. Por meio desta estrutura, o fator secundário (que é alocado na subparcela) é testado com maior precisão do que aquele que está na parcela. Comparações múltiplas de médias são realizadas quando o teste F da interação for significativo e se não for, quando o teste F para algum fator for. O teste Scheffé poderá ser utilizado para se comparar algum contraste de interesse. Entretanto, quando a interação é significativa e se tem que estudar as médias do efeito principal dentro de cada nível do efeito secundário é necessário calcular o resíduo combinado, bem como os graus de liberdade deste, o que torna este tipo de estudo bastante trabalhoso. Todavia, para os demais estudos, os cálculos são semelhantes ao fatorial duplo. Assim, este trabalho tem como objetivo apresentar uma rotina em linguagem R para realizar o teste de Scheffé em experimentos instalados segundo delineamentos inteiramente ao acaso, em blocos completos ou em quadrados latinos em que os tratamentos apresentem estrutura em parcela subdividida.

^{*} a14006@bcc.unifal-mg.edu.br, Universidade Federal de Alfenas - UNIFAL

Variáveis associadas à probabilidade de sobrevivência até os 60 anos na população da região metropolitana de Belo Horizonte

Bruna da Costa Silva* ; Luiz Otávio de O. Pala

O presente trabalho utiliza a metodologia de análise de regressão linear múltipla para identificar as relações de variáveis demográficas e socioeconômicas que se associam a variável em estudo, ou seja, a probabilidade de sobrevivência até os 60 anos. O conjunto de observações refere-se aos municípios que compõem a região metropolitana de Belo Horizonte, cujos dados são provenientes do Atlas do Desenvolvimento Humano do Brasil, compondo 37 municípios brasileiros. Além disso, os dados disponíveis têm como fonte o Censo Demográfico de 2010. Todas as análises foram desenvolvidas por meio do Software R (R CORE TEAM, 2015). Assim, com a aplicação da metodologia proposta, foram identificadas três variáveis significativas para o modelo, sendo elas: Porcentagem de domicílios com banheiro e água encanada, Taxa de mortalidade infantil e Índice de Desenvolvimento Humano (IDMH); apresentando um coeficiente de determinação de 0,8291. O modelo proposto apresentou resíduos Gaussianos, evidenciados pelo teste de normalidade de Shapiro-Wilk e, fatores de Variance Inflation Factor (VIF) inferiores a 4, ou seja, a regressão não é afetada por problemas de multicolinearidade. Por fim, aponta-se que a variável resposta é influenciada por fatores demográficos e sociais que tendem a elevar ou reduzir a probabilidade de sobrevivência dos indivíduos da região em análise.

* brunasilva208@hotmail.com, Universidade Federal de Alfenas - Campus Varginha

Análise de concordância no descarte de pintos de corte pelo critério de refugagem

Fabrício Pelizer de Almeida* ; Renner Ribeiro de Oliveira

O termo análise de concordância diz respeito à capacidade de aferir resultados idênticos em uma determinada avaliação, aplicados ao mesmo indivíduo ou atividade, por diferentes instrumentos, avaliadores ou momento, ou ainda por uma combinação desses episódios (LITTLE e RUBIN, 2002). O presente artigo analisou a concordância entre operadores quanto ao descarte de pintos de corte pelo 'critério de refugagem' ao nascer. Os dados foram obtidos avaliando os critérios adotados pelos técnicos agropecuários quanto à eliminação de aves refugadas (-1) ou em bom estado (+1). Foram considerados na pesquisa quatro técnicos (A; B; C; D), vinte aves escolhidas aleatoriamente e três repetições consecutivas. Os dados foram tabulados e submetidos à análise estatística utilizando-se se o pacote Quality Tool e a função Attribute Agreement Analysis do software MINITAB (v. 17). Os resultados foram analisados utilizando-se o índice de kappa (k), como parâmetro da avaliação de concordância para os testes. No experimento, logo antes de se formarem os lotes de pintos para os criadouros, o grau de concordância entre os técnicos nas três avaliações foi significativo ($p < 0,005$), com índice de kappa (k) elevado ("A"; k: 0,7942; "B"; k: 0,7333; "C"; k: 0,7257; "D"; k: 0,8000), sugerindo que o 'critério de refugagem' é consistente na avaliação desses operadores. Quanto à avaliação de cada técnico, comparando-se ao padrão adotado pela empresa, apenas um deles obteve um desempenho bastante discrepante ("D"; k: 0,0667). Todos os demais técnicos, individualmente demonstraram alinhamento significativo ($p < 0,005$) com o padrão operacional ("A"; k: 0,2952; "B"; k: 0,8666; "C"; k: 0,4271). Quanto à reprodutibilidade do teste, os técnicos conjuntamente concordam entre si nas três avaliações propostas (k: 0,2727), e em relação ao padrão operacional (k: 0,4139) significativamente ($p < 0,005$). Como desdobramento operacional do trabalho, sugere-se atuar no treinamento e acompanhamento de rotina do técnico "D", reiterando o padrão operacional e ajustando a concordância do critério de avaliação com os demais técnicos da empresa.

* fabricao.almeida@uniube.br, Universidade de Uberaba

Gerenciador de métricas de publicação (GMP)

Felipe de Melo Taveira^{*}; Lucas Oliveira Nabeto; Flavio Bittencourt;
Luiz Eduardo da Silva; Eric Batista Ferreira.

Nos dias atuais, a escassez de recursos de financiamento da pesquisa vem deixando o ambiente acadêmico cada vez mais concorrido. São comuns editais de diversas naturezas que ranqueiam os pesquisadores com base em alguma métrica indicadora de sua produção e/ou produtividade. Exemplos são os processos de progressão e promoção na carreira, editais de bolsas de iniciação científica com bolsas ou voluntária, iniciação à extensão, monitoria (com bolsa ou voluntária), estágio etc. Em instituições de ensino e pesquisa privadas é possível pensar em programas de acréscimo salarial por produção ou programa de metas, que devem ser cumpridas de forma mínima em um dado intervalo de tempo.

Para todas essas e outras situações, se faz necessário o gerenciamento de métricas que indiquem a produção/produtividade de um certo grupo de pesquisadores. Uma das formas gratuitas e confiáveis que se tem para fazer nos dias de hoje são os serviços fornecidos pelo Google Acadêmico (GA). O GA é capaz de monitorar e coleccionar os artigos publicados e suas respectivas citações, para qualquer pesquisador que assim permitir.

O Gerenciador de Métricas de Publicação (GMP) é um software desenvolvido pela Universidade Federal de Alfenas (Figura 1), capaz de se conectar ao GA, popular um banco de dados sobre produção de um dado conjunto de pesquisadores e ranqueá-los segundo algumas métricas indicadoras de produtividade, como o índice h, índice g, fator de impacto ou simplesmente número de artigos publicados em dado intervalo de tempo.

^{*} felipe.melo.taveira@gmail.com, Universidade Federal de Alfenas

O uso das redes neurais artificiais para prever o padrão de resposta da resistência a ferrugem em *coffea arábica*

Gabi Nunes Silva^{*}; Isabela de Castro Sant'Anna; Vinícius Quintão Carneiro;
Cosme Damião Cruz; Moysés Nascimento; Ithalo Coelho de Sousa;
Renato Domiciano Silva Rosado; Eveline Teixeira Caixeta

O café, em particular o *Coffea arabica*, representa uma das principais culturas agrícolas brasileira. A principal doença que afeta a cafeicultura é a ferrugem alaranjada, causada por *Hemileia vastatrix*, sendo a pior em abrangência e danos. Para minimizar os prejuízos causados por essa praga, os programas de melhoramento genético têm investido na obtenção de variedades resistentes, com o sequenciamento e identificação de marcadores moleculares associados à resistência. Nesse contexto, a seleção genômica ampla (GWS) tem sido amplamente utilizada. No entanto, algumas técnicas de GWS apresentam problemas relacionados à suposição de distribuição normal dos valores dos fenótipos em estudo. Assim, sua aplicação em variáveis categóricas, como resistência a ferrugem, torna-se inadequada. Visando contornar tais limitações, o objetivo deste trabalho foi utilizar as Redes Neurais Artificiais (RNA) para prever o padrão de resposta da resistência a ferrugem em *C. arabica* em comparação à metodologia de GWS sob o enfoque bayesiano. Foram utilizados dados de 245 indivíduos de uma população F2 (H511-1) oriundos da autofecundação do híbrido F1 (H511-1), resultante do cruzamento do cultivar suscetível Catuaí Amarelo IAC 64 e o genitor resistente híbrido de Timor. A taxa de coincidência na classificação das duas metodologias foi de 81.37%. Ao simular uma estratégia de seleção de marcadores baseada na fenotipagem, as taxas de erros de classificação obtidas pela RNA variaram entre 44.90% e 53.47%, evidenciando que a ação conjunta dos marcadores é mais apropriada para prever resistência à *H. vastatrix*. Os resultados indicaram que as RNAs representam abordagem promissora para prever a resistência à ferrugem em *C. arabica*, dada a baixa taxa de erro (1.6%) de predição na população de validação.

^{*}gabi.nuessilva@yahoo.com.br, Universidade Federal de Viçosa

Perfil econômico-financeiro das instituições bancárias adquiridas em processos de fusões e aquisições (F&A)

Gabriel Rodrigo Gomes Pessanha* ; Matheus Saraiva Alcino;
Poliana Maria Benelli; Matheus Henrique Ribeiro.

Diante da escassez de evidências teóricas e empíricas acerca dos determinantes dos alvos de estratégias de F&A, o objetivo deste trabalho é identificar os determinantes das F&A nas instituições bancárias brasileiras adquiridas no período de 1995 a 2015. Ademais, a utilização de uma metodologia estatística inovadora na análise de F&A das instituições adquiridas também se configura como objetivo deste trabalho. Os dados necessários para a realização deste trabalho foram obtidos a partir de fontes secundárias, foram elas: Economática, Banco Central (2015) e KPMG (2015). O período analisado está compreendido entre o primeiro trimestre de 1994 e o segundo trimestre de 2015. O pacote estatístico escolhido para fazer as análises deste trabalho foi o software SPSS (Statistical Package for Social Science). A metodologia de análise de sobrevivência foi empregada nos dados econômico-financeiros a fim de se obter uma função do risco de ocorrência de F&A. Os resultados foram, na sua maioria, coerentes com a teoria e com as expectativas iniciais, evidenciando que uma maior fragilidade econômico-financeira potencializa o risco da instituição ser alvo de uma instituição adquirente. Constatou-se também que a análise de sobrevivência apresentou uma satisfatória aderência aos dados podendo ser considerada com uma metodologia a ser utilizada em futuras pesquisas, já que é capaz de contemplar a censura de dados e contornar um dos principais problemas relatados pelos pesquisadores no que se refere à escassez de dados após a fusão ou até mesmo aquisição, pois, em muitos casos, a instituição comprada deixa de existir passando a integrar a compradora.

* gabrielrgp@yahoo.com.br, Universidade Federal de Alfenas - UNIFAL

Avaliação do impacto da não - normalidade multivariada no teste de hipótese de razão de verossimilhanças para o vetor de médias de processos multivariados sob amostragem sequencial

Graziele A. Diniz^{*}; Sueli Aparecida Mingoti

Os testes de hipótese para o vetor de médias como o de Hotelling e o de razão de verossimilhanças, tem como suposição que os dados são provenientes de uma amostra aleatória cujos valores são gerados pela distribuição normal multivariada. Entretanto, em muitas situações práticas a suposição de normalidade é violada. Além disso, em algumas áreas, como controle de qualidade, procedimentos diferentes de coleta de dados podem ser empregados como amostragem dupla e amostragem sequencial por exemplo, métodos que em algumas situações se tornam mais vantajosos que o de amostragem aleatória simples. Sendo assim, o objetivo desse trabalho é apresentar os resultados do desempenho do teste para o vetor de médias de uma população no caso em que os dados não são provenientes da distribuição normal multivariada e os dados são coletados via amostragem sequencial. O teste estatístico avaliado foi o de razão de Verossimilhanças e a avaliação do impacto não normalidade multivariada dos dados no Erro Tipo I e no poder do teste foi realizada através da simulação Monte Carlo. Vários cenários foram simulados com distribuições multivariadas diversas sendo possível mostrar o efeito de não-normalidade. Para cada cenário tem-se as estimativas tamanho de amostra necessário para a rejeição ou não da hipótese nula, do poder do teste e do Erro Tipo I, e outras informações adicionais.

^{*} grazi_alexandrina@hotmail.com, Universidade Federal de Minas Gerais (UFMG)

Ajuste de modelos geoestatísticos espaço-temporal para avaliar a variação diária de temperatura em Lavras – MG

Henrique José de Paula Alves* ; Elias Silva de Medeiros; Marcelo de Carvalho Alves;
Daniel Furtado Ferreira; Renato Ribeiro de Lima

Diversas áreas da ciência, tais como, hidrologia, demografia, epidemiologia, entre outras, apresentam fenômenos caracterizados por variações no espaço e no tempo. Analisar a variação diária da temperatura em uma região é de grande importância para o plantio de certas cultivares. O presente trabalho apresenta um estudo geoestatístico espaço-temporal com o intuito de prever e prever a variação diária da temperatura, medida em graus Celsius da cidade de Lavras, no estado de Minas Gerais. Foi utilizado o software estatístico R, sendo utilizada a biblioteca “gstat” para efetuar as análises. A classe de modelo de covariâncias espaço-temporais produto-soma foi utilizada para ajustar modelos teóricos de semivariograma espaço-temporal, com diferentes estruturas. Os dados, que foram gentilmente cedidos pelo professor Humberto Alves Barbosa, da Universidade Federal de Alagoas (UFAL), foram obtidos através de imagens via sensoriamento remoto, sendo processados em um único dia. As medições foram feitas entre 8 horas e 17 horas, em intervalos de 15 em 15 minutos. Os resultados obtidos mostraram como a variação espaço-temporal da temperatura terrestre pode afetar o plantio de certas cultivares.

* jpahenrique@gmail.com, Universidade federal de Lavras

Métodos bayesianos aplicados na estimação do mérito genético do comprimento de carcaça em suínos

Filipe Ribeiro Formiga Teixeira; Ithalo Coelho de Sousa* ; Moysés Nascimento; Camila Azevedo Barroso; Laís Mayara Azevedo Barroso; Gabi Nunes Silva; Márcia Elaine Guimarães Lana

A seleção genômica ampla (SGA), foi proposta com o intuito de acelerar o processo de melhoramento genético.. Este procedimento utiliza informações diretas do DNA na seleção e predição do mérito genético, de forma a permitir alta eficiência seletiva, grande rapidez na obtenção de ganhos genéticos com a seleção de baixo custo em comparação com a seleção tradicional baseada em dados fenotípicos. O presente estudo teve como objetivo comparar quatro metodologias bayesianas de seleção genômica (Bayes A, Bayes B, RR-BLUP e LASSO Bayesiano) para estimar o mérito genético do comprimento de carcaça de uma população de suínos. Para tanto, foram analisados dados fenotípicos e genotípicos de uma população F2 de suínos provenientes da Granja de Melhoramento de Suínos da Universidade Federal de Viçosa que se referem ao comprimento de carcaça 345 suínos, juntamente com 237 marcadores moleculares SNPs distribuídos nos seguintes cromossomos de *Sus scrofa*: SSC1 (56), SSC4 (54), SSC7 (59), SSC8 (31), SSC17 (25) e SSCX (12). Posteriormente a obtenção do mérito genético, visando à comparação das metodologias calculou-se as acurácias, as quais podem variar de 0 a 1, sendo o melhor modelo aquele que apresentar o maior valor. O LASSO Bayesiano obteve a melhor acurácia (0,63), enquanto o Bayes A e B obtiverem a mesma acurácia de 0,54 e o RR-BLUP com 0,46. Para avaliar os efeitos dos marcadores em cada cromossomo foi construído o Manhattan plot. Os maiores efeitos de SNP's para o comprimento de carcaça estão presentes nos cromossomos 1, 7 e 17 os quais apresentam QTL's relacionados à características de carcaça na literatura.

* ithalo.coelho@gmail.com, Universidade Federal de Viçosa

Imputação múltipla em regressão logística aplicada a dados de efetividade da terapia antirretroviral

Jéssica Luiza Ferreira Ramos^{*}; Edna Afonso Reis; Juliana de Oliveira Costa;
Maria das Graças Braga Ceccato; Francisco de Assis Acurcio

Um problema frequente nos levantamentos de dados de saúde são as perdas de dados. A maioria dos softwares estatísticos utiliza a exclusão dos indivíduos com dados faltantes como padrão para as análises. Uma alternativa é a utilização de métodos de Imputação Múltipla (IM). O objetivo do trabalho é avaliar a robustez do método IM comparando resultados dos modelos de regressão logística para a efetividade da terapia antirretroviral. Os dados foram obtidos por meio do projeto ECOART. O método de IM gerou estimativas consistentes para o modelo de regressão logística.

^{*} jessica.luiza297@gmail.com, Universidade Federal de Minas Gerais

Modelo multivariado para taxas e proporções espacialmente dependentes

Assis, A. N.; Cardoso Neto, J.* Lima, M. S.

Historicamente a temperatura na região amazônica é bastante elevada e combinada com níveis baixos de umidade relativa do ar, pode ocasionar problemas à saúde humana. Portanto, é de interesse modelar a umidade relativa do ar para auxiliar o planejamento de ações de prevenção da saúde. A distribuição generalizada de valores extremos tem sido utilizada para modelar valores de umidade relativa do ar. Como a medida dessa variável é uma porcentagem, propomos a utilização de um modelo multivariado para taxas e proporções espacialmente dependentes. O modelo é construído a partir da marginalização de uma distribuição Kumaraswamy G exponencializada condicionada a um efeito aleatório espacial estável tal que a distribuição marginal de Y é $F(y; \theta) = \exp$.

* jcardoso@ufam.edu.br, Universidade Federal do Amazonas

Acidentes de trânsito e condições de mobilidade nas capitais do Norte, Sul e Centro-Oeste do Brasil

Júlia Cristo Santos Marques^{*}; Ana Carolina Soares Bertho.

Introdução: No Brasil, em 2014, 43.780 pessoas morreram vítimas de acidentes de trânsito. Para reduzir essas mortes, é importante conhecer as características e os fatores de risco desse problema de saúde pública. A pesquisa analisa a mortalidade por acidentes de trânsito nas capitais das regiões Norte, Sul e Centro-Oeste do Brasil em 2010, caracterizando as vítimas por sexo, faixa etária e tipo. Também foram investigados o nível de instrução da população e as condições de mobilidade cotidiana, com o objetivo de verificar se em capitais com piores condições de mobilidade as taxas de acidentes eram mais elevadas. **Metodologia:** Foram usados dados do Sistema de Informações de Mortalidade (SIM) e do Censo Demográfico 2010 (IBGE). Foram calculadas taxas por sexo, tipo de vítimas e faixa etária. Do Censo, além do número de habitantes, foram obtidas informações sobre nível de instrução e tempo de deslocamento casa-trabalho. **Resultados:** Em todas as capitais, as taxas foram maiores para os homens, sendo a de Porto Velho (RO) a mais elevada (57,29 por 100 mil hab.). Quanto ao tipo de vítima, destacam-se os motociclistas de 15 a 29 anos e os pedestres idosos. Em todas as capitais, exceto Manaus (AM), a maior parte das pessoas gastava de 6 a 30 minutos entre casa e trabalho. Verificou-se que quanto maior o nível de instrução, maior é a porcentagem de pessoas que levavam até 30 minutos no deslocamento casa-trabalho. Não foi identificada relação direta entre maiores tempos de deslocamento e maiores taxas de mortalidade. **Conclusões:** As principais vítimas são homens, especialmente motociclistas jovens e os pedestres idosos. As capitais do Norte apresentam as maiores taxas de mortalidade e as capitais do Sul, as menores. Porto Velho e Boa Vista, duas capitais de baixa densidade demográfica, apresentam as maiores taxas de mortalidade por acidentes de trânsito.

^{*} juliacrsto10@hotmail.com, Escola Nacional de Ciências Estatísticas - ENCE

Análise discriminante e análise de cluster aplicadas em dados sobre Pokémon

Júlia dos Santos Porphirio^{*}; Luiz Fernando Guilhem Nassif Maia

Este trabalho tem como objetivo empregar técnicas de Análise Discriminante para criar uma regra matemática de classificação que tome possível a identificação de Pokémon lendários. Após a escolha da melhor regra de decisão, iremos prever onde um grupo peculiar de Pokémon conhecido como “Ultra Beasts”, cujo status de lendário ainda não foi divulgado oficialmente, melhor seria inserido. Também foi realizada uma Análise de Cluster para comparar os grupos resultantes dessa análise com os grupos utilizados para separação na Análise Discriminante.

juliaporphirio@gmail.com, ENCE - Escola Nacional de Ciências Estatísticas

Desempenho do gráfico de controle T^2 de Hotelling

Juliana Vieira Gomes*; Camila Rafaela Gomes Dias; José Ivo Ribeiro Júnior.

Com o aumento da exigência sobre a qualidade dos produtos e serviços, o controle estatístico de processos, de forma multivariada, tem se tornado mais frequente. O primeiro gráfico de controle multivariado foi desenvolvido por Hotelling para o monitoramento do vetor de médias. Com o objetivo de analisar os efeitos das escalas das variáveis, das correlações entre elas (0; 0,25; 0,5; 0,75; 1) e do número de repetições por grupo racional (2 e 3), sobre as probabilidades dos alarmes falsos e verdadeiros, foram realizadas simulações no software R em diferentes cenários, com 1.000 subgrupos racionais. Para o processo sob controle estatístico, foi considerada a seguinte distribuição de probabilidades: Y_1 e $Y_2 \sim N_2(\mu_0 = [0 \ 0], \Sigma = [1 \ \rho; \rho \ 1])$. Já para o processo fora de controle de estatístico, o vetor μ_0 sofreu um deslocamento para o vetor $\mu_1 = [\delta \ \delta]$, sendo $\delta = 0,5; 1; 2$ e 3 . Esse deslocamento representa o aumento da média em número de desvio-padrão. Após as construções dos gráficos de controle T^2 de Hotelling, concluiu-se que as escalas das variáveis não afetam o desempenho, que a probabilidade do alarme falso é baixa ($\hat{\alpha} \leq 0,003$) e independe das correlações e do tamanho do subgrupo racional, e que a probabilidade do alarme verdadeiro é alta ($P_{\hat{d}} > 0,9$) somente quando são avaliadas variáveis com altas correlações ($r \geq 0,75$) e quando o vetor de médias sofre grandes alterações ($\delta \geq 2$). Portanto, recomenda-se o gráfico de controle T^2 de Hotelling para detectar grandes variações no vetor de médias de variáveis altamente correlacionadas entre si.

* juliana-vieira70@hotmail.com, UFV

Zero-inflated mixed Poisson regression models

Jussiane Gonçalves da Silva^{*}; Wagner Barreto de Souza

A general regression model based on a class of zero-inflated mixed Poisson distributions is introduced in order to deal with overdispersion and the excess of zero in count data. The proposed class generalizes the mixed Poisson regression models introduced by Barreto-Souza and Simas(2016) and unify some existent models, such as zero-inflated negative binomial and zero-inflated Poisson-inverse Gaussian regression models, just as open the possibility to introduce new zero-inflated models. Computational resources were explored in the R program to obtain the model parameters estimates, in which an EM algorithm was proposed, a simulation study was performed in order to check the results obtained through the EM algorithm for finite samples, as well as residual analysis and diagnostic for assess global influence.

^{*} jussianegoncalves@gmail.com, Universidade Federal de Minas Gerais

Regressão quantílica na avaliação da adaptabilidade fenotípica de cultivares de feijão

Laís Mayara Azevedo Barroso* ; Moysés Nascimento;
Ana Carolina Campana Nascimento; Ithalo Coelho de Sousa; Leiri Daiane Barili;
Naine Martins do Vale; José Eustáquio de Souza Carneiro

Diferentemente das metodologias tradicionais de adaptabilidade e estabilidade baseadas em regressão (média condicional), a Regressão Quantílica (RQ) possibilita generalizar a relação funcional para toda a distribuição dos valores fenotípicos. Diante dessa possibilidade, objetivou-se caracterizar o efeito do parâmetro de adaptabilidade ao longo da distribuição de produtividade de cultivares de feijão preto lançadas por diferentes instituições nos últimos 50 anos. Foram utilizados dados de um experimento sobre produção de grãos de 40 genótipos de feijão preto desenvolvidos pelos programas brasileiros de melhoramento genético entre os anos de 1960 a 2013, avaliados em 4 ambientes utilizando o DBC com três repetições. Inicialmente foram ajustados aos dados modelos de RQ considerando 9 quantis. Posteriormente, por meio do teste de Wald, verificou-se que os modelos quantílicos diferem quanto ao parâmetro de adaptabilidade (β) e produtividade média (μ). Considerando estes modelos, foram formados dois grupos, e os cultivares foram agrupados em cada um deles por meio da distância Eudiana média. Foi possível observar que genótipos lançados recentemente foram alocados no grupo determinado pelo modelo quantílico β . As estimativas dos parâmetros deste modelo revelaram que os genótipos pertencentes a esse grupo possuem uma alta produtividade e seu plantio é recomendado a ambientes favoráveis. Por outro lado, os genótipos do grupo μ , apresentaram baixa produtividade e são recomendados para ambientes desfavoráveis. Desta forma, o processo de melhoramento dos últimos 50 anos, possibilitou um incremento tanto na produtividade média quanto no valor do coeficiente de adaptabilidade.

* lais.azevedobarroso@gmail.com, Universidade Federal de Viçosa

Modelo de regressão lognormal com censura intervalar e erros de classificação

Larissa Natany Almeida Martins* ; Magda Carvalho Pires

A análise de sobrevivência utiliza ferramentas estatísticas para estudar o tempo até a ocorrência de um evento de interesse. Quando não se sabe ao certo o tempo exato de ocorrência desse evento, mas sabe-se que ele ocorreu em um intervalo, tem-se que os dados são do tipo censura intervalar. Se o evento é silencioso (ou assintomático), sua detecção pode estar condicionada ao resultado de testes diagnósticos que, por sua vez, podem estar sujeitos a erros de classificação. Sendo assim, o processo de inferência sobre a distribuição dos tempos de falha será comprometido e produzirá estimativas viciadas se esses erros forem ignorados. Para lidar com esse problema, Silva (2012) desenvolveu o Modelo Weibull de Riscos Proporcionais com Erros de Classificação, que é aplicado quando o interesse é identificar os fatores que influenciam no tempo de ocorrência do evento quando esse tempo tem distribuição Weibull. Este trabalho, por sua vez, avalia aquele modelo utilizando a distribuição Lognormal, que não faz parte da família de riscos proporcionais. Isso quer dizer que se a distribuição é Lognormal quando todas as covariáveis são iguais a zero, a distribuição não será Lognormal para outros valores de covariáveis. Estudos de simulação demonstraram que a metodologia proposta apresentou resultados satisfatórios quando comparados ao modelo convencional (que desconsidera os erros de classificação), produzindo estimativas menos viciadas.

* larissa.natany@hotmail.com, Universidade Federal de Minas Gerais

Comparação do desempenho de modelos da teoria de resposta ao item para itens dicotômicos: um estudo de simulação

Leonardo Gonçalves Pinheiro Júnior* ; Ilka Afonso Reis

Recentemente, a Teoria de Resposta ao Item (TRI) ganhou notoriedade no Brasil como método para produzir índices de proficiência de alunos em testes educacionais como o ENEM. A TRI é centrada na estimação do traço latente, representado pela letra grega teta (θ), que pode ser definido como alguma característica ou conhecimento do indivíduo como, por exemplo, o nível de conhecimento em matemática, a intensidade de uma depressão ou a motivação de um indivíduo em praticar atividades físicas. Além disso, a TRI também pode ser útil no processo de validação de um instrumento de mensuração, pois consegue indicar a importância de cada um dos seus itens na composição da nota geral. Este estudo trata de uma análise de desempenho dos três modelos mais conhecidos da TRI para itens dicotômicos, Modelo de Rasch, Logístico de 2 parâmetros e Logístico de 3 parâmetros. Os modelos se diferenciam no número de parâmetros (dificuldade, discriminação e acerto ao acaso) e deseja-se analisar a influência do número de itens e do tamanho amostral na estimação do traço latente. As simulações são divididas em três níveis: simulação do traço latente pelos três diferentes modelos, número de itens (15, 30 e 60) e tamanho amostral (100, 250, 500 e 1000). É pressuposto que os modelos apresentam melhor desempenho para grandes tamanhos amostrais e que, pelo princípio da invariância, a estimação dos parâmetros dos itens independe de θ e que a estimação de θ independe dos itens utilizados.

* leogpjr@gmail.com, UFMG

Uma análise da inadimplência ocorrida nas operações de crédito para pessoas jurídicas

Leonardo Henrique de Andrade Bento do Nascimento^{*}; Ana Carolina de Oliveira;
Gabriel Rodrigo Gomes Pessanha.

Este trabalho busca analisar a relação existente das operações de crédito do sistema financeiro que geram inadimplência com a taxa básica de juros da economia (Selic) e o produto interno bruto (PIB). Inicialmente, foram adotadas 4 variáveis independentes. Além da taxa Selic e do PIB foram considerados também o nível total de concessões acumuladas nas operações de crédito e o imposto sobre a renda, ambos para pessoas jurídicas. No entanto, elas foram retiradas do modelo visto que a inclusão das mesmas gera colinearidade no modelo. Os dados utilizados compõem uma série temporal mensal de março de 2011 até maio de 2017, retirados da base de dados do Instituto de Pesquisa Econômica Aplicada - IPEADATA. Com o objetivo de considerar os valores reais das observações, utilizou-se do indexador inflacionário IGP-DI. Para o desenvolvimento da técnica estatística que foi aplicada ao conjunto de dados, adotou-se o software GRET. A abordagem escolhida para a análise dos dados foi de natureza quantitativa com a utilização de Regressão Linear Múltipla (RLM) ajustada pelo método Mínimos Quadrados Ordinários. Os pressupostos acerca dos resíduos gerados pela RLM foram testados e respeitados. Em relação ao modelo, o coeficiente de determinação encontrado foi de 0,6449, ou seja, das variações ocorridas nas operações de crédito que finalizaram em inadimplência, 64,49% delas são explicadas pelas oscilações das variáveis independentes. Essa estatística foi comprovada mediante a realização do teste F que se mostrou significativa ao nível de 1%. A variável PIB também foi significativa ao nível de 1% enquanto a variável Selic foi ao nível de 5%. O relacionamento entre o PIB e o nível de inadimplência é positivo, em contrapartida o relacionamento da variável dependente com a taxa Selic é negativo. Ressalta-se que há a possibilidade do nível de inadimplência do tempo t não ser influenciado imediatamente pelas variações do PIB e da taxa Selic no mesmo instante de tempo. Sugere-se que o incremento de defasagens nas variáveis independentes possa resultar em uma análise mais efetiva.

^{*} leonardoandrade.h@outlook.com, Universidade Federal de Alfenas

Modelos GLARMA para séries temporais

Glaura C. Franco; Leticia Moreira de Barros*

Avaliação da metodologia dos modelos lineares generalizados autorregressivos médias móveis (GLARMA) para séries reais de contagem que possuem dados correlacionados.

* Imbarros96@hotmail.com, Universidade Federal de Minas Gerais - UFMG

Métodos de simulação de dados geográficos sintéticos para bases confidenciais

Leticia Silva Nunes^{*}; Thaís Paiva Galletti

Este trabalho apresenta métodos estatísticos para preservação de sigilo de bases de dados confidenciais, em especial, a simulação de dados sintéticos. Este método, simula dados sintéticos que são gerados a partir de distribuições de probabilidade especificadas para reproduzir o máximo de relações possíveis contidas nos dados originais. A simulação de dados sintéticos é atrativa pois permite controlar a preservação da privacidade e utilidade, além de permitir que os pesquisadores utilizem ferramentas convencionais de análise estatística. Além disso, busca-se estender metodologia para simulação de coordenadas geográficas sintéticas utilizando modelos para dados de área que é proposta em Paiva et al. (2014).

^{*} snunesleticia@gmail.com, Universidade Federal de Minas Gerais

Estimação teórica do tamanho crítico amostral para variáveis com dependência espacial

Lucas Monteiro Bianchi* ; Calisto Manuel Máquina ; Gerson Rodrigues Dos Santos;
Nerilson Terra Santos; Adriana Maria Rocha Trancoso Santos

Um dos desafios do planejamento amostral da Geoestatística está a determinação do tamanho amostral. Apesar de ser uma área da Estatística Espacial criada desde a década de 50, este ainda um tema complexo e sem consolidação. Até mesmo o que figura na literatura especializada não consolida o assunto, pois algumas vezes encontra-se o mínimo de 50 pontos amostrais, outras o mínimo de 100 pontos, e ainda 150 pontos amostrais. Assim, o objetivo desse trabalho é o de verificar se a estimação do tamanho amostral através do Teorema de Nyquist é suficiente para representar a distribuição espacial das variáveis sobre física e química de um solo do interior do estado de São Paulo. De acordo com este teorema, variogramas experimentais modelados pelo modelo exponencial podem indicar o tamanho amostral ótimo levando em consideração uma amostragem prévia da região em estudo e do parâmetro alcance estimado. Assim, com base nos dados analisados, a saber, Na, Mg, Cu, pH, K, MO, Zn, PMP, Ca, CC e Mn, pode-se dizer que uma amostra de 49 pontos amostrais apresenta uma densidade amostral menor do que a necessária para a predição populacional. Logo, um segundo planejamento amostral com a densidade adequada se faz necessário.

* lucas_bianchi123@hotmail.com, Universidade Federal de Viçosa

Análise bayesiana na previsão de ventos extremos em Sorocaba-SP

Gisele Carolina Almeida; Fabricio Goecking Avelar; Luiz Alberto Beijo*

O conhecimento da ocorrência de ventos extremos é de suma importância para o planejamento de atividades humanas de forma a evitar ou diminuir os impactos destrutivos. Sendo assim, o objetivo deste trabalho foi avaliar a aplicação da Inferência Bayesiana na previsão da ocorrência de ventos máximos semestrais, em Sorocaba-SP, utilizando-se a distribuição generalizada de valores extremos (GEV). Foi utilizada como distribuição a priori, a Normal Trivariada proposta por Coles e Powell (1996), e as informações dos hiperparâmetros foram obtidas do trabalho de Bautista et al. (2004), que estudaram os ventos máximos de Piracicaba-SP. Além disso, foram criadas diferentes combinações de distribuição a priori, usando estruturas de variâncias multiplicando-se a matriz de covariâncias por 1, 2 e 4. Os dados de velocidades máximas de vento (km h⁻¹) foram obtidos junto ao Instituto Nacional de Meteorologia (INMET), referentes ao período de agosto de 2006 a dezembro de 2016. Os máximos observados nos últimos cinco semestres (2014/2 a 2016/2) foram usados para avaliar a acurácia e o erro médio de previsão para os tempos de retorno de 2, 3, 4 e 5 semestres. A técnica Monte Carlo via Cadeias de Markov (MCMC) foi empregada para inferência a posteriori de cada parâmetro, e para análise de convergência foram utilizados os critérios de Geweke, e de Raftery-Lewis. A aplicação da inferência Bayesiana levou a erros de previsão menores que 6%, sendo que a distribuição a priori com estrutura de variâncias, na qual a matriz de covariâncias foi multiplicada por 4, apresentou acurácia na previsão da ocorrência de ventos máximos em todos os tempos de retorno. A metodologia Bayesiana forneceu resultados acurados e precisos, apresentando-se como uma excelente metodologia para a previsão de ventos extremos.

* prof.beijo@gmail.com, Universidade Federal de Alfenas

Modelos de séries temporais na análise da série de produção de automóveis no Brasil

Luiz Matheus Barbosa Santos* ; Maria Imaculada Sousa Silva

O início da evolução automotora no Brasil incia-se após o fim da Segunda Guerra mundial. Na década de 30, fabricas estrangeiras como a Ford e General Motors inauguraram suas linhas de montagem em território nacional, porém, somente no governo de Juscelino Kubitschek, em 1956, que as multinacionais automotivas de fato empreenderam a produção automobilísticas no país. Nos últimos anos, o ramo de produção de automóveis é um mercado gerador de milhões de empregos em todo o Brasil, além de movimentar milhões de reais no impulso à economia. Segundo a Associação Nacional dos Fabricantes de Veículos Automotores (Anfavea), a produção acumulada de veículos no ano de 2017 representa um crescimento de 20.9% comparada ao mesmo período do ano anterior no país, após quatro anos de recessão. De janeiro a abril, as exportações tiveram alta de 64.2% em relação ao ano passado, o que auxilia na diminuição da ociosidade da indústria, ainda significativa no Brasil. Com o objetivo de obter previsões acerca da produção de veículos em 2017 no território nacional, fora utilizada uma Série Temporal da produção de automóveis no Brasil nos períodos de janeiro de 2000 a junho de 2016 disponível no sítio eletrônico do Instituto de Pesquisa Econômica Aplicada (IPEA). Na análise utilizada, foram ajustados modelos auto regressivos integrados de médias móveis (ARIMA) aos dados originais e aos dados transformados pelo logaritmo natural, acrescentando-se ao modelo o efeito de intervenção na observação 107, correspondente ao mês de novembro do ano de 2008. Os critérios usados na comparação e escolha do modelo mais adequado foram o critério de informação bayesiano, critério de informação de Akaike, e o erro quadrático médio de previsão. De acordo com os critérios utilizados, foi possível inferir que entre outros modelos ajustados, o modelo SARIMA(1,1,0)(1,0,0) com intervenção descreve de forma adequada a informação dos dados da série.

* luismatheussantos@hotmail.com, Universidade Federal de Uberlândia

Fatores associados à expectativa de vida ao nascer no Japão

Luiz Otávio de Oliveira Pala* ; Bruna da Costa Silva

A expectativa de vida ao nascer é um indicador analisado por diversas áreas de pesquisa de interesse social e de saúde pública. Estudos apontam a relacionam tal variável com a condição de saúde, demográfica e econômica no qual estão expostos os habitantes de determinada localidade ou país. De acordo com o IBGE (2016), o Japão é um país que apresenta uma das maiores expectativas de vida ao nascer, de 83,5 anos, acompanhado de um índice de desenvolvimento humano equivalente a 0,891. O país apresentou melhorias das condições de saúde e habitação com o seu desenvolvimento econômico após o período da Segunda Guerra Mundial, alavancando a expectativa de vida (SAKURAI, 2007). Dada situação analisada, este trabalho explorou variáveis que se associaram ao crescimento da expectativa de vida desse país, entre os anos de 1980 e 2011, incluindo fatores de qualidade de vida da população, como taxas de vacinação infantil, percentual da população em sobrepeso, fumantes ou alcoólatras. Como justificativa, entende-se que o alto desenvolvimento deste indicador no Japão incentive a criação de políticas públicas que visem as condições de saúde pública no Brasil, como o controle da obesidade e do consumo de tabaco, e a ampliação da política de vacinação infantil contra a difteria, tétano, coqueluche e sarampo. Vale ressaltar que os dados analisados neste trabalho são disponibilizados pela OECD (2017). Para tal objetivo, utilizou-se da metodologia de regressão linear múltipla, com auxílio do Software R. Desta forma, notou-se que medidas de saúde, como taxas de vacinação, impactam diretamente no aumento da expectativa de vida e, por outro lado, fatores de risco, como o percentual de fumantes, impactam inversamente o comportamento da expectativa de vida. Por fim, entende-se por necessário a ampliação de estudos relacionados às condições de saúde pública em países como o Brasil, conectando experiências presenciadas por outras regiões e países, garantindo melhores condições e políticas de saúde a população brasileira.

* luizotavio.oliveira@gmail.com, Universidade Federal de Alfenas

Consumo de energia elétrica na região Sudeste do Brasil: uma aplicação do modelo SARIMA

Luiz R. Nakamura^{*}; Meire Mezzomo; Andréa C. Konrath; Andreia Zanella; André W. Zibetti

A análise de séries temporais ocupa lugar de destaque nas metodologias existentes para a realização de previsões de uma determinada série ao longo do tempo, podendo ser utilizada com o intuito de modelar um grande número de observações, considerando a possível sazonalidade presente nos dados. Dentre os modelos sazonais, o modelo auto-regressivo integrado de médias móveis sazonal (SARIMA) possui grande importância, sendo um dos mais utilizados na literatura atual. Seguindo este conceito, neste trabalho, aplicamos o modelo supracitado em uma série histórica do consumo mensal de energia elétrica (TWh) na Região Sudeste do Brasil, no período de 1985 à maio de 2017. Para a validação das previsões obtidas a partir dos modelos ajustados foram desconsideradas as observações dos três últimos meses no processo de estimação de seus parâmetros. Alguns modelos foram aplicados aos conjunto de dados e, para efeito de comparação, os seguintes critérios de qualidade de ajuste foram utilizados: i) critério de informação de Akaike (AIC); ii) critério de informação de Schwarz (SBC); e iii) erro quadrático médio mínimo de precisão (EQMMP). Finalmente, utilizando os três critérios, o modelo melhor ajustado aos dados foi o SARIMA(1,1,1)(1,0,1)₁₂. Com o modelo escolhido, foi possível analisar quedas e ascensões no consumo de energia na região em estudo e as previsões realizadas podem ser utilizadas como base para possíveis políticas públicas e/ou privadas dos atores da região.

^{*} luiz.nakamura@ufsc.br, Universidade Federal de Santa Catarina

Inferential aspects for the INAR(1) process with zero-inflated Poisson innovations

Luiza Piancastelli^{*}; Wagner Barreto-Souza

In this work we investigate inferential aspects of the INAR(1) (Integer Autoregressive of order one) with zero-inflated Poisson innovations proposed by Jazi et. al. (2012).

This INAR process is denoted as ZIPINAR(1) and is able to capture greater incidence of zeros than expected when using the Poisson distribution. We propose estimation by Conditional Least Squares (CLS) and Yule-Walker (YW) methods, which were not considered by Jazi et al. (2012). In order to compare those methods Monte Carlo simulations are performed.

Results pointed advantages for the CLS approach. For the parameter α results were similar. However, for the parameter representing the proportion of extra zeros, ρ , CLS produced less biased results. Although calculation of CLS estimators take some extra work, results pointed those as more reliable than those produced by YW estimation.

^{*} luh.sette@gmail.com, UFMG

Extending JAGS for spatial data

Magno Tairone de Freitas Severino^{*}; Vinicius Diniz Mayrink; Fabio Nogueira Demarqui

Bayesian hierarchical modeling for spatial data is challenging for professionals from other areas than statistics. Setting the model prior distributions and the likelihood are the simplest part of the process. What makes it difficult is the computation of the posterior full conditionals. The BUGS (Bayesian inference Using Gibbs Sampling) family of statistical softwares reduces the effort of modeling, since the user must indicate only the prior distributions and the likelihood. JAGS, the C++ developed open-source version of the BUGS family, does not contain any function or distribution for spatial modeling. Based on the paper from Wabersich and Vandekerckhove (2014), this project aims to fill this gap through the implementation of an extension to the JAGS software, allowing users

to perform a spatial data modeling and analysis. In this study we take into account the point-referenced data case and we analyze several covariance structures. We developed a novel JAGS module that contains twelve covariance functions, allowing the user obtain the covariance matrix through a simple command line in the JAGS model script. We validated the proposed module using a simple data set from Diggle and Ribeiro (2007): we compared the parameter estimates with those obtained from the maximum likelihood method. During the development period, we have realized that JAGS modules can be created only with functions, i.e. there is no need to add a routine to sample from certain distribution in order to build a module. This can be considered a contribution of our work, since the tutorial in Wabersich and Vandekerckhove (2014) does not mention this aspect.

^{*} magnotairone@gmail.com, UFMG

Erros do modelo de regressão linear simples: e se a suposição variância constante não for válida?

Maiara Fernanda Oliveira de Miranda^{*}; Ilka Afonso Reis

Esse estudo sobre Regressão Linear Simples trata da violação da suposição de variância constante e como isso afeta a estimação dos parâmetros. Para isso, simulamos populações de um milhão de pares (X, Y) , com X seguindo distribuição Uniforme entre 10 e 20, β_0 sendo igual a 5, β_1 igual a 0,5, $Y = \beta_0 + \beta_1 * X + \epsilon$. Os erros foram criados seguindo distribuição Normal ($\mu=0, \sigma^2=1*g(x)$), com $g(x) = 1; \log(X); \sqrt{X}; X; e X^2$. Retiramos, de cada uma dessas populações, 1000 amostras de tamanho $n=10, 50, 100, 250, 500$ e 1000 elementos. O modelo de regressão linear simples foi ajustado aos dados de cada um das 1000 amostras e as estimativas dos parâmetros foram resumidas por meio de médias e desvio-padrões e visualizadas graficamente por meio de histogramas. Percebemos que a violação da suposição de Variância Constante dos Erros não afeta a estimação do β_1 e β_0 , exceto para casos em que a alteração na variância do erro é muito agressiva e a amostra é pequena. Porém, há problemas na estimação do parâmetro σ^2 , que fica muito superestimado, mesmo com alterações sutis e amostras grandes. Com isso, o poder do teste F também fica muito prejudicado.

^{*} maiaraoliveira.m@gmail.com, UFMG

Análise da volatilidade da taxa de retorno das ações da vale na Bovespa utilizando o modelo APARCH

Marcela de Marillac Carvalho; Manoel Vitor de Souza Veloso *

Estudos sobre volatilidade têm sido utilizado na análise de séries temporais econométricas, contribuindo para a compreensão de questões na área de finanças. O presente trabalho tem como objetivos modelar e analisar a volatilidade no mercado de commodities metálicas, que configura um dos principais itens da pauta de exportações do país. Esse estudo foi realizado com retornos diários, em semanas de cinco dias, das ações ordinárias da Vale do Rio Doce (VALE) no período de 03/jan/2000 a 21/jun/2017, totalizando 4314 observações. Um modelo AR incompleto foi ajustado na série dos retornos, com defasagens específicas (1, 2, 3, 4, 10, 14, 19). Foi realizado o teste LM nos resíduos quadrado para confirmar a existência do efeito ARCH. Daí, um modelo AP ARCH(1, 1) foi ajustado, usando a distribuição t-Student, com a finalidade de captar, principalmente, o efeito de alavancagem, referente aos diferentes impactos de choques positivos e negativos da volatilidade da série. A análise foi realizada por meio do software GRET. Verificou-se que o efeito de alavancagem foi significativo e positivo, indicando que choques negativos passados têm maior impacto sobre a volatilidade condicional no presente, sendo que aproximadamente 99% estes choques tendem a persistir na volatilidade dos retornos da VALE por vários períodos de tempo. O ajuste da série dos retornos apresentou uma baixa volatilidade incondicional e também captou a existência de caudas pesadas. Observou-se uma alta volatilidade no ano de 2008, devido a crise do subprime originada nos Estados Unidos, havendo, nesse período, uma tendência de queda nas ações na companhia VALE devido ao aumento do risco e à estagnação na demanda de minério de ferro, principalmente na Europa, afetando o preço e o volume de negociação neste mercado. A partir de 2015, a volatilidade volta a apresentar altos níveis, consequência da crise política e econômica que o país vem atravessando, adicionando incerteza ao mercado, e das constantes oscilações no desempenho dos contratos futuros de commodities metálicas, ocasionando uma forte pressão nas ações da companhia. O modelo ajustado captou um efeito de alavancagem significativo e também identificou as alterações no padrão da volatilidade dos retornos da VALE, videnciando os períodos de incerteza no mercado acionário.

* manoel.souza@unifal-mg.edu.br, UNIFAL-MG

Predição na presença de diferentes níveis de multicolinearidade por meio de métodos de aprendizado estatístico

Márcia Elaine Guimarães Lana* ; Ithalo Coelho de Sousa; Ivan de Paiva Barbosa;
Ana Carolina Campana Nascimento; Moysés Nascimento.

Avaliar diferentes métodos estatísticos, a saber - Regressão Linear Múltipla (RLM), Árvores de Regressão (Tree), Tree Bagging (TBAGG) e Random Forest (RDF) - para predição de variáveis em 5 cenários. Cada cenário é composto por 100 observações de 5 variáveis simuladas, os quais diferem quanto à estrutura de covariância entre as variáveis preditoras.

* marcia.guimaraesufv@gmail.com, Universidade Federal de Viçosa

Análise de série temporal da produção de petróleo no Brasil de 2000 a 2016

Mariana de Souza Gonçalves* ; Camila Moreira Ribeiro;
Tatiane de Fátima Claudiano; Letícia Lima Milani Rodrigues.

A produção de petróleo é de extrema importância econômica, e, no Brasil, foi demarcada por um crescimento de desempenho constante nos últimos anos. No cenário atual, o país já produz o suficiente para satisfazer o seu mercado interno. Como é um insumo que pode ser utilizado na produção e no consumo de energia elétrica, além de ser uma matéria-prima indispensável, os esforços para realizar sua exploração e refinamento ocorrem de forma gradativa ao longo dos anos. De acordo com dados da Petrobras (Petróleo Brasileiro S/A), o Brasil é o segundo maior produtor de petróleo da América Latina (precedido pela Venezuela), e o décimo sétimo maior do mundo, com uma produção diária de dois milhões de barris por dia. Dada sua relevância, propõe-se observar, no presente trabalho, a evolução da produção de petróleo no Brasil por barris, utilizando uma série de dados disponibilizada pela Agência Nacional do Petróleo (ANP). Com o uso da metodologia SARIMA, foi realizada a análise de série temporal. Os dados utilizados são de frequência mensal, do período de janeiro de 2000 a dezembro de 2016, disponibilizados pelo Instituto de Pesquisa Econômica Aplicada. Portanto, a finalidade deste artigo é estudar a produção brasileira de petróleo, através da análise do comportamento da série temporal. Através do menor valor do Critério de Akaike (AIC), foi selecionado o modelo SARIMA(1,1,2)(1,1,1)₁₂, e feita a previsão, após verificar tendência e sazonalidade, para os meses de janeiro, fevereiro e março de 2017. Em comparação com a série original, o modelo apresentou um bom resultado, visto que o mesmo acompanhou a série.

* mari.cbq@hotmail.com, Universidade Federal de Alfenas

Comparação de métodos para seleção de preditores

Mariana Tavares Pinto Vidigal* ; Larissa de Carvalho Alves

Um dos propósitos da estatística é estudar a relação entre variáveis. A análise de regressão emprega modelos para analisar a relação linear entre uma variável resposta e uma ou várias variáveis preditoras. Em muitos estudos que envolvem modelagem estatística tem-se a presença de uma numerosa quantidade de variáveis explicativas que ocasiona, por consequência, dificuldade de estimação e obtenção de modelos complexos. Uma alternativa em circunstâncias como estas é selecionar quais dessas variáveis do conjunto de regressores irão conter o máximo de informação significativa para o modelo em questão. Muitos são os métodos na literatura que propõem técnicas de seleção de variáveis, dois deles são o Lasso Bayesiano (Park e Casella, 2008) e o Elastic Net Bayesiano (Li e Lin, 2010). A proposta deste trabalho é fazer estudos de simulações com o objetivo de comparar a eficiência e o poder de seleção dos dois métodos. Quando comparado ao Elastic Net, o Lasso, segundo Li e Lin (2010), apresenta três principais desvantagens. Em primeiro lugar, o Lasso não é capaz de selecionar um número de regressores maior que o tamanho da amostra. Em segundo lugar, quando os preditores apresentam uma estrutura de grupo, o Lasso seleciona apenas um regressor do grupo, desprezando os demais. Por último, quando há alta correlação entre os preditores, o Lasso apresenta um desempenho insatisfatório. Visto isto, foram feitas simulações intensivas apresentando estruturas de grupos e alta correlação em modelos de regressão para verificar tais propriedades e analisar outros pontos relevantes.

* marianatavarespv@gmail.com, Escola Nacional de Ciências Estatísticas

Avaliação do risco de investimento em grandes empresas brasileiras: uma aplicação do modelo GARCH

Marina Amorim^{*}; Luiza Sette; Thais Menezes; Leonardo Gonçalves

No mercado financeiro, é importante analisar o risco de investimentos, já que o mercado de ações é muito volátil. O trabalho tem como objetivo trabalhar com as séries de retornos diários de três grandes empresas brasileiras: AMBEV, Gerdau e Petrobras, os dados foram extraídos do portal Yahoo Finances, do período de 01/01/2013 a 31/12/2016. Para as séries foram calculadas medidas de risco através do VaR (Value at Risk) histórico, o VaR Econométrico via modelos GARCH, todos GARCH (1,1). Como resultado, para a carteira observamos altos valores do VaR e ES (Expected Shortfall), indicando significativa probabilidade de investimentos arriscados. Dentre as empresas analisadas, a AMBEV apresentou a menor projeção de lucratividade, mas também é a que oferece menor risco de perda. Outro fato observado neste estudo foi que a distribuição t do modelo Garch foi mais adequada nas três séries, as três séries apresentaram curtose maior que a da distribuição normal, e a distribuição t geralmente se adequa melhor a distribuições de cauda pesada, o que é o caso nas séries modeladas.

^{*} m2013034835@gmail.com, Universidade Federal de Minas Gerais - UFMG

Germinação e crescimento inicial do feijão vulgar (*Vigna Unguiculata*), sob efeito extrato aquoso de Mecrusse/*Cimbirre (Androstachys Johnsonii)*

Mario Sebastião Tuzine^{*}; Timóteo Frezara; Eduardo Langane; Natalino Calegario

A alelopatia pode ser definida como um processo pelo qual produtos do metabolismo secundário de um determinado vegetal são liberados, impedindo a germinação e o desenvolvimento de outras plantas relativamente próximas. Pesquisou-se efeito do extracto aquoso das folhas de Mecrusse (*Androstachys Johnsonii*) na germinação e crescimento inicial do feijão vulgar (*Vigna unguiculata*). O ensaio foi conduzido com delineamento inteiramente casualizado (DIC) com cinco tratamentos (T4=10%; T3=7,5%; T2=5,0% T1=2,5% e testemunha T0=0%); com cinco repetições. A verificação da germinação foi feita no intervalo de 12 em 12 horas (08h e as 20h). O experimento de crescimento foi realizado com as mesmas concentrações do onde foram medidas as partes aéreas, raiz e a biomassa seca total. As médias de tratamentos foram comparadas pelo teste de Duncan a 5% de probabilidade no pacote estatístico The R Project for Statistical Computing. Observou-se que a germinação foi alta no tratamento T0 na ordem de 91% em contrapartida para as sementes submetidas a T4 de extracto que germinaram 35%, tendo afectado ainda na velocidade de germinação sendo estatisticamente maior para o tratamento de agua em relação aos demais. Quanto ao crescimento, a altura diferiu em relação aos demais com 8cm para T0 e 2 cm para tratamento de T4, o mesmo sucedeu com a parte radicular com 7,3cm para T0 e 2,9cm para T4 também para a biomassa o resultado foi o mesmo com 0,41g/planta para T0 , e 0,39g/planta para T1 este deferiram em relação ao demais seno a massa de 0,05g/planta para T4. No que pode-se concluir que esses efeitos podem ser devido a substancia, presentes no mecrusse .

^{*} tuzine@yahoo.com.br, Instituto superior Politenico de Gaza -UFLA

Avaliação do efeito espacial no mapeamento de mortalidade neonatal precoce em Minas Gerais nos triênios de 2000-02 a 2012-14 na presença de dados sub-registrados

Marriane dos Reis Silva^{*}; Rosângela Helena Loschi; Guilherme Lopes de Oliveira.

Uma das grandes dificuldades na análise de mortalidade infantil em Minas Gerais é o sub-registro na contagem de óbitos que ocorrem em várias regiões do estado. Sendo assim, métodos estatísticos, mesmo os mais elaborados e robustos, que não levem em consideração esse possível erro presente na contagem dos dados, levarão a inferências viciadas para o risco relativo (RR) as quais serão de pouca validade para a definição de políticas públicas para a saúde.

O objetivo deste trabalho é obter as estimativas a posteriori para o RR de mortalidade neonatal precoce (MNP) no Estado de Minas Gerais nos triênios de 2000-02 a 2012-14. Os dados foram agrupados em triênios para evitar contagens próximas de zero o que podem levar a instabilidade nas estimativas. Foram utilizados dados de mortalidade neonatal precoce (crianças nascidas vivas que faleceram antes dos 7 dias) registrados na rede hospitalar. Para tratar adequadamente o sub-registro, utilizamos o modelo Poisson censurado o qual considera as contagens de mortes neonatais precoces observadas em áreas suspeitas de sub-notificação como os limites inferiores do real número de casos, ou seja, as contagens observadas nestas áreas são consideradas como informação censurada. Avaliamos também o impacto da inclusão do efeito espacial na sua estimativa. Uma dificuldade na implementação do modelo Poisson censurado é a definição das regiões censuradas. Construímos um critério de censura baseado no índice de adequação da informação desenvolvido por França et al. (2006). Tal índice de adequação avalia a qualidade da informação sobre óbitos e nascimentos nos municípios e microrregiões de Minas Gerais em 2000-2002. A distribuição a priori para a taxa de mortalidade em cada região foi construída baseando-se na informação fornecida por um especialista em análise de saúde pública no estado de MG.

Resumidamente, concluímos que as estimativas para o RR de MNP nas regiões com alto nível de pobreza e pouca assistência hospitalar, como o Norte e Nordeste de Minas Gerais, são mais altas e ficaram mais próximas do esperado, ao incluirmos a informação de subnotificação. Ao longo dos anos, essas estimativas ficaram maiores para essas regiões com piores indicadores de qualidade de vida, evidenciando uma melhora na coleta dos dados. Já a inclusão do efeito espacial, de modo geral, não alterou significativamente as estimativas do RR.

^{*} marriexpert@yahoo.com.br, Universidade Federal de Minas Gerais

Identificação de múltiplos pontos de mudança em sequências multivariadas

Matheus Barros Castro *

A análise de séries temporais é prática comum do âmbito da estatística. Entre os problemas analisados, está a identificação de pontos de mudança, isto é, momentos em que a distribuição de probabilidade dos elementos da série muda. Os pontos de mudança são causados por fatores externos ao processo aleatório natural da série. A identificação da mudança na série pode ser interessante, por exemplo, para determinar quais tipos de acontecimentos provocam mudanças significativas em indicadores econômicos ou quais eventos naturais causam alterações em características climáticas de certas regiões do globo.

O modelo partição produto (MPP) proposto por Hartigan (CSTM, 1990) é uma ferramenta bastante usada para a identificação de múltiplos pontos de mudança. Neste trabalho, vamos considerar o MPP para a identificação de múltiplos pontos de mudança no vetor de médias e matriz de covariância de dados normalmente distribuídos (Nascimento, 2017 e Moura, 2004). Neste contexto, o modelo, às vezes, indica mudanças em parâmetros onde elas não ocorreram. Nossa meta é propor metodologias baseadas em comparações sequenciais para identificar corretamente os parâmetros onde as mudanças ocorreram.

O modelo considerado foi implementado no R e o método Gibbs Sampler foi utilizado para amostrar da distribuição a posteriori. Foi proposto um estudo de simulação utilizando dados gerados de uma distribuição normal bivariada e diferentes cenários foram considerados para avaliar o método proposto.

* matheusb@ufmg.br, Universidade Federal de Minas Gerais

Modelos de probabilidades da família Tweedie para o consumo de energia elétrica (per-capto) da região sudeste do Brasil.

Matheus Ferreira do Amaral* ; Ednaldo Carvalho Guimarães;
Maria Imaculada de Sousa Silva; José Waldemar da Silva;

O objetivo deste trabalho foi ajustar modelos de probabilidade da família Tweedie aos dados de consumo (per-capto) de energia elétrica da região sudeste do Brasil, para cada mês. A série utilizada se refere ao consumo de energia elétrica per-capto (MWh), na região Sudeste do Brasil, no período de janeiro de 2004 a setembro de 2016. As informações foram obtidas da base de dados oficiais da empresa de Pesquisa Energética (EPE). As variáveis foram submetidas à análise descritiva e utilizando-se o programa R foram efetuados os cálculos das estimativas e a classificação das distribuições de probabilidades da família Tweedie para cada mês. Os resultados indicaram, por meio da classificação do parâmetro (p), que o modelo Poisson Composto ajustou-se bem aos dados para os meses, março, abril, maio, agosto, outubro e dezembro e para os meses janeiro, fevereiro, junho, julho, setembro e novembro o modelo com melhor desempenho de ajuste foi o Normal Inversa. O procedimento de seleção se baseou no perfil de verossimilhança e mostrou-se eficiente na modelagem de dados contínuos sem a presença de zeros. Na análise descritiva verificou-se que o mês de janeiro destaca-se com - maior consumo médio (per-capto) de energia elétrica (0,183774 MWh) . Verificou-se que a metodologia de ajuste de modelos probabilísticos da família Tweedie pode contribuir com a descrição do comportamento do consumo de energia no Brasil.

* matheusamaral.ferreira@hotmail.com, Universidade Federal de Uberlândia

Caraterização do perfil dos registros de violência sexual contra mulheres mineiras de 2010 a 2014

Natália da Silva Antunes^{*}; Ilka Afonso Reis

Este trabalho tem como objetivo caracterizar o perfil dos registros de violência sexual entre os anos de 2010 a 2014. Esperamos, também, contribuir para abordar e promover discussão a respeito do tema Violência Contra mulher em Minas Gerais e no enfrentamento da violência contra a mulher com o auxílio da metodologia estatística.

^{*} natyantunes93@gmail.com, Universidade Federal de Minas Gerais

Comparação de medidas de distância em análise de clusters para determinação de autoria em textos literários

Otávio A. S. Lima^{*}; Ilka A. Reis; Adriana S. Pagano

Análise de Cluster é uma técnica que pode ser aplicada para determinação de autoria. Oito tipos de distâncias foram investigados utilizando-se dados de poemas de quatro autores. As distâncias consideradas foram Eudidiana, Manhattan, Classic Delta, Argamon's Linear Delta, Eder's Delta, Eder's Simple, Canberra e Cosine. Os autores considerados para a análise foram Fernando Pessoa -- induídos seus heterônimos Alberto Caieiros, Álvaro de Campos e Ricardo Reis--, Teixeira de Pascoaes, Camilo Pessanha e Cesário Verde. As matrizes de distâncias consideradas foram usadas para classificar se um texto era ou não de determinado autor. Para efeitos de determinação de autoria dos textos, a distância Canberra foi a que teve melhor desempenho.

^{*} otavioasl@gmail.com, Universidade Federal de Minas Gerais

Estudo sobre a evasão discente de uma universidade de Minas Gerais via regressão logística

Patrícia de Sousa Ilambwetsi* ; Rosane Soares Moreira Viana; Archange Michael Ilambwetsi; Matheus de Paula Ferreira; Gérson Rodrigues dos Santos; Ronaldo Tavares

Este estudo analisa a evasão dos alunos de graduação na área de Ciências Exatas e Biológicas, divididos em cinco departamentos, de uma universidade no estado de Minas Gerais via modelos de regressão logística. Para tanto, foram determinadas seis covariáveis (cidade do aluno, ano da evasão, semestre, departamento, tipo de moradia e política afirmativa) como possíveis causas dessa evasão e avaliadas no período de 2008 a 2012, inseridas num contexto de modelos lineares generalizados. Assim, os principais objetivos foram: o estudo da influência das covariáveis na evasão dos alunos; a construção de tabelas cruzadas para estudar a frequência observada entre a evasão e as outras variáveis em estudo; escolha do melhor modelo que se ajustou aos dados com a utilização do método Step AIC para seleção das covariáveis significativas no modelo; e o cálculo da razão de chances (odds ratio) que compara a probabilidade de sucesso de uma variável em relação à outra. Os resultados mostraram que a decisão do aluno em continuar ou não no curso estava sendo fortemente influenciada pelo fato de sua escolha inicial ter sido por um curso de ciências exatas. Além disso, a distância da região de origem para a cidade universitária com mais de 140 km foi outro fator complicador.

* jpatysousa@yahoo.com.br, Universidade Federal de Viçosa

O uso do Shiny para o ensino de matemática e estatística no ensino médio

Paula Ribeiro Santos* ; Izabela Regina Cardoso de Oliveira

O Shiny é um pacote criado pelos desenvolvedores do RStudio, que permite a fácil interface entre o R e um navegador web. É uma estrutura de aplicações que facilita a construção de aplicações interativas com o R. Isso significa, por exemplo, executar temas complexos e retornar a saída de uma forma interativa no navegador. O objetivo desse trabalho foi demonstrar o uso desse recurso computacional no ensino de Matemática e Estatística na educação básica. Dentre as diversas vantagens do uso dessa ferramenta, destaca-se a possibilidade dos estudantes terem um aprendizado continuado, fora do ambiente da sala de aula, usando smartphone e tablets, por exemplo.

Nos aplicativos para o ensino de Matemática serão exploradas diversas funções matemáticas, e os conceitos de assíntota, limite e ponto crítico. Nos aplicativos sobre Estatística o enfoque será dado aos métodos descritivos, em que serão explorados diferentes tipos de gráficos para variáveis quantitativas e qualitativas, medidas de posição (média, mediana e moda) e dispersão (amplitude, desvio padrão, entre outras), e conceitos de assimetria. Espera-se que os profissionais da área de ensino sintam-se incentivados a utilizar essa e outras ferramentas para tornar o processo de ensino-aprendizagem mais dinâmico e atrativo para os jovens estudantes da era da informação e dos grandes avanços tecnológicos.

* pauillasant_s@hotmail.com, Universidade Federal de Lavras

Aplicativo pedagógico interativo de regressão linear

Pedro Ladeira Lóes* ; Edna Afonso Reis

O aplicativo consiste numa ferramenta pedagógica interativa web, para ilustrar graficamente o comportamento dos betas e estatísticas da Regressão Linear quando as suposições dos erros são violadas. O usuário pode alterar os erros da regressão gerando histogramas de 1000 modelos em simulações Monte Carlo. Para cada suposição, normalidade, independência, homocedasticidade e colineariedade, o experimento pode ser realizado. Gerando assim resultados dos estimadores e estatísticas do modelo. O objetivo principal é fornecer uma ferramenta gráfica para ilustrar a teoria matemática que sustenta as suposições da regressão. O aplicativo foi desenvolvido na plataforma shiny do R.

* ladeiraloes@gmail.com, UFMG

Uma análise exploratória sobre saúde e alimentação

Pedro Pires Costa^{*}; Pedro H. Parreiras; Pedro M. de Araújo; Victor Freire

Trabalho desenvolvido na disciplina Descrição e Apresentação de Dados com o intuito de explorar descritivamente a saúde e alimentação dos estudantes.

^{*} pedro.pires97@hotmail.com, Universidade Federal de Minas Gerais

Estudo da rentabilidade do Itaú Unibanco através de regressão linear

Poliana Maria Benelli* ; Gabriel Rodrigo Gomes Pessanha

O estudo da rentabilidade de uma empresa é um recurso necessário e visado principalmente pelos acionistas, visto que têm significativa participação no lucro líquido das companhias. Segundo Vieira et al. (2011), os índices de rentabilidade servem para medir a capacidade econômica da empresa, ou seja, evidenciam o grau de êxito econômico obtido pelo capital investido. O objetivo deste trabalho é a contribuição com as discussões sobre rentabilidade bancária, utilizando o indicador de rentabilidade Retorno sobre o Patrimônio Líquido (RPL ou ROE) que busca medir a atuação da organização no mercado e reaver a capacidade de valor investido. Diante deste exposto, pretende-se analisar os efeitos que as variáveis macroeconômicas Inflação, Produto Interno Bruto (PIB), Taxa de Câmbio e Taxa de juros Selic causam na Rentabilidade sobre Patrimônio Líquido (ROE) da instituição Itaú Unibanco. Inicialmente foram formuladas as 4 hipóteses a seguir, testadas pela metodologia empírica adotada: H1: A inflação apresenta relação direta com a rentabilidade bancária; H2: O PIB é diretamente proporcional ao ROE, pois quanto maior o crescimento econômico de um país, maior a produção naquele período; H3: A taxa de juros Selic é diretamente proporcional ao ROE; H4: A variação cambial é diretamente relacionada com a rentabilidade da instituição bancária. Os dados foram coletados através dos endereços eletrônicos Ipeadata (2017) e do Banco Central do Brasil (2017) e compreendem o período de março de 2008 a dezembro de 2015, sendo estes dados trimestrais. O estudo das variáveis foi feito mediante técnica de Regressão Linear Múltipla (RLM), via modelo de Mínimos Quadrados Ordinários (MQO), através do software gratuito Gretl 1.9.90. O modelo explica, em média, 93,15% das mudanças ocorridas na variável ROE, os testes de aderência do modelo foram aplicados e validados. As variáveis Taxa de inflação e Taxa de Juros Selic apresentaram relação direta com a variável dependente, apontando que um aumento nas taxas de Inflação e Selic implica em uma elevação na rentabilidade dos acionistas da instituição. Porém, as variáveis Taxa de Câmbio e PIB, demonstraram relação inversa com a rentabilidade. Sugere-se para trabalhos futuros, uma análise da rentabilidade de instituições bancárias utilizando outros indicadores de rentabilidade.

* poli_benelli@hotmail.com, Unifal-MG

Estudo comparativo entre a faixa original do NIR de bancada com a faixa reduzida para dados de teor de sacarose aparente

Raphael Henrique Teixeira da Silva^{*}; Roberta de Amorim Ferreira; Luiz Alexandre Peternelli

A Espectroscopia no Infravermelho Próximo (NIR) é um tipo de espectroscopia vibracional que emprega energia de fótons. Algumas das principais vantagens de se utilizar esta técnica é o fato de não ser necessário preparação de amostra, rapidez no tempo de análise, não destrutivo e invasivo, permite alta penetração do feixe de radiação de sondagem e possui aplicação quase universal. Existem dois tipos de NIR, o de bancada e o portátil. A diferença entre ambos os equipamentos é com relação à precisão dos resultados obtidos, sendo o NIR de bancada considerado mais preciso. Em contra partida, o NIR de bancada é um equipamento mais caro e, além disso, restrito a receber

as amostras no laboratório onde se encontra. Nesse sentido, o intuito deste trabalho foi comparar a faixa de comprimento de onda completa do NIR de bancada e retirar uma faixa menor deste intervalo que é equivalente à faixa de um NIR portátil. Concomitantemente, avaliar se apresentaram efeitos significativos para os resultados obtidos das análises espectrais do conjunto de dados de teor de sacarose aparente (PC) da cana-de-açúcar. Avaliando a questão do custo dos equipamentos, e devido aos resultados de ajuste, o NIR portátil aparenta ser uma alternativa pertinente em substituição ao NIR de bancada. Porém mais estudos devem ser realizados em outros conjuntos de dados para garantir maior consistência nas comparações.

^{*} raphaelhts@hotmail.com, UFV

Comportamento temporal do consumo residencial de energia elétrica nas regiões Sudeste e Nordeste do Brasil

Raphael Sussia Silva* ; Ednaldo Carvalho Guimarães

O objetivo deste trabalho foi: analisar descritivamente o comportamento temporal das séries históricas do consumo de energia elétrica, por residência, nas regiões nordeste e sudeste; ajustar os modelos de séries para fazer previsões para o consumo da energia; e verificar se o consumo residencial segue o mesmo padrão nas duas regiões. Foi utilizado o banco de dados da Empresa de Pesquisa Energética – EPE. O consumo mensal total foi dividido pelo número de consumidores para a obtenção do consumo por residência. Fizeram-se as análises gráficas das séries e ajustaram-se modelos teóricos de séries temporais aos dados observados. Na validação dos modelos ajustados foram utilizados os consumos mensais do ano de 2015 sendo comparados os valores realizados com os valores estimados. Os dados observados apresentaram aderência adequada o modelo aditivo de Winters, sendo que, para a região nordeste, onde para a região obteve-se $R^2 = 0,957$ Indicando um ótimo ajuste, já para a região Sudeste o $R^2 = 0,873$ que mostra um bom ajuste. As estimativas dos parâmetros de nível, tendência e sazonalidade, para a região nordeste, foram de: 0,40; $6,082 \times 10^{-5}$ e 0,001, respectivamente. Já para o sudeste obteve-se o parâmetro de nível de 0,29; tendência de $4,847 \times 10^{-7}$ e sazonalidade de 0,001. A função auto correlação e auto correlação de resíduos mostrou um modelo adequado em ambas às análises. A análise de erros entre valores estimados e valores observados do ano de 2015 apresentou média do erro relativo de 0,0262 (2,62%) para o Nordeste e de 0,0891 (8,91%). O comportamento temporal do consumo residencial, nas duas regiões, foi descrito pelo mesmo tipo de modelo, entretanto com parâmetros diferentes, sendo observada uma tendência de crescimento do consumo residencial maior na região sudeste. Os resultados permitiram conduzir que o consumo de energia residencial nas regiões nordeste e sudeste apresentam componentes de sazonalidade e de tendência.

* raphaelsussia@gmail.com, Universidade Federal de Uberlândia

Bandas de confiança para quantis usando núcleo-estimador

Raquel Cáfaró Marinho^{*}; Gregorio Saravia Atuncar

As aproximações pelo Teorema Central do Limite (TCL) são muito comuns na estatística, mas não fornecem bons resultados em todos os casos, especialmente com tamanhos de amostra pequenos, e isso influencia os intervalos de confiança obtidos. Este trabalho teve como objetivo utilizar duas técnicas diferentes para melhor aproximar os intervalos e bandas de confiança para a função de distribuição e quantis de distribuições: o método de núcleo-estimador, que suaviza a função de distribuição estimada, e a expansão de Edgeworth, que utiliza mais termos na aproximação assintótica. Para utilizar esses métodos, foram utilizadas três abordagens de forma a corrigir os erros dos estimadores do método de núcleo. A união dessas técnicas apresentou intervalos com maior probabilidade de cobertura e amplitude semelhante à dos intervalos já utilizados (como pela função de distribuição empírica e pela distribuição binomial no caso dos quantis), obtendo bom desempenho para as diversas distribuições usadas nas simulações. Os intervalos foram estendidos para obter bandas de confiança.

^{*} raquelcafaró@hotmail.com, Universidade Federal de Minas Gerais

Modelo de previsão para a série do custo da cesta básica em São João del-Rei

Mariana Carolina da Silva; Tania Moura Ferreira;
Rejane Corrêa da Rocha^{*}; Luciane Teixeira Passos Giarola

O objetivo principal deste trabalho foi ajustar um modelo de previsão, via metodologia de Box & Jenkins, para série da variação do custo total da cesta básica deflacionada pelo Índice de Preço ao Consumidor Amplo (IPCA) no período de janeiro de 2000 a maio 2017, sendo as observações de janeiro de 2017 a maio de 2017 reservadas para serem comparadas com as previsões. O cálculo da Cesta Básica de São João Del Rei é feito pelo Núcleo de Extensão e Pesquisa em Economia (NEPE) do Departamento de Ciências Econômicas da UFSJ, seguindo as diretrizes utilizadas pelo DIEESE. Para as análises da série foi utilizado o software Gretl. Inicialmente foi feita a análise visual do gráfico da amplitude versus média, constatando a necessidade de transformação logarítmica para estabilizar a variância. Posteriormente, foi feita a análise do gráfico da série transformada, do correlograma e do periodograma, constatando a necessidade da primeira diferença para retirada da tendência e não existência da componente sazonal. Para a escolha do melhor modelo foi utilizado o Critério de Informação de Akaike (AIC) e o erro percentual médio absoluto (MAPE). Desta forma, dentre os modelos ajustados, o que melhor se ajustou aos dados foi o ARIMA (2,1,3), sendo o $AIC=-590,8391$ e $MAPE=0,59\%$. Concluiu-se, que a metodologia de Box & Jenkins é útil para fazer previsões série da variação do custo total da cesta básica de São João del Rei.

^{*} rejcrocha@gmail.com, UFSJ

Investigação com estudantes do ensino médio sobre o entendimento das questões relacionadas à realidade LGBT+

Thiago Bassi Leandro da Silveira; Roberta Almeida Pereira* ; Rodrigo Ferreira da Cruz Barbosa; Nadja Carvalho Pereira; Leonardo Ferreira de Araújo Sá; Fausto Marques Pinheiro Júnior; Gustavo da Silva Ferreira.

Resumo: Questões relativas à realidade de Lésbicas, Gays, Bissexuais, Travestis e Transexuais têm sido bastante debatidas na sociedade, em especial os temas relacionados à homo-lesbo-bi-transfobia, identidades de gênero, orientação sexual, comportamento e ainda o impacto de todas essas informações nas noções de família. A diversidade cultural das identidades que derivam de sexualidades e gêneros não-normativos é representada pela sigla LGBT+ (o “+” denota a permanente preocupação em agregar outras formas dessa não-norma, como por exemplo penssexuais, assexuais, intersexuais, não-binários, queer dentre outras variações). Para investigar quais as opiniões pessoais dos jovens estudantes sobre o tema e compará-las aos posicionamentos da militância LGBT+, realizou-se uma pesquisa com estudantes do Ensino Médio de uma escola pública no Rio de Janeiro. Numa população de 799 estudantes foram selecionados, por amostra aleatória simples com nível de confiança de 90% e margem de erro de 5%, 250 participantes que responderam a 36 questões fechadas acerca do tema. De modo geral, a pesquisa permitiu observar que a maioria dos participantes (pelo menos 60%) deram respostas que indicam que suas opiniões pessoais estão de acordo com os posicionamentos da militância LGBT+. As questões relacionadas à identidade de gênero foram maioria dentre as que apresentaram maior dispersão nas respostas seguidas pelas relacionadas à sexualidade. Isto pode ser um indicador de que, mesmo num ambiente onde há debates acerca de gênero e sexualidade, essas questões ainda geram confusão. Os resultados forneceram informações que podem ser utilizados por dirigentes e demais membros da comunidade escolar em debates sobre a temática LGBT+. Além de orientar a militância LGBT+ acerca dos temas que carecem de maior visibilidade.

* profroberta92@gmail.com, ENCE / IBGE

Partição de dados para uso em validação cruzada e sua influência sobre a qualidade de ajuste e seleção de variáveis em modelos de calibração

Roberta de Amorim Ferreira^{*}; Raphael Henrique Teixeira da Silva; Luiz Alexandre Peternelli; Mateus Teles Vital Gonçalves; Reinaldo Francisco Teófilo

Métodos estatísticos multivariados associados a procedimentos alternativos para medição das características tecnológicas, como a espectroscopia no infravermelho próximo (NIR), vem sendo utilizados para predição de indivíduos mais produtivos, visto que seu uso é simples, rápido e de baixo custo. O grande volume de variáveis (comprimentos de onda) obtidos de uma única amostra impõe dificuldades para construção de um modelo de calibração. Para superar esses problemas aplicam-se procedimentos de compressão de dados e de seleção de variáveis. Para construção e validação do modelo o método dos mínimos quadrados parciais (PLS) se mostrou eficiente. Uma etapa muito importante ao se lidar com PLS, é a validação cruzada (ou “cross-validation - CV”), que consiste em particionar as amostras do conjunto de dados, em dois subconjuntos: treinamento teste. Quanto à partição dos dados, podemos destacar o algoritmo Kennard e Stone e o 10-fold CV. Comparamos a qualidade de ajuste (coeficiente de correlação – R, entre valores observados e preditos) e seleção de variáveis nas diferentes alternativas de CV em espectros NIR obtidos de 168 amostras de teor de sacarose aparente de clones de cana-de-açúcar. Para diferentes partições, observou-se uma seleção significativa de variáveis (0,74 a 6,3%) dentre as 3113 originais. A qualidade de ajuste manteve-se próxima. Os melhores resultados obtidos foram $R = 0,7902$ para o modelo completo, e $R = 0,798$ no modelo reduzido com 23 variáveis selecionadas. Concluimos que a seleção de variáveis pouco alterou a qualidade preditiva do modelo. Porém, diferentes partições acarretaram diferentes resultados. Maiores estudos serão conduzidos para identificar regiões dos espectros NIR responsáveis por essas diferenças.

^{*} robertadeamorim@hotmail.com, Universidade Federal de Viçosa

Validação do inventário de motivações para uso de jogos online a partir de estudo exploratório com usuários do Pokémon Go

Rômulo Rafael da Silva^{*}; Leonardo Gonçalves Pinheiro Júnior; Marcela Mansur-Alves;
Clarindo Isaías Pereira da Silva e Pádua.

O modelo empírico de motivações de jogador de Nick Yee (2006), composto por 10 construtos, compreende um inventário psicométrico para investigação das motivações dos usuários por jogos online. A pesquisa exploratória tem por objetivo a validação e a medição da confiabilidade do modelo e de suas propriedades psicométricas, adequando-se o instrumento às características do Jogo Pokémon GO e posterior aplicação junto ao público universitário. A coleta dos dados ocorreu entre 25/10/2016 e 02/11/2016 por formulário online enviado por e-mail à comunidade acadêmica da UFMG. A amostra é composta por 192 alunos - 157 da graduação (81,3%) - com 95 do sexo feminino, outros 95 do sexo masculino e 2 não declararam e a idade média foi de 23,5 anos. Quanto às áreas de conhecimento, Ciências Exatas e da Terra, Ciências Humanas e Engenharias foram as principais da amostra. A análise descritiva indica que os respondentes atribuem muita relevância aos itens dos construtos “Exploração” e “Escapismo”, ou seja, maior interesse pela imersão no mundo virtual do jogo em oposição ao real. O construto “Mecânica” foi excludido por não ser possível a medição de somente um item. Aos demais 7 construtos - Progresso, Competição, Socialização, Relacionamento, Trabalho em Equipe, Encenação e Personalização - pouca relevância foi atribuída a eles pelos respondentes. A análise fatorial exploratória indicou relevância de todos os itens que compõem os construtos validados na análise descritiva, exceto os dois itens (Q25 e Q26) do constructo “Socialização”. A dimensionalidade, confiabilidade e validade convergente dos construtos foram analisadas pelos critérios de Variância Extraída, Alfa de Cronbach, Confiabilidade Composta, Medida de adequação da amostra de KMO e Dimensionalidade. A partir dos resultados desses critérios, afirmar-se que os 9 construtos finais são relevantes e válidos. O estudo permitiu confirmar a validade dos construtos do instrumento e sua confiabilidade.

^{*} romulorafa@ufmg.br, UNIVERSIDADE FEDERAL DE MINAS GERAIS

Análise do risco de evasão no curso de estatística UFOP considerando dados censurados

Rosielle da Costa Farias^{*}; Graziela Dutra Rocha Gouvêa; Helgem de Souza Ribeiro Martins

A evasão escolar é caracterizada quando o aluno deixa de frequentar a instituição de ensino sem adquirir o grau almejado, e esse fenômeno pode acarretar várias preocupações para o estudante, para os órgãos públicos, bem como para a sociedade. A evasão no curso de Estatística na UFOP é um problema decorrente de várias causas que podem ser colocadas em discussão, para solucioná-los e para evitar o aumento deste evento. Uma área da Estatística que pode ser utilizadas nos casos de evasão, é a Análise de Sobrevivência. Neste estudo, a variável resposta é o tempo até falha e sua característica é a presença de censuradas. O objetivo do trabalho é estudar os fatores que influenciam no tempo até a evasão dos alunos do Curso de Estatística do ICEB/UFOP por meio do modelo de Cox (Cox, 1972) e então, contribuir com o departamento para tentar minimizar este problema. Neste estudo utilizou dados referentes a 166 alunos do curso de estatística entre 2008 e 2013 no ICEB/UFOP. O tempo foi medido em horas cursadas e as covariáveis estudadas foram: ANO - Ano de admissão, AD - Modo admissão, COR - Cor da pele, OR - Estado de nascimento, ID - Idade (em anos); CID - Cidade de nascimento, SEX – sexo. Para a análise foi utilizado o modelo de Regressão de Cox. A suposição básica para a utilização deste modelo é que as taxas de falhas sejam proporcionais. A estimação dos parâmetros foi feita pelo método da máxima verossimilhança parcial. A seleção das covariáveis foi baseada na proposta de Collet (1994). As análises foram realizadas no software R 3.4.0. Verificou-se a suposição de proporcionalidade para as covariáveis e procedeu-se o ajuste do modelo. Foi realizada a análise de resíduos de Schoenfeld e verificou-se a adequabilidade do modelo. De acordo com os resultados, as covariáveis significativas foram sexo (o risco de evasão dos alunos do sexo masculino é 1,76 vezes maior do que alunas do sexo feminino) e cor da pele (alunos de cor de pele branca tem o risco de evasão de 1,78 vezes maior de alunos de outras cores de pele).

^{*} rosi-elle@hotmail.com, UFV

Estimação bayesiana de intensidade de tráfego em filas multiservidoras

Sandy P. Alves^{*}; Frederico R. B. Cruz

Neste artigo, o foco é no problema de inferência em sistemas de filas multiservidoras markovianas, isto é, em filas nas quais o número de itens que chegam para serviço segue um processo de Poisson e os tempos de serviço são exponencialmente distribuídos. Importantes situações práticas podem ser modeladas aproximadamente por esse tipo de filas, tais redes de computadores e telecomunicações, sistemas manufatura e de serviços, sistemas de saúde, entre outros problemas semelhantes da vida real. Para a estimação da intensidade de tráfego na fila, que é definida como a razão entre a taxa da chegada e a de serviço, utiliza-se uma técnica bayesiana, o método de amostragem/reamostragem de importância (ou SIR, do inglês *sampling/importance resampling*). A partir da intensidade de tráfego, outras medidas de desempenho importantes podem ser determinadas (tais como, tamanho médio da fila, número esperado de usuários no sistema, probabilidade de sistema vazio, dentre outras). Resultados numéricos são apresentados, para demonstrar a exatidão e a eficiência do SIR, perante o uso de amostras de tamanho finito.

^{*} dyllas28@hotmail.com, UFMG

Modelos lineares generalizados e suas aplicações em Ciências Atuariais sob a visão bayesiana e frequentista da inferência

Sérgio Luiz Moreira Júnior* ; Vinícius Diniz Mayrink

Neste trabalho, selecionamos alguns bancos de dados relacionados à área de Ciências Atuarias, onde a variável resposta está contextualizada em termos de seguros automotivos e pessoal. Desenvolvemos uma análise de regressão utilizando modelos lineares generalizados para explicar e interpretar a relação de diversas covariáveis com a variável resposta de interesse que são: o número de sinistros de envolvendo veículos e valores de indenizações pagas para seguros pessoais. A análise é realizada sob o ponto de vista Clássico (usando a função `glm` do software R, baseada em máxima verossimilhança) e Bayesiano (utilizando o software JAGS e sua interface com o R). O estudo é focado nos modelos Poisson, Logístico e Gama. Os resultados são muito semelhantes na comparação das estimativas Clássicas e Bayesianas, ou seja, a interpretação dos resultados é muito parecida sob as duas abordagens. O objetivo principal deste trabalho é, além de comparar as duas abordagens de inferência, interpretar os efeitos das covariáveis e explorar o ajuste de modelos lineares generalizados no JAGS; lembrando que este software é bastante atrativo para a implementação do MCMC Gibbs Sampling, pois requer apenas a especificação das distribuições a priori e da função de verossimilhança (cálculos complicados para determinar distribuições condicionais completas não são necessários, sendo realizados automaticamente pelo JAGS). Ressaltamos também que uma outra vantagem do JAGS é o fato de construir toda estrutura do MCMC (incluindo passos utilizando algoritmos extras como o Metropolis-Hastings que permitem a amostragem indireta das condicionais completas desconhecidas).

* sergiojunior@ufmg.br, Universidade Federal de Minas Gerais

Análise de variância multivariada aplicada a bebidas energéticas consumidas por não atletas

Simone Daniela Sartorio de Medeiros^{*}; Isabela Rótoli Micaroni;
Marta Regina Verruma-Bernardi; Ionele do Socorro Santos Ferreira

As bebidas energéticas hoje possuem um perfil bastante variado de consumidores, sendo utilizadas para diversos fins e suas particularidades regem a preferência de cada consumidor. A comum utilização da ANOVA para dados com variáveis dependentes pode causar distorções nos resultados uma vez que só é válida para situações de variáveis não correlacionadas. Neste contexto, o uso da análise de variância multivariada (MANOVA) torna-se essencial, pois é uma técnica que investiga diferenças significativas entre grupos considerando simultaneamente as múltiplas variáveis dependentes. O objetivo do trabalho foi estudar a semelhança entre vetores de médias de 5 diferentes bebidas energéticas (A, B, C, D e E), oferecidas a 50 provadores considerados não atletas, através MANOVA utilizando o software R. O teste de aceitação foi aplicado para 5 variáveis (Aceitação Global, Aparência, Aroma, Cor e Sabor) com escala hedônica estruturada de 9 pontos. Verificou-se correlação significativa entre as variáveis e pela MANOVA constatou-se que existe diferença entre os vetores de médias de bebidas para todos os testes realizados (Wilks, Hotteling-Lawley, Pillai e Roy). Os intervalos de confiança mostraram que existe diferença entre as médias dos grupos, em particular, indicaram que a bebida C se diferenciou das demais bebidas enquanto ao aroma, sabor e aceitação, sendo a favorita pelos provadores, pois obteve a maior nota para todos os atributos avaliados. Desta forma, os atributos que mais chamaram a atenção dos não atletas foram o aroma e especialmente o sabor, pois geraram maior aceitação, uma vez que estas variáveis estão correlacionadas positivamente.

^{*} sisartorio@cca.ufscar.br, CCA/UFSCar

Análise de desempenho de operadoras de saúde utilizando análise multivariada

Taylor Fidelis^{*}; William Santos; Reinaldo Marques; Márcio Pereira;

A análise dos indicadores econômicos e financeiros envolve comparar o desempenho de determinadas empresas com outras de mesmo ramo e porte, verificando sua evolução ao longo do tempo. Os indicadores econômicos e financeiros fornecem informações relevantes que condizem com a realidade em qual uma operadora se encontra. Tais indicadores podem ser calculados a partir do balanço patrimonial, demonstração de resultado do exercício, fluxo de caixa e indicadores operacionais de desempenho. Neste contexto, o presente trabalho se propõe a analisar e comparar as demonstrações financeiras das operadoras de saúde suplementar brasileiras de grande porte do ano de 2008 até 2016. Para a avaliação da performance das operadoras via componentes principais e aplicação de métodos de agrupamento, e por fim ranking das operadoras ao longo do tempo.

^{*} taylorfidelis@hotmail.com, Universidade federal de alfnas

Mortalidade por doença de Chagas no Brasil: análise e previsões bayesianas dos efeitos de idade, período e coorte de nascimento

Taynãna César Simões*, Karina Cardoso Meira, Fábio Rocha da Silva, Fabíola Bof de Andrade, Laiane Félix Borges, Auzenda Conceição Parreira de Assis, Maria Vitórias Silva

A doença de Chagas (DC) é uma infecção parasitária crônica causada pelo *Trypanosoma cruzi*, sendo considerada uma doença negligenciada e de grande interesse para a Saúde Pública, por ser evitável e ter complicações graves. No Brasil, tem-se evidenciado tendência descendente na mortalidade pela doença nas últimas décadas, porém, a maioria dos estudos não considera o efeito da coorte de nascimento na evolução das taxas de mortalidade. Este trabalho teve por objetivo analisar os efeitos da idade, período e coorte de nascimento (Efeitos APC) na evolução da mortalidade por esta doença no país, nas faixas etárias de 20-24 a 80 e mais anos, durante os últimos 35 anos, segundo sexo e grandes regiões brasileiras, sendo ainda avaliados termos de efeitos aleatórios. Previsões são importantes devido às mudanças demográficas, além de avanços em tratamento e diagnóstico, assim estimamos as taxas de mortalidade por DC até o ano de 2034. Como ferramentas analíticas, foram utilizados os modelos idade-período-coorte bayesianos (BAPC), e o método determinístico INLA (Integrated Nested Laplace Approximations) para inferência estatística. As maiores taxas de mortalidade em todo o período foram observadas entre os homens e na região Centro-Oeste. De forma geral, houve diminuição das taxas ao longo do período observado. O risco de morte por DC apresentou aumento progressivo com o avançar da idade. Em contrapartida, verificou-se redução do risco de morte pela doença nas coortes de nascimento mais jovens. Os modelos BAPC multivariados se mostraram

mais significativos que os modelos estratificados por sexo e região de residência, porém sexo não teve efeito significativo nos efeitos temporais. Efeitos de período e coorte variaram entre as regiões brasileiras. Acredita-se que a redução progressiva da mortalidade por essa causa, nas coortes de nascimento mais jovens, seja resultado do efeito de período da redução da transmissão vetorial e transfusional do protozoário, assim como do avanço no conhecimento da história natural da doença com consequentes melhorias na terapêutica.

* projetos@cpqrr.fiocruz.br, Fundação Osvaldo Cruz

Infecção pelo vírus HPV: um modelo logístico com abordagem bayesiana

Thaís Pacheco Menezes* ; Magda Carvalho Pires

O objetivo do presente trabalho foi ajustar um modelo de regressão logístico para identificar os fatores de risco para infecção pelo vírus HPV. O estudo desenvolvido na Universidade da Califórnia-São Francisco (UCSF) consistiu no acompanhamento de 104 mulheres entre 13 e 21 anos com diagnóstico negativo para HPV. Como existem muitos tipos de vírus HPV, o teste diagnóstico aplicado pode produzir alguns falsos negativos e, por isso, uma adaptação no modelo logístico foi realizada. Uma abordagem bayesiana se mostrou vantajosa e métodos MCMC para inferência a posteriori foram necessários. Foram definidas distribuições a prior normais para os coeficientes da regressão e distribuições a priori Betas para as probabilidades de que o teste forneça falsos positivos e falsos negativos. O software JAGS foi utilizado em conjunto com o software R através do pacote 'rjags'. Verificou-se que 'ter um novo parceiro sexual nos últimos dois meses' é fator de risco para infecção pelo HPV.

* thaispacheco1207@gmail.com, Universidade Federal de Minas Gerais

Análise da confiabilidade de um questionário para determinar a satisfação dos alunos utilizando o software R

Pimentel, U.*; Gorini, E.

A realização de pesquisas de satisfação é de grande importância para o desenvolvimento de um ensino de qualidade. No Brasil Carmo, Silva e Miranda (2011) obteve o grau de satisfação de alunos da Universidade Federal do Amazonas, por meio de um questionário analisando as condições físico-psicológicas oferecidas nos cursos da universidade referida.

Diante disso, a utilização de questionários confiáveis é de extrema importância para a execução de uma pesquisa. Segundo Almeida, Santos e Costa (2010), o coeficiente alfa de Cronbach é usado para medir a correlação entre respostas em um questionário por meio das respostas dadas pelos respondentes.

O objetivo deste trabalho consiste em determinar a satisfação e insatisfação dos alunos para a melhoria da qualidade, utilizando o software R para uma análise descritiva das respostas referente a satisfação e utilização de recursos gráficos, como gráfico de densidade e blox Plot. Também foi realizado uma análise de Confiabilidade do questionário, para que o mesmo possa ser melhorado e sirva de base para futuras pesquisas.

Com os dados desta pesquisa realizada em 2015 foi possível verificar a necessidade de reformulação do questionário para análise de dois grupos. Tal fato é de extrema importância, pois no próximo semestre (2017/2) uma nova coleta de dados será realizada, e com isso será possível realizar uma análise estatística mais elaborada.

De acordo com os dados apresentados no estudo, pôde-se concluir que o grau de satisfação dos acadêmicos em relação ao CCA/UFES está comprometido devido a algumas variáveis na infraestrutura, como a climatização e estrutura física das salas de aula e bebedouro, entanto, os aspectos relacionados a qualidade do ensino, proporcionaram uma avaliação positiva para o grau de satisfação dos alunos.

E também foi possível verificar a importância da utilização de recursos computacionais ao se trabalhar com um grande conjunto de dados, onde o R proporciona uma facilidade para a realização de cálculos estatísticos resumos e gráficos. Por outro lado, o excel 2016 serviu de base para refinar os dados originais (sem nenhum tratamento), por possuir funções de buscas especializadas e proporcionar a filtragem dos dados.

Palavras Chave: Satisfação, alunos, R, alfa de cronbach, universidade.

* uillapimenteleng@gmail.com, UFES

Uma aplicação de modelo linear generalizado sob enfoque bayesiano para dados aumentados

Vânia de Fátima Lemes de Miranda^{*}; Patricia Mendes dos Santos; Cristian Tiago Erazo Mendes; Luciano Antônio de Oliveira; Júlio Sílvio de Souza Bueno Filho

Em diversos contextos de pesquisa, a distribuição dos dados coletados não são provenientes de uma densidade normal e a média muitas vezes não representa uma combinação linear dos parâmetros, mas sim uma função dos mesmos. Para estes casos é possível ainda conduzir uma análise de regressão utilizando modelos lineares generalizados. Os modelos lineares generalizados surgiram nos anos 70 e impactaram significativamente o desenvolvimento da teoria estatística. Nessa abordagem, a superdispersão pode ser modelada por meio da estimação do parâmetro de escala ou pela incorporação de efeitos aleatórios, o que abre perspectivas interessantes para a utilização do enfoque bayesiano. Neste trabalho analisou-se um conjunto de dados referentes ao número de irmãos de estudantes de uma turma de pós-graduação do programa de pós-graduação em estatística e experimentação agropecuária da UFLA. Algumas hipóteses foram assumidas, dentre as quais destacamos, a presença do efeito de bancada, ou seja, que a posição dos estudantes na sala afetaria a variável quantificada. O principal objetivo foi mostrar a flexibilidade do método Bayesiano para o ajuste de um modelo linear generalizado aplicado a dados de contagem, utilizando a distribuição de Poisson e a função de ligação log. Em particular, ressalta-se como a introdução de um fator latente a abordagem pode facilitar os cálculos algébricos e principalmente otimizar o processo de amostragem evitando utilizar o algoritmo metropolis hastings para amostragem de todos parâmetros do modelo. Destaca-se aqui que o efeito de bancada foi utilizado apenas em uma perspectiva didática o que, de fato, foi evidenciado pelos resultados da análise.

^{*} vaniafamat@gmail.com, UFLA

Aprendizagem bayesiana e identificabilidade em modelos de regressão com erros nas covariáveis

Victor S. Sartório^{*}; Gustavo M. A. Rocha

Neste trabalho é estudado o problema de identificabilidade de um ponto de vista bayesiano, com um enfoque nos modelos de regressão linear estrutural com erro nas covariáveis. É realizada uma revisão bibliográfica de resultados já existentes na literatura sobre o problema em questão e são apresentados resultados de estudos de simulação onde discutimos maneiras de como lidar com o problema de identificabilidade através de medidas de aprendizagem bayesiana, sem ter de recorrer à imposição de novas premissas ao modelo.

^{*} vct.sartor@gmail.com, ENCE - Escola Nacional de Ciências Estatísticas

Um estudo da tendência temporal da taxa suicida no Brasil

Victor Ferreira da Silva^{*}; Kelly Pereira de Lima; Édipo Menezes da Silva;
Felipe Fernandes; João Domingos Scalon;

No Brasil, as taxas de suicídio apresentam níveis baixos quando comparadas a outros países do mundo. Nos países como Lituânia, Bielorrússia, Rússia, Cazaquistão, Hungria, Japão e Coréia do Sul que registram índices que variam entre 20 e 40 óbitos por 100 mil habitantes, o Brasil teve 5,8 em 2014 (MAIA, 2016). Com isso, o trabalho tem como objetivo analisar a tendência temporal da taxa de suicídio no Brasil. Os dados foram obtidos a partir do relatório da Organização Mundial de Saúde (WHO), no período de 1980 a 2014. As tendências foram estimadas, por meio, dos mínimos quadrados ordinários, ou seja, usou-se a regressão linear simples, para descrever o comportamento dados em relação a variável dependente e independente. As taxas de suicídios foram a variável dependente e ano das respectivas taxas de ocorrência, a variável independente. Além disso, o coeficiente de determinação (R^2) e análise de resíduos foram usados para verificar da validação do modelo. Para a tendência significativa, ou seja, quando o valor-p $< 0,05$. A análise de resíduos não mostrou problemas com as suposições do erro aleatório. O modelo ajustado mostra que o tempo explica 92% da variabilidade das taxas de suicídio ($R^2 = 0,9175$; $p < 0,001$). O coeficiente angular de 0.06531 do modelo indicou um crescimento positivo, estatisticamente significativo, nas taxas de suicídio, ou seja, no Brasil a tendência apresenta uma variação média 6,5% a cada ano.

^{*} victor.est.ufop@gmail.com, Ufla

Uma análise espacial da dinâmica da desigualdade de renda em São Paulo nos anos de 1991, 2000 e 2010

Walef Machado de Mendonça* ; Patrícia de Siqueira Ramos.

O Brasil está há décadas entre os países mais desiguais do mundo e medir essas desigualdades pode auxiliar na elaboração políticas públicas que melhorem a distribuição de renda. Esta pesquisa busca avaliar a dinâmica espacial da desigualdade de renda nos municípios do estado de São Paulo utilizando o índice de Palma como variável. Esse estado foi escolhido devido à sua grande diversidade socioeconômica. O interesse é verificar se municípios com maior desigualdade de renda tendem a estar próximos de outros municípios também desiguais, isto é, verificar se há dependência espacial em relação à desigualdade. Para isso foi aplicada a análise exploratória de dados espaciais (AEDE) para descrever os padrões de associação espacial por meio do coeficiente de autocorrelação espacial, o I de Moran. O estudo foi feito utilizando a linguagem de programação Python e os dados dos censos demográficos do IBGE, consultados por meio do Atlas de Desenvolvimento Humano no Brasil 2013. Foi possível concluir que existe uma autocorrelação positiva entre a desigualdade e suas respectivas defasagens espaciais. Os testes envolvendo o I de Moran indicam a existência de dependência espacial em 1991, 2000 e 2010, ou seja, municípios desiguais tendem a estar próximos de municípios desiguais. Além disso, os municípios mais desiguais tendem a se concentrarem nas mesorregiões do Vale do Paraíba Paulista, metropolitana de São Paulo e Litoral Sul. Também foi possível observar, pela observação dos valores do índice de Palma, que houve uma queda nos níveis de desigualdade entre os anos de 1991 e 2010 (embora tenha havido aumento da concentração espacial da desigualdade de renda).

* walefm2@gmail.com, Universidade Federal de Alfenas

Uma análise do programa Bolsa Família utilizando pareamento por escore de propensão

Wanessa Gabrielle Toledo Martins^{*}; Patrícia de Siqueira Ramos; Lincoln Frias

O objetivo desse trabalho foi realizar uma avaliação de impacto do Programa Bolsa Família (PBF) sobre a participação dos beneficiários no mercado de trabalho, utilizando o Pareamento por escore de propensão (PSM) pelos métodos de estratificação e de blocos. Após o pareamento, verificou-se que o PBF não influencia no nível de atividade das pessoas no mercado de trabalho.

^{*} wanessa_toledo@hotmail.com, Universidade Federal de Alfnas

Biblioteca 'Wordevok' do software R

Wesley Pereira^{*}; Gilvan Guedes; Denise Duarte; Rodrigo Ribeiro; Marina Amorim

Como legado do Relatório de Monografia apresentado pelo autor em 2017, a biblioteca 'wordevok' foi construída com o intuito de se tornar uma ferramenta capaz de transformar dados colhidos pela Técnica de Associação Livre de Palavras em dados relacionais para a construção de grafos. Além disso, a biblioteca possui um conjunto de funções que auxiliam na descrição e interpretação dos resultados obtidos pela metodologia aplicada neste trabalho, visando facilitar a utilização dos mesmos métodos para trabalhos futuros.

^{*} wesleyhspereira@gmail.com, Universidade Federal de Minas Gerais

Desenvolvendo uma TIC para o ensino de probabilidade

Wesley Velloso Marques^{*}; Adriana Dias; Flávio Bittencourt.

A estatística e a probabilidade tem sido assunto de discussão de diversos autores sobre a necessidade de ser inserida desde as séries iniciais. Estes conteúdos têm sido ampliados e cada vez mais têm-se notado que são tão importantes quanto à geometria, a álgebra e a aritmética. Entende-se que a sua importância esteja relacionada ao desenvolvimento de indivíduos mais críticos e com maior capacidade de argumentação. Isto porque na estatística e na probabilidade é permitido trabalhar com as incertezas e a aleatoriedade, o que põe fim à aceitação de que tudo é determinístico e que ao mesmo tempo permite ao indivíduo compreender que muitos acontecimentos são decorrentes de causas específicas os quais podem ser detectáveis com certeza científica. Entretanto, ensinar probabilidade ou mostrar a existência do acaso não é tão trivial, porém com o advento da informática e as Tecnologias da Informação e Comunicação (TIC's) acredita-se que pode-se promover mudanças tanto na dinâmica da aula quanto na forma de ensinar e aprender. Há muitos programas que são utilizados para a construção de material de apoio ao ensino (TIC's), um utilizado neste trabalho foi o FlashDevelop, que é um programa computacional livre e de código aberto, escrito em C# e construído sobre componente eficiente do editor Scintilla que apresenta como característica permitir a edição, compilação, depuração e publicação de um projeto Flash ActionScript. O objetivo deste trabalho é desenvolver uma TIC para o ensino de probabilidade de modo que o aluno perceba o acaso por meio de animações que utilizam lançamentos de dados, moedas e seleção de cartas de baralhos.

^{*} wmarques092@gmail.com, UNIFAL

Distribuições de probabilidade Fuzzy para severidade de sinistros

William Oliveira Santos^{*}; Leandro Ferreira; Reinaldo Antônio Gomes Marques;
Leonardo Henrique Costa; Luísa Pimenta Terra

Incertezas relacionadas a uma determinada variável aleatória podem ser incorporadas utilizando a teoria dos conjuntos fuzzy com suas funções de pertinência. O objetivo deste trabalho foi apresentar um estudo de distribuições de probabilidade fuzzy em modelos de severidade de sinistros individuais. Para tanto, algumas distribuições de probabilidade contínuas foram analisadas, sendo seus parâmetros representados por números fuzzy. Com base nos resultados obtidos, os modelos probabilísticos fuzzy de severidade de sinistros individuais, considerando incertezas nos parâmetros das distribuições, são uma alternativa viável no estudo de modelos probabilísticos de sinistros agregados para precificação de seguros, provisões técnicas e resseguro.

^{*} william.osantos96@gmail.com, Universidade Federal de Alfenas

Confecção de pictogramas no R

Willian Alves Lion*; Beatriz de Oliveira Rodrigues; Felipe de Melo Taveira; Flávio Bittencourt; Adriana Dias

Nos Parâmetros Curriculares Nacionais (PCN's) consta que a partir do ensino fundamental o aluno deverá ser capaz de utilizar diferentes linguagens como a matemática e a gráfica com o objetivo de produzir, expressar e comunicar suas ideias. A linguagem gráfica é aparentemente simples, porém a sua função é de suma importância pela possibilidade de representar a variável de interesse em uma forma de mais fácil visualização. Permite a quem observa verificar o comportamento da variável, como tendências, como valores mais frequentes, entre outros. Um gráfico pouco explorado nos livros didáticos das séries iniciais e até nos livros de graduação é o pictograma. Um pictograma é um gráfico em que se usa símbolos em substituição às colunas (ou barras) para representar a variável de interesse. Acredita-se que as representações de diferentes situações do cotidiano do aluno, como qual time os seus colegas torcem, por meio de imagens seja mais lúdico de se trabalhar e introduzir o conceito de gráfico, do que utilizar de técnicas padrões do ensino. No programa R é possível confeccionar uma gama de gráficos para representar quaisquer variáveis ou estudos de interesse, devido à essa versatilidade do R pensou-se em explorar este campo. Desta forma, o objetivo deste trabalho é apresentar alguns pictogramas que seja possível de se construir pelo R a fim de que tal produto possa ser utilizado por professores dos diferentes níveis de ensino.

* willianBr66@outlook.com, Universidade Federal de Alfenas